Chapter 19
The Arrow of Time and Information Theory

Vieri Benci

Abstract In this paper first we present the notions of Boltzmann entropy and Shannon entropy and some notions from information theory. Also we define some new concepts such as Combinatorial Entropy and Computable Information Content. In the second part, we argue that the mechanisms which determine the two arrows of time (the thermodynamic arrow and the evolution arrow) can be modeled and better understood using these concepts.

Keywords Shannon entropy · Boltzmann entropy · Information theory · Algorithmic Information Content · Irreversible systems · Maxwell devil

1 Introduction

1.1 The Arrows of Time

The problems related to time are as old as human thinking. One of the most fascinating and unsettling problem is the arrow of time. In modern science the most meaningful indicators of flowing of time and of its direction are essentially two:

- (*II law*) The second law of thermodynamics: the passing of time destroys information. Time is Kronos who devours his offspring. Everything is consumed by the flow of time, and, at the very end, the universe will be an undifferentiated mass where even light and darkness will be hopelessly mixed together.
- (*Evolution*) Historical, biological, cosmological evolution: the passing of time creates information. In cosmological evolution, light is separated from darkness; galaxies, stars, and planets take their form; in biological evolution life arise from mud and bacteria, protista, fungi, plants, and animals evolve in always more complex forms; and then intelligence appears and evolution continues in history and gives origin to more and more complex civilizations.

V. Benci (✉)
Dipartimento di Matematica Applicata "U. Dini", Università degli Studi di Pisa, Largo Bruno Pontecorvo 1/c, 56127 Pisa, Italy
e-mail: benci@dma.unipi.it

S. Albeverio, P. Blanchard (eds.), *Direction of Time*,
DOI 10.1007/978-3-319-02798-2_19,

These two aspects of the arrow of time are apparently contradictory with each other. But there is more to say: both of them are in contradiction with the fundamental laws of physics.

In fact the fundamental laws of physics are reversible: they do not distinguish past and future. From a mathematical point of view, the law of physics are expressed by differential equations (e.g. the equations of Hamilton, Maxwell, Schrödinger, Einstein, etc.). The "state" of any physical system at time t is described by a function $u(t)$ which solves the equations involved. A peculiarity of all these equations lies in the fact that if $u(t)$ is a solution, also $u(-t)$ is a solution. This is the meaning of the word "reversibility", at least in this paper. This fact, in the physical world has the following meaning: if $u(t)$ describes the evolution of a physical system, also $u(-t)$ represents a possible evolution of the same system (with different initial conditions). This fact is in evident contradiction with experience. Nobody is born old and gets younger until becoming a child and finally disappearing in an egg.

1.2 The Aim of This Paper

Today, these apparent contradictions are understood reasonably well. The relation between the second law and the reversibility of the fundamental law of physics has been explained by Boltzmann. His theory has received and still receives many objections, but it is essentially correct (at least in my opinion and in the opinion of most scientists). The evolution arrow and its apparent contradiction with the second law, in recent times, has been object of a lot of attention and the study of dissipative systems explains reasonably well the underlying mechanisms. Nevertheless, still there are many subtle questions to be settled from many points of view: philosophical, physical and mathematical. One of these questions is related to the meaning of "information" and its relation with the many notions of "entropy". In this paper, we propose new and more precise definitions of these concepts which help, we hope, to clarify some delicate points on this matter. My point of view arises from the consideration that these concepts are universal and should be applicable to many contexts and not only to physical systems. Moreover, in any model in which the existence and coexistence of the two arrows of time are present, the basic definitions must be relatively simple. These ideas are supported by the empirical experience which we get from computer simulations. If we simulate a dynamical system of mixing type (such as the cat map) and we start with an ordered distribution of the initial points, we experience a growth of disorder which we can assimilate to the growth of entropy in a physical systems. On the other hand, if we simulate an irreversible dynamical system (such as the Conway "game of life") we see complex structures to appear. We experience a sort of creation of order and information. If creation and destruction of information can be simulated so easily, the right mathematical definitions which can describe ad eventually explain these phenomena, must be relatively simple.

2 Review of the Main Notions

2.1 Clausius and Boltzmann Entropy

One of the classic formulations of the second law of thermodynamics is the following:

"The entropy of an isolated system grows until the thermodynamical equilibrium which corresponds to the maximum of entropy"

In this case, by entropy, we mean the thermodynamic entropy defined (by Clausius in 1850) as follows:

$$S = \int \frac{dQ}{T}$$

According to the theory of Boltzmann the thermodynamic entropy can be defined from the law of classical dynamics. Boltzmann's theory is based on the distinction between macrostate and microstate.

The macrostate is defined by the property accessible to the observer, namely by the quantity which can be experimentally measured; the microstate is given by its complete description, namely the position and the velocity of all the elementary components of the system. For example, if we are dealing with a perfect gas, the macrostate is described by volume, pressure and temperature; the microstate is described by the position and the velocity of each molecule with a given accuracy.

According to Boltzmann, the entropy of the system can be defined by the following equation:

$$S = k \ln W \tag{1}$$

where k is the Boltzmann constant and W is the number of microstates compatible with the given macrostate.

This theory is consistent with many theoretical and experimental facts and it is accepted by the majority of scientists.

2.2 Shannon Information and Entropy

Let $\mathcal{A}$ be an alphabet, namely a finite collection of symbols (letters).

Given a finite string σ (namely a finite sequence of symbols taken in our alphabet), the intuitive meaning of *quantity of information* $I(\sigma)$ contained in σ is the following one:

$I(\sigma)$ is the length of the smallest binary message from which you can reconstruct σ.

Thus, formally

$$I : \Sigma(\mathcal{A}) \to \mathbf{N}$$

I is a function from the set of finite strings in a finite alphabet $\mathcal{A}$ which takes values in the set of natural numbers. There are different notions of information and some of them will be discussed here. The first one is due to Shannon.

In his pioneering work, Shannon defined the quantity of information as a statistical notion using the tools of probability theory. Thus in Shannon framework, the quantity of information which is contained in a string depends on its context. For example the string $'pane'$ contains a certain information when it is considered as a string coming from a given language. For example this world contains a certain amount of information in English; the same string $'pane'$ contains much less Shannon information when it is considered as a string coming from the Italian language because it is much more common (in fact it means "bread"). Roughly speaking, the Shannon information of a string σ is given by

$$I(\sigma) = \log_2 \frac{1}{p(\sigma)}$$

where $p(\sigma)$ denotes the probability of σ in a given context. The logarithm is taken in base two so that the information can be measured in binary digits (bits).[1]

If in a language the occurrences of the letters are independent of each other, the information carried by each letter is given by

$$I(a_i) = \log \frac{1}{p_i}$$

where p_i is the probability of the letter a_i. Then the average information of each letter is given by

$$H = \sum_i p_i \log \frac{1}{p_i} \tag{2}$$

Shannon called the quantity H entropy for its formal similarity with the Boltzmann's entropy. Now, we will discuss the reason of this similarity.

2.2.1 Shannon Versus Boltzmann Entropy

Consider a set of n particles and suppose that the phase space X of each particle is divided in L small cells. We can label any cell by a letter a_i of an ideal alphabet. Then the microstate of the system (with the accuracy given by the grain of our

[1] From now on, we will use the symbol "log" just for the base 2 logarithm "$\log_2$" and we will denote the natural logarithm by "ln".

partition of the phase pace) can be represented by a string of letters. Namely the string

$$a_1a_2a_3\dots$$

represents the state in which the particle 1 is in the cell $a_{1,}$ the particle 2 is in the cell $a_{2,}$ and so on. At this point it is possible to compute the Boltzmann entropy of our microstate; it is given by $\log_2 W$ where W is the set of all the configurations of our system. Clearly two configurations must be considered as belonging to the same macrostate if they have the same number of particles np_i in the cell a_j. Then the number W takes the following form:

$$W = \frac{n!}{\prod_{i=1}^{N} n_i!} \tag{3}$$

where $n := \sum_{i=1}^{L} n_i$ is the number of particles, n_i is the number of particles in the cell a_i.

We can give a nice form to the number W using the following approximation given by the Stirling formula:

$$\log n! = n \log n + n \log e + O(\log n) \tag{4}$$

Using this formula we get

$$\begin{aligned}
\log W &= \log \frac{n!}{\prod_{i=1}^{N} n_i!} = \log n! - \sum_{i=1}^{L} \log n_i! \\
&= n \log n + n \log e - \sum_{i=1}^{L} n_i \log n_i - e \sum_{i=1}^{L} n_i + O(\log n) \\
&= n \log n - \sum_{i=1}^{L} n_i \log n_i + O(\log n) \simeq n \log n - \sum_{i=1}^{L} n_i \log n_i
\end{aligned}$$

Now, setting $p_i = n_i/n$, we get the equation

$$S = \log W \cong n \sum_{i=1}^{L} p_i \log \frac{1}{p_i} \tag{5}$$

where the number p_i can be interpreted as the probability that a particle lies in the cell a_i. The formal similarity between (2) and (5) is evident. The main difference consists of the factor n. This is because the entropy (5) represents the "information" necessary to describe the full microsystem, while the (2) represents average the information of each letter. The full information of a typical message of n letter is given by nH. In this comparison we can identify a language $\mathcal{L}$ with a macrostate, provided that we define a language as the set of all messages (strings) of length n which contain exactly np_j times the letter a_j. A more appropriate definition of language (or, information source) will be given in the next section. Anyhow, in this heuristic description, the Shannon entropy could be defined as

$$H = \frac{\log W}{n}$$

Thus we have the following scheme:

$$\begin{array}{lll} & \textbf{Boltzmann} & \textbf{Shannon} \\ \mathcal{A} & \textit{set of the cell} & \textit{alphabet} \\ n & \textit{number of particles} & \textit{length of the message} \\ \mathcal{L} & \textit{macrostate} & \textit{language} \\ \sigma & \textit{microstate} & \textit{string} \\ W & \textit{microstates in } \mathcal{L} & \textit{strings in } \mathcal{L} \\ S(\mathcal{L}) & \log W & - \\ H(\mathcal{L}) & - & \frac{1}{n}\log W \end{array} \tag{6}$$

Notice that the factor n which distinguish $S(\mathcal{L})$ from $H(\mathcal{L})$ makes $S(\mathcal{L})$ an "extensive" measure while $H(\mathcal{L})$ is an average measure. For example, in a particles gas, $H(\mathcal{L})$ is the average Boltzmann entropy for particle. This point is source of many misunderstanding when we use the word "entropy" in an interdisciplinary context. In the following we will define different "kind" of entropies; in order to avoid these misleading facts, we will use the letters S and H as extensive and average "measures", respectively.

If we assume that the strings are very long, the statistical properties of a language can be studied letting $n \to \infty$. This fact, in our comparison correspond in taking the thermodynamic limit.

2.2.2 A Mathematical Definition of Shannon Entropy

In this section we will give the exact mathematical definition of Shannon entropy. We will define Shannon entropy in a new way, which emphasizes its similarity with Boltzmann entropy and which will be useful later when we will introduce the notion of Computable Information Content. We refer to [1] for more details on this point.

Let σ be a finite string of length n. We set

$$S_0(\sigma) = \log W(\sigma) \tag{7}$$

where $W(\sigma)$ is the number of strings which can be obtained by σ permuting its letters. Notice that

$$S_0(\sigma) \leq |\sigma| \cdot \log|\mathcal{A}|$$

where $\mathcal{A}$ is the alphabet of σ, namely the set of letters which appear in σ. Moreover $S_0(\sigma) = 0$ iff σ is constant.

We will call a parsing of σ a partition of σ in shorter strings w which we will call words. For example if

$$\sigma = \text{“}betubetube\text{”}$$

two parsings of σ are given by

$$\alpha_1 = (be, tube, tube)$$
$$\alpha_2 = (bet, u, bet, u, e)$$

Given a parsing α, we will denote by $W(\alpha)$ the number of strings which can be obtained permuting the words of α. In our example we have

$$W(\alpha_1) = 3;$$
$$W(\alpha_2) = \frac{5!}{2! \cdot 2! \cdot 1!} = 30$$

Given a parsing α, we will call dictionary of α the set $V(\alpha)$ of words w (with $|w| > 1$) which appear in α. In our example we have

$$V(\alpha_1) = \{be, tube\}$$
$$V(\alpha_2) = \{bet\}$$

We define the *combinatorial entropy of* σ as follows:

$$S_{\text{com}}(\sigma) = \min_{\alpha}\left[\log W(\alpha) + \sum_{w \in V(\alpha)} S_0(w)\right] \tag{8}$$

Notice that $S_{\text{com}}(\sigma) \le S_0(\sigma)$. In fact, if α contains only one-letter words, we have $\log W(\alpha) = \log W(\sigma)$ and $V(\alpha) = \emptyset$. Since $S_{\text{com}}(\sigma)$ is obtained taking the minimum over all the partitions, it turns out that $S_{\text{com}}(\sigma) \le S_0(\sigma)$.

Given any string σ we denote by $\alpha(\sigma)$ the partition which gives the minimum in (8) and set $V(\sigma) = V(\alpha(\sigma))$; if two partitions give the same minimum value, we take the one which corresponds to the smaller dictionary.

In our example, we have $\alpha(\sigma) = \alpha_1$, $V(\sigma) = V(\alpha_1) = \{be, tube\}$

$$S_{\text{com}}(\sigma) = \log W\big(\alpha(\sigma)\big) + S_0(be) + S_0(tube)$$
$$= \log 3 + \log 2 + \log 4! \cong 7.169$$

We define the *average combinatorial entropy* of σ in the following way:

$$H(\sigma) = \frac{S_{\text{com}}(\sigma)}{|\sigma|}$$

Now let ω be an infinite string and let $\omega^n \in \mathcal{A}^n$ be the finite string obtained taking the first n digits of ω and set

$$H(\omega) = \max_{n\to\infty} \lim H\big(\omega^n\big)$$

Since $H(\sigma) \le \frac{S_{\text{com}}(\sigma)}{|\sigma|} \le \log|\mathcal{A}|$ the maximum limit is finite.

Let $\mathcal{A}^{\mathbf{N}}$ be the set of all the infinite string in the alphabet $\mathcal{A}$, let μ be a probability measure on $\mathcal{A}^{\mathbf{N}}$ and let $T : \mathcal{A}^{\mathbf{N}} \to \mathcal{A}^{\mathbf{N}}$ be the shift map (defined as follows: $(T\sigma)_i = \sigma_{i+1}$). μ is called invariant (or stationary) if for every $A \subset \mathcal{A}^{\mathbf{N}}$, $\mu(A) = \mu(T^{-1}(A))$. If μ is invariant the couple $(\mathcal{A}^{\mathbf{N}}, \mu)$ is called an information source.

Now, we can give a definition of Shannon entropy which can be proved to be equivalent to the usual one.

Definition 1 The Shannon entropy of $(\mathcal{A}^{\mathbf{N}}, \mu)$ is defined by

$$h_\mu = \int_{\mathcal{A}^{\mathbf{N}}} H(\omega)\, d\mu$$

It is also possible to prove that for μ-almost every string $\omega \in \mathcal{A}^{\mathbf{N}}$ the limit

$$H(\omega) = \lim_{n\to\infty} H\big(\omega^n\big) \tag{9}$$

exists (see [1]). Clearly, if μ is ergodic, $H(\omega) = h_\mu$ for μ-almost every string $\omega \in \mathcal{A}^{\mathbf{N}}$.

2.3 Information Content

As we have seen the Shannon notion of information relies strongly on the notion of probability and this is very disappointing for the aims of this paper for the following reasons:

- we think that from an epistemological point of view the definition of probability presents many problems and does not help to clarify the nature of notion such as "entropy" and irreversibility
- we think that the notion of information is primitive and that the notion of probability should be derived by it
- our goal is to give definition which can be applied also to cellular automata and to computer simulations and this objects are strictly deterministic; thus the notion of probability should be avoided at least as primitive concept.

Moreover the Shannon information is context dependent and also this fact is in contrast with our aims. However, there are measures of information which depend intrinsically on the string and not on its probability within a given context. We give a general definition of information content which apply to many different contexts.

Definition 2 Let

$$U : \Sigma(\mathcal{A}) \to \Sigma\big(\{0, 1\}\big)$$

be an injective map and set

$$I_U(\sigma) = \big|U(\sigma)\big|$$

The function

$$I_U : \Sigma(\mathcal{A}) \to \mathbf{N}$$

is called information function relative to U if, for any infinitely long string ω for which the limit (9) exists, we have

$$H_U(\omega) \le H(\omega) \tag{10}$$

where

$$H_U(\omega) = \min_{n\to\infty} \lim \frac{I_U(\omega^n)}{n} \tag{11}$$

The number $I_U(\sigma)$ will be called U-information content of σ.

$U(\sigma)$ can be thought as a coding of the string σ in binary alphabet. (10) relates the information content to the Shannon entropy. Actually, (10) represents a kind of *optimality* of the coding U. The U-information content I_U allows to define the U-entropy, $H_U(\omega)$ of a single infinite string ω by (11). $H_U(\omega)$ represents the average information content of the string ω and it des not depend on any probability measure.

$H_U(\omega)$ allows to give an exact relation between the Shannon entropy h_μ and the information content I_U:

Theorem 3 *Let* $(\mathcal{A}^{\mathbf{N}}, \mu)$ *be an information source with entropy* h_μ*; then*

$$h_\mu = \int H_U(\omega)\, d\mu$$

Proof See [1]. □

Of course there are many functions U and I_U which satisfy Definition 2; for infinitely long strings they are equivalent in the sense of Theorem 3. However, they can be very different from each other when we consider finite strings, particularly when these strings are generated by a non-stationary information source. In the following we will discuss some of them.

2.4 Algorithmic Information Content

One of the most important of the information functions is the Algorithmic Information Content (*AIC*). In order to define it, it is necessary to define the notion of partial recursive function. We limit ourselves to give an intuitive idea which is very close to the formal definition. We can consider a partial recursive function as a computer C which takes a program P (namely a binary string) as an input, performs some computations and gives a string $\sigma = C(P)$, written in the given alphabet $\mathcal{A}$, as an output.

The *AIC* of a string σ is defined as the shortest binary program P which gives σ as its output, namely

$$I_{AIC}(\sigma, C) = \min\{|P| : C(P) = \sigma\}$$

In this case the function $U(\sigma)$ of Definition 2 is just the shortest program which produces σ. We require that our computer is a universal computing machine. Roughly speaking, a computing machine is called *universal* if it can simulate any other machine. In particular every real computer is a universal computing machine, provided that we assume that it has virtually infinite memory. For a precise definition see e.g. [7] or [5]. We have the following theorem due to Kolmogorov.

Theorem 4 *If C and C' are universal computing machine then*

$$\left|I_{AIC}(\sigma, C) - I_{AIC}\left(\sigma, C'\right)\right| \leq K\left(C, C'\right)$$

where $K(C, C')$ is a constant which depends only on C and C' but not on σ.

This theorem implies that the *AIC*-information content of σ with respect to C depends only on σ up to a fixed constant and then its asymptotic behavior does not depend on the choice of C. For this reason from now on we will write $I_{AIC}(\sigma)$ instead of $I_{AIC}(\sigma, C)$.

The shortest program which gives a string as its output is a sort of *ideal* encoding of the string. The information which is necessary to reconstruct the string is contained in the program.

Unfortunately this coding procedure cannot be performed by any algorithm (Chaitin Theorem).[2] This is a very deep statement and, in some sense, it is equivalent to the Turing halting problem or to the Gödel Incompleteness Theorem. Then the Algorithmic Information Content is not computable by any algorithm.

This fact has very deep consequences for our discussion of the arrow of time as we will see later. For the moment we can say that the *AIC* cannot be used as a reasonable physical quantity since it cannot be measured nor computed.

3 Computable Information Content

3.1 The Idea of Computable Information Content

Suppose that we have some lossless (reversible) coding procedure $Z : \Sigma(\mathcal{A}) \rightarrow \Sigma(\{0, 1\})$ such that from the coded string we can reconstruct the original string (for example the data compression algorithms that are in any personal computer). Since

[2] Actually, the Chaitin theorem states a weaker statement: a procedure (computer program) which states that a string σ of length n can be produced by a program shorter than n must be longer than n.

the coded string contains all the information that is necessary to reconstruct the original string, we can consider the length of the coded string as an approximate measure of the quantity of information that is contained in the original string. We can define the information content of the string σ as the length of the compressed string $Z(\sigma)$, namely

$$I_Z(\sigma) = \left|Z(\sigma)\right|$$

The advantage of using a Compression Algorithm lies in the fact that, in this way, the information content $I_Z(\sigma)$ turns out to be a computable function and hence it can be used in computer simulations and it can be considered as a measurable physical quantity.

We will list the properties which the notion of computable information content must satisfy for our purposes.

A function

$$I_{CIC} : \Sigma(\mathcal{A}) \to \mathbf{N}$$

is called Computable Information Content if it satisfies the following properties:

- (i) it an information function in the sense of Definition 2.
- (ii) it is computable.
- (iii) $I_{CIC}(\sigma) = M_{CIC}(\sigma) + S_{CIC}(\sigma)$ where $S_{CIC}(\sigma)$ satisfy the following properties:
  - (S1) $S_{CIC}(\sigma) \leq \log W(\sigma)$
  - (S2) $S_{CIC}(\sigma\tau) \leq S_{CIC}(\sigma) + S_{CIC}(\tau)$
  - (S3) $S_{CIC}(\sigma) \geq I_{AIC}(\sigma) - const.$

The properties (i) and (ii) are satisfied by I_Z defined by any reasonable compression algorithm Z. The important peculiarity of the Computable Information Content lies in the possibility of decomposing the global quantity of information in two parts:

- $S_{CIC}(\sigma)$ which we will call computable entropy of σ and represents the *disordered* part of the information.
- $M_{CIC}(\sigma)$ which we will call macroinformation of σ which represent the *regular* part of the information.

The properties (S1), (S2) and (S3) of the entropy are chosen in order to fit our intuitive idea of measure of *disorder.* For example, by (S1), we deduce that a constant string has null entropy: no disorder. (S2) can be interpreted in the following way: the "disorder" of two string is additive unless the two strings are correlated with each other in some way. (S3) gives a lower bound to the quantity of disordered information of a string. Since the best program P which produces σ must be random (in the sense of Chaitin [5]), our string is forced to contain a "quantity of disorder" at least equal to $|P| = I_{AIC}(\sigma)$. The negative constant $-const$ is necessary to make a consistent theory. For example, the entropy of a constant string c is 0, but $I_{AIC}(c) > 0$.

These properties makes the computable entropy come close to the Boltzmann definition of entropy and this fact is very relevant for the interpretation of physical phenomena.

3.2 The Definition of CIC

Functions I_{CIC} which satisfy (i), (ii), and (iii) exist. We will give an example of it.

Suppose to have a string σ and to have computed $\alpha(\sigma)$ and $V(\sigma)$ as in Sect. 2.2.2. If you want to send a message from which a receiver can reconstruct the string σ, a possible strategy is the following one:

- (i) you send to the receiver the dictionary $V(\sigma)$ and the number $n(w,\sigma)$ which specifies the number of times that the word w appears in the parsing $\alpha(\sigma)$.
- (ii) you send another number which select α among all the $S_{\mathrm{com}}(\sigma)$ possible strings which have the same dictionary $V(\sigma)$ and the same numbers $n(w,\sigma)$.

In this way, the information content of the full message is divided in two parts: part (i) which specifies the "macroscopic features" of the string and part (ii) which specifies only a number $s \approx S_{\mathrm{com}}(\sigma)$ which selects σ among all the strings with the same features.

The above procedure makes possible the following definition of macrostate:

Definition 5 Given two strings σ_1 and σ_2, we say that they belong to the same macrostate if

- $V(\sigma_1) = V(\sigma_2)$
- for every word $w \in V(\sigma_1)$, $n(w,\sigma_1) = n(w,\sigma_2)$

Roughly speaking, the string σ_1 and σ_2 belong to the same macrostate if they can be described in the same way, namely if they have the same dictionary and the same occurrence of each word in the dictionary. So they have the same macroinformation and the same entropy.

4 Information and Dynamics

The various notions of information are useful in many problems. Here we will consider their application to dynamical systems and will investigate the implications relative to the arrows of time which is the main point of this meeting.

We assume to have a dynamical systems consisting of many particles; using the same construction as Sect. 2.2.1 (see table (6)), we can apply the previous results. Our discretized phase space will be given by $\Omega = \mathcal{A}^n$ where $\mathcal{A}$ is the alphabet which corresponds to the graining of the phase space X of a single particle. Notice that the notion of I_{CIC} makes sense also when the number of particles is low, but in this case S_{CIC} will be close to 0 and the statistical behavior is not interesting (unless we decide to study the statistics making the average over long times). For simplicity we assume that time is discrete. The transition map $f : \Omega \to \Omega$ must be considered as the evolution map at time 1. We will consider both Hamiltonian dynamics (reversible dynamics) and dissipative systems. Of course, we may think of dissipative systems as subsystems of an Hamiltonian systems of which the microscopic dynamic variables have been ignored.

Also, we remark that we are not interested in taking the thermodynamic limit. This limit will simplify the equations but will hide some interesting notion such as the notion of macroinformation.

4.1 Physical Systems

If you consider the discretization of a continuous Hamiltonian system of weakly interacting particles, you obtain the usual description of statistical mechanics. In this case it is possible to identify the *CIC*-entropy with the physical entropy via the Boltzmann equation (1). The concrete computations are the same and any possible difference is of the order of $\log n$ where n is the number of particles.

However, if the particles interact strongly with each other and give a rich structure to the system, our description cannot be reduced to the traditional one, both for the presence of macroinformation (which might become relevant) and for a different notion macrostate. In particular the *CIC*-entropy of a state is not equal to a probability measure of the macrostate deduced by the Liouville theorem.

At this point it is interesting to stress the differences between this approach and the Brillouin point of view [4]. Also for him, the physical entropy is information, namely the information which the observer does not have; in particular he writes an equation like this

$$I_{\mathrm{tot}} = I_{\mathrm{obs}} + S \tag{12}$$

where I_{tot} is the total information, I_{obs} is the information of the observer, and S is the entropy. In the above equation, he considered I_{tot} constant since the system is reversible and he gets the following equality:

$$\Delta S = -\Delta I_{\mathrm{obs}}$$

which can be interpreted as follows: an increase of entropy ΔS equals the increase of ignorance of the observer. Thus he identifies the entropy as negative information and he can call the information of the observer "negentropy".

In our approach, (12) is replaced by the following one:

$$I_{CIC} = M_{CIC} + S_{CIC} \tag{13}$$

where the macroinformation might be related to the information of the observer, but in no way can be identified with it. In fact in (13) the observer does not play any role. Moreover, in a real system, in general both M_{CIC} and S_{CIC} grow with time. A more detailed description of this scenario is done in next sections.

4.2 Chaotic Reversible Systems

First of all let us consider "chaotic" Hamiltonian system. Most of the states of a system belong to the same macrostate Σ_0 and have the maximum *CIC*-entropy which is of the order of $\log|\Omega|$ and of course, they have minimum macroinformation.

Thus, for most of the initial condition the system enters the macrostate Σ_0 and it will stay there for a time of the order of Poincaré time. Σ_0 can be considered as the state which corresponds to the thermodynamical equilibrium. If you start with an initial condition with low *CIC*-entropy and high macroinformation then the *CIC*-entropy will increase until the maximum entropy while the macroinformation will decrease until the minimum which is a value very small if compared with the value of S_{CIC}.

In this sense time destroys information: namely, the macroinformation of the initial conditions is lost, in the sense that it cannot be recovered by a computable algorithm. In fact, if you have a "disordered" configuration, in general, there is not a computable procedure to know if it is derived by an "ordered" situation or not.

Thus we have obtained the traditional point of view of Boltzmann. The use of *CIC* makes possible to give a precise sense to the sentences:

information is destroyed

and

the disorder increases.

In fact, in this contest, they simply mean that M_{CIC} decreases and S_{CIC} increases.

4.3 Gradient-Like Systems

A discrete dynamical system (Ω, f) is called gradient-like if it admits a Lyapunov function, namely a function $\mathcal{V} : \Omega \to \mathbf{R}$, such that

- $\mathcal{V}(f(x)) \leq \mathcal{V}(x),\ x \in \Omega$
- $\mathcal{V}(f(x)) = \mathcal{V}(x) \Leftrightarrow f(x) = x$

In dissipative physical system the Lyapunov function usually corresponds to the energy. Gradient-like systems evolve until reaching a stable equilibrium configuration x_0. Usually these configurations have low Information Content. Thus the evolution make to decrease both the entropy and the macroinformation; there is an absolute decrease of I_{CIC}. The system loses its memory and any kind of information is destroyed. This is obvious since the transition map f is not injective and different initial conditions lead to the same final configuration. If we embed this system in an invertible system, we will get a chaotic system and the consideration of the previous section apply.

4.4 Self-organizing Systems

If, in a physical dissipative system, there is an input of energy from the outside, in general, stable equilibrium configurations cannot exist. In this case many phenomena may occur; stable periodic orbits, stable tori or even strange attractors. Sometimes, very interesting spatial structures may appear. Analogous phenomena occur in non-reversible cellular automata. The most famous of them is the Conway game of life in which a lot of intriguing shapes appear in spite of the simplicity of the transition map.

From the point of view of Information theory, these are the systems which make the macroinformation to increase. If we start from an initial data with a low macroinformation content, the macroinformation will increase until reaching a limit value. If the system is infinite the macroinformation will increase for ever. For example you may think of the game of life in an infinite grid with initial conditions having only a finite number of black cells (and thus you start with an initial condition which has finite information).

As in the case of gradient-like system, we may embed these systems in a reversible system. In this situation, we will have also an increase of the entropy. Thus the two main arrows of time, described in the introduction, will be present. We believe that a sufficiently large system, in which many nonlinear interactions play a role, is very likely to present such a behavior.

From a general and qualitative point of view, these systems represent a good model for large natural system in which the appearance of complex structures occurs.

5 An Exorcism of the Maxwell Demon

The Maxwell demon acts on a pipe which joins two containers of particles, A and B. This pipe has a gate which can be kept open or closed by a demon. He opens the gate when he sees a particle coming from A and he closes it when he sees a particle coming from B. In this way, at some point all the particles will be in the container B (actually the "original" Maxwell demon made a distinction between slow and fast particles but the argument is the same).

At the end of this operation the entropy of the system of particles will be reduced since all the particles are in one container, namely in the container B.

This seems a violation of the second law of thermodynamics. Where is the catch?

Many different explanations have been proposed to exorcise this demon and to save the second law. We will give a brief sketch of some of them.

5.1 Szilard

The first important contribution to this problem was given by Leo Szilard in 1929. He thought that the measurement performed by the demon cause an increase of entropy in the environment which compensate the decrease of entropy in the containers. He was rather vague about the mechanism responsible of the entropy increase and the questions relative to this point were left open.

5.2 Brillouin and Gabor

The next important contribution came by Léon Brillouin (1956) and Dennis Gabor who saw in the Indetermination Principle of Quantum Mechanics the key point in the exorcism of the demon (see e.g. [4]). When the demon performs its measurement, he needs to send an energetic beam of light on the particle and this fact has an energetic cost which increases the entropy of the environment. We think that this explanation is wrong for at least two reasons: (1) you may imagine that this experiment is performed with big balls and in this case the perturbations of the photons are not relevant, or, to say it in a different way, the explanation should be independent on the scale, while every explanation which includes $\hbar$ depends on the scale; (2) Charles Bennet made a model in which the observation of the demon is independent of the presence of wuantum phenomena. Quantum mechanics has nothing to do with the Maxwell demon.

5.3 Landauer

A big step toward the right answer was made by Rolf Landauer (1961) who studied the constrains imposed to computation by physical laws (see e.g. [6]). He identified some operations which he "called" logically irreversible. These logically irreversible operations are also physically irreversible since they make the entropy of the environment to increase. One of them is the erasing of the memory of the computing machine, whatever its internal nature is. Clearly if you erase the memory you cannot make a time reversal and come back to the initial condition. This fact implies an increase of the entropy of the environment. The entropy balance is easy to calculate if you take in account the distinctions between *AIC*, *CIC* and *CIC*-entropy. If you assume that the computer plus the environment are ruled by reversible equations, then the *AIC* is preserved. However, this information is not contained in the computer after that its memory has been cleared. Thus, this information has been transferred to the environment, and since we may assume that it is contained in it in a random way, this information makes the *CIC*-entropy of the environment increase.

5.4 Bennet

The final step was made by Charles Bennet (1982) who gave the following explanation (see e.g. [3]). The demon needs a buffer to store the information that a particle is coming from A or from B in order to keep the door open or closed. Afterwards, he must store the analogous information relative to the next ball in a buffer. Thus he has two choices: or he uses some extra memory or he clears the buffer. According to Bennet, in order to perform a cycle, at some point, the demon needs to clear his memory, and this fact will make the entropy of the environment to increase.

5.5 Our Point of View

I think that this explanation is essentially right even if it presents some weak points which have been pointed out by the conference of David Albert. More or less Albert says the following: first of all, you do not need to consider a cycle; this has nothing to do with the second law which just states that the entropy of an isolated system does not decrease. Now, assume that the containers and the demon D constitute an isolated system and that the demon does not erase its memory. At the end of the process, the system $A + B + D$ will have a lower entropy, at least if you define the entropy as the measure of the final macrostate in the phase space. In fact it is not difficult to imagine a Hamiltonian for which this is true.

However, the Bennet point of view can be easily saved using the notion of *CIC*-entropy rather than a probability measure in the phase space. In fact, every thing becomes clear if we identify the physical entropy with the *CIC*-entropy. When the memory of the demon has stored all the past history of this process, it contains a string with a large content of disordered information (*CIC*-entropy). It is exactly the information which you need to reverse the process. Thus if you make a *CIC*-entropy balance, you discover that the *CIC*-entropy is the same. Thus any contradiction disappears.

Moreover, if you assume that our system is not is isolated, this description can say more. When you erase the memory, the *AIC* contained in the memory of the demon-computer will be discharged in the environment (since the system $A + B + D +$ [environment] is reversible). In this operation the global *CIC*-entropy will increase, since you cannot find an algorithm which is able to recover the information spread in the environment.

6 Conclusions

We think that the right description of the origin of irreversibility, complexity and the arrow of time lies in a good notion of "information content". A good notion must be independent of the notion of probability for the reasons described in Sect. 2.3.

Moreover, we think that it is very important to distinguish two different meanings of the notion of information:

- the general *abstract* notion of information (such as the *AIC*) which in reversible system is a constant of motion and exists only in the mind of God (but not in the mind of the demons, at least, if they are submitted to the laws of our universe).

- a computable notion of information (such as the *CIC*) which is related to physical quantities; it changes in time and can be used to describe the observed phenomena.

The distinctions between these two notions is marked by the Turing halting theorem which, we think, is one of the deepest theorems discovered in last century and whose consequences are not yet all completely understood.

Once, we have agreed to consider the *CIC* (or any other "epistemologically" equivalent notion of information) as the relevant physical quantity, it is important to have a mathematical method to separate the *CIC* in two different components:

- the entropy which corresponds to the old idea of "measure of disorder". From a physical point of view this information cannot be used to make exact deterministic predictions. It is the information dispersed in the chaos and it cannot be recovered without a violation of the Turing halting theorem.
- the macroinformation which is related to physical measurable quantities and can be used to make predictions. Moreover, the macroinformation is strictly related to various indicators of complexity.

Thus, in information theory, we have the distinction between macroinformation and *CIC*-entropy. This is similar to the distinction, in thermodynamics, between free energy $F = E - TS$ and bad energy. The *CIC*-entropy cannot be used to make predictions, while TS cannot be used to perform any work. However, it is very important to underline that, in isolated systems, free energy and macroinformation behave in a quite different way: free energy always decreases, while macroinformation might increase. The development of life, in all its forms, determines a decrease of free energy and an increase of macroinformation. Probably there is a deep mathematical relation between the evolution of these two quantities. The study of the interplay between macroinformation, entropy and the other physical quantities is a good way to investigate the origin and the evolution of complex structures.

In Sect. 3.2, we have proposed a mathematical model which makes a distinction between macroinformation and entropy. This is not the only possible model and probably is not the best. However, it seems to me that this is a good direction for investigating this kind of problems.

References

1. Benci, V., Menconi, G.: Some remarks on the definition of Boltzmann, Shannon and Kolmogorov entropy. Milan J. Math. **73**, 187–209 (2005)
2. Benci, V., Bonanno, C., Galatolo, S., Menconi, G., Virgilio, M.: Dynamical systems and computable information. Discrete Contin. Dyn. Syst. **4**, 935–960 (2004)
3. Bennet, C.: Demons, machines, and the second law. Sci. Am. **257**(5) 108–116 (1987)
4. Brillouin, L.: Scientific Uncertainty, and Information. Academic Press, New York (1964)

5. Chaitin, G.J.: Information, Randomness and Incompleteness, Papers on Algoritmic Information Theory. World Scientific, Singapore (1987)
6. Landauer, R.: Fundamental physical limitations of the computational process. Ann. N.Y. Acad. Sci. **426**, 161–170 (1985)
7. Li, M., Vitanyi, P.: An Introduction to Kolmogorov Complexity and Its Applications. Springer, Berlin (1993)