

# THE BOLTZMANN ENTROPY AND RANDOMNESS TESTS

PETER GÁCS

**ABSTRACT.** In the context of the dynamical systems of classical mechanics, we introduce two new notions called “algorithmic fine-grain and coarse-grain entropy”. The fine-grain algorithmic entropy is, on the one hand, a simple variant of the randomness tests of Martin-Löf (and others) and is, on the other hand, a connecting link between description (Kolmogorov) complexity, Gibbs entropy and Boltzmann entropy.

The coarse-grain entropy is a slight correction to Boltzmann’s coarse-grain entropy. Its main advantage is its less partition-dependence, due to the fact that algorithmic entropies for different coarse-grainings are approximations of one and the same fine-grain entropy. It has the desirable properties of Boltzmann entropy in a wider range of systems, including those of interest in the “thermodynamics of computation”. It also helps explaining the behavior of some unusual spin systems arising from cellular automata.

## 1. INTRODUCTION

The present paper is an elaboration of the ideas proposed first in [6]. In its present form, the intended audience is mainly computer-scientists interested in the relation of description complexity to thermodynamics. This determines the form of exposition somewhat: Section 8 gives a (very brief) introduction to some thermodynamical concepts and problems that physicists probably do not need. But we hope that the main technical content of the paper will ultimately be also considered a contribution to physics and dynamical system theory: mainly in providing some tools to extend thermodynamical reasoning to some new classes of systems.

The concept of computability is assumed to be known; for computability in continuous spaces, we refer to [5]. We also assume some familiarity with the notions of computability and description (Kolmogorov) complexity. A good survey of the latter is [10].

**1.1. Coarse-graining.** The physical model considered in the present paper is that of classical mechanics: it is defined by a phase space  $\Omega$  and a dynamics  $U^t$  giving the point  $U^t\omega$  on an orbit at time  $t$ , where the transformation  $U^t$  preserves the volume

---

*Date:* December 7, 2023.

*2020 Mathematics Subject Classification.* 60A99, 68Q30, 80A05, 82B03, 82C03.

*Key words and phrases.* entropy, description complexity, Kolmogorov complexity, algorithmic entropy, randomness test.

Partially supported by NSF grant CCR-920484 and by NWO through NFI Project ALADDIN under Contract number NF 62-376 and Scientific Visitor Award B 62-394. Part of the work was done while the author was visiting Rutgers University and DIMACS. .

measure  $L$  (like in Liouville's Theorem, see Section 8). In case of a container of “ideal” gas consisting of  $n$  simple molecules, the state space is the  $6n$ -dimensional Euclidean space given by the positions and moments of each molecule.

A thermodynamical system is characterized by a relatively small number of parameters (functions of the state) called *macroscopic* parameters  $u_1, \dots, u_m$ . The canonical example is a certain quantity of gas in a container, with the macroscopic parameters of volume, temperature, energy and pressure. A microscopic state is the completely specified state. A macroscopic state is determined by the macroscopic parameters and it determines only the (by far) most probable behavior and properties of the system and only when the system is in equilibrium.

We will assume that each macroscopical parameter  $u_i$  takes only a finite number of values: a macroscopical parameter that is originally a real number will only be taken to a certain precision agreed in advance. This gives a partition of the state space into *cells*

$$\Omega = \bigcup_u \Gamma_u$$

corresponding to *macrostates* where  $u = (u_1, \dots, u_m)$ . Finer and finer partitions  $\mathcal{P}_i$  of the phase space can be introduced by adding more macroscopic parameters and more precision. The partition interpretation of a macrostate is called *coarse-graining*.

Coarse-graining solves the *paradox of irreversibility*: in a mechanical system, any evolution seems just as possible as the corresponding reverse evolution, and at the same time, the world seems to be full of irreversible phenomena (examples omitted). To reconcile the two pictures when we say that a certain transformation from state  $a$  to state  $b$  is reversible this statement refers to *macrostates*; what is meant is that the reverse transformation exists for *most* microstates within the macrostate  $\Gamma_b$ , as measured by volume. Now asymmetries are quite possible. It is easy to imagine macrostates  $a, b$  such that in a time unit, 99.99% of the elements of  $\Gamma_a$  end up in  $\Gamma_b$  but only 0.01% of the elements of  $\Gamma_b$  end up in  $\Gamma_a$ . This is exactly what happens with the gas, except that the percentages are much more extreme.

Another possible interpretation of a macrostate is as a certain distribution  $\nu$  over microscopic states. It is possible (but not always done) to require  $\nu$  to be a probability distribution, given by a density function  $p(\omega)$  with  $\int p(\omega)L(d\omega) = 1$ . Gibbs called such a distribution an *ensemble*. The ensemble  $p_\Gamma(\omega)$  corresponding to a macrostate  $\Gamma$  is defined as  $1/L(\Gamma)$  for  $\omega \in \Gamma$  and 0 elsewhere.

**1.2. Coarse-grained entropy.** The Boltzmann entropy of a cell  $\Gamma$  is  $\log L(\Gamma)$ . This quantity depends on the choice of the partition: indeed, another digit of precision will decrease it by about  $\log 10$ . In the typical classical examples, this difference is negligible compared to the volumes in question. In nontypical systems, partition dependence can lead to paradoxes: we will show that the Maxwell demon paradox is one of these.

We propose a new quantity

$$S(\Gamma) = H(\Gamma) + \log L(\Gamma) \tag{1.1}$$

called *coarse-grained algorithmic entropy* to replace Boltzmann entropy. Here,  $H(\Gamma)$  is the description (Kolmogorov) complexity of the cell  $\Gamma$  (to be defined later). This quantity is closely related to the one introduced by Zurek, and we will return to their relation. The term  $H(\Gamma)$  is typically negligible compared to the Boltzmann entropy  $\log L(\Gamma)$ ; however, the new entropy is less partition-dependent since it is an approximation to a certain partition-independent quantity  $H(\omega) = H_L(\omega)$  which we call (fine-grained) algorithmic entropy.

The paper explores the basic properties of fine-grained and coarse-grained algorithmic entropy. Description complexity itself will be shown to be a special case algorithmic fine-grained entropy. The latter is a simple variant of the randomness tests of Martin-Löf (and others). Its integral over a Gibbs ensemble is close to the so-called Gibbs entropy. It possesses a simple conservation property that, together with coarse-grained algorithmic entropy, is helpful in explaining the Maxwell demon paradox as well as some other physical situations and models not handled well by Boltzmann entropy. Explanation of the entropy increase property for the new coarse-grained entropy does not rely on the huge volume differences in cells the way it does in the case of Boltzmann's entropy.

We hope that the new quantity extends the possibilities of the kind of reasoning involving entropy to a wider range of systems (in particular, large computers in which it is not clear whether the whole memory content should be considered macroscopic or microscopic information). Here is an overview the content of the paper.

Section 2 defines description complexity  $H(x)$  (the self-delimiting version) of a finite object  $x$  and shows its main properties.

Section 3 introduces randomness tests. The algorithmic (fine-grained) entropy  $H_\mu(\omega)$  of a state  $\omega$  of the system with respect to an underlying measure  $\mu$  (typically, the volume measure over a finite-dimensional space or the counting measure over a discrete space) is defined as a common generalization of the description complexity  $H(x)$  and the randomness test of Martin-Löf (as modified by Levin). It is the negative logarithm of the maximal (to within an additive constant) lower semicomputable function  $f(x) \geq 0$  with the property that

$$\int f(\omega)\mu(d\omega) \leq 1.$$

It is known that the description complexity  $H(x)$  is within constant distance from  $H_\#(x)$  where  $\#$  is the counting measure. The paper [12] has dealt with some statistical applications of this quantity.

Section 4 discusses an addition property for algorithmic entropy, which is the generalization of the information-theoretic addition property of complexity.

Section 5 expresses  $H_\mu(\omega)$  in terms of complexity, with the help of a formula

$$H(\Gamma|\mu) + \log \mu(\Gamma)$$

similar to (1.1).

In Section 6, we will prove a nondecrease property for algorithmic entropy. We give some examples that are better handled by our definition, partly due to the

fact that our entropy automatically satisfies the condition we call here the Noncompensation Condition. These examples do not belong to mainstream statistical mechanics, but are borderline cases that may become more important with further miniaturization of computers (“mesoscopic systems”?). It would also be interesting to see other chaotic systems in which the extension of the notion of entropy increase is useful.

Our definition resembles Zurek’s entropy defined in [13] but there are important differences. It is defined more generally (not just for systems consisting of a classical and a “demon” part). Its properties are proved with fewer assumptions. It connects randomness tests, fine-grain and coarse-grain entropy (these notions are not distinguished by Zurek) and complexity.

Heuristic arguments (and, of course, experience) show that in a nonequilibrium system, Boltzmann entropy can be expected to increase strongly. These arguments consist of an easy part, which we will call Noncompensation Condition and a difficult part, the Weak Mixing Condition which, however plausible, can be rigorously proved only in some special cases. With our algorithmic Boltzmann entropy, the Noncompensation Condition comes free. The nontrivial, mixing part of the argument remains just as difficult and conditional as for Boltzmann entropy but our framework allows at least to formulate the mixing property in a sharp way.

In Section 7, we will show an additivity property and will discuss its relation to the so-called Landauer thesis which says that erasing a bit of information from, say, computer memory (even if it is done indirectly) requires the dissipation of a certain minimal amount of heat. Maxwell’s demon paradox will also be alluded to in this context.

For the definitions of the various notions of computability in a continuous space (like the phase space of a system of particles), we refer to the lecture notes [5]. The concepts involved are quite technical, but it is possible to rely on an intuitive understanding of computability for the most part.

**1.3. Notation.** The sets of natural numbers, integers, rational numbers, real numbers will be denoted respectively by  $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ . The set of nonnegative real numbers will be denoted by  $\mathbb{R}_+$ . We use  $\wedge$  and  $\vee$  to denote min and max. In what follows  $\log$  is the logarithm to base 2. The quantities

$$\lfloor x \rfloor, \lceil x \rceil$$

are the largest integer  $\leq x$  and the smallest integer  $\geq x$  respectively. For a set  $E$ , we denote the function taking the value 1 in  $E$  and 0 outside by  $1_E(\omega)$ , and call it the *indicator function* of  $E$ . Borrowing from [11], for a two-variable function  $f(x, y)$  and a measure  $\mu$ , we will sometimes use the notation

$$\mu f = \int f(x) \mu(dx), \quad \mu^y f(x, y) = \int f(x, y) \mu(dy).$$

Let  $\mathbb{B} = \{0, 1\}$ . We introduce a special object  $\Lambda$  called the “empty string”, and define, as usual, the set of finite binary strings as

$$\mathbb{B}^* = \{\Lambda\} \cup \bigcup_{n=1}^{\infty} \mathbb{B}^n.$$

The set of infinite binary strings will be denoted by  $\mathbb{B}^\omega$ . For a string  $s \in \mathbb{B}^*$ , let

$$l(s)$$

denote its length.

## 2. COMPLEXITY

Given some computer  $F$ , let  $H_F(y)$  be the length of the shortest program (measured in bits) that causes  $F$  to output string  $y$ . We will require the program to be self-delimiting: no endmarker is allowed. The machine-dependence of this concept is limited since there is a machine  $G$  on which the function  $H_G(y)$  is optimal within an additive constant: for every other machine  $F$  there is a constant  $c_F$  such that for all  $x$  we have  $H_G(x) \leq H_F(x) + c_F$ . The notation

$$f(x) \stackrel{+}{<} g(x).$$

will mean that for some constant  $c$  and for all  $x$  we have  $f(x) \leq g(x) + c$ . The notation  $\stackrel{*}{<}$  will mean the same with a multiplicative constant. The notation  $\stackrel{+}{=}$  means that both  $\stackrel{+}{<}$  and  $\stackrel{+}{>}$  hold. With this notation, the invariance theorem’s formula can be written as  $H_G(x) \stackrel{+}{<} H_F(x)$ . The function  $H(x) = H_G(x)$  is called the *complexity* of the natural number  $x$  conditional on the information  $y$ . (This is the modified version of the Kolmogorov-Solomonoff complexity invented by Levin and Chaitin. The book [10] uses the notation  $K(x)$  for the same quantity.) The conditional complexity  $H(x | y)$  is defined by leaving  $y$  everywhere as a parameter. The elementary properties of the complexity function are discussed in several expositions, and we will not dwell on them. Let us just mention that  $H(n) \stackrel{+}{<} \log n + 2 \log \log n$ , and

$$H(f(x) | y) \stackrel{+}{<} H(x | g(y)) \stackrel{+}{<} H(x). \quad (2.1)$$

for any computable functions  $f, g$ .

The function  $H(x)$  is not computable, but it has a certain weaker property. Let  $\mathbf{Q}$  be the set of rational numbers. We call a function  $f(x)$  from natural numbers to real numbers *upper semicomputable* if there is a computable sequence  $f_n(x)$  with rational values such that  $f_n(x) \searrow f(x)$ . It is easy to see that  $H(x | y)$  is upper semicomputable. Thus, we can compute arbitrarily exact upper bounds on  $H(x)$  but we will not know how close we are to the limit. It is known that no nontrivial lower bounds can be computed for  $H(x)$ . However, there are strong statistical lower bound results. There is a single property from which these can be derived:

$$\sum_x 2^{-H(x|y)} \leq 1.$$

It is easy to prove this inequality by considering programs to our standard self-delimiting machine that are obtained by coin-tossing. We will need the following important theorem of Levin.

**Theorem 1** (Coding). *Let us consider the class of lower semicomputable functions  $f(x, y)$  with the property that  $\sum_x f(x, y) \leq 1$ . The function  $2^{-H(x|y)}$  is an element of this class and is maximal in it, to within a multiplicative constant. In other words, for each element  $f$  of this class, we have  $2^{-H(x|y)} >^* f(x, y)$ .*

### 3. RANDOMNESS

We assume that the reader is familiar with the treatment of randomness tests in [5]. As there, following [2], we define a computable metric space as a tuple  $\Omega = (\Omega, d, D, \alpha)$  where  $(\Omega, d)$  is a metric space, with a countable dense subset  $D$  and an enumeration  $\alpha$  of  $D$ . It is assumed that the real function  $d(\alpha(v), \alpha(w))$  is computable. As  $x$  runs through elements of  $D$  and  $r$  through positive rational numbers, we obtain the enumeration of a countable basis

$$\beta = \{ B(x, r) : x \in D, r \in \mathbb{Q} \}$$

(of balls or radius  $r$  and center  $x$ ) of  $\mathbf{X}$ , giving rise to a constructive topological space  $\tilde{\Omega}$ . Whenever we will have a transformation group  $U^t$  we will assume that it as well as the invariant volume measure  $L$  are computable.

**3.1. Randomness tests.** A nonnegative lower semicomputable function  $f_\mu(\omega, y)$  over the space  $\mathcal{M}(\Omega) \times \Omega \times Y$  is a (parametrized) *test of randomness* or, shortly, test with respect to a parameter  $y$ , if for all  $\mu, y$  we have

$$\int f_\mu(\omega, y) \mu(d\omega) \leq 1.$$

Here is some motivation for the case of probability measures. For a moment, forget the parameter  $y$ . If a certain casino claims that it draws elements from  $\Omega$  according to the distribution  $\mu$  then it must accept the following deal:

- (1) I prove that  $f_\mu(\omega)$  is a test of randomness;
- (2) I offer two dollars for a game, and ask for  $\omega$ ;
- (3) my payoff is  $f_\mu(\omega)$ .

If the casino owner indeed draws according to  $\mu$  then the test property implies that my expected payoff is at most a dollar, so she even makes more than a dollar of profit on average. My strategy is to try to find some nonrandomness in  $\omega$ , (without seeing  $\omega$  first) by making an acceptable test function  $f_\mu(\omega)$  as large as possible. The following theorem is proved in [5], relying also on [7]:

**Theorem 2.** *There is a universal test, that is a test  $\mathbf{t}_\mu(\omega)$  with the property that for every other test  $f_\mu(\omega)$  there is a constant  $c_f > 0$  with  $c_f f_\mu(\omega) \leq \mathbf{t}_\mu(\omega)$ .*

The theorem can easily be generalized to include an extra parameter  $y$  in  $f_\mu(\omega, y)$ . Then it says that among all randomness tests, there is a certain one, denoted by  $\mathbf{t}_\mu(\omega \mid y)$  and called a *universal test*, that takes only values of the form  $2^n$  for

(possibly negative) integers  $n$  and is maximal to within a multiplicative constant: it has the property that for all other tests  $f_\mu(\omega, y)$ , we have

$$f_\mu(\omega, y) \stackrel{*}{<} \mathbf{t}_\mu(\omega | y).$$

This test is a close relative of the universal tests of Martin-Löf and Levin and can be used as a criterion of randomness. (The property that  $\mathbf{t}$  takes only values  $2^n$  is only for convenience.)

In the gambling interpretation, the universal test is, in some sense, optimal. Its existence is surprising: for example, if  $\omega$  is supposed to be a sequence of coin tosses then I could make my test function large for those  $\omega$ 's in which the frequency of 1's is at least 60% (since their probability is small). This way, I would profit from a certain kind of cheating the casino might attempt. Alternatively, I can make  $f$  larger on many other improbable sets of sequences. The universal test anticipates and combines all these strategies.

The *algorithmic (fine-grained) entropy* of  $\omega$  with respect to  $\mu$  is defined as

$$H_\mu(\omega | y) = -\log \mathbf{t}_\mu(\omega | y).$$

We will delete  $\mu$  from the subscript when it is obvious from the context.

$H$  can take arbitrarily large negative values, even  $-\infty$ . In other words, an object can be infinitely nonrandom, though the measure of such objects has probability 0. For a finite measure  $\mu$ , the function  $H_\mu(\omega)$  is bounded from above. For infinite measures, it can also take arbitrarily large positive values; but it will never be  $\infty$ .

Let  $H_\mu(\omega) = H_\mu(\omega | 0)$  where  $Y$  is chosen as the one-element set  $\{0\}$ . If both  $\Omega$  and  $Y$  have measures  $\mu, \nu$  then we define

$$H_{\mu,\nu}(\omega, y) = H_{\mu \times \nu}((\omega, y))$$

where  $\mu \times \nu$  is the product measure.

To give some idea of how  $H_\mu(\omega)$  depends on  $\omega$  and  $\mu$  we give an upper bound. For  $\omega \in \Omega_n \setminus \Omega_{n-1}$ , let  $m(\omega) = \mu(\Omega_n)$ . Then it is easy to show that

$$H_\mu(\omega) \stackrel{+}{<} \log(m(\omega) + 1) + 2 \log(\log(m(\omega) + 1) + 1).$$

Here are some additional easy properties of  $H(\omega)$ .

### Proposition 3.1.

$$\mu\{\omega : H_\mu(\omega) < m\} < 2^m \quad (-\infty < m < \infty), \tag{3.1}$$

$$H_\nu(y | \mu) \stackrel{+}{<} -\log \int 2^{-H_{\mu,\nu}(\omega, y)} \mu(d\omega). \tag{3.2}$$

If  $f$  is a computable function then

$$H_\mu(\omega | y) \stackrel{+}{<} H_\mu(\omega | f(y)).$$

The first inequality states that  $H(\omega)$  is large only with small probability. The second one is needed for the addition theorem, stated later.

For the volume measure  $L$  in the phase space of a dynamical system, we define  $H(\omega) = H_L(\omega)$  as the fine-grained algorithmic entropy of a state  $\omega$ . Let

$$E_m = \{ \omega : H(\omega) < m \}.$$

In (3.1), we have shown  $L(E_m) < 2^m$  for all  $m$ . For finite space volume  $L(\Omega)$ , this implies that

$$L(E_{\log L(\Omega)-m})/L(\Omega) < 2^{-m}, \quad (3.3)$$

that is, the proportional volume of the set of those points where algorithmic (fine-grained) entropy is not close to its maximum is very small.

Let us note that the Coding Theorem 1 says

$$H(x | y) \stackrel{+}{=} H_{\#}(x | y)$$

where  $\#$  is the counting measure.

The following statement is similar to the universal uniform test theorem, but for arbitrary measures:

**Proposition 3.2.** *Let  $(x, y, \nu) \mapsto f_\nu(x, y)$  be a nonnegative lower semicomputable function with  $F_\nu(x) = \log \nu^y f_\nu(x, y)$ . Then for all  $x$  with  $F_\nu(x) > -\infty$  we have*

$$H_\nu(y | x, \lfloor F_\nu(x) \rfloor) \stackrel{+}{<} -\log f_\nu(x, y) + F_\nu(x).$$

The proof is in [5].

#### 4. ADDITIVITY

**4.1. Additivity of information.** Complexity has the following additivity property, due in various forms to Kolmogorov, Levin, Gacs, Chaitin:

$$H(x, y) \stackrel{+}{=} H(y) + H(x | y, H(y)).$$

This property, in an appropriate form, generalizes for algorithmic entropy, as shown (in a different and slightly less general context) by Vovk and Vyugin:

**Theorem 3** (Addition).

$$H_{\mu, \nu}(x, y) \stackrel{+}{=} H_\nu(y | \mu) + H_\mu(x | y, H_\nu(y | \mu), \nu).$$

Notice that this theorem differs from the original addition theorem only by having  $\mu, \nu$  everywhere in the appropriate conditions and subscripts. (When  $\mu$  is in the subscript it does not have to be in the condition.) The proof is in [5]. A simple corollary is

$$H_{\mu, \nu}(x, y) \stackrel{+}{<} H_\nu(y) + H_\mu(x | y).$$

The function  $H_\mu(\omega)$  behaves quite differently for different kinds of measures  $\mu$ . For example, inequality (2.1) implies

$$H(y) \stackrel{+}{<} H(x, y).$$

In contrast, if  $\mu$  is a probability measure then

$$H_\nu(y) \stackrel{+}{>} H_{\mu, \nu}(\omega, y).$$

This comes from the fact that  $2^{-H_\nu(y)}$  is a test for  $\mu \times \nu$ . Due to these considerations, the following relation does not follow easily from the addition property (it would be if  $H_\mu(\omega) \stackrel{+}{<} H_{\mu \times \#}(\omega, z)$  would hold): For  $z$  from a countable set  $Y$ , we have

$$H_\mu(\omega) \stackrel{+}{<} H_\mu(\omega | z) + H(z). \quad (4.1)$$

This relation can be generalized slightly:

**Lemma 4.1.** *For a computable function  $f(y, z)$  on a countable set  $Y$ , we have*

$$H_\mu(\omega | y) \stackrel{+}{<} H_\mu(\omega | f(y, z)) + H(z).$$

The proof is in [5].

Taking  $f_\mu(x, y) = 1$  in Proposition 3.2 gives the inequality

$$H_\mu(x | \lfloor \log \mu(X) \rfloor) \stackrel{+}{<} \log \mu(X),$$

with a physical meaning when  $\mu$  is the phase space measure. Using (4.1), this implies

$$H_\mu(x) \stackrel{+}{<} \log \mu(X) + H(\lfloor \log \mu(X) \rfloor). \quad (4.2)$$

The symmetric quantity  $I(x, y) = H(x) + H(y) - H(x, y)$  is called the *mutual information* between the objects  $x, y$ . It generalizes to

$$I_{\mu, \nu}(x, y) = H_\mu(x | \nu) + H_\nu(y | \mu) - H_{\mu, \nu}(x, y).$$

(with respect to the measures  $\mu, \nu$ ). A related quantity is

$$I_\mu(y : x) = H_\mu(x) - H_\mu(x | y),$$

the *information* that  $y$  carries about  $x$  with respect to  $\mu$ . When  $\mu$  is obvious from the context we omit it from the subscript. The complexity addition theorem is equivalent to the following relation between the two kinds of information:

$$I(x, y) \stackrel{+}{=} I(x, H(x) : y)$$

which also generalizes to algorithmic entropy. The right-hand side can be interpreted as the information that the pair  $(x, H(x))$  carries about  $y$ . Thus,  $I(x : y)$  is itself not symmetric but is “almost” symmetric since  $I(x, H(x) : y)$  is equal to a symmetric quantity. More generally,

$$I_{\mu, \nu}(x, y) \stackrel{+}{=} I_\mu(x, H_\mu(x | \nu) : y | \nu).$$

On a countable set  $Y$ , Lemma 4.1 implies

$$I_\mu(y : x) \stackrel{+}{<} H(y). \quad (4.3)$$

This says that an object cannot carry more information about a string than its own complexity.

## 5. RANDOMNESS, COMPLEXITY, COARSE-GRAINED ENTROPY

**5.1. Cells.** As pointed out in [5], it is not convenient to define a measure  $\mu$  constructively starting from  $\mu(\Gamma)$  for open cells  $\Gamma$ . The reason is that no matter how we fix  $\Gamma$ , the function  $\mu \mapsto \mu(\Gamma)$  is typically not computable. It is better to work with bounded computable functions, since for such a function  $f$ , the correspondence  $\mu \mapsto \mu f$  is computable. We therefore introduce a certain fixed, enumerated sequence of (Lipschitz) functions that will be used frequently. Let  $\mathcal{F}_0$  be the set of functions of the form  $g_{u,r,1/n}$  where  $u \in D$ ,  $r \in \mathbb{Q}$ ,  $n = 1, 2, \dots$ , and

$$g_{u,r,\varepsilon}(x) = |1 - |d(x, u) - r||^+ / \varepsilon|^+$$

is a continuous function that is 1 in the ball  $B(u, r)$ , it is 0 outside  $B(u, r + \varepsilon)$ , and takes intermediate values in between. Let

$$\mathcal{F}_1 \tag{5.1}$$

be the smallest set of functions containing  $\mathcal{F}_0$  and the constant 1, and closed under the Boolean operations:  $\vee$ ,  $\wedge$ , and the complementation  $f \mapsto 1 - f$ . Each element of the set  $\mathcal{F}_1$  is a continuous (actually, Lipschitz) function with values between 0 and 1. We could also call the set  $\mathcal{F}_1$  the set of *fuzzy cells*: (cells whose boundaries are somewhat fuzzy ...). Let  $\mathcal{E}$  be the smallest set of functions containing  $\mathcal{F}_1$  and closed under the Boolean operations and also rational linear combinations. The following completeness property holds:

**Proposition 5.1.** *All bounded continuous functions can be obtained as the limit of an increasing sequence of functions from the enumerated countable set  $\mathcal{E}$  of bounded computable Lipschitz functions introduced in (5.1).*

The proof is routine.

Under some special conditions, we will still get “sharp” cells. Let  $f$  be a bounded computable function over  $\Omega$ , let  $\alpha_1 < \dots < \alpha_k$  be rational numbers, and let  $\mu$  be a computable measure with the property that  $\mu f^{-1}(\alpha_j) = 0$  for all  $j$ . In this case, we will say that  $\alpha_j$  are *regular points* of  $f$  with respect to  $\mu$ . Let  $\alpha_0 = -\infty$ ,  $\alpha_{k+1} = \infty$ , and for  $j = 0, \dots, k$ , let  $U_j = f^{-1}((j, j+1))$ . The sequence of disjoint r.e. open sets  $(U_0, \dots, U_k)$  will be called the *partition generated by  $f, \alpha_1, \dots, \alpha_k$* . If we have several partitions  $(U_{i0}, \dots, U_{ik})$ , generated by different functions  $f_i$  ( $i = 1, \dots, m$ ) and different regular cutoff sequences  $(\alpha_{ij} : j = 1, \dots, k_i)$ , then we can form a new partition generated by all possible intersections

$$V_{j_1, \dots, j_n} = U_{1, j_1} \cap \dots \cap U_{m, j_m}.$$

A partition of this kind will be called a *regular partition*. The sets  $V_{j_1, \dots, j_n}$  will be called the *cells* of this partition. It is easy to see that the values  $\mu V_{j_1, \dots, j_n}$  are computable from the names of the functions  $f_i$  and the cutoff points  $\alpha_{ij}$ .

From now on, we will assume that a computable sequence of functions  $b_1(\omega), b_2(\omega), \dots$  over  $\Omega$  is given, with the property that for every  $\omega_1, \omega_2 \in \Omega$  there is a  $j$  with  $b_j(\omega_1) < 0$ ,  $b_j(\omega_2) > 0$ . Let us give the correspondence between the set  $\mathbb{B}^\omega$  of infinite binary sequences and elements of the set

$$\Omega^0 = \{\omega \in \Omega : b_j(\omega) \neq 0, j = 1, 2, \dots\}.$$

For a binary string  $s_1 \dots s_n = s \in \mathbb{B}^*$ , let

$$\Gamma_s$$

be the set of elements of  $\Omega$  with the property that for  $j = 1, \dots, n$ , if  $s_j = 0$  then  $b_j(\omega) < 0$ , otherwise  $b_j(\omega) > 0$ . This correspondence has the following properties.

- (a)  $\Gamma_\Lambda = \Omega$ .
- (b) For each  $s \in \mathbb{B}$ , the sets  $\Gamma_{s0}$  and  $\Gamma_{s1}$  are disjoint and their union is contained in  $\Gamma_s$ .
- (c) For  $\omega \in \Omega^0$ , the set  $\{\Gamma_s : \omega \in \Gamma_s\}$  forms a basis of its neighborhoods.

If  $s$  has length  $n$  then  $\Gamma_s$  will be called a *canonical n-cell*, or simply canonical cell, or  $n$ -cell. From now on, whenever  $\Gamma$  denotes a subset of  $\Omega$ , it means a canonical cell. We will also use the notation

$$l(\Gamma_s) = l(s).$$

The three properties above say that if we restrict ourselves to the set  $\Omega^0$  then the canonical cells behave just like binary subintervals: they divide  $\Omega^0$  in half, then each half again in half, etc. Moreover, around each point, these canonical cells become “arbitrarily small”. It is easy to see that if  $\Gamma_{s1}, \Gamma_{s2}$  are two canonical cells then they either are disjoint or one of them contains the other. If  $\Gamma_{s1} \subset \Gamma_{s2}$  then  $s_2$  is a prefix of  $s_1$ . If, for a moment, we write  $\Gamma_s^0 = \Gamma_s \cap \Omega^0$  then we have the disjoint union  $\Gamma_s^0 = \Gamma_{s0}^0 \cup \Gamma_{s1}^0$ . For  $\omega \in \Gamma_s$ , we will write

$$s = \omega_1 \dots \omega_n, \quad \omega^n = \omega_1 \dots \omega_n.$$

Thus, for elements of  $\Omega^0$ , we can talk about the  $n$ -th bit  $\omega_n$  of the description of  $\omega$ : it is uniquely determined. The  $2^n$  cells (some of them possibly empty) of the form  $\Gamma_s$  for  $l(s) = n$  form a partition

$$\mathcal{P}_n$$

of  $\Omega^0$ . Let  $\mathcal{M}(\Omega)$  be the set of nonnegative measures  $\mu$  over  $\Omega$  with the property that  $\mu(B(x, r)) < \infty$  for all  $x, r$ . Let  $\mathcal{M}^0(\Omega)$  be the set of measures  $\mu \in \mathcal{M}(\Omega)$  with the property that for each  $n$ ,  $\mu b_n^{-1}(0) = 0$ . For these measures  $\mu$ , the functions  $b_j$  ( $j \leq n$ ) and the cutoff point 0 form a regular partition of  $\Omega$  as defined above.

*Examples 5.2.*

1. If  $\Omega$  is the set of infinite binary sequences with its usual topology, the functions  $b_n(\omega) = \omega_n - 1/2$  generate the usual cells, and  $\Omega^0 = \Omega$ .
2. If  $\Omega$  is the interval  $[0, 1]$ , let  $b_n(\omega) = -\sin(2^n \pi \omega)$ . Then cells are open intervals of the form  $(k \cdot 2^{-n}, (k+1) \cdot 2^{-n})$ , the correspondence between infinite binary strings and elements of  $\Omega^0$  is just the usual representation of  $\omega$  as the binary decimal string  $0.\omega_1\omega_2\dots$ .

□

When we fix canonical cells, we will generally assume that the partition chosen is also “natural”. The bits  $\omega_1, \omega_2, \dots$  could contain information about the point  $\omega$  in decreasing order of importance from a macroscopic point of view. For example, for a container of gas, the first few bits may describe, to a reasonable degree of precision, the amount of gas in the left half of the container, the next few bits may describe

the amounts in each quarter, the next few bits may describe the temperature in each half, the next few bits may describe again the amount of gas in each half, but now to more precision, etc. From now on, whenever  $\Gamma$  denotes a subset of  $\Omega$ , it means a canonical cell. From now on, for elements of  $\Omega^0$ , we can talk about the  $n$ -th bit  $\omega_n$  of the description of  $\omega$ : it is uniquely determined.

**5.2. Characterizing tests via complexity.** In this section, we restrict ourselves to finite measures. Let  $\mathcal{M}_R(\Omega)$  be the set of measures  $\mu$  with  $\mu(\Omega) = R$ .

Theorem 1 (the Coding Theorem) implies that if  $x$  runs over a discrete space then  $H_\#(x) \stackrel{+}{=} H(x)$ . More generally, it is shown in [5] that

$$H_\mu(x) \stackrel{+}{=} H(x | \mu) + \log \mu(x).$$

Here, the function  $H(x | \mu)$  must be defined appropriately, as a lower semicomputable function of the pair  $(x, \mu)$ . There is a similar characterization of tests over arbitrary spaces. Let us denote

$$H_\mu(\Gamma) = H(\Gamma | \mu) + \log \mu(\Gamma)$$

for canonical cells  $\Gamma$ . The following theorem is proved in [5]:

**Theorem 4.** *Suppose that the space  $X$  is compact. Then for all computable measures  $\mu \in \mathcal{M}_R^0(X)$ , we have*

$$H_\mu(\omega) \stackrel{+}{=} \inf_n (\log \mu(\omega^{\leq n}) + H(\omega^{\leq n} | \mu)). \quad (5.2)$$

Here, the constant in  $\stackrel{+}{=}$  depends on the computable measure  $\mu$ .

The  $\stackrel{+}{<}$  part of the statement is valid in a more general space, and without assuming computability or compactness. Assume that a separating sequence  $b_1, b_2, \dots$  is given as defined in Subsection 5.1, along with the set  $\Omega^0$ . For each  $x \in X^0$ , the binary sequence  $\omega_1, \omega_2, \dots$  has been defined. Let  $\mathcal{M}_R^0(\Omega)$  be again the set of those measures  $\mu$  with  $\mu(\Omega \setminus \Omega^0) = 0$ . The following statement has also been proved in [5]:

**Proposition 5.3.** *For all measures  $\mu \in \mathcal{M}_R^0(\Omega)$ , we have*

$$H_\mu(\omega) \stackrel{+}{<} \inf_n (\log \mu(\omega^{\leq n}) + H(\omega^{\leq n} | \mu)). \quad (5.3)$$

*Remarks 5.4.*

1. Since we assumed  $\mu$  computable, we can actually delete it from the condition in the complexity in (5.2). We left it there only since we would really like to prove that this equation holds uniformly over  $\mathcal{M}(\Omega^0)$  and not only for computable measures.
2. An interesting application where a related formula is used is the “minimum description length” principle (MDL) theory of statistics. There, instead of the description complexity  $H(x)$ , often the codeword length  $C(x)$  of some other coding (universal over some class of measures) is considered, and the quantity  $C(\Gamma) + \log \mu(\Gamma)$  is called the *redundancy*. In these statistical applications, the presence or absence of  $\mu$  in the condition makes a difference.

□

The Test Characterization Theorem, in the form

$$H_L(\omega) \stackrel{+}{=} \inf_n (H(\omega^n) + \log L(\Gamma_{\omega^n})) \quad (5.4)$$

says that the fine-grained algorithmic entropy  $H(\omega) = H_L(\omega)$  with respect to the invariant volume measure  $L$  can be essentially expressed as  $H(\omega^n) + \log L(\Gamma_{\omega^n})$  for a certain  $n$ : so, it is the sum of the Boltzmann entropy for the partition  $\mathcal{P}_n$  plus the description complexity of the macroscopic description  $\omega^n$ . For the systems and partitions of interest in physics, the additive term  $H(\omega^n) \stackrel{+}{<} 2n$  is typically negligible compared to the other one since the total number of macroscopic cells is typically small compared to the volume of the large cells.

In “practice”, to find the “right”  $n$  we should keep increasing it, include (the program for) more and more bits of  $\omega$  into the macroscopic description as long as the complexity increase buys greater decrease in the Boltzmann entropy  $\log L(\Gamma_{\omega^n})$  (our *a priori* uncertainty about  $\omega$ ).

The following theorem says that, for most elements  $\omega$  of a cell  $\Gamma$ , the value of  $H_\mu(\omega)$  cannot be much higher than  $H_\mu(\Gamma)$ .

**Theorem 5** (Stability).

$$\mu\{\omega \in \Gamma : H_\mu(\omega) < H_\mu(\Gamma) - H(l(\Gamma)) - m\} \stackrel{*}{<} 2^{-m}\mu(\Gamma).$$

We can also interpret this theorem as saying that if some elements of the cell are (sufficiently) random then most of them are (sufficiently) random. Note that the difference  $H(l(\Gamma))$  is less than  $2\log n$  for  $\Gamma_{\omega^n}$ .

*Proof.* Let  $f(\Gamma) = 2^{-H(l(\Gamma))} \int_\Gamma 2^{-H(\omega)} \mu(d\omega)$ . This function is semicomputable and  $\sum_\Gamma f(\Gamma) \leq 1$ . Therefore  $f(\Gamma) \stackrel{*}{<} 2^{-H(\Gamma)}$ . Rearranged, this gives:

$$\mu(\Gamma)^{-1} \int_\Gamma 2^{-H(\omega)} \mu(d\omega) \stackrel{*}{<} 2^{-H(\Gamma)+H(l(\Gamma))}.$$

From here, we conclude with Markov’s inequality. □

Let us define

$$H^n(\omega) = \min_{i \leq n} S(\Gamma_{\omega^i}). \quad (5.5)$$

Let  $i_n(\omega)$  be the  $i$  where the minimum is achieved. As a function of  $n$ , for certain  $\omega$ , this may make a sudden jump from a much smaller value to  $n$ , suggesting instability in the coarse-grained quantity. However, there is another interpretation. In general, what is given is a canonical cell  $\Gamma$ . Theorem 5 (the Stability Theorem) implies

$$L\{\omega \in \Gamma : H(\omega) < S(\Gamma) - m - H(l(\Gamma))\} \stackrel{*}{<} 2^{-m}L(\Gamma).$$

In other words, for most elements  $\omega$  of the canonical cell  $\Gamma$ , the value  $H(\omega)$  is not much less than  $S(\Gamma)$ . In a sense, a string  $\omega^n$  describing a canonical cell never has too few bits since it is always a nearly optimal description for most states in it. But it may have too many bits in the sense that deleting some of the last ones results

in a significant decrease of the entropy of the canonical cell: the deletion decreases the complexity with only a smaller increase in  $\log L$ .

## 6. ENTROPY INCREASE PROPERTIES

**6.1. Fine-grain nondecrease.** As stated in the Introduction, we are considering an isolated physical system with state space  $\Omega$  whose development is described by a transformation group  $U^t$ . We also assume that  $U^t\omega$  is computable as a function of the pair  $\omega$  and  $t$ . We assume the existence of a computable invariant measure  $L$  (the “Liouville measure”): it has the property that  $L(U^tA) = L(A)$  for all  $t$  and all measurable sets  $A$ . Under suitable conditions, the existence, computability and even uniqueness of  $L$  can be proven.

Since  $U^t$  is measure-preserving, the function  $2^{-H(U^t\omega)}$  is a parametrized randomness test. From Proposition 3.1, we obtain therefore

$$H(U^t\omega) \stackrel{+}{>} H(\omega | t) = H(\omega) - I(t : \omega).$$

This is, in essence, our entropy nondecrease formula since we will see that the term  $I(t : \omega)$  is generally very small. It can also be regarded as a special case of a more general randomness-conservation property formulated by L.A. Levin in several ways, see its latest form in [9]. (Other forms can be found in [4], [5].) Alas, it is also an entropy nonincrease formula. Indeed, the same inequality between  $H(\omega)$  and  $H(U^t\omega)$ , can also be used between  $H(U^t\omega)$  and  $H(\omega)$ . We get therefore

$$-I(t : \omega) \stackrel{+}{<} H(U^t\omega) - H(\omega) \stackrel{+}{<} I(t : U^t\omega). \quad (6.1)$$

According to this, the only amount of decrease we will ever see in  $H(U^t\omega)$  is due to the information that the value of the time  $t$  may have on  $\omega$ , which is very small for all simple moments of time. But the amount of increase is also only due to the information that  $t$  may have on  $U^t\omega$ . Let us explore the nondecrease property.

**Theorem 6** (Entropy Nondecrease). *Let  $\lambda$  be the length (Lebesgue) measure, and let  $T$  be a rational value of time. We have*

$$\lambda\{t \in [0, T] : H(U^t\omega) < H(\omega) - H(T) - m\} \stackrel{*}{<} 2^{-mT}.$$

This theorem follows, by an application of Markov’s Inequality, from the following lemma.

**Lemma 6.1.**

$$T^{-1} \int_0^T 2^{I(t:\omega)} dt \stackrel{*}{<} 2^{H(T)}.$$

*Proof.* The function  $f(\omega) = T^{-1} \int_0^T 2^{-H(t,\omega)} dt$  is a randomness test and therefore

$$-\log f(\omega) \stackrel{+}{>} H(\omega | T) \stackrel{+}{>} H(\omega) - H(T)$$

(using Lemma 4.1). Hence

$$T^{-1} \int_0^T 2^{H(\omega|t)} dt \stackrel{*}{<} 2^{-H(\omega)+H(T)}$$

which by rearrangement gives just what we want.  $\square$

**6.2. Gibbs ensembles.** Another accepted model of a macrostate is a certain distribution  $\nu$  over microscopic states given by a density function (ensemble)  $p(\omega)$  with respect to the volume measure  $L$ . Let us require  $\int p(\omega)L(d\omega) = 1$ . We can ask what is the probability density to find it in state  $\omega$  at time  $t + t_0$ ? Let us call this new ensemble  $p^t$ .

The classical definition of Gibbs entropy of a probability distribution with density function  $p(\omega)$  over  $L$  is

$$G(p) = - \int p(\omega) \log p(\omega) L(d\omega).$$

In the special case when  $p$  is the macrostate-ensemble we have  $G(p_\Gamma) = \log L(\Gamma)$ , the Gibbs entropy is the same as the Boltzmann entropy. Liouville's Theorem implies  $G(p^t) = G(p)$ , that is the Gibbs entropy of an ensemble does not change at all in an isolated system during evolution. This shows that in case of the evolution of isolated nonequilibrium systems, the evolution of a Gibbs ensemble does not express adequately what we consider thermodynamic behavior. The problem is that even if at the starting time  $t_0$  the Gibbs ensemble was something simple, it can develop in time  $t$  into a very complicated density function that does not correspond to any reasonable macroscopic description. Ensembles that are invariant in time retain their usefulness, however, for equilibrium systems.

Let us relate Gibbs entropy to the average of algorithmic entropy. Let  $\mu$  be a measure and  $p(\omega) \in C_B(\Omega)$  a computable (hence continuous) nonnegative function (this condition could be relaxed somewhat but the result given here is sufficient to see the ideas). Then

$$\nu(f) = \int f(\omega)p(\omega)\mu(d\omega)$$

defines a finite measure. The following theorem expresses the randomness test for  $\nu$  by the randomness test for  $\mu$ .

**Theorem 7.**

$$H_\nu(\omega) \stackrel{+}{=} H_\mu(\omega) + \log p(\omega)$$

where the constant in  $\stackrel{+}{=}$  depends on the definition of the function  $p(\omega)$ .

*Proof.* For an arbitrary lower semicomputable function  $f(\omega, \nu)$  we have

$$\int f(\omega, \nu)\nu(d\omega) = \int f(\omega, \nu)p(\omega)\mu(d\omega).$$

Therefore  $f$  is a test for  $\nu$  if and only if  $fp$  is a test for  $\mu$ . The maximal  $f$  will therefore be  $\stackrel{*}{=} 2^{-H_\mu(\omega)+\log p(\omega)}$ .  $\square$

Notice that the theorem does not allow to compute  $H_\mu(\omega)$  from  $H_\nu(\omega)$  in the places where  $p(\omega) = 0$ .

The uniform bound (4.2) on  $H_\nu(\omega)$  implies

$$H_\nu(\omega) \stackrel{+}{<} \log \nu(\Omega) + H(\lfloor \log \nu(\Omega) \rfloor).$$

Combining these, we see that  $\log \nu(\Omega) - \log p(\omega)$  is nearly an upper bound to  $H_\mu(\omega)$ : more exactly,

$$H_\mu(\omega) \stackrel{+}{<} -\log p(\omega) + \log \nu(\Omega) + H(\lfloor \log \nu(\Omega) \rfloor).$$

The last terms disappear, of course, if  $\nu$  is a probability measure. On the other hand, Markov's inequality implies that  $H_\mu(\omega)$  is, with overwhelming  $\nu$ -probability, close to this upper bound. From this, it is easy to see the following relation to Gibbs entropy:

$$\begin{aligned} \int H_\mu(\omega) \nu(d\omega) &\stackrel{+}{<} - \int -p(\omega) \log p(\omega) \mu(d\omega) \\ &\stackrel{+}{<} \int H_\mu(\omega) \nu(d\omega) + H(\lfloor \log \nu(\Omega) \rfloor). \end{aligned}$$

This is what we mean by saying that the Gibbs entropy is close to the average of algorithmic entropy.

**6.3. The increase of coarse-grained entropy.** Let us now consider the much more speculative problem of approach to equilibrium, the argument that the algorithmic Boltzmann entropy  $H^n(U^t\omega)$  must indeed increase fast if it is far from its upper bound  $\log L(\Omega)$ .

There is a classical argument to show that in a nonequilibrium system, Boltzmann entropy can be expected to increase fast until it almost reaches its upper bound  $\log L(\Omega)$ . The argument relies on two properties that I will try to formulate now. Let us look at systems in which

$$M = \log L(\Omega)$$

is a parameter, the maximal value of the entropy. We will not try to require convergence of the entropy exactly to  $M$ , only to  $M - \alpha(M)$  where  $\alpha(M)$  is some function with

$$\lim_{M \rightarrow \infty} \alpha(M)/M = 0.$$

The function  $\alpha(M)$  is typically of the order of  $\log M$  and it gives us a little freedom to simplify expressions by ignoring logarithmic terms. Let  $C(d, M)$  be the union of all cells  $\Gamma$  with

$$B(\Gamma) < M - d.$$

We are looking for an entropy increase property saying the following. If we start from any cell  $\Gamma$  then as time passes, the proportion of those of its elements that do not end up in a cell with entropy  $M - d$  becomes smaller than  $\varepsilon_d$  where  $\lim_{d \rightarrow \infty} \varepsilon_d = 0$ . In the formulation, we write  $\varepsilon_d$  as  $2^{-f_1(d)}$ .

*Condition 6.2* (Boltzmann Entropy Increase Property). There is a constant  $a_1 > 0$ , a function  $f_1(d)$  with  $\lim_{d \rightarrow \infty} f_1(d) = \infty$  and a function  $f_2(t, d, e)$  with  $\lim_{t \rightarrow \infty} f_2(t, d, e) = f_1(d)$  such that for all  $d, e$  with  $a_1 \alpha(M) < d < e$  and  $\Gamma \not\subset C(e, M)$  we have

$$L(U^t\Gamma \cap C(d, M))/L(\Gamma) < 2^{-f_2(t, d, e)}.$$

□

Two properties will provide a satisfactory condition. The first one says the union  $C(d, M)$  of all “small” cells taken together is small. We call it the Noncompensation Condition to suggest that the smallness of small cells is not compensated by their quantity. The condition holds for typical systems simply because the total number of cells is small.

*Condition 6.3* (Noncompensation Condition). There are constants  $b_1, b_2$  such that for  $b_1\alpha(M) < d$  we have

$$L(C(d, M))/L(\Omega) \leq 2^{-b_2 d}.$$

□

In Example 8.2, the total number of cells is the total number of partitions of  $n$  into  $\sum_{i=1}^m n_i$ , which is less than  $n^m$ , therefore  $L(C(d, M))/L(\Omega) \leq 2^{-(d-m\log n)}$ . Typically, we will have  $M > n$  and we agreed that  $m < n^{1/2}$ . So if we choose  $\alpha(M) = M^{1/2} \log M$  then if  $d > 2\alpha(M)$  we have  $d - m\log n > d - n^{1/2} \log n = d - \alpha(M) > d/2$ , so the condition holds with  $b_1 = 2, b_2 = 1/2$ .

*Example 6.4.* The enormous differences in cell sizes seem to be typical for statistical mechanical systems even without reference to the small number of cells. Let for example, in a container of a mole of gas, our macroscopic variable be the binary number  $0.0\omega_1\omega_2$  giving approximately the relative quantity of gas in the left half of the container. Then the cells  $\Gamma_{0.000}$  and  $\Gamma_{0.011}$  are absolutely negligible in volume compared to the cells  $\Gamma_{0.001}$  and  $\Gamma_{0.010}$ . □

Algorithmic coarse-grained entropy satisfies the Noncompensation Condition automatically, with  $b_2 = 1$ . Indeed, the set  $\{\omega : H^n(\omega) < m\}$  is contained in  $E_m = \{\omega : H(\omega) < m\}$ , which, according to (3.3), has volume at most  $2^{-m}L(\Omega)$ .

The second condition says that if the system starts from a state in a not very small cell then after a time  $t$ , it is unlikely to end up in any small union of cells.

*Condition 6.5* (Weak Mixing Condition). There is a constant  $c_1 > 0$ , a function  $h_1(d)$  with  $\lim_{d \rightarrow \infty} h_1(d) = \infty$  and a function  $h_2(t, d, e)$  with  $\lim_{t \rightarrow \infty} h_2(t, d, e) = h_1(d)$  such that for all  $d, e$  with  $c_1\alpha(M) < d < e$ , if  $\Gamma \not\subset C(e, M)$  and  $D$  is a union of cells with  $L(D)/L(\Omega) \leq 2^{-d}$  then

$$L(U^t\Gamma \cap D)/L(\Gamma) < 2^{-h_2(t, d, e)}.$$

□

This property says that as  $t$  grows the transformation  $U^t$ , while preserving the volume of the cell  $\Gamma$ , will distribute its content thinly over a large area, so that only a small fraction of it can intersect the small union of cells  $D$ . In other words, after a while, the cells will be “mixed”. The property is in general difficult to prove but is plausible in typical physical systems. The following theorem only serves to check out the consistency of the above concepts.

**Theorem 8.** *Theorem The Weak Mixing Condition and the Noncompensation Condition imply the Boltzmann Entropy Increase Property.*

*Proof.* Assume that the Weak Mixing Condition and the Noncompensation Condition with the appropriate constants and functions  $a_1, b_1, b_2, h_1, h_2$ , and let  $\Gamma \not\subset C(e, M)$ . Let  $D = C(d, M)$  and  $d > b_1\alpha(M)$ , then the Noncompensation Condition implies  $L(D)/L(\Omega) \leq 2^{-b_2d}$ . The Weak Mixing condition implies

$$L(U^t\Gamma \cap D)/L(\Gamma) < 2^{-h_2(t, b_2d, e)}$$

for  $d$  such that  $c_1\alpha(M) < b_1d$ . Thus, the Boltzmann Entropy Increase Condition holds with  $a_1 = c_1/b_1$  and  $f_2(t, d, e) = h_2(t, b_2d, e)$ .  $\square$

The following examples illustrate the insufficiency of the traditional Boltzmann entropy  $\log L(\Gamma)$ .

*Example 6.6.* Take a large container filled with ideal gas and a few large balloons. At start, the balloons are fixed. Then we release them. They gain energy from collisions with the gas molecules until they achieve the average energy appropriate to their number of degrees of freedom. It is reasonable to count the positions of the balloons to the macrostate of the system. The volume of the cell will be essentially determined by the energy of the system consisting of the gas alone. This energy, and hence the Boltzmann entropy, becomes smaller by the amount transferred to the balls.  $\dashv$

This example becomes less ridiculous if we replace balloons by the memory of a computer. For a while, it will still be reasonable to count the content of the memory as part of the macroscopic description: it is given by specifying the “global” charges and magnetizations of all the tiny areas on the disks and the silicon memories. However, as the size of a site storing an individual bit decreases, there will come a point where it is not reasonable to consider the memory state as part of the macroscopical description. The communication of two computers, one with “macroscopic” memory and the one with “microscopic” memory, leads to the Maxwell demon paradox. This shows the necessity of the smooth transition between macroscopic and microscopic transitions exhibited by algorithmic entropy.

In terms of our scheme, we are talking about increasing  $n$  (refining the partition). The additive term  $H(\omega^n)$  which is so insignificant for small values of  $n$ , gains in significance in this process and makes the transition continuous. Ignoring it, by defining entropy just as  $\log L(\omega^n)$ , is bound to lead to paradoxes.

The following system also defies Boltzmann entropy but submits to algorithmic coarse-grain entropy.

*Example 6.7* (The baker’s map). Let  $\Omega$  be the set of doubly infinite binary sequences  $\omega = \dots \omega_{-1}\omega_0\omega_1\omega_2 \dots$  with the shift transformation  $(U^t\omega)_i = \omega_{i+t}$  over discrete time. Let us write  $\omega^n = \omega_{-\lfloor n/2 \rfloor} \cdots \omega_{\lceil n/2 \rceil - 1}$ . The  $n$ -cells are, of course, cells of the form  $\Gamma_{\omega^n}$ . Let the volume  $L$  be such that all  $n$ -cells have the same volume  $2^{-n}$ . Since all  $n$ -cells have the same measure no matter what fixed precision we choose, the Boltzmann entropy of  $U^t\omega$  does not increase with  $t$ . The quantity  $H^n(U^t\omega)$ , however, will be shown below to increase fast for all typical sequences, between times time 0 and  $n$ , linearly from  $-n$  to 0.

Consider a typical example of an infinite sequence  $z$  which has  $z^n = 0 \cdots 0$  and whose other bits are random (this has now a precise meaning but let us just use the informal understanding). Then  $H(z^n) \leq 2 \log n$ ,  $\log \Gamma_{z^n} = -n$ , therefore

$$H^n(z) \leq 2 \log n - n.$$

Now, for  $t > n$ , the string  $(U^t z)^n$  consists of random bits, so essentially,  $H^n(U^t z) \geq 0$ . Between time 0 and  $n$ , this algorithmic coarse-grained entropy increases linearly from  $-n$  to 0. The fine-grained algorithmic entropy  $H(U^t z)$  will, on the other hand, only increase slowly since in our choice of the precision  $n$ , we can follow the increase of  $t$ . We have

$$\begin{aligned} H(U^t z) &\stackrel{+}{<} H^{t+n}(U^t z) \leq 2 \log n + 2 \log t + t - (t + n) \\ &= 2 \log n + 2 \log t - n. \end{aligned}$$

On the right-hand side of the first equation, the first two terms upper-bound the complexity of  $n$  and  $t$ , the second term is the length of the nonzero part of the string  $(U^t z)^{t+n}$ . Therefore  $H((U^t z)^{n+t})$  is bounded by the sum of the first three terms, while the last term gives the logvolume of the cell. This inequality shows that  $H(U^t z)$  can only grow as slowly as  $\log t$ . The general inequality (6.1) gives  $H(U^t \omega) - H(\omega) \stackrel{+}{<} I(t : (U^t \omega))$ . For rational values of  $t$ , the inequality (4.3) gives  $I(t : (U^t \omega)) \stackrel{+}{<} H(t)$ . Hence

$$H(U^t \omega) - H(\omega) \stackrel{+}{<} H(t).$$

This means that for some simple rational values of  $t$ , the algorithmic entropy hardly increases at all. For example, if  $t$  is an integer of the form  $2^n$  then the increase is at most  $2 \log \log t$ .

We can also use independent biased coin tossings for the measure, where the probability of 1 is  $p = 0.3$ . If  $\omega^n = 0 \cdots 0$  then  $B(\omega^n) = n \log 0.7$ . But for a random  $\omega^n$ , this value will be  $\approx n(0.3 \log 0.3 + 0.7 \log 0.7)$  which is considerably smaller. Therefore Boltzmann entropy decreases strongly. The algorithmic Boltzmann entropy is, on the other hand, negative for the all 0 starting cell and grows to  $\approx 0$  as expected.

We can even use any ergodic stationary process for the measure, and obtain similar results. For such processes, there is an “asymptotic equidistribution property” guaranteeing that most volume will be taken up by  $n$ -cells of about the same size  $2^{-hn}$  (where  $h$  is the so-called “entropy rate”).  $\square$

In typical physical systems, the partitions given by the canonical cells have no simple connection with the computable transformation group  $U^t$  of our dynamical system. In particular, they are not “generated” from the first partition into  $\Gamma_0$  and  $\Gamma_1$  by  $U^t$  the way they are in the baker’s map.

*Remark 6.8* (Kolmogorov-Sinai entropy). An arbitrary ergodic computable stationary measure could be considered in place of the coin tossing as well. Let us point out the connection of our quantities to the Kolmogorov-Sinai entropy of such a measure. For a random sequence (and hence for almost all sequences)  $\omega$ , the algorithmic Boltzmann entropy  $H(\omega^n) + \log \mu(\Gamma_{\omega^n})$  remains bounded. This quantity

is the difference of two quantities that increase therefore equally fast:  $H(\omega^n)$  and  $-\log \mu(\Gamma_{\omega^n})$ . According to the Shannon-McMillan-Breiman Theorem, with probability 1, we have  $\lim_{n \rightarrow \infty} -\log \mu(\Gamma_{\omega^n})/n = h$  where  $h$  is the Kolmogorov-Sinai entropy of the process  $\mu$ . Thus, the Kolmogorov-Sinai entropy shows the rate of increase of the Boltzmann entropy of a stationary process as the partition is being refined. Due to the boundedness of the quantity  $H(\omega^n) + \log \mu(\Gamma_{\omega^n})$ , we arrive to (a version of) Levin's theorem from [14], saying that this is also the rate of increase of  $H(\omega^n)$ .  $\square$

The reason that Boltzmann entropy does not increase to  $\log L(\Omega)$  in the baker's map with any stationary measure is just that, in case of the uniform distribution, the  $n$ -cells have the same volume, and in the more general case still, there is an “asymptotic equidistribution property” guaranteeing that most volume will be taken up by  $n$ -cells of about the same size  $2^{-hn}$  (where  $h$  is the so-called “entropy rate”). Therefore the Noncompensation Condition (the more trivial of the two conditions) is not satisfied. If for this same map we use  $n$ -cells of different enough volumes to satisfy this condition then Boltzmann entropy will increase. In fact, when  $H^n(\omega)$  is high this can be treated as a concise expression of the fact that in every “simple” partition with extremely different cell sizes, the point  $\omega$  would end up in a large cell.

The example of the shift transformation suggests that actually, in typical chaotic systems, the parameter  $n$  can be made a function of  $t$  as long as it grows slower than linearly with  $t$ . Thus, if  $\lim_{t \rightarrow \infty} n(t)/t = 0$  then in the baker's map with the uniform distribution,

$$H^{n(t)}(U^t \omega)$$

will approach  $\log \mu(\Omega)$  almost as fast as if we held  $n$  constant. The growth of  $n(t)$  seems a good measure of the mixing of  $U^t$ .

In conclusion, we suggest that the new quantity  $H_L^n(\omega)$  extends the idea of entropy increase to a wider class of chaotic systems than the one in which it has originally worked, and can also serve as a useful tool for formulating conjectures concerning the nature of chaoticity and its extent.

**6.4. The paradox of typicality.** The notion of a typical object is an informal one, and the present remark calls attention to the fact that our intuition concerning the properties of typical objects may be misleading. Consider the space of infinite 0-1 sequences obtained by tossing a biased coin, with probabilities  $1/3, 2/3$ . We would consider typical those sequences in which the relative frequency of 0's tends to  $1/3$ . On more reflection, we would consider those sequences  $\omega$  typical that also satisfy all other criteria or randomness, and are random according to Martin-Löf's definition, or, equivalently, which have  $H_p(\omega) > -\infty$  where  $p$  is the appropriate coin-tossing measure.

Consider now a dynamical system with the volume measure  $L$ . The (fine-grained) entropy nondecrease property, which can also be considered a randomness-conservation property, guarantees that the above defined typicality is “conserved”: the evolution of a system takes typical states into typical ones.

There is, however, another, similarly attractive idea of typicality, which we will call “local typicality”, which is not conserved. Consider a given partition  $\mathcal{P}_n$  and a given cell  $\Gamma$  in this partition. Let us call those states  $\omega$  of the cell  $\Gamma$  “locally typical” whose fine-grained entropy  $H(\omega) = H_L(\omega)$  is close to the coarse-grained entropy  $H^n(\omega) = \log L(\Gamma) + H(\Gamma)$ . We know from the Stability Theorem 5, that most points of each cell (in terms of the measure  $L$ ) are typical in this sense. However, local typicality is not conserved. Indeed, assume that coarse-grained entropy increases for most points, and that the coarse-grained entropy of  $\Gamma$  is low. Then for most points  $\omega$  in  $\Gamma$  (and hence for most locally typical points), their coarse-grained entropy increases, so  $U^t\omega$  belongs to a cell  $\Gamma'$  with much higher coarse-grained entropy. At the same time, the fine-grained entropy  $H(U^t\omega)$  does not change too much with respect to  $H(\omega)$ . So, the locally typical state  $\omega$  turns into a locally nontypical state  $U^t\omega$ . The reason is that  $\omega$  was a locally typical point of a “nontypical” cell, and it is still carrying this history. However, as long as only the macroscopic information embodied in the cell  $\Gamma'$  is available for inspection and manipulation, this history is inaccessible to later observers.

The non-conservation of local typicality eliminates a potentially attractive “principle”: namely that the state we have at present is locally typical. Consider the present state of a container of gas after a wall was removed that had confined the gas to one half. In a usual macroscopic description (partition), the coarse-grained entropy of the present state will be much larger than what it was before the wall removal. Since the fine-grained entropy is approximately the same (since it did not change much), the present state is actually highly nontypical.

The refuted principle is attractive since, together with the Boltzmann Entropy Increase Property (or its counterpart using algorithmic coarse-grained entropy) it could be used to prove that entropy increase is not only likely to occur but will occur. One possible substitute of the principle is the introduction of probabilistic perturbations, see e.g. [3]. We prefer to say the following:

The entropy increase property relates strictly only to the present macroscopic state of our system, and does not assert directly anything about the present microscopic state.

## 7. MAXWELL’S DEMON

**7.1. Entropy balance.** Let  $\mathcal{X}$  and  $\mathcal{Y}$  be two systems where  $\mathcal{Y}$  is considered to be the environment from which  $\mathcal{X}$  is temporarily isolated. In order to “do something” to  $\mathcal{X}$ , we couple it with  $\mathcal{Y}$ , giving rise to a joint Hamiltonian, and a joint transformation  $U^t(\xi, \eta)$ . Let us assume that, being in classical mechanics, the impulses and momenta of the joint system are simply the impulses and momenta of the two subsystems, therefore the Liouville measure on  $X \times Y$  is, even in the coupled system, the product of the original Liouville measures  $L_X, L_Y$  in the subsystems. Let  $(\xi_t, \eta_t) = U^t(\xi, \eta)$ , and

$$\Delta H(\xi) = H(\xi_t) - H(\xi).$$

Notice now that for our measure,

$$H(\xi) + H(\eta) = H(\xi, \eta) + I(\xi, \eta).$$

**Theorem 9** (Entropy Balance).

$$\Delta H(\xi) + \Delta H(\eta) \stackrel{+}{>} I(\xi_t, \eta_t) - I(\xi, \eta) - I(t : \xi, \eta).$$

*Proof.* According to (6.1) applied to the joint system we have  $\Delta H(\xi, \eta) \stackrel{+}{>} -I(t : \xi, \eta)$ . Note that this formula cannot be applied now to the parts of the system since they do not have their own transformations now. Using this, we have

$$\begin{aligned} H(\xi_t) + H(\eta_t) &= H(\xi_t, \eta_t) + I(\xi_t, \eta_t) \\ &\stackrel{+}{>} H(\xi, \eta) - I(t : \xi, \eta) + I(\xi_t, \eta_t) \\ &= H(\xi) + H(\eta) + I(\xi_t, \eta_t) - I(\xi, \eta) - I(t : \xi, \eta), \end{aligned}$$

which gives the statement by rearrangement.  $\square$

Since the last term is generally negligible this theorem says that if the two systems were originally independent ( $I(\xi, \eta) \approx 0$ ) then a decrease in the entropy of  $\xi$  must be accompanied by an increase in the entropy of  $\eta$ . The entropy balance theorem is not new, of course, for Boltzmann entropy. But its present form makes it useful for the treatment of Maxwell's demon.

**7.2. Maxwell's demon and Landauer's thesis.** Maxwell's demon is a being sitting at a tiny door between two compartments of gas and letting the molecules through selectively with the goal of entropy decrease in the container. Principles of thermodynamics seem to contradict to the possibility of such a demon, so it seems paradoxical, and demanding explanation. The typical explanations assume either that the door will heat up and begin to work randomly after a while, or that in order to make its observations, the demon must descend into this world more than she cares to and interact energetically with the molecules; this heats her up, making it harder and harder for her to concentrate. These explanations introduce additional physical assumptions which are alien to the general mathematical nature of the second law (increase of disorder). Several such explanations are refuted by more refined models (see [1]).

A convincing modern solution emerged in a principle announced by Landauer (see [1]). Let us model the demon as some computer-controlled device interacting with the gas. She seems to be able to decrease the Boltzmann entropy of the gas only at the expense of the increase of her own information content. Landauer introduced a principle saying that in order to erase a bit of information, a certain minimal amount ( $kT \log 2$ ) of heat dissipation into the environment (and, of course, investment of the corresponding amount of work into the system) is necessary.

*Remark 7.1.* In order to prove that the erasure results in heat dissipation, Landauer argues that the erasure must be a general operation that decreases the phase space of the computer memory. I find it difficult to interpret the increase of  $H(\eta)$  universally as heat dissipation. Consider a memory  $\xi$  consisting of a row of pendulums swinging

transversally. The bit 1 means that the pendulum swings while the bit 0 means it does not. Let the “environment”  $\eta$  be an identical row of pendulums, each of which hangs motionless originally. Now we can use  $\eta$  to erase the memory by moving it next to  $\xi$  in the right moment. The change in  $\eta$  is obviously reversible, so there is no heat dissipation.  $\square$

Using our framework, the demon paradox occurs since the state of demon’s memory was implicitly considered part of the macroscopic description of the joint system gas-demon: the quantity of information in it failed to contribute to the classical Boltzmann entropy. Since this device is able to decrease the Boltzmann entropy of the gas at the expense of the increase of the information content of its memory (without increasing its own Boltzmann entropy), it is able to decrease the Boltzmann entropy of the total system. Our solution eliminates the paradox by including the information content (complexity) of the macroscopic description into the expression for entropy. It should be considered as the rigorous formulation of the more special and informal principles of Landauer and Zurek.

Zurek [13] saw that the Maxwell’s demon paradox and Landauer’s thesis are two sides of the same interaction between an information-processing machine (the demon) and a classical thermodynamic system. The demon turns entropy into information, the information-erasure operation turns information into entropy. Zurek constructed an entropy-like quantity specifically for this situation and argued that it is non-increasing. He created a special macroscopic variable  $d$  (without actually distinguishing macroscopic and microscopic), whose value is equal to the demon’s memory state. He defined then a quantity called “physical entropy” associated with such a system that is essentially  $Z(a, d) = B(a) + H(d)$  where  $B(a)$  is the Boltzmann entropy of the classical part. This can be seen as essentially the same as  $B(a, d) + H(a, d)$ . Indeed,  $B(a, d) = B(a)$  since the demon’s macroscopic and microscopic states are the same. Also, we can delete  $a$  from  $H(a, d)$  since we are interested in situations in which  $d$  contains much more information than  $a$ .

Zurek argues that if, at constant temperature  $T$ , the system is brought from state  $(a_1, d_1)$  to state  $(a_1, d_2)$  then the amount of work obtained is at most  $Z(a_2, d_2) - Z(a_1, d_1)$ . For this, he tacitly assumes that the work gained from the operation of the system can be separated into the work obtained from bringing the classical machine from  $a_1$  to  $a_2$  and into the work bringing the memory from  $d_1$  to  $d_2$ . With this assumption, the second law indeed implies the upper bound  $T(B(a_2) - B(a_1))$  on the first kind of work and Landauer’s principle implies the bound  $T(H(d_2) - H(d_1))$  for the second kind of work.

Formally, our coarse-grained algorithmic entropy looks similar to Zurek’s but is defined more generally, and has many connections to various different definitions of entropy (for ensembles as well as cells) and also to the theory of randomness. In particular, the above bound can be proven without the tacit assumptions.

Let  $\xi$  be the gas whose entropy the demon is trying to decrease. We also count the whole state  $\eta$  of the demon into her macroscopic description. As it is usual with classical machines, we can assume that there is an  $n$  such that  $H^n(\xi)$  is close to  $H(\xi)$  but  $n$  (and therefore  $H(\xi^n)$ ) is still negligibly small with respect to  $H(\xi)$ , and

therefore

$$H(\xi) \approx \log L(\Gamma_{\xi^n}).$$

The Entropy Balance Theorem guarantees that going from  $(\xi, \eta)$  to  $(\xi^t, \eta^t)$ , the decrease in the sum  $H(\xi) + H(\eta)$  will be small. Since  $H^n(\xi^t) > H(\xi^t)$  this implies that any decrease in  $H^n(\xi)$ , the Boltzmann entropy of the machine, must be compensated by an increase in  $H(\eta) \approx H(\eta^t)$ , the information content of the demon's memory (this is the Maxwell's demon direction) and vice versa (this is the Landauer thesis direction).

## 8. TEN MINUTES ON CLASSICAL THERMODYNAMICS

**8.1. Thermodynamical systems.** Entropy was first introduced in classical, “phenomenological” thermodynamics. This theory, which is also the form of thermodynamics most widely used in engineering, is concerned with a physical system when the latter is in a state called “equilibrium”.

An *equilibrium state* of the system is characterized by the fact that for all practical purposes, its properties relevant for interaction with the rest of the world are determined by a relatively small number of parameters (functions of the state) called *macroscopic* parameters  $u_1, \dots, u_m$ . The simplest system to consider is a certain quantity of gas in a container, with just a few macroscopic parameters: volume, temperature energy and pressure. Two of these can actually be deleted, since they are a function of the other two, but it is not necessary for our purposes to minimize the number of parameters.

Let us agree that energy will always be included among the parameters. The *first law of thermodynamics* is a consequence of a more general law of physics: it says that the energy of an isolated system does not change. The interaction of the system with the outside world involves some exchange of energy.

**8.2. Dynamics and volume.** According to the laws of classical mechanics, an isolated system undergoes an evolution described by a transformation group  $U^t$ . If at time  $t_0$ , the system was in state  $\omega$  then at time  $t + t_0$  it will be in state  $U^t\omega$ . The group  $U^t$  is generally given by a system of differential equations which, at least in the example of ideal gas given with coordinates and impulses, are called the Hamiltonian equations. For this case, Liouville's Theorem holds saying that the volume of a domain remains invariant under transformation under  $U^t$ . In most other cases also, a natural measure is found on  $\Omega$  that remains invariant under  $U^t$ ; we will call this measure the *volume*, and denote the volume of a set  $A$  by  $L(A)$ .

The law of energy conservation says that during the evolution of an isolated system, it is confined to a surface of the state space determined by the requirement that the energy is equal to a certain value. Therefore the volume measure to use will be actually obtained by restricting the original volume measure to a thin layer determined by the requirement that the value of energy is in a certain small interval, and normalizing. This measure, in the limit, is called the *microcanonical ensemble*.

If the present state is  $\varphi$  then the state at time  $t$  is  $U^t\varphi$  where the unitary operator  $U^t$  has the form  $e^{\frac{Ht}{i\hbar}}$ , and  $H$  is the Hamilton (energy) operator.

**8.3. The paradox of irreversibility.** If a mechanical system has a certain trajectory  $t \mapsto U^t\omega$  of its evolution from state  $\omega = a$  to state  $U^{t_1}\omega = b$  then there is also a trajectory of evolution from state  $b$  to state  $a$ . It is sufficient to reverse all the final velocities of all particles and the system will trace its trajectory backward. Therefore any evolution seems just as possible as the corresponding reverse evolution.

At the same time, the world seems to be full of irreversible phenomena. Imagine a container  $A$  of gas separated from an empty container  $B$  by a wall. After the removal of the wall much of the gas will occupy container  $B$  and the reverse physical process, when the gas collects itself spontaneously in part  $A$ , will never be seen. The reversibility of the equations seems to be in contradiction to irreversibilities of this kind.

In the main part of the paper, we explain how coarse-graining resolves the paradox.

**8.4. Equilibrium.** For an isolated system, we may want to call *equilibrium states* those macrostates that are relatively stable: the values of the macroscopic variables will undergo only small fluctuations in time. On reflection, however, this requirement must be weakened to hold only for *most* microstates within the macrostate (as measured by volume).

How can an equilibrium state be transformed into a different equilibrium state at all? Suppose that our system is a container of gas. The system can be combined with some other systems like a heat reservoir or a piston connected to a lever. Then some constraint is removed (e.g. an insulation is removed or a piston is allowed to move) changing the nature (the equations of motion) of the joint system by making its parts interdependent. When a new equilibrium is reached the constraints can be restored.

A useful formal way to describe a non-isolated equilibrium system is via an *ensemble* (probability distribution) invariant in time. Such a system does not have a deterministic evolution describable by a transformation  $U^t$  since the evolution of the system  $\omega$  is only the projection of the (possibly deterministic) evolution of a larger system  $(\omega, \xi)$ . The most popular ensemble for such cases is the Gibbs *canonical ensemble* (a generalization of the so-called Boltzmann distribution) defined by a density proportional to  $e^{-E(\omega)/kT}$ , with the energy  $E(\omega)$ , the Boltzmann constant  $k$  and the temperature  $T$ . It can be shown to be invariant if the system  $\omega$  is part of a large system  $(\omega, \xi)$  where  $\xi$  is a heat reservoir of temperature  $T$ .

**8.5. Boltzmann entropy.** Boltzmann defined the *entropy* of a macroscopic state  $a$  as the logarithm of the volume of  $\Gamma_a$ :

$$B(a) = \log L(\Gamma_a). \quad (8.1)$$

The most obvious problem with this definition seems to be its dependence on the particular choice and number of macroscopic variables and the precision with which we want to determine them. Indeed, another digit of precision will decrease the Boltzmann entropy of most states by about  $\log 10$ . The volumes in question are,

however, in the typical classical examples, so large, that such a small difference is negligible (especially, if we take into consideration that the actual definition also multiplies  $B(a)$  by the very small Boltzmann constant  $k$ ). We will also use the notation

$$B(\Gamma_a) = B(a)$$

since it is harmless here to identify a cell with its description.

*Example 8.1.* Let us consider  $n$  identical, indistinguishable molecules of gas in a container with rigid walls, and the following cell of phase space. The positions of the molecules are restricted to the container, of volume  $V$ . The total kinetic energy of the molecules is confined between the values  $K$  and  $K - \Delta K$ . The kinetic energy is  $\frac{1}{2m} \sum_{i=1}^{3n} p_i^2$  where  $p_i$  are the impulses  $mv_i$  and  $m$  is the mass of a molecule. The entropy for this “cell” consists of two terms. The first term is coming from the amount of freedom in choosing the positions of the molecules and is  $n \log V$ . The second term is the log volume of the set of points  $p = (p_1, \dots, p_{3n})$  such that  $K - \Delta K \leq \frac{1}{2m} \sum_i p_i^2 < K$ . For “reasonable” values of  $\Delta K$  this is about the same as the log volume of the set of impulse points  $p$  with total energy  $< K$ , because most of the volume of a high-dimensional ball of  $p$  with  $\sum_i p_i^2 < 2mK$  is near its surface. This volume is  $C_{3n}(2mK)^{3n/2}$  where  $C_n$  is the volume of an  $n$ -dimensional unit ball. This gives for entropy the value

$$n(\log V + \frac{3}{2} \log(2mK)) + \log C_{3n}. \quad (8.2)$$

It turns out that the correct volume measure to use is  $n!$  times smaller due to the fact that the molecules are not distinguishable. Therefore  $\log n!$  must be subtracted from here to get the correct value. This can be best seen when we consider the act of mixing two quantities of gas, with the same pressure and temperature. Depending on whether the two gases are of different kind or not the entropy of the mixture will or will not be greater than the sum of the two original entropies. (This is the so-called Gibbs paradox.)  $\square$

*Example 8.2.* Consider a container of ideal gas consisting of  $n$  molecules of the same kind which is partitioned, in our mind, into  $m$  compartments  $C_1, \dots, C_m$  where  $m$  is much smaller than  $n$ : say,

$$m < \sqrt{n}.$$

Ignore velocities, for simplicity. Let the macro variable  $n_i$  give the number of molecules in compartment  $C_i$ . Let  $\mathbf{n} = (n_1, \dots, n_m)$ . Let us also consider a second set of variables, the numbers  $i_1, \dots, i_n$  telling that molecule  $j$  is in compartment  $i_j$ .

The description  $\mathbf{i} = (i_1, \dots, i_n)$  is, of course, more detailed than the description  $\mathbf{n}$ , which is a function  $\mathbf{n}(\mathbf{i})$  of  $\mathbf{i}$ . Therefore  $\Gamma_{\mathbf{n}} = \bigcup_{\mathbf{n}=\mathbf{n}(\mathbf{i})} \Gamma_{\mathbf{i}}$ .

Since the molecules are all of the same kind, the dynamics, and therefore the measure  $L$  will certainly be invariant with respect to the exchange of molecules. Therefore if  $\mathbf{n}(\mathbf{i}) = \mathbf{n}(\mathbf{i}')$  then  $L(\Gamma_{\mathbf{i}}) = L(\Gamma'_{\mathbf{i}'})$ . Let us consider first the molecules distinguishable. Then  $L(\Gamma_{\mathbf{n}}) = N(\mathbf{n})p_{\mathbf{n}}$  where  $p_{\mathbf{n}}$  is the common volume of all  $\Gamma_{\mathbf{i}}$

with  $\mathbf{n}(\mathbf{i}) = \mathbf{n}$ , and

$$N(\mathbf{n}) = \frac{n!}{n_1! \cdots n_m!}.$$

If we also assume that all the compartments have the same shape and size then we get that  $p_{\mathbf{n}}$  does not depend on  $\mathbf{n}$ . Indeed,  $p_{\mathbf{n}} = V_m^n$ , the  $n$ -th power of the volume of a single compartment since after  $\mathbf{i}$  is given what is left to determine is only the exact position of each molecule in its compartment. The Boltzmann entropy of state  $\mathbf{n}$  is equal to

$$H(\mathbf{n}) = \log L(\Gamma_{\mathbf{n}}) \approx n(-\sum_j f_j \log f_j + \log V_m)$$

where  $f_j = n_j/n$ . Again, if the molecules are indistinguishable it is necessary to subtract  $\log n!$ .  $\square$

Let us call a system *volume-defined* (this is the kind of system given preference in textbooks) if the two macroscopic variables  $V$  (volume) and  $E$  (energy) determine the rest. Here,  $V$  can be given exactly, and  $E$  is given with some precision. The entropy

$$S(V, E)$$

is given as the phase volume of the set of phase points corresponding to coordinates in volume  $V$  and energy between  $E - \Delta E$  and  $E$ . Example 8.1 shows that this is approximately the volume of the set with the simple limitation that the energy is  $< E$  (the volumes of an  $n$ -dimensional ball and a shell of it are close).

Consider a system consisting of two parts 1, 2, such that both parts have separate descriptions, both microscopic and macroscopic. Then a cell corresponding to the joint system is the Cartesian product of its projection cells in systems 1 and 2, and its volume is also the product of the corresponding volumes. This way, the entropy of the joint system is the sum of the two entropies, so entropy is an *additive* function of the state.

Entropy was first defined in a so-called “phenomenological” model, using only the notions of temperature, heat and reversibility, without reference to atoms. We will point out the connection more closely a little later, but the last example already gives some indication.

**8.6. The growth of entropy.** Since entropy measures the amount of phase space occupied by a macrostate it sounds plausible that a system tends to be in states with high entropy but not obvious without further discussion. Anyway, this is exactly how classical thermodynamical systems are known to behave: according to one formulation of the *second law* of thermodynamics, entropy in isolated systems cannot decrease. There is also a “phenomenological” formulation, saying that heat cannot be transformed into work (leaving everything else unchanged). There are some other related entropy increase properties, so let us list all of those interesting us:

- in an isolated system, entropy cannot decrease;
- an isolated system undergoes an irreversible transformation exactly when its entropy actually increases;

- an equilibrium state of an isolated system is a maximum (or at least a local maximum) of entropy.

Just as the precise notion of reversibility had to be formulated statistically, the same can be expected for these entropy increase properties for Boltzmann entropy.

Formulation and proof are, of course, different and rigorous proof is unavailable for most realistic systems. The paper analyzes the heuristic arguments for the increase of Boltzmann entropy; here, we want to understand the ordinary physical consequences.

**8.7. Temperature and pressure.** Suppose that two volume-defined systems are brought into contact and are allowed to come into thermodynamic equilibrium without a change to their volume. Then according to the maximum entropy principle, the joint entropy is maximal, so a small reversible energy exchange between the two containers should not change it. If at this time therefore the amount of energy  $dE = dE_1 = -dE_2$  was transmitted reversibly from system 2 to system 1 and the entropy of system  $S_i$  increased by  $dS_i$  then  $dS_1 + dS_2 = 0$ ; it follows that

$$\frac{dS_1}{dE_1} = \frac{dS_2}{dE_2}.$$

We know from experience that two systems are in equilibrium only if their temperatures are the same, so we conclude that  $\frac{dS}{dE}$  depends only on the temperature. In Example 8.1 of an ideal gas, all its energy is kinetic and therefore, using (8.2), we have  $dS/dE = dS/dK = 3n/2K$ . The quantity  $K/n$  is the average kinetic energy of a molecule, and  $\frac{2}{3}\frac{K}{n}$  is sometimes used to define the *temperature*  $T$  of the (ideal, monatomic) gas. More generally, one can define temperature for volume-defined systems as follows. First we express energy via entropy and volume as  $E(S, V)$ . Then we define

$$T = \frac{\partial E(V, S)}{\partial S}.$$

We can also introduce *pressure* for a volume-defined system by

$$p = \frac{-\partial E(V, S)}{\partial V}$$

getting the equation

$$dE = TdS - pdV$$

for reversible changes in volume-defined systems. Here, the first term is the heat part  $dQ$  of the energy received by the system, and the second one is the work part. The equation turns into inequality for irreversible changes.

**Acknowledgement.** The article [8] by Joel Lebowitz and discussions with Shelly Goldstein provided my absolutely indispensable instruction on Boltzmann entropy without implying, however, the agreement of these persons with the ideas expressed here. The paradox of typicality was recognized due to a remark by V. G. Vovk.

## REFERENCES

- [1] C.H. Bennett, *Demons, engines and the second law*, Scientific American **255** (1987), no. 11, 108–116. [7.2](#)
- [2] Vasco Brattka and Gero Presser, *Computability on subsets of metric spaces*, Theoretical Computer Science (2002). [3](#)
- [3] Carlton M. Caves, *Information, entropy, chaos*, Physical Origins of Time Asymmetry (J. Pérez-Mercader and W. H. Zurek, eds.), Cambridge University Press, Cambridge, England, 1993. [6.4](#)
- [4] Peter Gács, *Exact expressions for some randomness tests*, Z. Math. Log. Grdl. M. **26** (1980), 385–394, Short version: Springer Lecture Notes in Computer Science 67 (1979) 124–131. [6.1](#)
- [5] ———, *Lecture notes on descriptional complexity and randomness*, Tech. report, Boston University, Computer Science Dept., Boston, MA 02215, 1988, <http://www.cs.bu.edu/~gacs/papers/ait-notes.pdf>. [1](#), [1.2](#), [3](#), [3.1](#), [3.1](#), [4.1](#), [4.1](#), [5.1](#), [5.2](#), [5.2](#), [6.1](#)
- [6] ———, *The Boltzmann entropy and randomness tests*, Proceedings of the Workshop on Physics and Computation, IEEE Computer Society Press, 1994, Extended abstract., pp. 209–216. [1](#)
- [7] Mathieu Hoyrup and Cristóbal Rojas. Computability of probability measures and Martin-Löf randomness over metric spaces. *Information and Computation*, 207(7):830–847, 2009. [3.1](#)
- [8] Joel L. Lebowitz, *Boltzmann's entropy and time's arrow*, Physics Today (1993), 32–38. [8.7](#)
- [9] L. A. Levin, *Randomness conservation inequalities: Information and independence in mathematical theories*, Information and Control **61** (1984), no. 1, 15–37. [6.1](#)
- [10] M. Li and P. M. B. Vitányi, *Introduction to Kolmogorov complexity and its applications (second edition)*, Springer Verlag, New York, 1997. [1](#), [2](#)
- [11] David Pollard, *A user's guide to measure-theoretic probability*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, U.K., 2001. [1.3](#)
- [12] V. G. Vovk and V. V. Vyugin, *On the empirical validity of the Bayesian method*, Journal of the Royal Statistical Society B **55** (1993), no. 1, 253–266. [1.2](#)
- [13] W. H. Zurek, *Algorithmic randomness and physical entropy*, Physical Review A **40** (1989), no. 8, 4731–4751. [1.2](#), [7.2](#)
- [14] A. K. Zvonkin and L. A. Levin, *The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms*, Russian Math. Surveys **25** (1970), no. 6, 83–124. [6.8](#)

COMPUTER SCIENCE DEPARTMENT, BOSTON UNIVERSITY

Email address: [gacs@bu.edu](mailto:gacs@bu.edu)