

Research



Cite this article: Corominas-Murtra B, Seoane

LF, Solé R. 2018 Zipf's Law, unbounded complexity and open-ended evolution.

J. R. Soc. Interface **15**: 20180395.

<http://dx.doi.org/10.1098/rsif.2018.0395>

Received: 29 May 2018

Accepted: 19 November 2018

Subject Category:

Life Sciences – Mathematics interface

Subject Areas:

biocomplexity, evolution, biophysics

Keywords:

complexity, algorithmic complexity, open-ended evolution, Zipf's Law

Authors for correspondence:

Bernat Corominas-Murtra

e-mail: bernat.corominas-murtra@ist.ac.at

Ricard Solé

e-mail: ricard.sole@upf.edu

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.4324055>.

Zipf's Law, unbounded complexity and open-ended evolution

Bernat Corominas-Murtra¹, Luís F. Seoane^{2,3,4} and Ricard Solé^{3,4,5}

¹Institute of Science and Technology Austria, Am Campus 1, 3400, Klosterneuburg, Austria

²Department of Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

³UPF-PRBB, ICREA-Complex Systems Lab, Dr Aiguader 88, 08003 Barcelona, Spain

⁴Institute Evolutionary Biology, UPF-CSIC, Pg Maritim Barceloneta 37, 08003 Barcelona, Spain

⁵Santa Fe Institute, 1399 Hyde Park Road, 87501 Santa Fe, NM, USA

BC-M, 0000-0001-9806-5643; LFS, 0000-0003-0045-8145; RS, 0000-0001-6974-1008

A major problem for evolutionary theory is understanding the so-called *open-ended* nature of evolutionary change, from its definition to its origins. Open-ended evolution (OEE) refers to the unbounded increase in complexity that seems to characterize evolution on multiple scales. This property seems to be a characteristic feature of biological and technological evolution and is strongly tied to the generative potential associated with combinatorics, which allows the system to grow and expand their available state spaces. Interestingly, many complex systems presumably displaying OEE, from language to proteins, share a common statistical property: the presence of Zipf's Law. Given an inventory of basic items (such as words or protein domains) required to build more complex structures (sentences or proteins) Zipf's Law tells us that most of these elements are rare whereas a few of them are extremely common. Using algorithmic information theory, in this paper we provide a fundamental definition for open-endedness, which can be understood as *postulates*. Its statistical counterpart, based on standard Shannon information theory, has the structure of a variational problem which is shown to lead to Zipf's Law as the expected consequence of an evolutionary process displaying OEE. We further explore the problem of information conservation through an OEE process and we conclude that statistical information (standard Shannon information) is not conserved, resulting in the paradoxical situation in which the increase of information content has the effect of erasing itself. We prove that this paradox is solved if we consider non-statistical forms of information. This last result implies that standard information theory may not be a suitable theoretical framework to explore the persistence and increase of the information content in OEE systems.

1. Introduction

Life has been evolving on our planet over billions of years, undergoing several major transitions along with multiple events of both slow and rapid change affecting structure and function [1–4]. Life seems to be indefinitely capable of increasing in complexity. This is illustrated, as an instance, by the trend towards larger genomes and diverse cell types exhibited by multicellular organisms. Moreover, the emergence of high neuronal plasticity and complex communication provided the substrate for non-genetic modes of adaptation. A key concept that pervades many of these innovations is the idea that evolution is 'open-ended'. Following [5], open-ended evolution (OEE) can be defined as follows: 'a process in which there is the possibility for an indefinite increase in complexity.' What kind of systems can exhibit such unbounded growth in complexity [6]? What are the conditions under which the complexity—and thus, the information content of the system—can increase and what are the footprints of such an open-ended increase of complexity? Which kind of information is encoded in an OEE system? The aim of this paper is to give hints to the these questions.

Open-ended evolutionary change needs a dynamical behaviour allowing complexity to grow in an unbounded way [5,7]. This requires a very large exploration space but this is only a necessary requirement. For example, as noticed in [8] mathematical models used in population genetics involving infinite alleles—using Markov models—do not display OEE. Previous attempts to address the problem of OEE involved different approximations and degrees of abstraction. John von Neumann was one of the early contributors to this issue [5,9,10]. In all these studies, some underlying mechanism is assumed to be operating, and arguments are made concerning the presence of self-replication, genotype–phenotype mappings, special classes of material substrates and physico-chemical processes [5,11]. On the other hand, a theory of OEE might demand a revision of the role of novel niches and abiotic changes, as well as refining what we understand as the open-endedness of a system [12,13]. Special suitable candidates for OEE systems are complex systems exhibiting generative rules and recursion. The best-known case is human language. Thanks to recursion, syntactic rules are able to produce infinite well-formed structures and thereby the number of potential sentences in a given language is unbounded [14]. In another example, Darwinian evolution proceeds through tinkering [15,16], continuously reusing existing parts. These are first copied—hence bringing in some redundancy into evolving systems—but are later on modified through mutation or recombination. Despite the obvious differences existing between Darwinism in biology and human-guided engineering [15], this process of tinkering appears to be common too in the growth of technological systems, thus indicating that copy-and-paste dynamics might be more fundamental than expected [17].

These systems are very different in their constitutive components, dynamics and scale. However, all share the presence of a common statistical pattern linked to their diversity: fat-tailed distributions. Four examples are provided in figure 1. In all these cases, the frequency distribution of the basic units decays following approximately Zipf's Law. Zipf's Law was first reported for the distribution of city sizes [21], and then popularized as a prominent statistical regularity widespread across all human languages: in a huge range of the vocabulary, the frequency of any word is inversely proportional to its rank [22,23]. Specifically, if we rank all the occurrences of words in a text from the most common word to the less common one, Zipf's Law states that the probability $p(s_i)$ that in a random trial we find the i th most common word s_i (with $i = 1, \dots, n$) falls off as

$$p(s_i) = \frac{1}{Z} i^{-\gamma}, \quad (1.1)$$

with $\gamma \approx 1$ and Z the normalization constant, i.e. $Z = \sum_{i \leq n} i^{-\gamma}$. Stated otherwise, the most frequent word will appear twice as often as the second most frequent word, three times as often as the third one, and so on. This pattern is found in many different contexts and can emerge under different types of dynamical rules (see [23–27] and references therein).

The examples shown in figure 1 involve: (a) LEGO models, (b) human language, (c) proteins, and (d) evolved electronic circuits. The first example provides an illustration of structures emerging through copy–paste and combination in a non-biological setting. This toy system allows exploitation of the intrinsic combinatorial explosion associated with

the multiple ways in which different bricks can be interlinked. In figure 1a, we plot the number of times that each type of brick occurred within a very large dataset of LEGO models [18]. The rank plot reveals that some simple bricks—as those shown in figure 1a, right—are extremely common whereas most bricks, having more complex shapes and larger size, are rare. The analysis showed that the statistical distribution can be well fitted using a generalized form of equation (1.1) known as the Pareto–Zipf distribution. This reads

$$p(s_i) = \frac{1}{Z} (i + i_0)^{-\gamma}, \quad (1.2)$$

where Z is again the corresponding normalization and i_0 a new parameter that allows us to take into account the curvature for small i -values. This picture is similar to the one reported from the study of large written corpora, as illustrated in figure 1b [28]. Our third example is given by the so-called protein domains, which are considered the building blocks of protein organization and an essential ingredient to understand the large-scale evolution of biological complexity [29–32]. Here each protein domain—or *fold*—is characterized by its essentially independent potential for folding in a stable way and each protein can be understood as a combination of one, two or more domains. In figure 1c, the rank distribution of observed folds from a large protein database is displayed. Domains define the combinatorial fabric of the protein universe and their number, although finite, has been increasing through evolution [31]. The fourth example gives the frequency of use of four-element modules within complex circuits [20].

The repertoire of LEGO bricks, words, protein domains and circuit modules provide the raw materials to combinatorial construction, but they also share the underlying presence of grammar, to be understood here as the compact description of a language. As indicated in [18], if we treat pieces of LEGO as words and models as utterances, LEGO appears as a class of artificial language and the resulting structures are passed from generation to generation through cultural transmission. This is of course a largely metaphoric picture, since the final outcome of the combinatorics is usually a non-functional design, unbounded by the potential combinations but not by functional constraints. This might actually be the reason why its statistical distribution, described by equation (1.2) deviates from equation (1.1). Protein domains too exhibit a grammar in which a set of generative rules for combining the available folds provides an explanatory mechanism for the observed repertoire of protein structures [19,33,34]. In summary, these systems—and others like electronic circuits or genomes, molecular networks [35–37] and complex circuits [38] and even evolved technology [39]—are characterized by a growth process that is expanding their inventories over time, the presence of generative rules allowing new structures to emerge, and a common statistical pattern described by Zipf's Law.

In this paper, we provide a general definition, or *postulates* of OEE based on algorithmic information theory (AIT), and we show that the common presence of Zipf's Law in these seemingly disparate systems may be deeply connected to their potentially open-ended nature. Furthermore, we explore the consequences that OEE has for the conservation of information, identifying the *information loss paradox* in OEE systems. This paradoxical situation, in which the system loses all its past information in the long run, even though

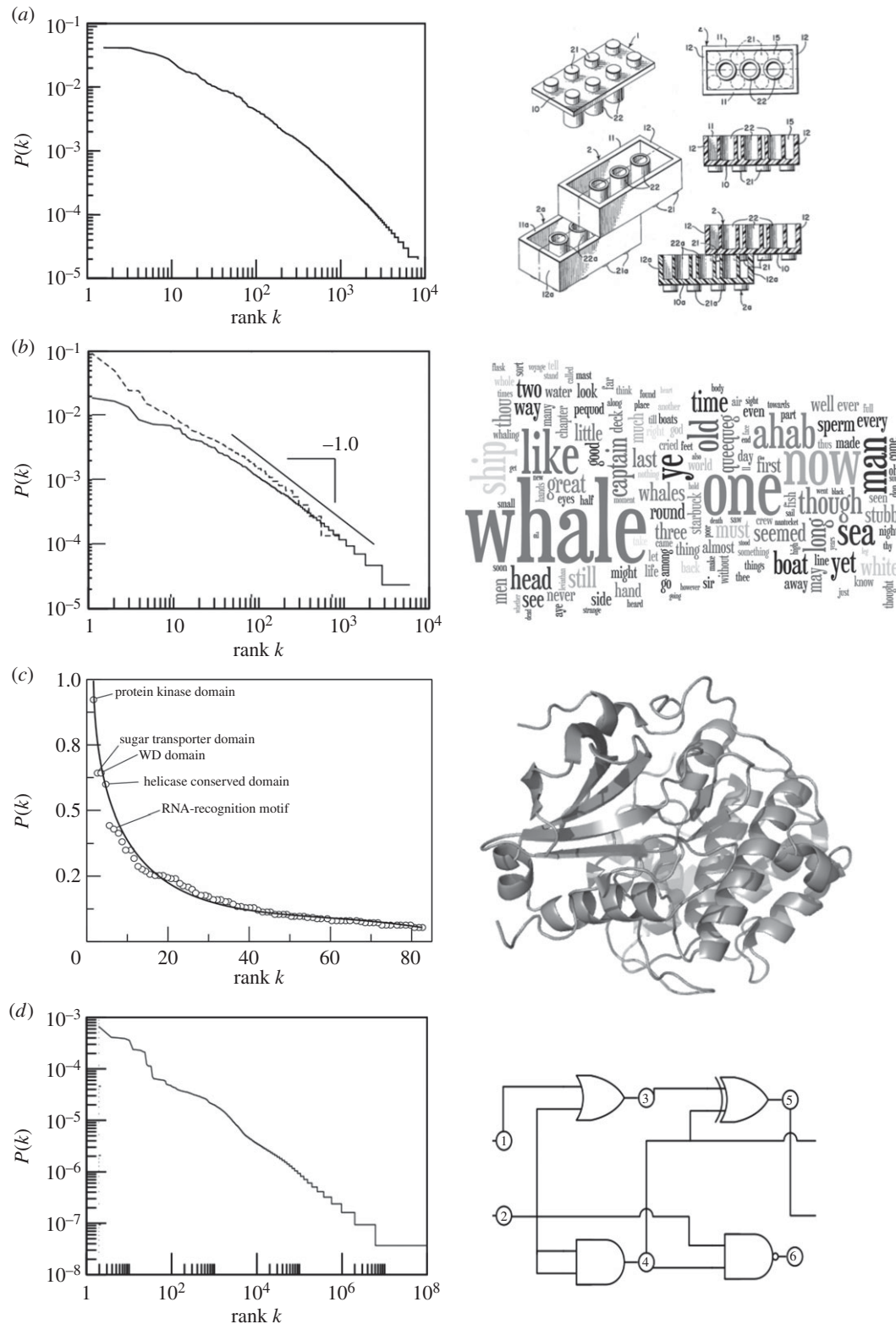


Figure 1. Zipf's Law distributions are commonly found in very different systems candidates to display open-endedness. Here, we show several examples of scaling behaviour involving (a) LEGO systems, (b) written language and (c) proteins. In (a), we display (in log scale) the probability of finding the i th most abundant type of LEGO brick within a very large number of systems (see details in [18]). In (b), the log-scale rank-size distribution of Herman Melville's *Moby Dick* is displayed. The dashed line shows the frequency versus rank for words having length 5, which is the average length of words in this particular book. The plot displayed in (c) shows, with linear axes, the corresponding rank distribution of protein folds in a large protein database (redrawn from [19]). The line is a power-law fit. Here the names of some of the domains, which are associated with particular functional traits, are indicated. (d) Zipf's Law in the frequency of logic modules used in evolved complex circuits (adapted from [20]).

the step-by-step information transmission is maximized, is shown to be a problem of the statistical nature of Shannon information theory. Indeed, we prove that, in the general setting of AIT, information can be conserved and systems can grow without bounds without removing the traces of

its past. Therefore, the general study of OEE systems must be framed in a theoretical construct not based on standard information theory, but in a much more general one, inspired in non-statistical forms of information content. We finally observe that the connection of fundamental results of

computation theory, and even Gödel’s incompleteness theorem, with general problems of evolutionary theory has been approached before in [8,40,41].

2. Algorithmic information theory

AIT [42–52] is a natural framework to address the problem of OEE. It incorporates powerful (still unexplored) tools to model the complexity of living systems, which, for example, has often been associated with information storage in the genome [40,48]. Such information results from the growth of genome complexity through both gene duplication and the interactions with the external world and is (by definition) a path-dependent process. Here we consider that we *encode* our evolving system into strings of symbols. We assume that, as long as the system evolves, such descriptions can grow and change, in a path-dependent way. As we shall see, the derived abstract framework is completely general, and applies to any system susceptible of displaying OEE.

A natural question arises when adopting such an abstract framework: why are we using Kolmogorov Complexity for our approach to OEE? The first reason is that it is based on strings obtained from a given alphabet, which naturally connects with a representation based on sequences [48] such as those in some of our examples from figure 1. Second, it connects with information theory (which is the most suitable coarse-grained first approximation to biology [53]) resulting in a more fundamental framework. Third, it consistently distinguishes predictable from unpredictable sequences in a meaningful way, and how these scale with size. Finally, the algorithmic definition based on the use of a *program* matches our intuition that evolution can be captured by some computational picture.

Let us first introduce a key concept required for our analysis: Kolmogorov—or algorithmic—complexity, independently developed by Kolmogorov [45], Solomonoff [46] and Chaitin [47]. Roughly speaking, if a given process can be described in terms of a string of bits, the complexity of this string can be measured as the shortest computer program capable of generating it [49,50]. The underlying intuition behind this picture—see figure 2—is that *simple*, predictable strings, such as 101010101010... can be easily obtained from a small piece of code that essentially says ‘write “10” followed by ‘repeat’ as many times as needed. This would correspond to a regular system, such as a pendulum or an electronic oscillator—see figure 2*a,b*—and the simple dynamical pattern is reproduced by a *short* program. Instead, a random string generated by means of a coin toss (say 010011001110110101010...) would only be reproduced by using a program that writes exactly that sequence and is thus as long as the string itself—figure 2*c,d*. Other stochastic processes generating fluctuations—figure 2*e,f*—and represented as strings of n bits can be similarly described, and their complexity shall lie somewhere between both extremes.

The stochasticity inherent to the most algorithmically complex strings (e.g. a coin toss, as introduced above) invites us to think in terms of *statistical* or *information entropy*. But the Kolmogorov complexity is, conceptually, a more fundamental measure of the complexity of such processes [51,52]. A formal definition follows. Let x and p be finite binary strings of length $\ell(x)$ and $\ell(p)$, respectively. Let \mathcal{T}_u be a universal Turing machine. Note that a finite binary string p can

define the computations that a universal Turing machine [54] will implement when p is fed as an input—i.e. it can define *programs* executed by the Turing machine. We will consider a set of *prefix free* programs: in such a set of programs, no program is the prefix of another program. This property is crucial for most of the results of AIT or even standard information theory [51,52]. Let $\mathcal{T}_u(p)$ denote the output of the computer \mathcal{T}_u when running the program p . Considering now all possible programs p that produce x as an output when fed into \mathcal{T}_u , the (prefix free) Kolmogorov complexity $K_{\mathcal{T}_u}(x)$ of the string x with respect to the universal computer \mathcal{T}_u is defined as [51]

$$K_{\mathcal{T}_u}(x) = \min_{p:\mathcal{T}_u(p)=x} \{\ell(p)\}. \tag{2.1}$$

This quantity is computer independent up to an additive constant [51,52] so we will omit the subindex when referring to it. If x is a random string, we would have a simple relation:

$$K(x) = \ell(x), \tag{2.2}$$

since all $\ell(x)$ bits need to be included, and we say that the sequence x is *incompressible*.

In addition, and as it happens with the statistical entropy, one can define the *conditional* algorithmic complexity as follows: let x , y and p be finite binary strings again and let \mathcal{T}_u^y be a universal Turing machine to which a description of y has already been made available. The Kolmogorov complexity of x given y is the length of the shortest program p that, when applied to a universal Turing machine, modifies y to display x as an output

$$K(x|y) = \min_{p:\mathcal{T}_u^y(p)=x} \{\ell(p)\}. \tag{2.3}$$

Notice that even though $K(x)$ can be arbitrarily large, $K(x|y)$ accounts for the minimum program that knows the differences between x and y and amends them.¹

3. General conditions for open-ended evolution: postulates

We shall concern ourselves with dynamic systems whose description can be made in terms of finite binary strings σ_t at each time step t over evolutionary time. The *complexity* of such an object at time t is given by $K(\sigma_t)$. This object shall evolve through intermediate steps in a path-dependent manner; thus the quantities $K(\sigma_t)$, $K(\sigma_{t+\Delta t})$ and $K(\sigma_{t+\Delta t}|\sigma_t)$ and the relationships between them will play a paramount role.

Let σ_t be the description of the system at time t . Let the sequence $\Sigma(t) \equiv \{\sigma_1, \sigma_2, \dots, \sigma_t\}$ be the *history* of the system until time t in arbitrary time units. We want the process that builds σ_t to be an open-ended evolutionary one, hence we turn our attention to the complexity of its evolutionary history $\Sigma(t)$. A minimal condition that this historical process has to obey to be called open-ended is that its complexity (properly normalized) always increases:

Axiom 3.1 Open-endedness. We say that the process that generates σ_t is open-ended if

$$\frac{K(\Sigma(t))}{t} \leq \frac{K(\Sigma(t+1))}{t+1}, \tag{3.1}$$

for all $t = 1, \dots, \infty$. Of all open-ended processes that obey

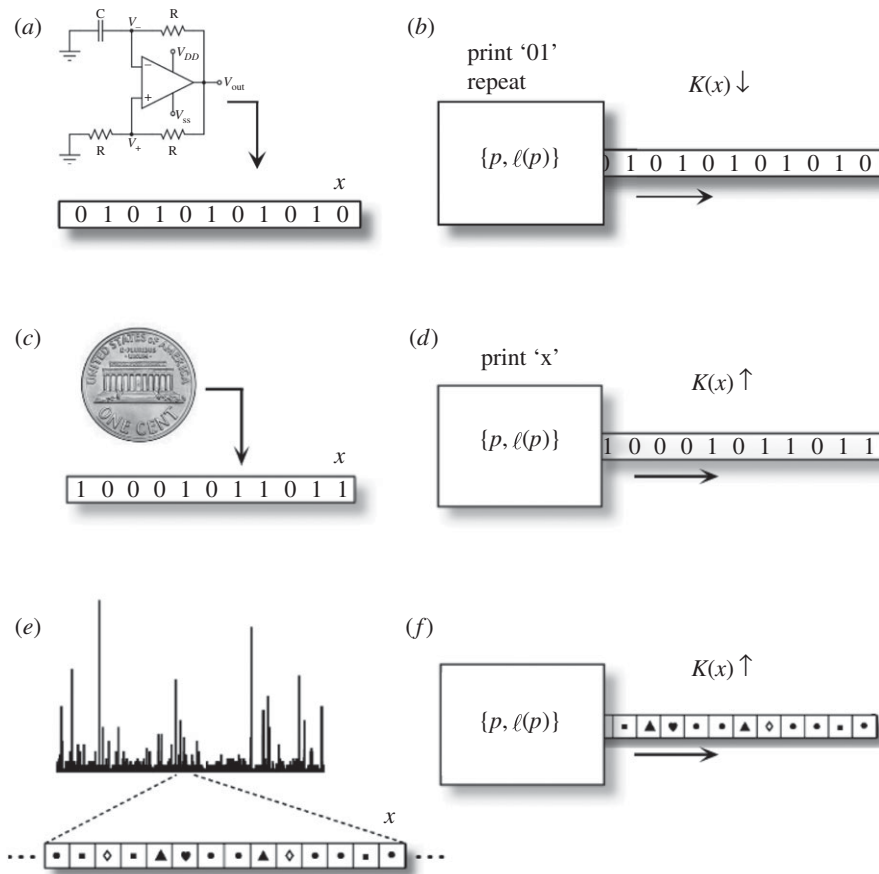


Figure 2. Measuring string complexity. If $\ell(p)$ is the length of the program, the Kolmogorov complexity $K(x)$ is defined as the smallest program able to write x . Simple dynamical systems (a) such as oscillators produce predictable, simple strings (b) thus having a low complexity. On the other extreme (c), a coin toss creates a completely random sequence and the program (d) is such that $K(x) = \ell(x)$. A system exhibiting broad distributions (e–f) and a large set of states is also likely to display high $K(x)$.

equation (3.1), we are interested in those whose complexity is not bounded:

Axiom 3.2 Unboundedness. We say that the process generating σ_t has an unbounded complexity if for any natural number $N \in \mathbb{N}$ there is a time t such that

$$\frac{K(\Sigma(t))}{t} > N. \quad (3.2)$$

These two axioms imply that information is always being added by the generative process in the long term—hence more bits are needed to describe later stages of the evolutionary history. The knowledge of the history up to time t is not enough to predict what will happen next. If it were, the description of later stages of the evolutionary history would be implicit in the description of the history at time t , and axiom 3.1 would be violated. Equation (3.2) also implies that the information of the processes we are interested in will never converge, eventually diverging for large times. These equations do not impose any condition on the complexity of the system at a given time step. Notably, (i) they admit a situation in which the description of the system—but not of its history—drops ($K(\sigma_t) > K(\sigma_{t+1})$, which might happen in biology [55], see also figure S1) and (ii) they do not imply any connection between states σ_t and σ_{t+1} . This second point is possible because we have not requested yet that this is an evolutionary process. We would hardly call a process ‘evolutionary’ if its successive steps are completely unrelated, hence:

Axiom 3.3 Heredity principle. Evolutionary processes attempt to minimize the action

$$S(\Sigma(t) \rightarrow \Sigma(t+1)) \equiv K(\Sigma(t+1) | \Sigma(t)). \quad (3.3)$$

That is, evolutionary processes try to minimize the amount of operations implemented to move the system from one state to the next, under whichever other constraints might apply. In the case of open-ended evolutionary systems, they try to minimize the number of operations needed to unfold in time while always increasing the informational content of the evolutionary history (as equations (3.1) and (3.2) demand). We could apply the same axiom, say, to Darwinian evolutionary processes saying that they attempt to minimize equation (3.3) subjected to random mutation and selection. (Note that this has no saying on whether Darwinian processes are inherently open-ended or not.) Axiom 3.3 defines an AIT-based least action principle that imposes that the information carried between successive steps is maximized as much as other constraints allow, thus turning the generative process into a path-dependent one. Without the Heredity principle, we could end up with a sequence of totally unrelated objects—i.e. a purely random, unstructured process hardly interpretable as an *evolving* system.

We take these axioms as our most general postulates of OEE. Note how they capture a very subtle, blurry tension between memory, to preserve past configurations and the system’s ability to innovate, thus always adding new

information that can upset established structures. As other authors have hinted at [5,12,13], OEE phenomenology seems to emerge out of this conflict, which is also familiar in other aspects of complex and critical systems, designs and behaviours [56–58]. These are often described by a compromise, the ‘edge of chaos’, between unchanging, ordered states and structureless, disordered configurations.

In a nutshell, our working definition of open-endedness implies that the size of the algorithm describing the history of the system does not converge in time. Therefore, even if every evolutionary stage accepts a finite algorithm as a description, the evolutionary path is asymptotically uncomputable. These postulates are assumed to be satisfied by all open-ended systems. However, they turn out to be too generic to extract conclusions of how OEE systems may behave or which kind of observable footprints are expected from them. To gain a greater insight about the effects of OEE we can study a *strong* version of these postulates that applies not to evolutionary histories, but to objects themselves. Hence we demand that:

$$K(\sigma_t) \leq K(\sigma_{t+1}), \quad (3.4)$$

at any $t = 1, \dots, \infty$ and that for every natural number $N \in \mathbb{N}$ there is a time t such that

$$K(\sigma_t) > N. \quad (3.5)$$

Also, in the strong version of OEE the action:

$$S(\sigma_t \rightarrow \sigma_{t+1}) \equiv K(\sigma_{t+1} | \sigma_t) \quad (3.6)$$

is minimized, constrained by equations (3.4) and (3.5). As before, $S(\sigma_t \rightarrow \sigma_{t+1})$ denotes an informational entropy—or missing information—when inferring σ_{t+1} from σ_t . We know from [49] that this quantity is bounded by

$$S(\sigma_{t+1} \rightarrow \sigma_t) \geq |K(\sigma_{t+1}) - K(\sigma_t)| + \mathcal{O}(1). \quad (3.7)$$

As discussed before, the general OEE postulates—equations (3.1)–(3.3)—allow for the complexity of σ_t to drop, so such processes are not necessarily OEE in the strong sense defined by equations (3.4)–(3.6). However, it can be proved that every unbounded OEE process in the general sense must contain an unbounded OEE process in the strong sense—see the electronic supplementary material. That is, the strong version of OEE still can teach us something about the footprints of open-ended evolution.

4. Statistical systems: a variational approach to open-ended evolution

We will explore now the consequences of the definition stated above for systems that accept a description—possibly partial—in terms of statistical ensembles. The aim is to write the three conditions for OEE described by equations (3.4)–(3.6) in the language of statistical information theory. We will assume now that the statistical properties of this very finite string σ_t are themselves accurately accounted for by a random variable X_t . In other words: we consider that the string σ_t as a sequence of observations at the system at time t . This will provide a description of the system, in terms of observable states, at time t . We further consider that such a description of the system is the outcome of a random variable X_t . Then, the algorithmic complexity $K(\sigma_t)$

is of the order of the Shannon entropy $H(X_t)$ associated with the random variable X_t [51,52]:

$$K(\sigma_t) = H(X_t) + \mathcal{O}(1).$$

Recall that this is the minimal information required to describe the behaviour of a single outcome of X_t , not a *sequence of trials* of the random variable X_t .² This random variable will represent an observation or realization of the system. Assume that we discretize the time, so we use the subscript n or m instead of t , and that we label the states $i = 1, \dots, n$. Now let us define the following family of nested subsets:

$$\begin{aligned} \Omega_1 &= \{1\} \\ \Omega_2 &= \Omega_1 \cup \{2\} = \{1, 2\} \\ &\dots = \dots \\ \Omega_{n+1} &= \Omega_n \cup \{n+1\} = \{1, \dots, n+1\}. \end{aligned}$$

The open-ended evolutionary process will traverse the above family of nested subsets, adding a new state per evolutionary time step. We now define a sequence of different random variables

$$X_1, \dots, X_n,$$

such that X_k takes values over the set Ω_k and follows the probability distribution $p_k(1), \dots, p_k(k)$, with $\sum_{i \leq k} p_k(i) = 1$. Then

$$H(X_n) = - \sum_{i \leq n} p_n(i) \log p_n(i).$$

The variational principle derived from the path-dependent process implies now the minimization of the conditional entropy of the random variable X_{n+1} given the random variable X_n , namely

$$H(X_{n+1} | X_n) = - \sum_{i \leq n} p_n(i) \sum_{k \leq n+1} \mathbb{P}_n(k | i) \log \mathbb{P}_n(k | i),$$

where $\mathbb{P}_n(k | i) \equiv \mathbb{P}(X_{n+1} = k | X_n = i)$. We will finally assume (without loss of generality) that the probability distributions p_2, \dots, p_n are sorted in decreasing order, i.e.:

$$p_k(1) > p_k(2) > \dots > p_k(k).$$

In the electronic supplementary material, we discuss the conditions under which the consecutive achievement of ordered probability distributions is possible.

Therefore, for statistical systems, the previous constraints for open-endedness from equations (3.4) and (3.5) must now be rewritten as follows: first,

$$H(X_n) \leq H(X_{n+1}), \quad (4.1)$$

and, for any $N \in \mathbb{N}$, there will be a n such that

$$H(X_n) > N. \quad (4.2)$$

In addition, the path dependence condition stated in equation (3.6) implies that:

$$\text{minimize } H(X_{n+1} | X_n). \quad (4.3)$$

In summary, we took a set of conditions, described by equations (3.4)–(3.6), valid in the general AIT framework, and we have re-written them in terms of statistical entropy functions through equations (4.1)–(4.3). We finally observe

that the condition that the probability distribution must be strictly ordered leads to

$$H(X_n) < \log n.$$

Accordingly, the case of total randomness (fair coin toss) is removed.

4.1. Minimizing the differences between shared states

Condition (4.3) is difficult to handle directly. Nevertheless, it can be approached as follows: we first find a minimum by extremalizing a given Kullback–Leibler (KL) divergence, and then we will prove that this solution indeed converges to the absolute minimum of $H(X_{n+1}|X_n)$.

Let us define the distribution \hat{p}_{n+1} as

$$\hat{p}_{n+1}(k) \equiv p_{n+1}(k|k < n+1) = \frac{p_{n+1}(k)}{\sum_{i < n+1} p_{n+1}(i)}.$$

$\hat{p}_{n+1}(k)$ is the probability that $k < n+1$ appears when we draw the random variable X_{n+1} excluding the outcomes = $n+1$. Clearly, \hat{p}_{n+1} and p_n are defined over the set Ω_n , whereas p_{n+1} is defined over the set Ω_{n+1} . Since the support sets for both \hat{p}_{n+1} and p_n are the same, one can use the KL divergence defined as the relative entropy (or information gain) between p_n and \hat{p}_{n+1} :

$$D(p_n \| \hat{p}_{n+1}) = \sum_{k \leq n} p_n(k) \log \frac{p_n(k)}{\hat{p}_{n+1}(k)}.$$

Now we impose the condition of path dependence as a variational principle over the KL divergence and then we write the following Lagrangian which defines the evolution of our system:

$$\begin{aligned} \mathcal{L}(\hat{p}_{n+1}(1), \dots, \hat{p}_{n+1}(n); \theta_{n+1}) \\ = D(p_n \| \hat{p}_{n+1}) + \theta_{n+1} \left(\sum_{k \leq n} \hat{p}_{n+1}(k) - 1 \right). \end{aligned}$$

The minimization of this Lagrangian with respect to the variables upon which it depends imposes that

$$\hat{p}_{n+1} = p_n,$$

which implies that

$$\text{and } \left. \begin{aligned} p_{n+1}(k) &= \theta_{n+1} p_n(k) \quad \forall k \leq n \\ p_{n+1}(n+1) &= 1 - \theta_{n+1}. \end{aligned} \right\} \quad (4.4)$$

By construction, $0 < \theta_{n+1} < 1$. Equation (4.4) imposes that the conditional probabilities between X_n and X_{n+1} read as

$$\begin{aligned} \mathbb{P}_n(X_{n+1} = i | X_n = k) &= \delta_{ik} \theta_{n+1}, \quad \text{for } i \leq n; \\ \mathbb{P}_n(X_{n+1} = n+1 | X_n = k) &= 1 - \theta_{n+1}, \quad \text{for } k \leq n. \end{aligned}$$

This defines a channel structure that leads to

$$H(X_{n+1}|X_n) = H(\theta_{n+1}), \quad (4.5)$$

being $H(\theta_{n+1})$ the entropy of a Bernoulli process having parameter θ_{n+1} , i.e.:

$$H(\theta_{n+1}) = -\theta_{n+1} \log \theta_{n+1} - (1 - \theta_{n+1}) \log (1 - \theta_{n+1}).$$

In the electronic supplementary material, it is proven that

$$H(\theta_{n+1}) \rightarrow \min H(X_{n+1}|X_n).$$

We thus have found the specific form of the conditional entropy governing the path dependency of the OEE system, imposed by equation (4.3).

We finally remark some observations related to the flow of information between past and present states. First, we note that, from equation (4.5), the relationship between the entropies of X_n and X_{n+1} satisfies the following Fano's-like equality:

$$H(X_{n+1}) = \theta_{n+1} H(X_n) + H(\theta_{n+1}). \quad (4.6)$$

Finally, from the definition of mutual information between X_n and X_{n+1} , one obtains:

$$I(X_{n+1} : X_n) = H(X_{n+1}) - H(X_{n+1} | X_n),$$

and from equations (4.5) and (4.6) we arrive at the amount of information transmitted from the time step n to $n+1$:

$$I(X_{n+1} : X_n) = \theta_{n+1} H(X_n). \quad (4.7)$$

This is a good estimate of the maximum possible information transmitted per evolutionary time step. Nevertheless, even in this case, we shall see that the statistical information transmitted along time in an open-ended system has to face a paradoxical behaviour: the total loss of any past history in the long run—see §4.3.

4.2. Zipf's Law: the footprint of OEE

As discussed at the beginning, a remarkably common feature of several systems known to exhibit OEE is the presence of Zipf's Law. We will rely now on previous results [24,25] to show that the solution to the problem discussed above is given precisely by Zipf's Law. We first note that, thanks to equation (4.4), the quotient between probabilities:

$$\frac{p_n(i+j)}{p_n(i)} = f(i, i+j),$$

remains constant for all n as soon as $p_n(i+j) > 0$. In the electronic supplementary material, following [25], we provide the demonstration that, in a very general case, the solution of our problem lies in the range defined by

$$\left(\frac{i+1}{i} \right)^{(1-\delta)} > f_n(i, i+1) > \left(\frac{i+1}{i} \right)^{(1+\delta)}.$$

It can be shown that $\delta \rightarrow 0$ if the size of the system is large enough. Therefore,

$$f_n(i, i+1) = \frac{p_n(i)}{p_n(i+1)} \approx \frac{i+1}{i},$$

which leads us to the scaling distribution:

$$p_n(i) \propto i^{-1}. \quad (4.8)$$

In other words, Zipf's Law is the only asymptotic solution, which immediately suggests a deep connection between the potential for open-ended evolutionary dynamics and the presence of this particular power law. Note that Zipf's Law is a necessary footprint of OEE, not a sufficient one: other mechanisms might imprint the same distribution [23]. We emphasize the remarkable property that this result is independent of the particular way the evolving system satisfies the OEE conditions imposed by equations (4.1)–(4.3).

4.3. The loss of information paradox in OEE

The above description of the evolution of open-ended statistical ensembles leads to an unexpected result: statistical systems displaying OEE lose any information of the past after a large period of complexity growing. Indeed, although

information is conserved in a huge fraction step by step, it is not conserved at all if we compare large periods of evolution. Therefore, the capacity to generate an ensemble encoding an unbounded amount of information through evolution results in a total erasure of the past, even if a strong path dependency principle is at work (figure 3).

To see what happens with information along the evolutionary process in the limit of large n 's, we first rewrite mutual information between X_n and X_m , $m < n$ as follows:

$$I(X_m : X_n) = \sum_{i \leq n} p_n(i) \sum_{k \leq m} \mathbb{P}(k|i) \log \frac{\mathbb{P}(k|i)}{p_m(k)},$$

where, in this case, $\mathbb{P}(k|i) \equiv \mathbb{P}(X_m = k | X_n = i)$. Then we define the following constant C_m :

$$C_m = \prod_{2 \leq k \leq m} (\theta_k)^{-1},$$

where the θ_k 's are the ones arising from equation (4.4). From here, one can prove—see the electronic supplementary material—that:

$$p_m(1) = \frac{1}{C_m},$$

Now, observe that we can generalize equation (4.7) as follows:

$$I(X_m : X_n) \leq \theta_{m+1} \cdot \dots \cdot \theta_n H(X_m).$$

This allows us to obtain the following chain of inequalities:

$$\begin{aligned} I(X_m : X_n) &\leq \prod_{m < i \leq n} \theta_i H(X_m) \\ &= \frac{1}{C_n} \prod_{2 \leq k \leq m} \theta_k^{-1} H(X_m) \\ &= \frac{C_m}{C_n} H(X_m). \end{aligned} \quad (4.9)$$

The above inequalities have an interesting consequence. Indeed, from equation (4.9), if $C_n \rightarrow \infty$, then

$$\lim_{n \rightarrow \infty} I(X_m : X_n) \leq \lim_{n \rightarrow \infty} \frac{C_m}{C_n} H(X_m) = 0. \quad (4.10)$$

In the electronic supplementary material, it is proven that, in OEE statistical systems, indeed we have that $C_n \rightarrow \infty$. Thus, $I(X_m : X_n) \rightarrow 0$: *no statistical information is conserved in open-ended systems in the long term.*

4.4. Solving the paradox: algorithmic information can be maintained

We have shown above that *statistical information* cannot be maintained through arbitrarily long evolutionary paths if the evolution is open-ended. The emphasis is on the word *statistical*. As we shall see, using a rather informal reasoning, other types of information based on the general setting of AIT can be maintained. Let σ_n be a description, in bits, of an object at time n and σ_N its description at time $N > n$. Let us assume that σ_N , in its most compressed form, can only be written as a concatenation of two descriptions, to be indicated with symbol ' \oplus ':

$$\sigma_N = \sigma_n \oplus \sigma_{N-n}.$$

Now assume that $K(\sigma_N) = \mu N$, $K(\sigma_n) = \mu n$ and $K(\sigma_{N-n}) = \mu(N - n)$, with $0 < \mu < 1$. If π_n is the minimal program that

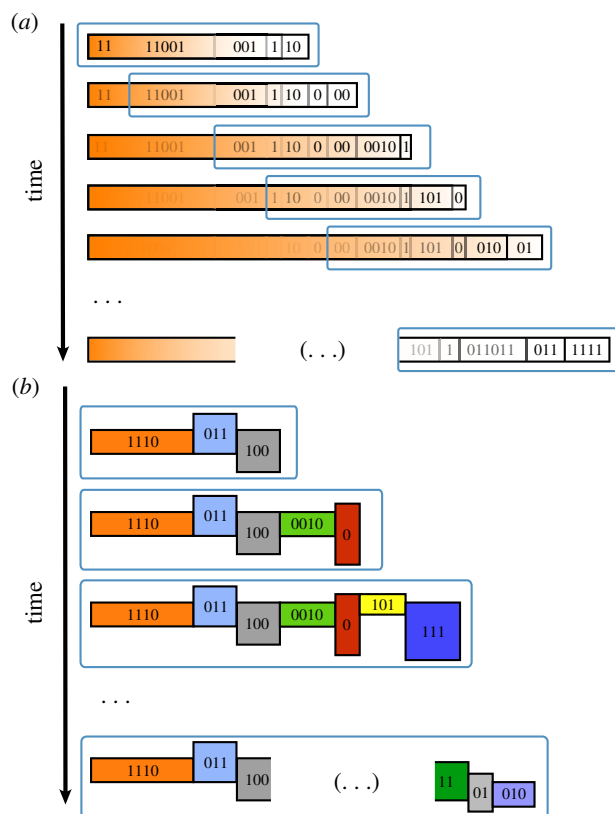


Figure 3. The paradox of information loss: (a) a statistical description of the system displays OEE. This means that, through time, the entropy of the ensemble grows without bounds. The consequence of that is that the information about the past history is totally erased as time goes by. Only a fraction of the history closest to the present survives. Therefore, there is no conservation of the information—see §4.3. (b) If information is encoded in a non-statistical way, such as bit strings, it can be preserved. The historical information survives through evolutionary stages, even if the system displays OEE.

prints σ_n and π_{N-n} is the minimal program that prints σ_{N-n} . Then, there is a program π_N defined as

$$\pi_N = \pi_n \oplus \pi_{N-n},$$

such that, when applied to a universal Turing machine, gives σ_N , i.e. $\mathcal{T}_u(\pi_N) = \sigma_N$. If we already know π_n , it is clear that

$$K(\sigma_N | \sigma_n) = K(\sigma_{N-n}) + \mathcal{O}(1).$$

We observe that, under the assumptions we made

$$|K(\sigma_N) - K(\sigma_n)| = K(\sigma_{N-n}),$$

so $K(\sigma_N | \sigma_n) \approx |K(\sigma_N) - K(\sigma_n)|$ close to the bound provided by Zurek [49], already used in equation (3.7). As we shall see, the immediate consequence of that is that the algorithmic mutual information between σ_N and σ_n does not depend on N . Let $I(\sigma_N : \sigma_n)$ be the algorithmic mutual information between σ_N and σ_n :

$$I(\sigma_N : \sigma_n) = K(\sigma_N) - K(\sigma_N | \sigma_n).$$

Then, one has that:

$$\begin{aligned} I(\sigma_N : \sigma_n) &= K(\sigma_N) - K(\sigma_N | \sigma_n) \\ &\approx K(\sigma_N) - K(\sigma_{N-n}) \\ &\approx K(\sigma_n), \end{aligned}$$

we thus have

$$\lim_{N \rightarrow \infty} I(\sigma_N : \sigma_n) \approx K(\sigma_n). \quad (4.11)$$

Within the AIT framework this implies that information of previous stages of the evolution can be maintained.

The result reported above has an important consequence: in an OEE system in which information is maintained, the *information* is encoded by generative rules that cannot be captured by simple statistical models. Therefore, Shannon information theory is of little use to understand the persistence of the memory of past states in OEE systems.

5. Discussion

In this paper, we have considered a new approach to a key problem within complex systems theory and evolution, namely, the conditions for *open-ended evolution* and its consequences. We provided a general formalization of the problem through a small set of *postulates* summarized by equations (3.1)–(3.6) based on the framework of AIT. Despite the high degree of abstraction—which allows us to extract very general results—important specific conclusions can be drawn: (i) in statistically describable systems, Zipf's Law is the expected outcome of OEE. (ii) OEE systems have to face the *statistical information loss paradox*: Shannon information between different stages of the process tends to zero, and all information of the past is lost in the limit of large time periods. (iii) This paradoxical situation is solved when considering non-statistical forms of information, and we provided an example where algorithmic information between arbitrary time steps is maintained. This result, however, does not invalidate previous approaches of statistical information theory concerning the study of *flows* of information within the system [59], since our result refers to the structural complexity of the evolving entity. It is important to stress that *information* may unfold in several meanings or formal frameworks when talking about evolving systems. Moreover, further explorations should inquire into the role of information flows in keeping and promoting the increase of structural complexity of evolving systems. In addition, it is worth emphasizing that, at the current level of development, our framework might fail to incorporate some processes, such as exaptation or abiotic external drives, that are not fully algorithmic but identified as key actors in evolutionary systems [13]. All these issues are relevant in order to understand and eventually *build* OEE systems, as it is the case within the context of artificial life [12,60] by considering the possibility of building a system able to evolve under artificial conditions and maintain a constant source of *creativity* [61,62].

Since Zipf's Law is the outcome of a statistical interpretation of the OEE postulates given in equations (3.1)–(3.6), one may be tempted to conclude that information is not conserved in those systems exhibiting Zipf's Law in its statistical patterns. Instead, in line with the previous paragraph, it is important to stress that the statistical ensemble description can be just a partial picture of the system, and that other mechanisms of information prevalence, not necessarily statistical, are at work. Therefore, if our system exhibits Zipf's Law and we have evidence of information conservation, the statistical pattern may be interpreted as the projection of other types of non-statistical information to the statistical observables.

Biological systems exhibit marked potential capacity for OEE resulting from their potential for growing and exploring new states and achieving novel functionalities. This open-

endedness pervades the apparently unbounded exploration of the space of the possible. The two biological systems cited in the Introduction, namely human language and the protein universe, share the presence of an underlying grammar, which both enhances and constrains their combinatorial potential. Analogously, the example provided by models of evolution through gene duplication or tinkering revealed that scaling laws and other properties displayed by protein networks emerge from the amplification phenomena introduced by growth through copy-and-paste dynamics [63–65]. One way of doing this is provided by the tinkered nature of evolutionary change, where systems evolve by means of extensive reuse of previous parts to explore novel designs [15,16,32]. This mechanism fully matches our assumptions: generative rules that enable expansion of the state space, while the redundant nature of the process enables most of the previous structures to be kept. Again, our axioms capture this delicate balance between memory and innovation, order and disorder, that OEE systems seem to exploit as they unfold.

We reserve a final word for a general comment on the role of OEE in the theory of biology. Postulates described by equations (3.1)–(3.6) explicitly relate OEE to unpredictability. This, according to classic results like the *no free lunch theorem* [66], puts a question mark on the possibility of a theory of evolution in the sense of classical physics. This issue, discussed also in [8], may exclude the possibility of a predictive theory in terms of the explicit evolutionary innovations that will eventually emerge. Nevertheless, in this paper we prove that this is not an all-or-nothing situation: interestingly, the postulates of OEE, which rule out the existence of a predictive theory, are precisely the conditions that allow us to identify one of the possible statistical regularities—Zipf's Law—governing such systems and thereby make *predictions* and, eventually, propose *physical* principles for them, adding a new, unexpected ingredient to the debate on predictability and evolution [67]. According to that, these principles would predict the statistical observables, but not the specific events that they represent.

Data accessibility. This article has no additional data.

Authors' contributions. B.C.-M., L.F.S. and R.S. contributed to the idea, development, and mathematical derivations in this paper. All three authors contributed to the writing and elaboration of figures.

Competing interests. We declare we have no competing interests.

Funding. This work has been supported by the Botín Foundation, by Banco Santander through its Santander Universities Global Division, a MINECO FIS2015-67616 fellowship, the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya (R.S. and L.F.S.) and the Santa Fe Institute (R.S.).

Acknowledgements. We thank Jordi Piñero, Sergi Valverde, Jordi Fortuny, Kepa Ruiz-Mirazo, Carlos Rodríguez-Caso and the members of the Complex Systems Lab for useful discussions. B.C.-M. thanks Stefan Thurner, Rudolf Hanel, Peter Klimek, Vittorio Loretto and Vito DP Servedio for useful comments on previous versions of the manuscript.

Endotes

¹This quantity has been used as a conditional complexity within the context of evolved symbolic sequences [48]. In this case, $K(s|e)$ referred to the length of the smallest program that gives the string s from a given environment e , also defined as a string.

²Rigorously speaking, one should say that, if σ is the description in bits of the outcomes of N trials of X_i , then $K(\sigma)/N \rightarrow H(X_i)$.

Reference

- Maynard-Smith J, Szathmáry E. 1995 *The major transitions in evolution*. Oxford, UK: Oxford University Press.
- Schuster P. 1996 How does complexity arise in evolution? *Complexity* **2**, 22–30. (doi:10.1002/(ISSN)1099-0526)
- Lane N. 2009 *Life ascending: the ten great inventions of evolution*. London, UK: Profile Books.
- Solé R. 2016 Synthetic transitions: towards a new synthesis. *Phil. Trans. R. Soc. B* **371**, 20150438. (doi:10.1098/rstb.2015.0438)
- Ruiz-Mirazo K, Umerez J, Moreno A. 2008 Enabling conditions for ‘open-ended’ evolution. *Biol. Philos.* **23**, 67–85. (doi:10.1007/s10539-007-9076-8)
- Bedau MA, McCaskill JS, Packard NH, Rasmussen S, Adami C, Green DG, Ikegami T, Kaneko K, Ray TS. 2000 Open problems in artificial life. *Artif. Life* **6**, 363–376. (doi:10.1162/106454600300103683)
- Nehaniv CL. 2000 Measuring evolvability as the rate of complexity increase. In *Artificial Life VII Workshop Proc* (eds CC Maley, E Boudreau), Portland, OR, pp. 55–57. New York, NY: MIT Press.
- Day T. 2012 Computability, Gödel’s incompleteness theorem, and an inherent limit on the predictability of evolution. *J. R. Soc. Interface* **9**, 624–639. (doi:10.1098/rsif.2011.0479)
- Von Neumann J, Burks AW. 1966 Theory of self-reproducing automata. *IEEE Trans. Neural Networks* **5**, 3–14.
- McMullin B. 2000 John von Neumann and the evolutionary growth of complexity: looking backward, looking forward. *Artif. Life* **6**, 347–361. (doi:10.1162/106454600300103674)
- Ruiz-Mirazo K, Peretó J, Moreno A. 2004 A universal definition of life: autonomy and open-ended evolution. *Origins Life Evol. Biosph.* **34**, 323–346. (doi:10.1023/B:ORIG.0000016440.53346.dc)
- Taylor T *et al.* 2016 Open-ended evolution: perspectives from the OEE workshop in York. *Artif. Life* **22**, 408–423. (doi:10.1162/ARTL_a_00210)
- de Vladar HP, Santos M, Szathmáry E. 2017 Grand views of evolution. *Trends Ecol. Evol.* **32**, 324–334. (doi:10.1016/j.tree.2017.01.008)
- Hauser MD, Chomsky N, Fitch WT. 2002 The faculty of language: what is it, who has it and how did it evolve? *Science* **298**, 1569–1579. (doi:10.1126/science.298.5598.1569)
- Jacob F. 1977 Evolution as tinkering. *Science* **196**, 1161–1166. (doi:10.1126/science.860134)
- Solé RV, Ferrer R, Montoya JM, Valverde S. 2002 Selection, tinkering, and emergence in complex networks. *Complexity* **8**, 20–33. (doi:10.1002/cplx.v8i1)
- Solé RV, Valverde S, Casals MR, Kauffman SA, Farmer D, Eldredge N. 2013 The evolutionary ecology of technological innovations. *Complexity* **18**, 15–27. (doi:10.1002/cplx.v18i4)
- Crompton A. 2012 The entropy of LEGO. *Environ. Planning B* **39**, 174–182. (doi:10.1068/b37063)
- Searls DB. 2002 The language of genes. *Nature* **420**, 211–217. (doi:10.1038/nature01255)
- Raman K, Wagner A. 2011 The evolvability of programmable hardware. *J. R. Soc. Interface* **8**, 269–281. (doi:10.1098/rsif.2010.0212)
- Auerbach F. 1913 Das Gesetz der Bevölkerungskonzentration. *Petermanns Geographische Mitteilungen* **59**, 74–76.
- Zipf G. 1949 *Human behavior and the principle of least-effort*. Cambridge, MA: Addison-Wesley.
- Newman MEJ. 2004 Power laws, Pareto distributions and Zipf’s law. *Contemp. Phys.* **46**, 323–351. (doi:10.1080/00107510500052444)
- Corominas-Murtra B, Solé RV. 2010 Universality in Zipf’s law. *Phys. Rev. E* **82**, 011102. (doi:10.1103/PhysRevE.82.011102)
- Corominas-Murtra B, Solé RV. 2011 Emergence of Zipf’s law in the evolution of communication. *Phys. Rev. E* **83**, 036115. (doi:10.1103/PhysRevE.83.036115)
- Corominas-Murtra B, Hanel R, Thurner S. 2015 Understanding scaling through history-dependent processes with collapsing sample space. *Proc. Natl Acad. Sci. USA* **112**, 5348–5353. (doi:10.1073/pnas.1420946112)
- Corominas-Murtra B, Hanel R, Thurner S. 2016 Extreme robustness of scaling in sample space reducing processes explains Zipf’s law in diffusion on directed networks. *New J. Phys.* **18**, 093010. (doi:10.1088/1367-2630/18/9/093010)
- Ferrer R, Solé RV. 2002 Zipf’s law in random texts. *Adv. Complex Syst.* **5**, 1–6. (doi:10.1142/S0219525902000468)
- Koonin EV, Wof YI, Karev GP. 2002 The structure of the protein universe and genomic evolution. *Nature* **420**, 218–223. (doi:10.1038/nature01256)
- Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA. 2004 Structure, function and evolution of multidomain proteins. *Curr. Opin. Struct. Biol.* **14**, 208–216. (doi:10.1016/j.sbi.2004.03.011)
- Kim KM, Caetano-Anollés G. 2012 The evolutionary history of protein fold families and proteomes confirms that the archaeal ancestor is more ancient than the ancestors of other superkingdoms. *BMC Evol. Biol.* **12**, 13. (doi:10.1186/1471-2148-12-13)
- Rodríguez-Caso C, Medina MA, Solé RV. 2005 Topology, tinkering and evolution of the human transcription factor network. *FEBS J.* **272**, 6423–6434. (doi:10.1111/EJB.2005.272.issue-24)
- Przytycka T, Srinivasan R, Rose GD. 2002 Recursive domains in proteins. *Prot. Sci.* **11**, 409–417. (doi:10.1110/ps.24701)
- Gimona M. 2006 Protein linguistics—a grammar for modular protein assembly? *Nature Rev. Cell. Biol.* **7**, 68–73. (doi:10.1038/nrm1785)
- Schuster P, Fontana W, Stadler PF, Hofacker IL. 1994 From sequences to shapes and back: a case study in RNA secondary structures. *Proc. R. Soc. Lond. B* **255**, 279–284. (doi:10.1098/rspb.1994.0040)
- Huynen MA, Stadler PF, Fontana W. 1996 Smoothness within ruggedness: the role of neutrality in adaptation. *Proc. Natl Acad. Sci. USA* **93**, 397–401. (doi:10.1073/pnas.93.1.397)
- Nimwegen E, Crutchfield JP, Huynen M. 1999 Neutral evolution of mutational robustness. *Proc. Natl Acad. Sci. USA* **96**, 9716–9720. (doi:10.1073/pnas.96.17.9716)
- Fernandez P, Solé RV. 2007 Neutral fitness landscapes in signaling networks. *J. R. Soc. Interface* **4**, 41–47. (doi:10.1098/rsif.2006.0152)
- Arthur WB, Polak W. 2006 The evolution of technology within a simple computer model. *Complexity* **11**, 23–31. (doi:10.1002/(ISSN)1099-0526)
- Adami C. 1998 *Introduction to artificial life*. New York, NY: Springer.
- Hernández-Orozco S, Hernández-Quiroz F, Zenil H. 2018 Undecidability and irreducibility conditions for open-ended evolution and emergence. *Artif. Life* **24**, 56–70. (doi:10.1162/ARTL_a_00254)
- Solomonoff RJ. 1960 A preliminary report on a general theory of inductive inference. Technical Report ZTB-138, Zator Company, Cambridge, MA.
- Chaitin GJ. 1977 Algorithmic information theory. *IBM J. Res. Dev.* **21**, 350–359. (doi:10.1147/rd.214.0350)
- Chaitin GJ. 1990 *Information, randomness and incompleteness: papers on algorithmic information theory*, vol. 8. Singapore: World Scientific.
- Kolmogorov AN. 1965 Three approaches to the quantitative definition of information. *Problems Inform. Transm.* **1**, 1–7.
- Solomonoff R. 1964 A formal theory of inductive inference part I. *Inform. Control* **7**, 1–22. (doi:10.1016/S0019-9958(64)90223-2)
- Chaitin GJ. 1969 On the simplicity and speed of programs for computing infinite sets of natural numbers. *J. ACM* **16**, 407–422. (doi:10.1145/321526.321530)
- Adami C, Cerf NJ. 2000 Physical complexity of symbolic sequences. *Physica D* **137**, 62–69. (doi:10.1016/S0167-2789(99)00179-7)
- Zurek WH. 1989 Thermodynamic cost of computation, algorithmic complexity and the information metric. *Nature* **341**, 119–124. (doi:10.1038/341119a0)
- Chaitin GJ. 1982 *Algorithmic information theory*. New York, NY: John Wiley & Sons, Inc.
- Li M, Vitányi PMB. 1997 *An introduction to Kolmogorov complexity and its applications*, 2nd edn. Berlin, Germany: Springer.
- Cover TM, Thomas JA. 1991 *Elements of information theory*. New York, NY: Wiley.
- Hopfield JJ. 1994 Physics, computation, and why biology looks so different. *J. Theor. Biol.* **171**, 53–60. (doi:10.1006/jtbi.1994.1211)
- Turing A. 1936 On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc.* **42**, 230–265. (doi:10.1112/plms/s2-42.1.230)

55. Neef A, Latorre A, Peretó J, Silva FJ, Pignatelli M, Moya A. 2011 Genome economization in the endosymbiont of the wood *Cryptocercus punctulatus* due to drastic loss of aminoacid synthesis capabilities. *Genome Biol. Evol.* **3**, 1437–1448. (doi:10.1093/gbe/evr118)
56. Kauffman SA. 1993 *The origins of order: self-organization and selection in evolution*. Oxford, UK: Oxford University Press.
57. Mora T, Bialek W. 2011 Are biological systems poised at criticality? *J. Stat. Phys.* **144**, 268–302. (doi:10.1007/s10955-011-0229-4)
58. Muñoz MA. 2018 Colloquium: criticality and dynamical scaling in living systems. *Rev. Mod. Phys.* **90**, 031001. (doi:10.1103/RevModPhys.90.031001)
59. Tkačik G, Bialek W. 2016 Information processing in living systems. *Annu. Rev. Condens. Matter Phys.* **7**, 12.1–12.29. (doi:10.1146/annurev-conmatphys-031214-014803)
60. Solé RV, Valverde S. 2013 Macroevolution *in silico*: scales, constraints and universals. *Palaeontology* **56**, 1327–1340. (doi:10.1111/pala.2013.56.issue-6)
61. Kauffman SA. 2000 *Investigations*. Oxford, UK: Oxford University Press.
62. Nguyen A, Yosinski J, Clune J. 2016 Understanding innovation engines: automated creativity and improved stochastic optimization via deep learning. *Evol. Comput.* **24**, 545–572. (doi:10.1162/EVCO_a_00189)
63. Pastor-Satorras R, Smith E, Solé RV. 2003 Evolving protein interaction networks through gene duplication. *J. Theor. Biol.* **222**, 199–210. (doi:10.1016/S0022-5193(03)00028-6)
64. Solé RV, Pastor-Satorras R, Smith E, Kepler TB. 2002 A model of large-scale proteome evolution. *Adv. Complex Syst.* **5**, 43–54. (doi:10.1142/S021952590200047X)
65. Dorogovtsev SN, Mendes JFF. 2003 *Evolution of networks: from biological nets to the Internet and WWW*. Oxford, UK: Oxford University Press.
66. Wolpert DH, Macready WG. 1997 No free lunch theorems for optimization. *IEEE Trans. Evol. Comp.* **1**, 67–82. (doi:10.1109/4235.585893)
67. Hordijk W. 2016 Evolution: limited and predictable or unbounded and lawless? *Biol. Theory* **11**, 187–191. (doi:10.1007/s13752-016-0251-5)