Séminaire Poincaré 2 (2003) 13 – 27

Entropy, a Protean Concept

Roger BALIAN
Service de Physique Théorique
CEA - Saclay
F-91191 Gif-sur-Yvette Cedex

Abstract. We review at a tutorial level the many aspects of the concept of entropy, in thermodynamics, information theory, probability theory and statistical physics. The consideration of relevant entropies and the identification of entropy with missing information enlighten the paradoxes of irreversibility and of Maxwell's demon.

The concept of entropy, invented one and a half century ago, has given rise to an immense literature. Under various guises it appears in many branches of physics, of mathematics, and even of most other sciences. Like the Greek divinity Proteus, it uses to change its shape so as to escape anyone who tries to grasp it, and it also presides prevision and deceit. To catch its meaning, we need to recognize it through its metamorphoses. We shall therefore review with an introductory scope some of its aspects, focusing on those which are relevant to statistical physics.

1 Macrophysics

1.1 Entropy in thermostatics

Entropy has first been introduced as a mathematical tool in the framework of thermodynamics. This science, born in the first half of the XIXth century, deals with the general laws that govern the transformations of systems. Actually, what is usually called "the laws of thermo*dynamics*" are physical constraints about the transformations allowed by physics that lead from one equilibrium state to another. These laws do not pertain to the dynamics of the processes, that is, their time-dependence, but only refer to their *initial and final state.* We therefore prefer here to speak of "thermostatics".

In the modern formulation of thermostatics [1], a physical or mechanical or chemical isolated system is analyzed into homogeneous subsystems referred to by the index a. The equilibrium state of each one is characterized by a set of *extensive variables* A_j such as volume, energy, number of constitutive particles. These quantities can be transformed or transferred but are *conserved* (the First Law and its extensions). For the composite system, the state variables are denoted as A_i, where $i = (j, a)$ is a double index indicating the nature j of the variable and the subsystem a. In the initial state the exchanges between subsystems are blocked and the system lies in a constrained equilibrium state with fixed A_i's. If we release some of these blockings by letting the subsystems interact, only some constraints on the set $\{A_i\}$ remain, in particular those imposed by the conservation laws. After some time, we reach a global equilibrium state. The extensive variables of each subsystem take new values, and the *Second Law*, in the following formulation, determines them. There exists for each subsystem a at equilibrium a function $S_{\rm th}(\{A_j\})$ of the extensive variables that characterize its state, its *thermodynamic entropy.* It is concave, extensive and additive: the entropy of the whole system is the sum of those of its subsystems. The overall equilibrium state is then the one that *maximizes the total entropy* $S_{\rm th}(\{A_i\})$, $i = (j, a)$ subject to the constraints imposed on the variables A_i by the initial state, the conservation laws and the allowed exchanges.

By stating that the entropy of an isolated system cannot decrease when it goes from a constrained equilibrium state to a less constrained one, the Second Law expresses the *irreversibility* of macroscopic processes. It also provides an upper bound for the efficiency of thermal machines, as was first shown by Carnot (1824). However its field of application covers not only thermal processes (as suggested by the word "*thermo*dynamics") but also other types of transfers, for instance, mixing of substances or chemical reactions: any such spontaneous process raises the entropy.

As a consequence of the Second Law, the partial derivative γ_i of the entropy $S_{\rm th}(\{A_i\})$ with respect to one of its variables A_i takes the same value for two systems that are in relative equilibrium with respect to the considered quantity A_i: if the two *intensive variables* γ_i's are equalized, no transfer of this quantity occurs although exchanges are allowed. The differential

$$dS_{\rm th} = \sum_i \gamma_i \, dA_i \tag{1}$$

exhibits γ_i and A_i as conjugate variables with respect to the entropy. In particular $\beta = \partial S_{\rm th}/\partial E$ takes the same value for two systems which have been brought to relative equilibrium by thermal contact. The *Zeroth Law* follows: relative thermal equilibrium between pairs of systems is a relation of equivalence, implemented by the existence of relative temperatures, which are any functions of β. In particular $\beta^{-1} = T$ is the absolute temperature.

1.2 Entropy in thermodynamics

Thermodynamics proper describes the time-dependence of physical, mechanical or chemical processes, in conditions of local quasi-equilibrium [2]. As above the system is analyzed into a set of subsystems a, each of which is at any time nearly at equilibrium. For continuous media, the subsystems a may be infinitesimal; more precisely each one should be a volume element large on the molecular scale but sufficiently small so that the densities of conserved extensive quantities (energy, particle, or momentum densities are nearly constant within it.

The instantaneous state is thus characterized, for continuous media, by the densities $\rho_j(\mathbf{r}, t)$ of the conserved quantities A_j. These variables are labelled by a continuous index $\mathbf{r}$ and a discrete index j. The *conservation laws* involve fluxes of local current densities $\mathbf{J}_j(\mathbf{r}, t)$ for each A_j, which satisfy

$$\frac{\partial \rho_j}{\partial t} + \mathrm{div}\mathbf{J}_j = 0 \ . \tag{2}$$

Entropy is defined at each time as in thermostatics. In an isolated system, its time-dependence is obtained from (1) and (2) as

$$\frac{dS_{\rm th}}{dt} = \int d^3\mathbf{r} \sum_j \mathbf{J}_j \,.\, \nabla\gamma_j \ , \tag{3}$$

which exhibits the *fluxes* $\mathbf{J}_j$ and the *affinities* $\nabla\gamma_j$ as conjugate variables with respect to the *dissipation* $dS_{\rm th}/dt$. One of the laws of the thermodynamics of non-equilibrium processes is the *Clausius–Duhem inequality*, which expresses the fact that entropy cannot be destroyed in any circumstance. For a continuous medium, one can associate with the entropy density ρ_S a current density of entropy $\mathbf{J}_S$, and dissipation is locally expressed by

$$\frac{\partial \rho_S}{\partial t} + \mathrm{div}\mathbf{J}_S \geq 0 \ , \tag{4}$$

which implies that (3) is never negative. This inequality sets constraints on the response equations that relate the fluxes $\mathbf{J}_j$ to the affinities $\nabla\gamma_j$, and hence through (2) on the dynamical equations for the densities ρ_j.

The Clausius–Duhem inequality should not be confused with the Second Law although both express an entropy increase. The inequality (4) holds *locally* and at *each time*, but it applies only to systems evolving sufficiently slowly, in a local equilibrium regime. The Second Law compares merely the *global* entropies of the *initial* and the *final* state, both being in equilibrium (but with more constraints in the initial state); during intermediate times, the system may well be far from equilibrium, for instance, if an explosive chemical reaction takes place, and its entropy need not be defined.

2 Information and probability

2.1 Entropy in communication theory

The theory of communication, founded in 1948 by Shannon and Weaver, aims at improving the transmission of signals. At first sight, its only analogy with thermodynamics seems to be the search for an optimum efficiency for the considered process, but we shall see that the entropies introduced in the two fields of science are actually related to each other. The basic idea, here, is that the amount of information transmitted by a message can be measured. One can then compare the performance of different transmission devices or of different coding systems, so as to minimize the duration of a transmission or the size of a memory.

Information is a concept related to *probability*. Indeed the information brought along by some message to a receiver is meaningful only if this message is extracted from a set of messages m that might a priori have been emitted. Before transmission the receiver ascribes to each message m a probability p_m. His *surprisal*, that is, the amount I_m of information that he gains by getting knowledge of some message m among the possible set should be a decreasing function $I_m = I(p_m)$ of the probability p_m: we gain very little information when being told about something we were practically certain of, whereas nearly unexpected messages are very informative. Moreover, a compound message mn consisting in two uncorrelated submessages m and n with respective probabilities p_m and q_n should carry an amount of information equal to the sum $I_m + I_n$. Since probabilities are then multiplicative while information is additive, the continuity and t! he! decrease of the function $I(p_m)$ imply

$$I_m = -\log p_m \ , \tag{5}$$

within a positive factor which defines the unit of information. This unit is the bit if the logarithm in (5) has base 2.

While I_m measures the amount of information *gained* by reception of the message m, Shannon's entropy measures the amount of information which is *missing before reception*, and which on average will be gained through reception. Since each message m has the probability p_m to reach the receiver, Shannon's entropy is defined by weighting I_m by the probability p_m:

$$S_{\mathrm{Sh}}(\{p_m\}) = \sum_m p_m I_m = -\sum_m p_m \log p_m \ . \tag{6}$$

For a number W of equally probable messages, for which $p_m = 1/W$ $(m = 1, 2, \cdots W)$, Shannon's entropy reduces to the celebrated expression of Boltzmann within a multiplicative factor:

$$S_{\mathrm{Sh}} = \log W \ . \tag{7}$$

Shannon's entropy thus characterizes the perplexity of the receiver before transmission of some message among the set $\{m\}$, or the average missing information.

A direct proof of (6) was given by Shannon who postulated a strong additivity property of $S_{\mathrm{Sh}}(\{p_m\})$: its expression should be additive, not only for compound events mn such that m and n are uncorrelated, but also if information is gained by steps. Suppose the events m are grouped in bunches, and suppose we are first informed about the occurrence of one among these bunches. The entropy should contain a corresponding first contribution, which is a function of the probabilities of the bunches. In addition, it should contain, weighted by the probability of each bunch, contributions associated with the conditional probability of each event among the considered bunch. The identity thus satisfied by $S_{\mathrm{Sh}}(\{p_m\})$ for different numbers of events then implies the form (6).

Shannon's entropy is a powerful tool for optimizing the amount of information involved in the transmission of messages or the storing of data. In both cases, the messages should be coded. Shannon and Weaver proved the existence of an *optimum coding*, which depends on the probabilities p_m and on a possible noise that may destroy part of the messages.

2.2 Entropy of a probability distribution

Shannon's expression (6) associates with any discrete probability set $\{p_m\}$ a number S, whether the index m labels messages or any other set of events. In the latter case S characterizes the *uncertainty* associated with this probability law, or its *spreading.* If the weights p_m are concentrated on some events, the uncertainty is lower than when they are spread. If all events are equally probable the uncertainty increases with their number. A quantitative evaluation of such an uncertainty is provided by (6) and (7).

Many properties of the function (6) enforce this interpretation of entropy as a measure of uncertainty. For a given number W of events, it is *minimum* and equal to 0 when one event is certain while all other ones have probability zero. It reaches its *maximum* (7) in the most random situation where all probabilities p_m are equal (to $1/W$). For compound events mn with joint probabilities P_{mn}, the separate probabilities of the m's and the n's are $p_m = \sum_n P_{mn}$ and $q_n = \sum_m P_{mn}$, respectively. The entropy is then *additive* for independent events, *subadditive* for correlated ones:

$$S(\{p_m\}) + S(\{q_n\}) = S(\{p_m q_n\}) \geq S(\{P_{mn}\}) \ . \tag{8}$$

(The equality holds only when $P_{mn} = p_m q_n$.) This inequality expresses that correlations carry some information. As another property of the entropy, consider a single set of events m, to which two different sets of probabilities p_m and q_m can be ascribed for two different statistical ensembles. If these two ensembles are mixed with non-vanishing weights λ and $(1-\lambda)$ into a single one, the new ensemble is characterized by probabilities $P_m = \lambda p_m + (1-\lambda) q_m$. The *concavity* property of entropy,

$$S(\{P_m\}) > \lambda S(\{p_m\}) + (1-\lambda) S(\{q_m\}) \ , \tag{9}$$

ensures that the uncertainty, as measured by S, is raised by mixing of populations.

2.3 Continuous probabilities

A difficulty arises for continuous distributions of probability. Consider a random real variable x, governed by a continuous probability density $p(x)$. In order to extend to this situation the definition (6), we split the x-axis into intervals x_m , $x_m + \Delta_m \equiv x_{m+1}$ and define p_m as the probability for x to lie between x_m and x_{m+1}. It would be natural to define $S(\{p(x)\})$ as the limit of $S(\{p_m\})$ when all Δ_m's tend to zero, but this quantity diverges. However, if all Δ_m's are equal, adding the constant $\log \Delta$ to $S(\{p_m\})$ provides a finite limit

$$S(\{p(x)\}) = -\int dx \ p(x) \log \ p(x) \ , \tag{10}$$

which defines the entropy of the probability distribution $p(x)$.

This quantity is not invariant under a change of the variable x. Whereas a linear transformation simply adds a constant to it, a non-linear transformation or equivalently a choice of unequal Δ_m's may modify (10) arbitrarily. Additional hypotheses are therefore needed to define unambiguously the entropy associated with a continuous distribution. For instance, translational invariance over x of the phenomenon characterized by a probability law $p(x)$ justifies the choice (10), which arises from a uniform partition $\Delta_m = \Delta$.

More generally, for continuous variables x lying on some manifold, the existence of an *invariance group* or a metric is necessary to define unambiguously $S(\{p(x)\})$. Actually such a group already existed implicitly for the discrete probabilities p_m involved in $S(\{p_m\})$, since all the events m were treated on the same footing: the very construction of Shannon's entropy implies that it is invariant under permutation of theses events.

3 Statistical physics

3.1 Von Neumann's entropy

Twenty years before Shannon, von Neumann introduced a similar expression, in the quite different context of quantum theory. There, probabilities are unavoidable as exemplified by the Heisenberg uncertainty relations. This irreducible intrusion of probabilities arises from the non-commutative nature of the observables, the algebraic objects that represent the physical quantities. To this replacement of usual random variables $A(m)$ by *non-commuting observables* $\hat{A}$ corresponds the replacement of probability distributions p by *density operators* $\hat{D}$, which are represented by matrices in the Hilbert space associated with the considered system. The expectation value of $\hat{A}$ is given by

$$<\hat{A}> = \mathrm{Tr}\ \hat{D}\hat{A}\ , \tag{11}$$

where the trace, taken on the Hilbert space, replaces the summation over the elementary events m.

Since any quantity at a given time can be represented in quantum mechanics under the form (11), our whole information at the considered time is represented in a probabilistic way by the density matrix $\hat{D}$. Anticipating the idea of Shannon, who associated the missing information (6) with the set of probabilities p_m, von Neumann associated the entropy

$$S_{\mathrm{vN}}(\hat{D}) = -\mathrm{Tr}\ \hat{D}\ln\hat{D} \tag{12}$$

with the density operator $\hat{D}$. We thus interpret (12) as a measure of the uncertainty associated with the description by $\hat{D}$ of the state of the system. When written in terms of the eigenvalues p_m of $\hat{D}$, the expression (12) is identical with (6). It can be constructed directly, starting from some natural axioms [3]. The invariance of S_{Sh} under the group of permutations of the events m is replaced here by the unitary invariance in the Hilbert space: all representations by matrices of quantum mechanics, deduced from one another by unitary transformations, are equivalent. Through its diagonalization, $\hat{D}$ behaves, for a finite quantum system, more like a discrete than like a continuous probability distribution, in spite of the continuity of the underlying unitary group. The definition (12) of its entropy thus does not involve the difficulties of (10).

The properties of von Neumann's entropy are similar to those of Shannon's entropy: additivity, subadditivity, concavity. They enforce the interpretation of (12) as a measure of the lack of information associated with the density operator $\hat{D}$. The minimum, $S_{\mathrm{vN}} = 0$ of (12) is reached when $\hat{D}$ is a pure state, that is, a projection onto a single wavefunction. Contrary to what happens for discrete probabilities, such pure states still involve uncertainties but are the best defined states allowed by quantum mechanics.

States $\hat{D}$ with non zero (positive) entropy occur in *quantum statistical physics* because the wave functions of systems made of a large number of particles cannot be fully determined. The entropy $S_{\mathrm{vN}}(\hat{D})$ then measures the uncertainty about such a state. The von Neumann entropy is also of interest in the framework of *quantum measurement theory.* Even for a quantum system having very few degrees of freedom, the measurement of some of its observables implies interaction with a macroscopic apparatus, which can be described only by means of quantum statistical mechanics. The equivalence between negentropy and information that we shall discuss in §5.2 below explains why von Neumann's entropy allows us both to measure the dispersion of a state $\hat{D}$, which will be identified with the thermodynamic entropy, and to evaluate the amount of information gained through a quantum measurement.

3.2 Entropy in classical statistical mechanics

In classical statistical physics, a state is described by the *density in phase* $D(\mathbf{r}_1, \mathbf{p}_1, \cdots \mathbf{r}_N, \mathbf{p}_N, t)$, which is a density of probability in the phase space of the N considered particles. If we were to define for the entropy of such a state an expression similar to (10) with integration over the $3N$-dimensional phase space, this entropy would depend on the measure of integration and thus would not be defined unambiguously. However, regarding a density in phase as a limit of a state $\hat{D}$ of quantum statistical physics, one can show that the trace in (11) and (12) tends to an integral over phase space (as would be directly introduced in classical statistical mechanics), but with the well-defined measure

$$\frac{1}{N!}\prod_{n=1}^{N}\frac{d^3\mathbf{r}_n\, d^3\mathbf{p}_n}{h^3},$$

where h is Planck's constant, and where the factor $1/N!$ arises from Pauli's principle about indistinguishability of particles.

As we shall see in § 4.2 the entropy $S_{\rm th}$ of thermostatics can be identified with von Neumann's entropy for equilibrium states. Two problems arising in classical statistical mechanics are then solved. On the one hand the *Gibbs paradox*, according to which the entropy of classical statistical mechanics does not seem extensive for a set of identical particles, is elucidated owing to the factor $1/N!$. On the other hand the limiting process which starts from quantum statistical mechanics generates the *absolute entropy*, without any additive constant, which satisfies the *Third Law*, or Nernst Law: the limit towards the zero absolute temperature of the absolute entropy vanishes.

3.3 The entropy as measure of disorder

Both Shannon's entropy (6) for a probability set and von Neumann's entropy (12) have been introduced as a *measure of missing information.* They do not appear as properties of the object under study in itself, but rather characterize the knowledge about it of its observers, who describe it by means of probabilities. These entropies thus have a partly subjective character, since they numerically characterize the *uncertainty of the observers.* Such a concept fits with the subjective interpretation of probabilities [4]. They should be regarded as mathematical tools for making consistent predictions, starting from the available information. The entropy (6) measures the quality of such predictions. In fact, since all observers placed in the same conditions should attribute the same probabilities to a set of possible events, probabilities are intersubjective rather than subjective.

Likewise, in statistical mechanics, a state represented by a density operator collects our information on some system. It does not describe this system in itself, but as a sample chosen among an *ensemble* of systems all prepared by the same procedure. This ensemble may be real, or may just be a set of thought similar copies, not completely identical but all having the same known features.

We may alternatively interpret the von Neumann entropy as a *measure of disorder* of the state described by the (probabilistic) density operator $\hat{D}$. Actually the concept of disorder should be identified with that of uncertainty: when we say that a fully mixed pack of cards is disordered, it only means that we know nothing about their ordering; for a conjurer who is aware of this ordering, there is no disorder in the pack. Maxwell already wrote: "Confusion, like the correlative term order, is not a property of material things in themselves, but only in relation to the mind who perceives them." Entropy allows us to make this idea quantitative.

4 Maximum statistical entropy and applications

4.1 The maximum entropy criterion

According to this interpretation, the assignment of probabilities to a set of events should depend on the data available to the observer, but should be made in a consistent way so that any other observer makes the same inferences starting from the same data. In statistical physics the problem is the same: which density operator $\hat{D}$ should we assign to describe the state of a system belonging to some ensemble characterized by a set of macroscopic data?

When nothing is known but the list of the W possible events m, it is natural to resort to Laplace's *principle of indifference* or of insufficient reason, and to assign the same probability $p_m = 1/W$ to all these events, as is done in the theory of games. Likewise, if no information is available about the spin $\frac{1}{2}$ of a particle, the obviously unbiased choice for its density operator is $\hat{D} = \frac{1}{2}\hat{I}$. This state describes an unpolarized spin, the expectation value of which vanishes in any direction. This principle relies on the idea that any other probability distribution would introduce bias, by favourizing without any reason the prediction of some events to which larger probabilities would have been assigned.

The maximum entropy principle [5] extends this idea to situations where some probabilistic information is given, in the form of expectation values. Suppose that, for instance in statistical mechanics the expectation values $A_i \equiv < \hat{A}_i >$ of some observables $\hat{A}_i$ are known. From this information we wish to infer the expectation values of other quantities. To this aim we need to assign a density operator $\hat{D}$ to the system. It should of course satisfy the constraints

$$\text{Tr } \hat{D}\hat{A}_i = A_i \tag{13}$$

about the known quantities, but they do not suffice to determine $\hat{D}$. Whithin this allowed set, consider two density operators $\hat{D}_1$ and $\hat{D}_2$ such that $S_{\text{vN}}(\hat{D}_1) < S_{\text{vN}}(\hat{D}_2)$. This inequality means that $\hat{D}_1$ contains more information that $\hat{D}_2$, by an amount $S_{\text{vN}}(\hat{D}_2) - S_{\text{vN}}(\hat{D}_1)$. However the density operator that should describe the state of the considered system (or rather of the considered ensemble of systems) should not carry more information than what is contained in the data A_i. Thus $\hat{D}_1$, which contains more information than $\hat{D}_2$, is certainly biased. Hence we are led to select for $\hat{D}$ the density operator that maximizes $S_{\text{vN}}(\hat{D})$, subject to the constraints (13) for the set $\hat{A}_i$. This maximum entropy criterion means that we describe the situation by means of the *least biased* density operator, or the *mo! st! uncertain*, or the *most disordered*, among the set compatible with the available data.

The result of this procedure is found by introducing *Lagrange multipliers* γ_i for each constraint on A_i and ζ for the normalization. We thus have to express that

$$\delta \left[S_{\text{vN}}(\hat{D}) - \sum_i \gamma_i \text{Tr } \hat{D}\hat{A}_i - \zeta \text{ Tr } \hat{D} \right]$$

vanishes for any Hermitean variation $\delta\hat{D}$ of $\hat{D}$. Letting $\zeta = \ln Z - 1$ and regarding Z as a function of the variables γ_i, we thus obtain for $\hat{D}$ the generalized Gibbsian distribution

$$\hat{D} = \frac{1}{Z} \exp\left(- \sum_i \gamma_i \hat{A}_i \right) , \qquad Z(\{\gamma_i\}) \equiv \text{Tr} \exp\left(- \sum_i \gamma_i \hat{A}_i \right) , \tag{14}$$

where the multipliers γ_i are related to the data A_i through

$$A_i = -\frac{\partial}{\partial \gamma_i} \ln Z , \tag{15}$$

a consequence of (13) and (14). The value of the von Neumann entropy (11) of the state (14), larger than the entropy associated with any other $\hat{D}$ compatible with the data A_i, is given by

$$S_{\text{vN}}(\{A_i\}) = \ln Z + \sum_i \gamma_i A_i . \tag{16}$$

Altogether, if the von Neumann or the Shannon entropy, which measures missing information or disorder, is also regarded as a measure of bias, the least biased inferences based on the knowledge of the set A_i should rely on the probability distribution (14), (15). The unicity of this distribution is ensured by the concavity of entropy.

The validity of the maximum entropy criterion has been questioned. For Shannon's entropy, its use can be justified by direct approaches where requirements on the consistency of the inference procedure [6] lead to the same result as the maximization of S_{Sh}, provided we deal with discrete events m. In quantum statistical physics, an alternative justification for (14) was given, based on the idea that the expectation value A_i of $\hat{A}_i$, expressed by (13) in terms of the density operator $\hat{D}$ of the considered system, may be identified with the mean value $\sum_\alpha A_i^\alpha/\mathcal{N}$ over an ensemble of $\mathcal{N}$ analogous systems $\alpha = 1, 2, \ldots \mathcal{N}$ described by $\hat{D}$, in the limit of large $\mathcal{N}$ [3, 7]. This identification is consistent with the equivalence of the two interpretations of probabilities, either a tool for predictions or a set of frequencies [4].

The maximum entropy criterion is currently used in various contexts. We resume below some of its outcomes in statistical physics.

4.2 Equilibrium statistical physics

Equilibrium statistical physics underlies thermostatics at the microscopic level. It allows in particular to derive the Laws of thermostatics as statistical consequences of microphysics. It is based on a description of the state of a system by means of the density operator $\hat{D}$ constructed as above from the same macroscopic equilibrium data A_i as in thermodynamics.

Consider first a homogeneous piece of material contained in a volume V. Its macroscopic equilibrium states are characterized by its number N of particles (we assume for simplicity that they are all of the same kind) and by their energy. At the microscopic level these two quantities are the only data (13); they are identified with the expectation values of the particle number operator $\hat{N}$, and of the Hamiltonian $\hat{H}$, respectively. We therefore describe this piece of material by the grand canonical density operator (14), where the observables $\hat{A}_i$ are here $\hat{N}$ and $\hat{H}$. From (15), (16) we can derive for S_{vN} the same equation (1) as for S_{th}.

Consider now more generally a compound system as for the Second Law of thermostatics, but described from the viewpoint of statistical physics. In the initial state the data A_i (where $i = (j, a)$ now denotes both the nature j of the variable and the subsystem a) are the extensive conserved variables of each subsystem; they may take arbitrary values since the exchanges are blocked. The overall equilibrium density operator factorizes as a product of contributions associated with each subsystem, e.g., grand canonical distributions, with multipliers γ_i ($i = (j, a)$) taking independent values for each subsystem a. Hence its von Neumann entropy is the sum of the entropies (16) of all parts. The determination of the final equilibrium density operator may then be performed by maximizing S_{vN} in *two steps.* The first step, that we just described, leads to a function $S_{\mathrm{vN}}(\{A_i\})$ of the extensive conservative variables of all the subsystem! s,! which we have identified with those entering the entropy S_{th} of thermostatics. Then the second step is exactly the same as in the Second Law of thermostatics; to wit, this function $S_{\mathrm{vN}}(\{A_i\})$ is maximized as function of the variables A_i, subject to the remaining constraints. Since the *entropy of thermostatics* S_{th} is defined in a unique fashion within a multiplicative constant, we can altogether identify it with the sum of the *von Neumann entropies* of the subsystems, each one being evaluated *at equilibrium.* The Second Law thus appears simply as this second step of the maximum entropy criterion.

For a homogeneous system, the Legendre transformation (15,16), where the expectation values $<A_j>$ are identified with the extensive variables A_j of thermostatics such as E and N and S_{vN} with S_{th}, justifies the identification of the *Lagrange multipliers* γ_j in (14) with the *intensive variables* defined by (1). In particular the multiplier associated with $\hat{H}$, noted β, is identified with the inverse of the absolute temperature.

Actually, the Second Law alone defines the entropy S_{th} not only within a multiplicative constant, but also within an additive constant. This arbitrariness is lifted by the Third Law. The above definition $S_{\mathrm{vN}}(\{A_i\})$ of entropy issued from statistical physics involves no additive constant, and it *implies the Third Law* since (16) tends to zero (for usual materials) in the zero-temperature limit $\beta \to \infty$.

As regards the multiplicative constant it depends on the choice of units for S_{th}. The natural choice, where $S_{\mathrm{th}}(\{A_i\}) = S_{\mathrm{vN}}(\{A_i\})$ is dimensionless, leads to temperatures measured in energy units and taking very small values, whereas entropies are very large since the uncertainty S_{vN} increases as the number N of elementary constituents of the material. On the other hand, measuring temperatures in kelvin, a unit defined from the triple point of water, provides $S_{\mathrm{th}} = kS_{\mathrm{vN}}$ where $k \simeq 1.38 \times 10^{-23} JK^{-1}$ is Boltzmann's constant.

The idea of reducing thermostatics to equilibrium statistical mechanics, and of interpreting the entropy as a measure of the disorder or the uncertainty at the microscopic scale after introduction of probabilities, can be traced back to Boltzmann who dealt with the kinetic theory of classical gases. As we already noted, the advent of quantum mechanics has paradoxically brought up simplifications: Owing to the introduction of discreteness, S_{vN} is defined as unambiguously as S_{th}. Accordingly, even for a classical system like a monatomic gas, Planck's constant enters the expression of the thermostatic entropy.

4.3 Relevant entropies and dissipation

A typical problem of non-equilibrium statistical physics is the prediction of the expectation value of specified quantities at time t from a set of initial data. The initial state $\hat{D}(t_0)$ is assigned from these data as in (14). The observables $\hat{A}_i$ are, however, no longer constants of the motion as in thermostatics, so that $\hat{D}(t_0)$ does not commute with the Hamiltonian $\hat{H}$. The density operator $\hat{D}(t)$ at the time t is then found by solving the Liouville–von Neumann equation

$$i\hbar\frac{d\hat{D}}{dt} = [\hat{H}, \hat{D}] \ , \tag{17}$$

and its construction provides the required quantities. However, the detailed description by means of $\hat{D}(t)$ involves a huge amount of variables both without interest and unpracticable. We therefore select some set of *relevant observables* $\hat{A}_i$ which are expected to govern the evolution. Their choice is guided by phenomenological macroscopic approaches; we shall give two examples below. Most often they are the slowest variables, whereas the remaining irrelevant variables evolve on a much shorter time scale. Their elimination may be performed, at least formally, by the projection method of Nakajima and Zwanzig [8] that we now sketch.

The relevant variables $A_i(t)$ are deduced at each time from $\hat{D}(t)$ through (13). The observables initially given belong to the relevant set $\{\hat{A}_i\}$, so that the initial density operator $\hat{D}(t_0)$ has the form (14); if the set $\{\hat{A}_i\}$ contains observables whose initial value $A_i(t_0)$ is not specified, the corresponding multipliers vanish. However, at later times, the solution $\hat{D}(t)$ of eq. (17) has no reason to retain the same exponential form involving only the relevant observables $\hat{A}_i$. Starting then from the set $A_i(t) = \mathrm{Tr}\ \hat{D}(t)\hat{A}_i$ of relevant variables at the time t, we can associate with them by means of the maximum entropy criterion a *reduced density operator* $\hat{D}_0(t)$ of the form (14), where the multipliers $\gamma_i(t)$ depend on time. Their value is determined by the conditions $A_i(t) = \mathrm{Tr}\ \hat{D}_0(t)\hat{A}_i$ where the $A_i(t)$ are deduced ! fr! om $\hat{D}(t)$, and where the right-hand side satisfies eq. (15). Thus $\hat{D}(t)$ and $\hat{D}_0(t)$ are equivalent as regards the relevant variables, but $\hat{D}_0(t)$ has the *maximum entropy*, given by (16) in terms of the set $\{A_i(t)\}$. In other words both states $\hat{D}(t)$ and $\hat{D}_0(t)$ account for the information $A_i(t)$, but $\hat{D}_0(t)$ involves no more information. The difference $S_{\mathrm{vN}}[\hat{D}_0(t)] - S_{\mathrm{vN}}[\hat{D}(t)]$ measures an extra amount of information about the irrelevant variables, included in $\hat{D}(t)$ which keeps full track of our knowledge of the initial data $\{A_i(t_0)\}$.

The quantity $S_{\mathrm{vN}}[\hat{D}_0(t)]$ defines the *relevant entropy* $S(\{A_i\})$ associated with the quantities $A_i(t)$. It depends on the chosen set $\hat{A}_i$ of observables, and characterizes the information which is missing at the time t when one follows only the evolution of their expectation values. By construction we have $S(\{A_i\}) \geq S_{\mathrm{vN}}[\hat{D}(t)]$. On the other hand, we find from (12) and (17) that

$$i\hbar\frac{dS_{\mathrm{vN}}(\hat{D})}{dt} = -\mathrm{Tr}\ [\hat{H}, \hat{D}] \ln \hat{D} = 0 \ , \tag{18}$$

and hence that $S_{\mathrm{vN}}[\hat{D}(t)]$ is constant. Thus the relevant entropy satisfies $S(\{A_i(t)\}) \geq S(\{A_i(t_0)\})$, which means that some information about the relevant variables is *lost during the evolution.* On the other hand the constancy of $S_{\mathrm{vN}}[\hat{D}(t)]$ implied by (18) means that our information about *all* possible observables, which is described by $\hat{D}(t)$, is conserved by the Hamiltonian evolution (17).

¿From a geometric viewpoint, this *reduction of the description*, which associates with $\hat{D}(t)$ the less detailed distribution $\hat{D}_0(t)$, is in the space of states $\hat{D}$ a *projection* onto the *manifold of reduced sates* (14), in the direction of constant relevant variables A_i. Actually, the space of states $\hat{D}$ has not only a structure of vector space, but can also be regarded as a Riemannian space, where a natural *metric generated by entropy* [8] according to

$$ds^2 = -d^2 S_{\mathrm{vN}} \tag{19}$$

allows to define distances between states and angles. The projection from $\hat{D}(t)$ to $\hat{D}_0(t)$ then appears as an *orthogonal projection.*

Our dynamical problem of non-equilibrium statistical mechanics is then reduced to the search for the projection of the detailed trajectory of $\hat{D}(t)$, which is generated by (17) and by an initial condition $\hat{D}(t_0) = \hat{D}_0(t_0)$. The *projection method* provides for $\hat{D}_0(t)$ an integro-differential equation that relates $d\hat{D}_0/dt$ to $\hat{D}_0$ at the same time and also at earlier times. The corresponding *memory term* results from the *elimination of the irrelevant variables.* We shall not write here this equation of motion which expresses $\hat{D}_0(t)$ in terms of its past history between t_0 and t; its explicit form can be found in ref. [8].

If the set $\{\hat{A}_i\}$ has been suitably chosen, the memory time is short so that $\hat{D}_0(t)$ is governed within a good approximation by a mere differential equation. The memory term is replaced by a *dissipative term*, which prevents $d\hat{D}_0/dt$ from being generated by an effective Hamiltonian, and which thus allows for a time-dependence of the relevant entropy $S(\{A_i\})$.

In physical situations where the selection of relevant variables thus leads to a differential equation for $\hat{D}_0(t)$, $\hat{D}_0(t+\Delta t)$ depends only on $\hat{D}_0(t)$, for Δt small but large compared to the memory time associated with the evolution of the irrelevant variables. We would therefore have found the same $\hat{D}_0(t+\Delta t)$ by starting the evolution from $\hat{D}(t) = \hat{D}_0(t)$, letting $\hat{D}$ evolve according to (17), then projecting $\hat{D}(t+\Delta t)$ on the set of relevant states. Hence we have $S(\hat{D}_0(t+\Delta t)) \geq S(\hat{D}(t+\Delta t)) = S(\hat{D}(t)) = S(\hat{D}_0(t))$. Altogether, for a choice of the set $\{\hat{A}_i\}$ such that the memory time is short, we find $S(\{A_i(t+\Delta t)\}) \geq S(\{A_i(t)\})$; this inequality holds along the motion of $\hat{D}_0(t)$: the *relevant entropy cannot decrease*, whereas the von Neumann entropy remains constant. This means that *information about the relevant observables is continuously lost* towards the irrelevant ones during the evolution, and this loss is *irretrievable.* In other words the *dissipation* $dS(\{A_i\})/dt \geq 0$ measures the rate at which a *loss of order* takes place in the relevant variables $A_i(t)$.

We illustrate below this general approach by two examples.

4.4 The thermodynamic entropy as a relevant entropy

For the processes described by non-equilibrium thermodynamics, the variables A_i governed by the macroscopic equations are the conservative variables of each subsystem or of each volume element. For instance, for a fluid, they are the densities $\rho_j(\mathbf{r})$ of particles, of energy and of momentum at each point. They are identified at the microscopic level with the relevant variables A_i $(i = j, \mathbf{r})$ of the projection method. It is the conservative nature of these variables which ensures that they evolve over much longer time scales than the other variables which are discarded. Indeed, without couplings between subsystems or volume elements, the thermodynamic variables A_i or $\rho_j(\mathbf{r})$ would remain constant. For sufficiently weak couplings and for small affinities $\nabla\gamma_j$, the currents $\mathbf{J}_j$ are weak and from (2) it appears that the motion of the variables A_i is slow.

The parameters γ_i of the reduced density operator $\hat{D}_0(t)$ given by (14) are then identified at each time with the local intensive variables $\gamma_j(\mathbf{r})$ $(i = j, \mathbf{r})$ associated with the set $\rho_j(\mathbf{r})$, which are directly related to the local temperature, chemical potential and hydrodynamic velocity.

Since $\hat{D}_0$ is factorized into a product of contributions from each subsystem, the relevant entropy associated with it is the sum over all subsystems of the entropy of equilibrium statistical mechanics, that we already identified with the entropy of thermostatics. The relevant entropy relative to the present choice of variables A_i is therefore the same as the entropy of macroscopic non-equilibrium thermodynamics. Its increase gives a microscopic justification for the Clausius–Duhem inequality (4) if the time-scale of the thermodynamic variables is longer than that of the microscopic variables that have been projected out.

4.5 Boltzmann's entropy

Boltzmann's equation for gases describes their dynamics in terms of the density of particles $f(\mathbf{r}, \mathbf{p}, t)$ in the *single-particle* phase space: $f(\mathbf{r}, \mathbf{p}, t)\ d^3\mathbf{r}\ d^3\mathbf{p}$ is the expectation value of the number of particles lying in the volume element $d^3\mathbf{r}\ d^3\mathbf{p}$ of this phase space at time t. This quantity evolves according to Boltzmann's equation

$$\frac{\partial f(\mathbf{r}, \mathbf{p}, t)}{\partial t} + \frac{\mathbf{p}}{m} \cdot \nabla_{\mathbf{r}} f(\mathbf{r}, \mathbf{p}, t) = \mathcal{I}\{f\}\ , \tag{20}$$

where the left-hand side accounts for the drift of f due to the free motion of the particles (with mass m). The right-hand side is the collision integral; it is quadratic in f and describes the change in f due to interparticle collisions. Since the collisions are nearly local in space and in time, $\mathcal{I}$ involves $f(\mathbf{r}, \mathbf{p}', t)$ at the same point and time as in the left-hand side, but involves integrations over the two momenta entering the two f's. The establishment of (20) requires that $f(\mathbf{r}, \mathbf{p}, t)$ as function of $\mathbf{r}$ and t varies slowly compared to the sizes of the particles and the duration of the collisions.

Boltzmann proved that his equation obeys the *H-theorem*: the quantity

$$H(t) \equiv \int d^3\mathbf{r}\, d^3\mathbf{p}\ f(\mathbf{r},\mathbf{p},t) \ln f(\mathbf{r},\mathbf{p},t) \tag{21}$$

is a decreasing function of time.

In order to recover Boltzmann's equation and H-theorem from microphysics, we start from the most detailed description, in terms of the *density in phase*, that is the probability density $D(\mathbf{r}_1, \mathbf{p}_1, \cdots \mathbf{r}_N, \mathbf{p}_N, t)$ in the N-particle phase space. Its time-dependence is generated by the Liouville equation, the classical limit of (17), where collisions are described by the interparticle potential. We take as the set of relevant variables the values of the single-particle density $f(\mathbf{r},\mathbf{p},t)$, obtained by integrating $D(\mathbf{r}_1, \mathbf{p}_1, \cdots \mathbf{r}_N, \mathbf{p}_N, t)$ over the phase space of $N-1$ particles. The index i stands here for $\mathbf{r}, \mathbf{p}$. In the reduced density $D_0(\mathbf{r}_1, \mathbf{p}_1, \cdots \mathbf{r}_N, \mathbf{p}_N, t) \propto f(\mathbf{r}_1, \mathbf{p}_1, t) \times \cdots \times f(\mathbf{r}_N, \mathbf{p}_N, t)$, all correlations that possibly exist in D are eliminated. (These correlations are generated at each collision.) ! Th! e projection method then produces an equation of motion for D_0 or equivalently for $f(\mathbf{r},\mathbf{p},t)$, in which the memory-time is the duration of a collision. The neat separation of time scales allows us to neglect this duration, and f is thus governed by a differential equation that reduces to Boltzmann's equation with its instantaneous collision term.

The relevant entropy S_1 associated with the relevant variables $f(\mathbf{r},\mathbf{p},t)$ is identified with Boltzmann's entropy, which is easily shown to equal $-H(t)$ within a multiplicative and an additive constant. The H-theorem is recovered as a special case of the increase of relevant entropies in regimes where the memory is short. Here the increase of the Boltzmann entropy S_1 is interpreted as a loss of information which results from the fact that, although correlations are created by each collision, these correlations have no effect on the subsequent evolution because two particles which underwent a collision have little chance to meet again. The evolution of $f(\mathbf{r},\mathbf{p},t)$ is the same as if all these *correlations are forgotten.*

Boltzmann's entropy S_1 related to (21) should not be confused with the thermodynamic entropy of subsection 4.4: Even for a gas of non-interacting particles, S_1 coincides with S_{th} only when at each point $f(\mathbf{r},\mathbf{p},t)$ behaves as a Gaussian in $\mathbf{p}$; otherwise S_{th} is larger than S_1 for the same values of the thermodynamic variables $\rho_j(\mathbf{r},t)$. Nor should the H-theorem be confused with the Second Law or with the Clausius–Duhem inequality. On the one hand, Boltzmann's equation deals with gases only. On the other hand, it holds beyond local equilibrium, in ballistic regimes that cannot be described in terms of the thermodynamic variables. Indeed, in a thermodynamic or hydrodynamic regime, local equilibrium implies that $f(\mathbf{r},\mathbf{p},t)$ has at each point $\mathbf{r}$ a Maxwellian form, that is, behaves as a Gaussian in $\mathbf{p}$. The reduced description provided by $f(\mathbf{r},\mathbf{p},t)$, a function of 6 variables, is! m! ore detailed than the reduced hydrodynamic description in terms of 5 functions of 3 variables, the densities of energy, particles and momentum in ordinary space (or equivalently the local temperature, chemical potential and hydrodynamic velocity).

Boltzmann's entropy accounts for the uncertainty associated with the sole knowledge at each time of the single-particle density $f(\mathbf{r},\mathbf{p},t)$ of a gas enclosed in a vessel. For sufficiently large times, a local equilibrium, then the global equilibrium corresponding to the initial values of the total particle number and energy are attained. At these stages the Boltzmann entropy grows so as to reach the entropy of non-equilibrium thermodynamics, then of thermostatics. Before the thermodynamic regime is settled, starting from the full density in phase $D(\mathbf{r}_1, \mathbf{p}_1, \cdots, \mathbf{r}_N, t)$, we can derive at the microscopic scale two-particle, three-particle, ... correlation functions. More and more detailed reduced descriptions are thus obtained by dropping all the correlations of more than n particles [9]. Boltzmann's description corresponds to $n = 1$, with Boltzmann's entropy S_1, and we find a *hierarchy of relevant entropies* $S_2, \cdots S!_n! \cdots$ such that

$$S_{\text{th}} \geq S_1 \geq S_2 \geq \cdots \geq S_n \geq \cdots \geq S_{\text{vN}}(\hat{D}) \ . \tag{22}$$

Starting all from $S_{\text{vN}}(\hat{D}(t_0)) = S_{\text{vN}}(\hat{D}(t))$, the entropies S_N all increase as a function of time, but later and later; all of them finally reach the thermostatic entropy, larger than the whole set.

5 Two paradoxes

5.1 The paradox of irreversibility

Soon after the birth of the kinetic theory of gases, Zermelo and Poincaré pointed out the paradox of irreversibility: although the microscopic evolution (17) is invariant under time-reversal, the macroscopic evolutions are not. As regards thermodynamic variables, viscosity and thermal conduction are irreversible phenomena. In the dynamics of gases, the H-theorem also exhibits and *"arrow of time"*. We gave a general explanation to such irreversible behaviours: whenever the memory-time is short, the dissipative term in the reduced description generates an increase of the relevant entropy, the signature of irreversibility. This increase is a *statistical* property; it measures a loss of information towards the irrelevant variables.

This argument relies on the irretrievable nature of this loss. In principle, the equation of motion (17) does not prevent, for a finite system, some order hidden within the irrelevant degrees of freedom to surge back into the relevant ones, and to show off as a decrease of the relevant entropy. If we forget about the observer and consider an initial microstate *completely defined* at the initial time, it remains completely defined at all times so that nothing seems to prevent this complete order from showing off. For instance, if the N particles of a gas are all grouped at the initial time in the left half of a container, they will fill the full vessel after some delay. However the equations of motion are reversible and allow the converse evolution, where the particles occupying the full container spontaneously come together in its left half. Why do we never observe such a behaviour in practice? The reason why it does not occur is that we deal with systems compos! ed! with an extremely large number N of particles. Poincaré's *recurrence time*, after which a system governed by an equation such as (17) returns in the vicinity of its initial states, is then immensely large, even compared to the age of the Universe.

Boltzmann's explanation of the paradox of irreversibility is again probabilistic although we consider here the single trajectory of a well defined configuration. It relies on an analysis of the initial state. In the above example of the expansion of a gas, let us discretize the positions and momenta of the particles so as to count the configurations (in agreement with quantum mechanics). Denote as W the number of compressed states, such that the N particles lie in the left half of the vessel. The total number of states in the full vessel is $2^N W$, and in nearly all of them the particles are spread all over. Among the latter states, those for which, after some time τ, the particles are gathered in the left half are in one-to-one correspondence with the W compressed states. Hence only a very tiny proportion 2^{-N} of the spread states gives rise to the anomalous grouping process. Since N is of the order of the Avogadro number 6×10^{23}, we have no chanc! e ! whatsoever to observe such a process.

More generally, consider an irreversible thermodynamic process for which the entropy increases by ΔS_{th}. In the above example $\Delta S_{\text{th}} = kN \ln 2$. The corresponding increase of the relevant von Neumann entropy is $\Delta S_{\text{th}}/k$ where k is Boltzmann's constant $1.38 \times 10^{23} JK^{-1}$ and where ΔS_{th} has the order of JK^{-1}. Noting from (7) that the ratio between the numbers W of initial and final microstates is of the order of $e^{-\Delta S_{\text{th}}/k}$, we see that anomalous processes with reversed time are not forbidden in principle but that the *initial configurations* from which they might arise *are completely improbable*, as an exponential of -10^{23}, among the whole set of configurations involved in the most disordered thermodynamic state.

Nevertheless, experiments exist in which many microscopic variables that appear as irrelevant at first sight can actually be controlled. The most celebrated ones are *spin echo* experiments, in which only the total magnetic moment of a material containing N spins is observed, and only an applied magnetic field can be controlled. Under normal circumstances, a relaxation of the total magnetic moment occurs; the magnitude of this moment decreases, so that the relevant disorder increases as expected. However, some time after the total moment has vanished, it is possible through suitable pulses of the applied field to manipulate the individual spins in such a way that the order hidden in their correlations manifests itself by an increase of the magnetic moment. The conceptual interest of such experiments (which also have practical applications in NMR) is to demonstrate that some initial information which is apparently lost within microscopic degrees of freedom may in some e! xc! eptional cases be retrieved, and that the choice of relevant variables should depend on the circumstances.

5.2 Equivalence of information and negentropy

Another paradox has fed for one century many discussions about the meaning of entropy, the thought experiment of *Maxwell's demon* (1867). Two vessels A and B with the same volume are filled with a gas, initially at the same density and temperature. They communicate through a hole that the demon may open or close at will with negligible work. Whenever a particle arrives from A towards B, the demon lets it pass, but he stops the particles arriving from B towards A. The density thus increases on the side A, that eventually all the N particles reach. The Second Law seems violated, since the entropy $S_{\rm th}$ has decreased by $kN \ln 2$.

However, in order to operate, the demon must *know* on which side each particle lies. He must therefore have gained an amount of information equal according to (5) or (7) to $S_{\rm Sh} = N \log 2$. If entropy and information are measured in the same unit, it can be shown [10] that information may be transformed into negentropy, with possible losses. One may let the entropy of a system decrease by some amount, provided one uses to this aim at least the same amount of information.

Conversely, how is this information acquired? It can be shown that the measurements required in this purpose involve macroscopic physical devices which undergo observable transformations, and that in these transformations the entropy must increase by an amount at least equal to the information gained. We indicated that the changes of the entropy (12) during measurement processes were among the incentives of von Neumann when he introduced this expression. The equivalence between negentropy and information which is exhibited by transformations in either direction enforces the interpretation of the thermodynamic entropy in terms of $S_{\rm vN}$ issued from the maximum entropy criterion.

Thus, altogether, if the demon is automatized and if we do not consider the gain of information in the intermediate steps of the process, the whole system including the vessels, the mechanism of gate opening and closing and the measuring device that governs this operation, obeys the Second Law. The decrease of entropy of the gas is compensated for by at least the same increase of entropy in the measuring device. Proteus has exorcized Maxwell's demon, owing to his two shapes, entropy and information.

6 Other entropies

We list below various other entropies [3] which, in one way or another, measure disorder or missing information.

6.1 Relative entropy

The relative entropy of Kullback and Leibler,

$$S(p|q) = \sum_m p_m \log(\frac{p_m}{q_m}) \ , \tag{23}$$

characterizes the gain of information when prior probabilities q_m are replaced by an actual probability set $\{p_m\}$. Its introduction is natural in the continuum limit when no invariance exists to support (10); in this case a prior distribution $q(x)$ replaces the translationally invariant integration measure. Its quantum equivalent

$$S(\hat{D}_2|\hat{D}_1) = \mathrm{Tr}\ \hat{D}_2(\ln \hat{D}_2 - \ln \hat{D}_1) \tag{24}$$

is currently used to build mean-field approximations in which the exact $\hat{D}_1$ is replaced by a simpler density operator $\hat{D}_2$. Minimization of (24), which is positive for $\hat{D}_1 \neq \hat{D}_2$, provides the best $\hat{D}_2$, closest to $\hat{D}_1$.

6.2 Rényi's entropy

By releasing some among the conditions that characterize Shannon's entropy, other entropies can be defined. Rényi's α-entropies

$$S_\alpha(\{p_m\}) = \frac{1}{1-\alpha} \log \sum_m p_m^\alpha \tag{25}$$

are additive for a pair of uncorrelated events, but not subadditive (eq.(8) is violated) except in the limit as $\alpha \to 1$ where Shannon's entropy (6) is recovered from (25). They are useful in the context of fractality. The many attempts to build non extensive thermodynamics from the Tsallis entropy

$$S_q(\{\hat{D}\}) = \frac{1}{1-q}[\mathrm{Tr}\ \hat{D}^q - 1] \ , \tag{26}$$

directly related to (25), run counter to the Zeroth and Second Laws, because maximization of (25) or (26) introduces correlations between non-interacting subsystems [11].

6.3 Quantum information

Many recent works have been devoted to quantum information theory, in which a bit of information taking values 0 or 1 is replaced by a q-bit, that is, by the quantum state of a two-level system. In this context, it is useful to evaluate the *intricacy* of two systems, that is the purely quantum contribution to their order of correlations, for which several expressions are being proposed. (The von Neumann entropy takes into account both ordinary and quantums correlations.)

6.4 Kolmogorov's entropy

In the theory of *deterministic chaos*, the evolution of a system is characterized by a non-linear differential equation which generates a flow in the space of dynamical variables. Kolmogorov's entropy is a measure of the more or less disordered character of this *flow*, while all the entropies considered above referred to the probabilistic description of a state at a given time.

Chaotic dynamics have been proposed to explain, even for systems with few degrees of freedom, the irreversibility paradox. Indeed, prediction becomes hazardous if the motion is chaotic. However, the loss of information that the equations of motion (17) entail in statistical physics is due to the complexity associated with the large number of variables, not to non-linearity.

6.5 Algorithmic complexity

All the types of entropy reviewed so far characterize statistical ensembles of systems, or for Kolmogorov's entropy the full family of trajectories. The concept of algorithmic complexity has been proposed to measure the disorder existing in an *individual* message or in the configuration of a *single system*. The idea is the following. The considered configuration is first represented numerically by coding all its specific features. One then imagines how the resulting number can be constructed by means of algebraic operations in a Turing machine, that is, in an ideal computer. The algebraic complexity is the logarithm of the number of steps of the shortest program needed to realize this task. For a family of messages or of systems, its average can be identified with Shannon's missing information. This is still another face of the concept of entropy.

I wish to thank B. Duplantier for his careful reading.

The literature about entropy is immense, and specific searches should be made for each of its very many aspects. We quote below only a few books or articles, either for their tutorial nature or because they contain extensive bibliographies.

References

[1] H.B. Callen, Thermodynamics (Wiley, 1975).

[2] S.R. de Groot and P. Mazur, Thermodynamics of irreversible processes (North Holland, 1962). For an elementary introduction, see http://e2phy.in2p3.fr/2003.

[3] W. Thirring, Quantum mechanics of large systems (Springer, 1983).

[4] R.T. Cox, *Amer. J. Phys.* **14**, 1 (1946); B. de Finetti, Theory of probability (Wiley, 1974).

[5] E.T. Jaynes, *Phys. Rev.* **106**, 620 (1957); **108**, 171 (1957). An introduction aimed at a general audience is included in R. Balian, Université de Tous les Savoirs, vol. 4 (O. Jacob, 2001) p. 947.

[6] J. Uffink, *Studies in History and Phil. of Modern Phys.* **26B**, 223 (1995).

[7] R. Balian, N.L. Balazs, *Ann. Phys. (NY)* **179**, 97 (1987).

[8] A tutoral introduction to this subject is given in R. Balian, *Amer. J. Phys.* **67**, 1078 (1999). Many references can be found in R. Balian, Y. Alhessid and H. Reinhardt, *Phys. Reports* **131**, 1 (1986), where the entropy-based metric is introduced, and in J. Rau and B. Müller, *Phys. Reports* **272**, 1 (1996).

[9] J.E. Mayer and M.G. Mayer, Statistical mechanics, 2nd edition (Wiley, 1977) pp. 145–154.

[10] L. Brillouin, Science and information theory, (Academic Press, 1956).

[11] M. Nauenberg, *Phys. Rev. E* **67**, 036114 (2003) and references therein; R. Balian, to be published.