Complex Systems **4** (1990) 491–507

Partitions, Rational Partitions, and Characterization of Complexity

Mario Casartelli
Dipartimento di Fisica dell'Università di Parma,
Sezione Teorica, CNR, INFM, Italy

Abstract. The quantitative characterization of complexity for several dynamic processes is approached by a mathematical scheme based on the metric and combinatorial properties of the space $\mathcal{Z}$ of finite measurable partitions. Precisely, $\mathcal{Z}$ is embedded in a larger space $\mathcal{R}$ of new objects, "rational partitions," describing the antisimilarity between two probabilistic experiments. Rational partitions consist of properly reduced couples of ordinary partitions, and their main features are briefly reviewed. An extended entropy functional in $\mathcal{R}$ allows the introduction of indicators sensitive to different possible aspects of complexity for cellular automata, shifts, mappings, etc. These indicators appear to be accessible to numerical experiments in many nontrivial situations.

1. Introduction

It seems natural that there does not exist a uniquely defined approach to quantitative estimates of complexity, nor a single functional mapping the degree of complexity into a real number. Complexity, indeed, is in itself a complex concept, in the sense that it presents many aspects that are irreducible to one another. Updated reviews stress the matter [1–4]. Neither the Kolmogorov metric entropy, nor Kolmogorov or Chaitin idea of algorithmic complexity, for instance, seems to distinguish adequately the character of processes that behave very differently, during finite times, for a local observer. It would be useful therefore to get indices sufficiently flexible to reflect, through a proper tuning of their parameters, possible different aspects of local complexity.

Our purpose here consists in introducing a class $\mathcal{R}$ of mathematical objects, "rational partitions," extending the ordinary concept of partitions in probability spaces, whose properties fit a simplified characterization of complexity for several nontrivial models. In a number of interesting situations (e.g., all one-dimensional cellular automata, strange attractors, shifts, etc.) indices based on rational partitions are effectively computable quantities.

Moreover, rational partitions result to be endowed with peculiar probabilistic and combinatorial properties, which seem to deserve attention in themselves, independently from the complexity problem.

Some elementary facts and notations have to be recalled [5–7]. In a probability space $(\mathbf{M}, \mathcal{M}, \mu)$ ($\mathcal{M}$ being a σ-algebra of subsets of $\mathbf{M}$ and μ a probability measure on $\mathcal{M}$), a collection $\alpha \equiv \{A_1, \ldots, A_n\} \in \mathcal{M}$ is said to be a *measurable partition* of $\mathbf{M}$ if

$$\cup_{i=1}^{n} A_i = \mathbf{M} \quad \text{and} \quad A_j \cap A_k = \emptyset, \quad j \neq k \tag{1.1}$$

(all relations are to be intended mod 0). Sets A_i are called *cells* or *atoms* of α. The trivial partition consisting of a single atom (necessarily $\mathbf{M}$) will be denoted ν. The relation $\alpha \leq \beta$ (sometimes read as "α contained in β," with an abuse that does not make confusion) means that β refines α. Such a relation introduces a partial order in the set $\mathcal{Z}$ of all partitions.

If $\alpha \equiv (A_1, \ldots, A_n)$ and $\beta \equiv (B_1, \ldots, B_r)$ are two partitions, their *product* or *composition* $\gamma = \alpha \vee \beta$ is the minimal partition such that $\alpha \leq \gamma$ and $\beta \leq \gamma$. Atoms C_{ij} of γ are the nonempty intersections of A_i and B_j. The operation "$\vee$" is commutative and associative. For brevity, we shall often write $\alpha\beta$ instead of $\alpha \vee \beta$.

Another operation, $\gamma = \alpha \wedge \beta$, defines the *intersection* or *decomposition* γ between α and β, as the maximal partition such that $\gamma \leq \alpha$ and $\gamma \leq \beta$. Operations "$\vee$" and "$\wedge$" in $\mathcal{Z}$ may be seen as analogous to the minimal common multiple and maximal common divisor of entire numbers. In particular, for every α the *unit* partition ν trivially satisfies $\alpha \vee \nu = \alpha$, $\alpha \wedge \nu = \nu$.

The *Shannon entropy* of a partition α is the functional $H(\alpha)$:

$$H(\alpha) \stackrel{\text{def}}{=} -\sum_{i=1}^{n} \mu(A_i) \ln(\mu(A_i)) \tag{1.2}$$

A *conditional entropy* $H(\alpha|\beta)$ of α with respect to β is also defined:

$$H(\alpha|\beta) \stackrel{\text{def}}{=} -\sum_{i=1}^{n} \sum_{j=1}^{r} \mu(A_i \cap B_j) \ln(\mu(A_i|B_j)) \tag{1.3}$$

where, as usual, $\mu(a_i|B_j) = \mu(A_i \cap B_j)/\mu(B_j)$ whenever $A_i \cap B_j \neq \emptyset$. Main properties of $H(\alpha)$ and $H(\alpha|\beta)$, such as $H(\alpha|\beta) = H(\alpha\beta) - H(\beta)$, are listed in the references [5–7].

The set $\mathcal{Z}$ of all measurable finite partitions may be made a metric space introducing for every α and β the distance $\rho(\alpha|\beta)$:

$$\rho(\alpha|\beta) \stackrel{\text{def}}{=} H(\alpha|\beta) + H(\beta|\alpha) \tag{1.4}$$

All these quantities admit a natural probabilistic interpretation: a partition α may be seen as an abstract mathematical scheme for an *experiment*, where atoms represent single events. $H(\alpha)$ is the *a priori* incertitude about

experiment α, while $H(\alpha|\beta)$ is a measure of the residual incertitude of α when experiment β is assumed to be known. In this perspective, ρ gives an estimate of the degree of similitude between two experiments, and product $\vee$ gives the simplest experiment containing all informations from its factors. Nevertheless, $\vee$ is not a product in the usual sense, because all partitions $\gamma \leq \alpha$ behave on α as the unit partition: $\alpha\gamma = \alpha$; moreover, since for every α and β, $\alpha\beta \geq \alpha$ and $\alpha\beta \geq \beta$, there does not exist in $\mathcal{Z}$ any partition *inverse* to a given partition, i.e., such that their product is ν.

We shall prove (1) that $\mathcal{Z}$ may be embedded in a larger space $\mathcal{R}$ where the extended product $\vee$ becomes invertible, and (2) that $\mathcal{R}$ enjoys interesting probabilistic features, nontrivial extensions of those of $\mathcal{Z}$, which naturally apply to the study of complex systems. In analogy with the embedding of entire numbers in the rational field, where the product becomes invertible, the elements $\mathbf{a} \in \mathcal{R}$ will be called *rational partitions*.

In section 2, definitions introducing rational partitions are given, and their main properties are discussed with particular reference to the meaning of the extended entropy functional. In section 3, rational partitions are applied to the analysis of complexity in various processes. Final comments and conclusions are given in section 4.

2. Rational partitions

The aforementioned analogy with rational numbers consists in the following: consider the set $\mathcal{Z}^2 = \mathcal{Z} \times \mathcal{Z}$ of all ordered couples of finite measurable partitions. For couples (α_1, α_2) and (β_1, β_2), a product could be defined in a natural way by

$$(\gamma_1, \gamma_2) = (\alpha_1, \alpha_2) \vee (\beta_1, \beta_2) = (\alpha_1\beta_1, \alpha_2\beta_2) \tag{2.1}$$

Assuming moreover an equivalence relation such that for every η,

$$(\alpha_1\eta, \alpha_2\eta) \equiv (\alpha_1, \alpha_2) \tag{2.2}$$

it would follow, for every α, that $(\alpha, \alpha) \equiv (\nu, \nu)$, and consequently that

$$(\alpha, \beta) \vee (\beta, \alpha) = (\alpha\beta, \alpha\beta) \equiv (\nu, \nu) \tag{2.3}$$

In product (2.1) (ν, ν) is the unit couple; moreover, for couples of the form (α, ν), an obvious isomorphism holds with usual partitions and product. It seems therefore that $\mathcal{Z}^2$ may constitute the natural extension of $\mathcal{Z}$, where the "inverse" exists for every couple (and, in particular, $(\alpha, \nu)^{-1} = (\nu, \alpha)$). Nevertheless, such a naive extension is unsatisfactory in many respects. This essentially depends on the fact that a partition does not admit a uniquely defined decomposition into "prime" factors. For instance, it may happen

that $\gamma = \alpha_1\beta_1 = \alpha_2\beta_2$, and consequently that the reduction (2.2) of common factors may give both

$$(\gamma, \alpha_1\alpha_2) = (\beta_1, \alpha_2); \qquad (\gamma, \alpha_1\alpha_2) = (\beta_2, \alpha_1)$$

We shall introduce now an equivalence criterion allowing to overcome, as far as possible, such an ambiguity.

Definition 2.1. For every partition α with atoms $\{A_1, \ldots, A_N\}$, we define "simple factors" of α as the partitions $\alpha_k \equiv \{A_k, \bar{A}_k\}$, where $\bar{A}_k$ is the complementary set $M - A_k$. $S(\alpha)$ will denote the collection $\{\alpha_k\}$ of all simple factors of α, $S(\alpha, \sigma)$ the subcollection of those α_k such that $\alpha_k \wedge \sigma = \nu$, i.e., simple factors of α that are prime with σ. Therefore, $S(\alpha, \alpha) \equiv \nu$, $S(\alpha, \nu) \equiv S(\alpha)$, and $\beta_1 \leq \beta_2 \Rightarrow S(\alpha, \beta_2) \subseteq S(\alpha, \beta_1)$. $S(\alpha)$ generates α in the sense that $\alpha = \vee_k \alpha_k$.

Definition 2.2. For every α and β, let $\sigma = \alpha \wedge \beta$ be their maximal common factor. Classes $S(\alpha, \sigma)$ and $S(\beta, \sigma)$ are then defined. An operation $\pi : \mathcal{Z}^2 \to \mathcal{Z}^2$:

$$\pi(\alpha, \beta) = (\alpha', \beta') \tag{2.4}$$

may be introduced by defining α' and β':

$$\begin{cases} \alpha' = \vee_k \alpha_k, & \alpha_k \in S(\alpha, \sigma) \\ \beta' = \vee_j \beta_j, & \beta_j \in S(\beta, \sigma) \end{cases} \tag{2.5}$$

Clearly,

$$\alpha' \leq \alpha, \qquad \beta' \leq \beta \tag{2.6}$$

Example 2.3. Let $\mathbf{M} = (1, 2, 3, 4, 5, 6)$ be a set of six points. Let α and β be respectively

$$\begin{aligned} \alpha &\equiv \{(1), (2), (3), (4, 5, 6)\} \\ \beta &\equiv \{(1, 2), (3), (4, 5), (6)\} \end{aligned}$$

with obvious notation for subsets. Then

$$\sigma \equiv \alpha \wedge \beta = \{(1, 2), (3), (4, 5, 6)\}$$

and the classes $S(\alpha, \sigma)$ and $S(\beta, \sigma)$ are composed by

$$\begin{aligned} \alpha_1 &= \{(1), (2, 3, 4, 5, 6)\} \\ \alpha_2 &= \{(1, 3, 4, 5, 6), (2)\} \\ \beta_1 &= \{(1, 2, 3, 4, 5), (6)\} \\ \beta_3 &= \{(1, 2, 3, 6), (4, 5)\} \end{aligned}$$

therefore

$$\begin{aligned} \alpha' &= \{(1), (2), (3, 4, 5, 6)\} < \alpha \\ \beta' &= \{(1, 2, 3), (4, 5), (6)\} < \beta \end{aligned}$$

and

$$\sigma' = \alpha' \wedge \beta' = \{(1,2,3,4,5,6)\} \equiv \nu$$

Theorem 2.4. *π is a projection into an invariant subset $\mathcal{R} \subset \mathcal{Z} \times \mathcal{Z}$, i.e. $\pi^2 = \pi$.*

Proof. Since α' is generated by $S(\alpha, \sigma)$ it results that, between its simple factors, there are at least all generating factors $\alpha_k \in S(\alpha, \sigma)$, and therefore $S(\alpha', \sigma) \supseteq S(\alpha, \sigma)$. Iterating the process π, one should consider α'' generated by $S(\alpha', \sigma')$, where $\sigma' = \alpha' \wedge \beta'$. It holds that $\alpha'' \leq \alpha'$, as in general from 2.6, but since $\sigma' \leq \sigma$ we have

$$S(\alpha', \sigma') \supseteq S(\alpha', \sigma) \supseteq S(\alpha, \sigma)$$

and therefore $\alpha'' \geq \alpha'$, which implies $\alpha'' = \alpha'$. In the same way, $\beta'' = \beta'$. ∎

Note 2.5. Differently from rational numbers, it may happen that reduced partitions α' and β' are not relatively prime, i.e., that $\sigma' = \alpha' \wedge \beta' \neq \nu$, as another example shows:

Example 2.6. With $\mathbf{M} = (1,2,3,4,5,6,7,8,9,10)$, let

$$\begin{aligned} \alpha &= \{(1,2,3),(4,5),(6,7),(8,9,10)\} \\ \beta &= \{(1,2,3),(4,6),(5,7),(8,9,10)\} \end{aligned}$$

then

$$\begin{aligned} \sigma &= \{(1,2,3),(4,5,6,7),(8,9,10)\} \\ \alpha_1 &= \{(4,5),(1,2,3,6,7,8,9,10)\} \\ \alpha_2 &= \{(6,7),(1,2,3,4,5,8,9,10)\} \\ \beta_1 &= \{(4,6),(1,2,3,5,7,8,9,10)\} \\ \beta_2 &= \{(5,7),(1,2,3,4,6,8,9,10)\} \end{aligned}$$

and finally

$$\begin{aligned} \alpha' &= \{(4,5),(6,7)(1,2,3,8,9,10)\} < \alpha \\ \beta' &= \{(4,6),(5,7),(1,2,3,8,9,10)\} < \beta \\ \sigma' &= \alpha' \wedge \beta' = \{(4,5,6,7),(1,2,3,8,9,10)\} < \sigma \end{aligned}$$

The following theorem precisates the nature of σ'.

Theorem 2.7. *Let $\sigma' = \alpha' \wedge \beta' \neq \nu$. Atoms of σ' are either those atoms of σ that are composed with more than one atoms of α and β, or they are complementary sets to such atoms.*

Proof. If $\sigma = \alpha$ then $\alpha' = \nu$, and $\sigma' = \nu$ (the same with respect to β). If $\sigma \neq \alpha$, then there exists at least one atom in σ composed by two or

more atoms of α. Let $S^{(m)}$ be one such atom: $S^{(m)} = \cup_1^{k_m} A_j^{(m)}$; $\alpha_j^{(m)}$ will denote the simple partition corresponding to $A_j^{(m)}$. Clearly, $\alpha_j^{(m)} \wedge \sigma = \nu$, therefore $\alpha_j^{(m)} \in S(\alpha, \sigma)$ and $S^{(m)}$ is still a subset in α'. Clearly, also the complementary set to such sets stays in α' (and this is the only possibility for an atom of σ coinciding with a single atom of α). Repeating the same with respect to β, and taking into account that σ' is made up of subsets common to α' and β', the theorem follows. ∎

We remark that simple partitions are irreducible, and in this sense they play the role of prime factors for numbers. But every dicotomic partition (i.e., a partition composed of two atoms) is also irreducible. Therefore there may exist other classes of dicotomic (not simple) factors that, provided they satisfy theorem 2.4, would lead to a different reduction process $\bar{\pi}$: it seems that this reintroduces the ambiguity we wanted to cancel. The point is that the very choice of classes of irreducible factors constitutes a tool to fit peculiar features of objects to describe. The reduction process introduced above is the most natural for abstract partitions spaces, where only measurability is required (therefore it is always reproducible, and by default we shall refer to it). But atoms may enjoy further features (connection, order, etc.) suggesting a different and more convenient choice of irreducible factors [14]. In other words, classes of factors alternative to $S(\alpha)$, $S(\alpha, \sigma)$, etc. may be correlated to a definite similarity criterion between partitions. In this sense the possibility of a different reduction process is a richness, not an ambiguity of the method.

Definition 2.8. A rational partition $\mathbf{a} = (\alpha_1, \alpha_2)$ is an element of the set $\mathcal{R} \subset \mathcal{Z} \times \mathcal{Z}$ of the ordered couples such that $\pi(\alpha_1, \alpha_2) = (\alpha_1, \alpha_2)$.

To prove that $\mathcal{R}$ constitutes an extension of $\mathcal{Z}$, a consistent definition of the product is needed.

Definition 2.9. Let $\mathbf{a} = (\alpha_1, \alpha_2)$ and $\mathbf{b} = (\beta_1, \beta_2)$ be in $\mathcal{R}$. Their product $\mathbf{c} = \mathbf{a} \vee \mathbf{b} = \mathbf{ab} = (\gamma_1, \gamma_2)$ is

$$\mathbf{c} = \mathbf{ab} = (\gamma_1, \gamma_2) = \pi(\alpha_1\beta_1, \alpha_2\beta_2) \tag{2.7}$$

In the particular case when $\gamma_1 = \alpha_1\beta_1$, $\gamma_2 = \alpha_2\beta_2$, $\mathbf{a}$ and $\mathbf{b}$ will be called "relatively independent" rational partitions (or r.i. r-partitions).

Note 2.10. Product (2.7) is clearly commutative and idempotent in the whole $\mathcal{R}$, where $\mathbf{u} = (\nu, \nu)$ is the unit element. Moreover, r-partitions of the form (α, ν) constitute a subset $\mathcal{R}^\nu$ isomorphic to $\mathcal{Z}$.

Note 2.11. For $\mathbf{a} = (\alpha_1, \alpha_2) \in \mathcal{R}$, let $\bar{\mathbf{a}} = (\alpha_2, \alpha_1)$. Then

$$\mathbf{a}\bar{\mathbf{a}} = \pi(\alpha_1\alpha_2, \alpha_1\alpha_2) \equiv \mathbf{u}$$

and $\bar{\mathbf{a}}$ may be noted $\mathbf{a}^{-1}$. In particular, r-partitions of the form (ν, α) are inverse to those in $\mathcal{R}^\nu$. However, the inverse element is not uniquely defined in general, because equation $\mathbf{ax} = \mathbf{u}$ may have further solutions besides $\mathbf{x} = \mathbf{a}^{-1}$.

Note 2.12. Since π and product in $\mathcal{Z}$ do not commute, product in $\mathcal{R}$ is no longer associative, and equation $(\mathbf{ab}^{-1}) \vee (\mathbf{bc}) = \mathbf{ac}$ (which is true in $\mathcal{R}^\nu$) does not hold in general. Therefore definition 2.9 must be completed. For $n > 2$ we can distinguish, for instance, between a nonordered product

$$\vee_{k=1}^{n} \mathbf{a}_k = \pi\left(\vee_{k=1}^{n} \alpha_1^{(k)}, \vee_{k=1}^{n} \alpha_2^{(k)}\right)$$

and an ordered one defined iteratively through 2.7:

$$\vee_{k=1}^{n} \mathbf{a}_k = \mathbf{a}_n \vee \left(\vee_{k=1}^{n-1} \mathbf{a}_k\right)$$

Note 2.13. Partial order in $\mathcal{R}^\nu$ is extended in $\mathcal{R}$ by the relation

$$(\alpha_1, \alpha_2) \leq (\beta_1, \beta_2) \Leftrightarrow \alpha_1 \leq \beta_1, \alpha_2 \leq \beta_2 \tag{2.8}$$

but, as a consequence of the reduction process, the relation $\mathbf{a} \leq \mathbf{b} \Rightarrow \mathbf{a} \leq \mathbf{bc}$ is no longer true for every $\mathbf{c}$. It holds, however, for r.i. r-partitions. Distance ρ is also extended from $\mathcal{Z}$ to $\mathcal{R}$ defining ρ_r:

$$\rho_r(\mathbf{a}, \mathbf{b}) = \rho(\alpha_1, \beta_1) + \rho(\alpha_2, \beta_2) \tag{2.9}$$

The obvious meaning of ρ_r is that two r-partitions are near when both their first and second terms are near.

So far, apart from the extension of the metric, only combinatorial properties have been considered. The fact that we deal with measurable partitions enters with the probabilistic interpretation of $\mathcal{R}$: as a partition is a scheme for an experiment, a rational partition may be seen as a representation of the difference or "antisimilarity" between two experiments, since the reduction process eliminates common factors as far as possible. If $\mathbf{z} = (\alpha, \beta) \in \mathcal{R}$ and $\sigma = \alpha \wedge \beta \neq \nu$, this means that there is an irreducible similarity between α and β. On the contrary, it may happen that $\sigma = \nu$ even if every atom A_k in α is very similar to an atom B_j in β (in the sense that the symmetric difference of A_k and B_j is very small). So, there are two different concepts of similarity between partitions: a combinatorial one, which is taken into account by the reduction process π, and a metric one, which may regard also relatively prime or irreducible partitions.

To pass from these qualitative considerations to a quantitative estimate of antisimilarity, a suitable extension of entropy from $\mathcal{Z}$ to $\mathcal{R}$ is needed:

Definition 2.14. The entropy of a rational partition, or r-entropy, is a functional h_r in $\mathcal{R}$ with the following properties:

1. h_r coincides in $\mathcal{R}^\nu$ with the ordinary entropy H in $\mathcal{Z}$.
2. $h_r(\mathbf{a}) \geq 0$, and $h_r(\mathbf{a}) = 0 \Rightarrow \mathbf{a} = \mathbf{u}$

3. if $\mathbf{a}$ and $\mathbf{b}$ are r.i. r-partitions, $h_r(\mathbf{ab}) \leq h_r(\mathbf{a}) + h_r(\mathbf{b})$. A further axiom will be considered separately:
4. let α_n be a sequence of partitions irreducible with respect to β (i.e., $\mathbf{a}_n \equiv (\alpha_n, \beta) \in \mathcal{R}$), and such $\alpha_n \to \beta$ in the metric of $\mathcal{Z}$; then, $h_r(\mathbf{a}_n) \to 0$.

Considering for the moment properties (1)–(3), the only nontrivial point is the restriction of the "subadditivity" (3) to r.i. r-partitions. We shall return later to this point, which is deeply involved with the features of $\mathcal{R}$. Two important and, in a sense, natural examples of r-entropies are presented in the following theorem.

Theorem 2.15. *Functionals H^0 and H^1 defined for every r-partition $\mathbf{a} = (\alpha_1, \alpha_2)$ by*

$$H^0(\mathbf{a}) = H(\alpha_1) + H(\alpha_2) \tag{2.10}$$

$$H^1(\mathbf{a}) = \rho(\alpha_1, \alpha_2) \tag{2.11}$$

are r-entropies in $\mathcal{R}$.

Proof. Requirements (1) and (2) of definition 2.15 are easily checked for both H^0 and H^1, by elementary properties of entropy H and distance ρ in $\mathcal{Z}$. To prove (3) for H^0 we have

$$\mathbf{ab} = \pi(\alpha_1\beta_1, \alpha_2\beta_2) = (\gamma_1, \gamma_2)$$

for every $\mathbf{a}, \mathbf{b} \in \mathcal{R}$, and we may write $\alpha_1\beta_1 = \sigma\gamma_1$ and $\alpha_2\beta_2 = \sigma\gamma_2$. Now

$$\begin{aligned} H^0(\mathbf{ab}) &= H(\gamma_1) + H(\gamma_2) \leq H(\sigma\gamma_1) + H(\sigma\gamma_2) \\ &\leq H(\alpha_1) + H(\alpha_2) + H(\beta_1) + H(\beta_2) = H^0(\mathbf{a}) + H^0(\mathbf{b}) \end{aligned}$$

Thus, for H^0, the subadditivity is not even restricted to r.i. r-partitions, but holds in the whole $\mathcal{R}$.

Now let $\mathbf{a}$ and $\mathbf{b}$ be relatively independent, i.e., $\mathbf{ab} = (\alpha_1\beta_1, \alpha_2\beta_2)$. Then, using known properties of the conditional entropy in $\mathcal{Z}$,

$$\begin{aligned} H^1(\mathbf{ab}) &= \rho(\alpha_1\beta_1, \alpha_2\beta_2) = H(\alpha_1\beta_1|\alpha_2\beta_2) + H(\alpha_2\beta_2|\alpha_1\beta_1) \\ &= H(\alpha_1|\alpha_2\beta_2) + H(\beta_1|\alpha_1\alpha_2\beta_2) + H(\alpha_2|\alpha_1\beta_1) + H(\beta_2|\alpha_1\alpha_2\beta_1) \\ &\leq H(\alpha_1|\alpha_2) + H(\alpha_2|\alpha_1) + H(\beta_1|\beta_2) + H(\beta_2|\beta_1) \\ &= \rho(\alpha_1, \alpha_2) + \rho(\beta_1, \beta_2) = H^1(\mathbf{a}) + H^1(\mathbf{b}) \end{aligned}$$

∎

Note 2.16. The proof of (3) for H^1 does not work in the general case, when $\mathbf{ab} = (\gamma_1, \gamma_2)$ with $\alpha_1\beta_1 = \sigma\gamma_1$, $\alpha_2\beta_2 = \sigma\gamma_2$ and $\sigma \neq \nu$, because $H^1(\mathbf{ab}) = \rho(\gamma_1, \gamma_2) \geq \rho(\sigma\gamma_1, \sigma\gamma_2) = \rho(\alpha_1\beta_1, \alpha_2\beta_2)$. However, notwithstanding the failure of the proof, one may conjecture that H^1 too enjoies subadditivity in the whole of $\mathcal{R}$. The answer to this conjecture is negative. We shall

briefly indicate how to build a counterexample: let $\mathbf{x} = (\xi_1, \xi_2)$ be such that $H^1(\mathbf{x}) < \varepsilon$. Let $\mathbf{z} = (\xi_1, \xi_2)$ be relatively independent with respect to both $\mathbf{x}$ and $\mathbf{x}^{-1}$, and

$$H^1(\mathbf{z}) > \delta, \qquad H^1(\mathbf{z}\mathbf{x}^{-1}) < \varepsilon$$

then if $\mathbf{q} = \mathbf{z}\mathbf{x}^{-1}$ and $\mathbf{q}\mathbf{x} = \mathbf{z}\mathbf{x}\mathbf{x}^{-1} \equiv \mathbf{z}$,

$$H^1(\mathbf{q}\mathbf{x}) = H^1(\mathbf{z}) > \delta$$

$$H^1(\mathbf{q}) + H^1(\mathbf{x}) < 2\varepsilon$$

and a good choice of ε and δ leads to the result. It is not difficult to provide examples of r-partitions with the required features.

Note 2.17. As a corollary of theorem 2.15 we have that

$$H^p(\mathbf{a}) = (1-p)H^0(\mathbf{a}) + pH^1(\mathbf{a}) \tag{2.12}$$

for $0 \leq p \leq 1$, are also r-entropies (the proof is immediate). The converse is not true: there exist r-entropies that are not in the form 2.12 (for instance $h_r(\mathbf{a}) = \max\{H(\alpha_1), H(\alpha_2)\}$).

Note 2.18. Even if both H^0 and H^1 satisfy (1)–(3) in definition 2.14, their meaning is quite different: H^0 simply gives the entropy of a r.p. $\mathbf{z} = (\alpha, \beta)$ as the sum of entropies of α and β, while H^1 is a measure of their distance. Now, dealing with partitions of a Lebesgue space, it is possible that, for fixed M and ε, $H^0 > M$ and $H^1 < \varepsilon$. It is the same as with rational numbers $r = P/Q$, where rational "height" $P + Q$ may be larger than any fixed quantity, and the value of r as close as one wants to 1. Taking into account that the combinatorial similarity has been reduced by π, H^1 is therefore more representative of the metric similarity between its components, giving, with respect to H^0, a better account of the true difference between experiments.

Note 2.19. From elementary properties of ρ in $\mathcal{Z}$, $\rho(\alpha\gamma, \beta\delta) \leq \rho(\alpha, \beta)$ for every α, β, γ, δ. Therefore, if $\mathbf{a} = \pi(\alpha_1, \alpha_2)$, $H^1(\mathbf{a}) \geq \rho(\alpha_1, \alpha_2)$, while obviously $H^0(\mathbf{a}) \leq H(\alpha_1) + H(\alpha_2)$. In other words, a large common factor may mask the antisimilarity of two partitions; by eliminating this common part, reduction π amplifies the metric distance H^1 of the couple and reduces on the contrary its rational height H^0.

Note 2.20. The choice of H^1 as the "true" r-entropy would follow straight from the supplementary axiom (4), not satisfied by H^0. We are inclined to think, indeed, that such a metric continuity is important, e.g., for the problems of the next section. Nevertheless, the weaker definition with axioms (1)–(3) could turn out to be useful in contexts where such properties play a role as the one described in the following theorem.

Theorem 2.21. *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *in* $\mathcal{R}$. *Then*

$$\mathbf{a} \leq \mathbf{b} \Rightarrow H^0(\mathbf{a}) \leq H^0(\mathbf{b}) \tag{2.13}$$

while the same property does not hold for H^1

Proof. Since $\mathbf{a} \leq \mathbf{b}$, we write $\mathbf{a} = (\alpha_1, \alpha_2)$, $\mathbf{b} = (\gamma_1\alpha_1, \gamma_2\alpha_2)$. The first part of the theorem is quite obvious:

$$H^0(\mathbf{a}) = H(\alpha_1) + H(\alpha_2) \leq H(\gamma_1\alpha_1) + H(\gamma_2\alpha_2) = H^0(\mathbf{b})$$

To prove that 2.13 does not hold for H^1, we choose a particular $\mathbf{b} \equiv \mathbf{b}_\varepsilon = (\alpha_1, \gamma_\varepsilon\alpha_2) \geq \mathbf{a}$, where, for $\varepsilon > 0$, γ_ε is such that

$$\rho(\alpha_1, \gamma_\varepsilon) = H(\alpha_1|\gamma_\varepsilon) + H(\gamma_\varepsilon|\alpha_1) < \varepsilon \tag{2.14}$$

Define now A and B by

$$\begin{aligned} A &= H^1(\mathbf{a}) = H(\alpha_1|\alpha_2) + H(\alpha_2|\alpha_1) \\ B &= H^1(\mathbf{b}_\varepsilon) = H(\alpha_1|\gamma_\varepsilon\alpha_2) + H(\gamma_\varepsilon\alpha_2|\alpha_1) \end{aligned}$$

Let $A' = A - Z$, $B' = B - Z$, where $Z = H(\alpha_2|\alpha_1)$. Then

$$\begin{aligned} A' &= H(\alpha_1|\alpha_2) \\ B' &= H(\alpha_1|\gamma_\varepsilon\alpha_2) + H(\gamma_\varepsilon|\alpha_1\alpha_2) \end{aligned}$$

Since $H(\alpha_1|\gamma_\varepsilon\alpha_2) \leq H(\alpha_1|\gamma_\varepsilon)$ and $H(\gamma_\varepsilon|\alpha_1\alpha_2) \leq H(\gamma_\varepsilon|\alpha_1)$, from 2.14 it follows that $B' < \varepsilon$, and, choosing $\varepsilon < H(\alpha_1|\alpha_2)$, $B' < A'$. Therefore $B < A$, i.e., $H^1(\mathbf{a}) > H^1(\mathbf{b}_\varepsilon)$. ∎

A consistent extension of entropy in $\mathcal{R}$ is also expected to generalize the fundamental relation of the conditional entropy:

$$H(\alpha|\beta) = H(\alpha\beta) - H(\beta) \tag{2.15}$$

which, in $\mathcal{Z}$, has an immediate and meaningful interpretation in terms of "residual information." Then, whatever h_r is from definition 2.14, we pose

$$h_r(\mathbf{a}|\mathbf{b}) = h_r(\mathbf{ab}) - h_r(\mathbf{b}) \tag{2.16}$$

which reduces to 2.15 in $\mathcal{R}^\nu$. It is possible to choose $\mathbf{p}, \mathbf{q} \in \mathcal{R}$ in such a way that, for $\mathbf{a} = \mathbf{pq}^{-1}$ and $\mathbf{b} = \mathbf{q}$

$$\begin{aligned} h_r(\mathbf{a}|\mathbf{b}) &= h_r(\mathbf{pqq}^{-1}) - h_r(\mathbf{q}) \\ &= h_r(\mathbf{p}) - h_r(\mathbf{q}) \end{aligned}$$

Therefore, when $h_r(\mathbf{p}) < h_r(\mathbf{q})$, $h_r(\mathbf{a}|\mathbf{b}) < 0$. We have proven the following statement: in $\mathcal{R}$, the conditional r-entropy 2.16 is not necessarily nonnegative. In particular, $h_r(\mathbf{a}|\mathbf{a}^{-1}) = -h_r(\mathbf{a})$. Such a conclusion does not fit the

usual concept of conditional entropy as residual information. It excludes, for instance, that a distance in $\mathcal{R}$ may be based on a formula like 1.4 (it would give $\tilde{\rho}(\mathbf{a}, \mathbf{a}^{-1}) = -2h_r(\mathbf{a})$).

In conclusion, apart from the problem of choosing the right entropy h_r, there are in $\mathcal{R}$ peculiar features that cannot be seen as trivial extensions of those of $\mathcal{Z}$. They derive indeed from nonassociativity and reduction or annihilation of factors by the process π. For instance, subadditivity (3) in definition 2.14 is limited to r.i. r-partitions just to take into account these features. In this context, the fact that quantity 2.16 may become negative is neither surprising nor in contrast with assuming it as an extended conditional entropy: it is the very idea of conditioning, indeed, to be deeply revised.

3. Rational partitions and complexity

We shall provide here some examples showing that the concepts of section 2 naturally apply to the quantitative characterization of complexity. Such examples will be successively considered within a common frame, which seems to be quite general. The first example regards cellular automata (see [8–12] for general references). A cellular automaton (CA) is a triple $\{K, L, f\}$ consisting of an alphabet K, a lattice L, and a map $f : K^L \to K^L$ that, for every configuration $q \in K^L$, gives a new configuration $f(q)$ depending only on the values of a finite neighborhood of each point on the lattice. To regain the concepts of section 2, to a particular configuration $q \in K^L$, we associate a partition α, e.g., in the following way: let $\mathbf{M}$ be a finite subset of L, say a hypercube. Then, the collection of homogeneous connexe subsets of $\mathbf{M}$ (where vertices are homogeneous if labeled by the same symbol of K, connexe if they may be joined by a homogeneous walk) is a partition of $\mathbf{M}$. If $\mathcal{Z} = \mathcal{Z}(\mathbf{M})$ is the space of such partitions, the correspondence $K^L \to \mathcal{Z}$ will be denoted ϕ, i.e., $\alpha = \phi(q)$. Such a correspondence is noninvertible, the associated partition being invariant both for permutations in K and for different configurations in K^L that coincide in $\mathbf{M}$.

Starting from a configuration q_0, the evolution produces a sequence of configurations $q_{k+1} = f(q_k)$; a transformation T_f in $\mathcal{Z}(\mathbf{M})$ is then naturally defined through the relation $\alpha_{k+1} = T_f(\alpha_k) = \phi(f(q_k))$. Provided that a probability measure μ_M is given in $\mathbf{M}$, the triple $\{\mathbf{M}, T_f, \mu_M\}$ constitutes a dynamical system (note however that T_f is not measure-preserving for subsets in $\mathbf{M}$). If $\mathbf{M}$ is sufficiently large, complexity of the automaton is related to the problem of giving a reasonable estimate of the creation of newness in the sequence $\alpha_0, \alpha_1, \alpha_2, \ldots$, with $\alpha_k \equiv T_f^k \alpha_0$. We claim that the concepts of section 2 may be useful to this task. Indeed, the point is now to establish how different α_{k+1} is from α_k. Defining more in general

$$\mathbf{a}_{k,n} \equiv \pi(\alpha_{k+n}, \alpha_k) \tag{3.1}$$

the entropy $h_r(\mathbf{a}_{k,n})$ results in being a measure of the antisimilarity created between the kth and $(k+n)$th steps. With $n = 1$, in particular, one considers antisimilarity produced step by step. Of course, the choice of a definite h_r

is important, and the discussion of section 2 will be taken into account. By default we shall use $h_r \equiv H^1$, not only for its continuity (note 2.21) but also because π amplifies the distance between partitions of a couple (note 2.19); this fact may be extremely useful in numerical experiments.

Note that a periodic evolution immediately reflects in the periodicity of $h_r(\mathbf{a}_{k,n})$ as a function of "time" k. The time behavior (periodic or not) of $h_r(\mathbf{a}_{k,n})$ and its absolute value are both significant in order to analyze the complexity of the automaton, while the parameter n is related to the "memory" of the process. The most immediate complexity indicator for generic CA is therefore the entropy itself:

$$I_n(k, \mathbf{M}, \alpha) = h_r(\mathbf{a}_{k,n}) \tag{3.2}$$

which as a function of k shows the variety of time behavior of the process through the amplitude and regularity of oscillations. Provided that the (absolutely nontrivial) problem of an explicit computation has been solved, the study of the time complexity of the automaton is mapped, independently of the dimension of L, in the one-dimensional analysis of the function, $I_n(k, \mathbf{M}, \alpha)$ versus k. Usual tools in the analysis of time series (e.g., power spectrum obtained by Fourier Transform) may be immediately applied.

Consider, for definiteness, an automaton where K has two symbols, say 0 and 1, and $L = \mathbf{Z}$, the relative integers. Let $\mathbf{M}$ be a string of N cells. Consider moreover the simple case of the rule exchanging the symbols 0 and 1. Then, for every initial configuration $q \in \{0,1\}^{\mathbf{Z}}$, the corresponding partition α of $\mathbf{M}$ is stable: $T_f^k \alpha \equiv \alpha$. The couple $(T_f^k \alpha, T_f^{k+1} \alpha)$, seen as a rational partition, corresponds to the unit $\mathbf{u}$, consistently with the fact that this particular rule does not create any novelty in the evolution. Whatever h_r is, $h_r(\mathbf{u}) = 0$. With another rule, $T_f^{k+1} \alpha$ is different, in general, from $T_f^k \alpha$ and the quantity (3.2) says how different they are.

Another indicator is the time average of the previous one:

$$I_n(\mathbf{M}, \alpha) = \lim_N \frac{1}{N} \sum_{k=1}^{N} I_n(k, \mathbf{M}, \alpha) \tag{3.3}$$

giving a time-independent estimate of the capacity of the automaton in producing new configurations at distance n. The dependence on α and $\mathbf{M}$ may be eliminated in an obvious way:

$$I_n(\mathbf{M}) = \sup_\alpha I_n(\mathbf{M}, \alpha) \tag{3.4}$$

$$I_n = \lim_M I_n(\mathbf{M}) \tag{3.5}$$

More correctly, in (3.4) the sup operation should be written with respect to configurations $a \in K^L$, because there are infinite configurations (giving rise to different evolutions in $\mathbf{M}$) that are projected by ϕ in the same partition α. The limit in (3.5) means, of course, that the size of $\mathbf{M}$ goes to ∞. We stress, however, that the analysis of the complexity in a dynamic process

should reflect the kind of interest the observer has, and that a finite time or a finite window **M** can be as significant, in certain contexts, as the limit quantities (3.4) and (3.5). It is also evident that from indicators (3.2) and (3.3) other useful parameters may be obtained, such as the variance of the values appearing in the time series.

All these considerations and formalism apply immediately to a second example, constituted by shifts [5], i.e., a source of letters of an alphabet K giving outcomes $\omega_j \in K$ at discrete times $j = \ldots, -1, 0, 1, 2, \ldots$. If the source is a purely probabilistic one, and the outcome $\omega_j = c$ occurs with a given *a priori* probability $\mu(c)$, then this process is a Bernoulli Shift B. Its dynamic (or Kolmogorov–Sinai) entropy $h(B)$ is the Shannon entropy of the alphabet: $h(B) = -\sum_{c \in K} \mu(c) \log \mu(c)$. If, on the contrary, the source was ruled by a periodic law repeating a string $\omega_1, \ldots \omega_P$, then $h(B) = 0$, independently of the length P. Indeed, after its definition, the Kolmogorov entropy takes into account the whole process in time, and the "local" complexity of a string of finite length is a minor feature, in a global evaluation, with respect to periodicity.

Note that Kolmogorov entropy in itself is an index of complexity, e.g., in the sense that it distinguishes periodic from probabilistic shifts and classifies processes with different entropies into distinct classes of isomorphism. However, as noticed, besides the complexity of evolution law in the whole, the complexity of local evolution of configurations during finite times can also be of a certain importance, both practical and conceptual. Consider, for example, the random number generator (RNG) of a computer as a process (the alphabet being in this case the extremely large but finite set of possible outcomes). For practical purposes, inasmuch the time considered in calculations is short with respect to the periodicity of the RNG, it appears and works indeed as a genuine random source. Nevertheless, in principle, due to the recurrence of the algorithm, its ideal Kolmogorov entropy would be 0, exactly as in the trivial alternate sequence $1, 0, 1, 0, 1, 0, \ldots$. In this sense, Kolmogorov entropy seems to give a too drastic estimate of the complexity of the process. It would be useful to develop rigorously defined indices giving account (not only from a practical point of view in numerical approximation) of such local features. As for cellular automata, indices based on the entropy of rational partitions offer a natural answer to this exigence.

Previous formalism is immediately regained: we consider a "window" of length M in the sequence $\underline{\omega} \equiv \{\omega_j\}$, for instance $\omega_{k+1}, \ldots, \omega_{k+M}$. As time goes on, after n steps the string will be $\omega_{k+n+1}, \ldots, \omega_{k+n+M}$. Every string may be partitioned into homogeneous substrings. For example, to the string $AABCCCAB$ we associate the partition $\{(AA), (B), (CCC), (A), (B)\}$. The rational partition (3.1), $\mathbf{a}_{k,n} \equiv \pi(\alpha_{k+n}, \alpha_k)$, is then defined, and formulas defining indices (3.2)–(3.3) still work for shifts.

A third example is connected to the complexity of geometrical sets. Let S be a subset of $\mathbf{R}^n$. For simplicity, suppose $S \subset \mathbf{Q}^n$, $\mathbf{Q}^n$ being the unit hypercube. For a given lattice in $\mathbf{Q}^n$, the "black" cells containing at least a point of S may be distinguished from the empty "white" cells. Once again,

connected subsets of white or black cells determine a partition α_0 of $\mathbf{Q}^n$. An enlargement of S by a scale factor g (in symbols: $S' = f_g(S)$), determines a new partition α_1. Then $\mathbf{a} \equiv \pi(\alpha_0, \alpha_1)$ is an r-partition, and the functional $h_r(\mathbf{a})$ gives a measure of the antisimilarity created in the enlargement. This operation may be iterated, and complexity indices corresponding to (3.2)–(3.3) may be easily introduced, provided that in this case the enlargement size g, a continuous parameter, plays the role of the discrete memory n. Usual considerations about the complexity of fractal sets with respect to ordinary sets may be easily rephrased in this way. It is almost obvious, for instance, that a rectifiable curve, after a certain number of enlargements, produces a sequence of identical partitions, and therefore a sequence of null r-partitions with 0 entropy; in a Cantor triadic set, the case when $g = 3^k$ is clearly singular, etc. So, with a proper tuning of the different parameters, indices based on r-partitions can give account of several aspects of geometrical complexity: the difference between fractal and nonfractal sets, the peculiarity of an exact internal self-similarity, the mean amount of antisimilarity created in iterated enlargements, and so on.

As a last example, in the same spirit, one could examinate the onset and evolution of fractality in attractors: consider a large but finite sequence $x_1, \ldots, x_n$ of points evolving in an interval divided into M subintervals. Then, subintervals containing at least one x_k are "black" and distinct from white, empty subintervals. The usual correspondence ϕ will produce a first partition α_1, the following sequence $x_{n+1}, \ldots, x_{2n}$ will give α_2, and so on. The analysis, at this point, goes on as in previous examples. We stress that both n and M should be large, and this makes the reduction process π important in order to eliminate a spurious similarity between partitions.

In all of these examples there are several points to be clarified and developed: (1) elaboration of explicit procedures for numerical experiments; (2) a possible utilization of indicators in the problem of isomorphism; (3) applications of method shadowed here for shifts to the analysis of general time series, or functions, with a discretized range; (4) possible relations with other concepts of complexity (Chaitin–Kolmogorov, Bennet, etc.).

4. Comments and conclusion

All examples of section 3 may be unified in a simple mathematical scheme: let S be a space of "states" $q, r, s, \ldots$, with a transformation $f : S \to S$, and let ϕ be a correspondence $S \to \mathcal{Z}$, where $\mathcal{Z}$ is a space of finite partitions. Then f induces a transformation T in $\mathcal{Z}$:

$$T(\phi(q)) = \phi(f(q))$$

If $\alpha = \phi(q)$, for a fixed n we define

$$\mathbf{a}_n = \pi(\alpha, T^n\alpha)$$

Then, as indicated by 3.2, the functional $h_r(\mathbf{a}_n)$ in $\mathcal{R}$ gives a proper estimate (which crucially depends on ϕ and the space $\mathcal{Z}$ of partitions) of the antisimilarity created by f^n in S. With small obvious adaptations in the notation, definitions 3.2–3.3 follow quite naturally and generally.

To be effective, the procedure above requires some cautions. First, as remarked in section 2, it may be convenient to consider a reduction process based on a class of irreducible factors different from simple factors. Second, for continuous problems, the relevance of the reduction process π depends on the possibility of a significant discretization (a discussion of this and related arguments may be found in [13]). Third, both the partition space $\mathcal{Z}$ and the connection $\phi : S \to \mathcal{Z}$ are, in principle, arbitrary, and their meaning should be discussed case by case. For instance, consider two types of shift sequences:

(a) $\ldots A, B, A, B, A, B, A, B, A, B, A, B, A, B, A, B, A, B, A, B, \ldots$

(b) $\ldots A, B, A, C, A, D, A, C, A, C, A, E, A, C, A, B, A, D, A, C, \ldots$

(i.e., a regularly alternate sequence and a sequence where odd elements are random and different from the even, fixed elements): the simple correspondence ϕ described in section 3 would produce identical partitions for (a) and (b), and the analysis would give a misleading result. However, introducing another correspondence ϕ', based for instance on the assumption of couples of letters (independently of their order) as basic alphabet, one would obtain in case (a) the identical sequence:

$$\ldots AB, AB, AB, AB, AB.AB.AB, AB, AB, AB, \ldots$$

with the associate null partition ν, and in case (b) a highly irregular sequence

$$\ldots AB, AC, AD, AC, AC, AE, AC, AB, AD, AC, \ldots$$

with a very refined partition.

As a time average of the local production of antisimilarity, $I_n(\mathbf{M}, \alpha)$ may be seen as an analogous quantity to the Lyapunov exponents measuring the rate of divergence of nearby orbits in the phase space of a dynamic system. In this spirit, it is possible to define other indicators, say $L(T)$, making the analogy with Lyapunov exponents even sharper.

Taking into account that $\mathcal{Z}$ is a metric space, for every given orbit $\{T^n, \alpha\}$, it is sufficient indeed to compute the rate of divergence of a second near-partition β, which evolves freely for a short time and then is drawn near to the first orbit. In such a way, the usual computing procedure of Lyapunov exponents may be replicated in $\mathcal{Z}$ or $\mathcal{R}$ as faithfully as possible. More explicitly:

$$L(T) = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \log(R_k / R_k^0) \tag{4.1}$$

where $R_k^0 = h_r(\pi(T^k\alpha, \beta_k))$ and $R_k = h_r(\pi(T^{k+1}\alpha, T\beta_k))$, β_k being the second partition chosen near to $T^k\alpha$ and independent of it. Here we explicitly pose $h_r \equiv H^1$. Of course, the same procedure can be applied both in $\mathcal{Z}$ or in $\mathcal{R}$, the only difference being in the reduction π (but the difference between divergences in $\mathcal{Z}$ or $\mathcal{R}$ is in itself an interesting quantity, since it measures the persistence in evolution of the common part of two partitions). Even if there are experimental evidences that this approach may prove fruitful [14], some considerations justify the particular attention devoted to index 3.2 in the previous section: (1) The evolution of a single partition regards the whole of S. In this sense, it is intrinsically an averaged quantity, deeply different in this respect from the evolution of a phase point for a dynamic system. (2) Indicators 3.2 and 3.4 give estimates of a sort of autocorrelation, in the evolution of the complexity, and this appears to be in any case a meaningful quantity. To look at simultaneous evolutions of two partitions, as in 4.1, could be justified and interesting for particular problems (e.g., evolution of patterns), and to discriminate in cases which appear ambiguous in a simpler approach.

Both indicators 3.2 and 4.1 are accessible to explicit numerical calculations, also in nontrivial cases. Experiments referring for example to one-dimensional CA and to the "onset of fractality" in geometrical sets seem to confirm the relevance and reliability of indicators based on r-partitions (for instance, improving by quantitative measurements the classification of CA proposed by Wolfram in [10]). They will be fully reported elsewhere [14].

Acknowledgments

I thank A. Albrigi and P. Celli for useful discussions and for criticism on an early version of this work.

References

[1] P. Grassberger, *Physica*, **140A** (1986) 319.

[2] D. Zambella and P. Grassberger, *Complex Systems*, **2** (1988) 269.

[3] C. Lindgren and M.G. Nordahl, *Complex Systems*, **2** (1988) 409.

[4] L. Peliti and A. Vulpiani, eds., *Measures of Complexity* (Springer, Berlin, 1988).

[5] P. Billingsley, *Ergodic Theory and Information* (Wiley, New York, 1965).

[6] V.I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968).

[7] Ya.G. Sinai, *Introduction to Ergodic Theory* (Princeton University Press, Princeton, NJ, 1976).

[8] S. Wolfram, *Review of Modern Physics*, **55** (1983) 601.

[9] S. Wolfram, *Physica*, **10D** (1984) 1.

[10] S. Wolfram, *Comm. Math. Phys.*, **6** (1984) 15.

[11] G.Y. Vichniac, *Physica*, **10D** (1984) 96.

[12] T. Toffoli, *Cellular Automata Mechanics*, Technical Report 208, CCS Department, University of Michigan (1977).

[13] C.S. Hsu, *Cell-to-Cell Mapping* (Springer Verlag, New York, 1987).

[14] A. Albrigi and M. Casartelli, in preparation.