Probability, typicality and emergence in statistical mechanics

Probability, Typicality and Emergence in Statistical Mechanics

Sergio Chibbaro, Lamberto Rondoni and Angelo Vulpiani

January 23, 2021

Abstract

The relevance of probability theory is obvious in a subject called "Statistical Mechanics" (SM). On the other hand, SM arose as a microscopic description of single objects made of many (invisible) parts, thus justifying from an atomistic point of view the laws of thermodynamics. As a matter of fact, experimental measurements of thermodynamic quantities are conducted on a single system of interest, hence a fundamental problem arises in connecting probabilistic computations, *e.g.* the averages over ensembles of identical objects, with experiments.

One of the most evident aspects of macroscopic phenomena is that they are characterized by a clear trend in time, that cannot be reverted. On the other hand, our understanding of microscopic dynamics is that they are reversible in time. With the aid of analytical computations on stochastic systems, and of numerical simulations of deterministic Hamiltonian systems, we illustrate basic features of macroscopic irreversibility, thus of the microscopic foundations of the second principle of thermodynamics, along the lines of Boltzmann's kinetic theory. It will be evidenced that in systems characterized by a very large number of degrees of freedom, irreversibility concerns single realizations of the evolution processes, in the sense of the vast majority of the far-from-equilibrium initial conditions. That the vast majority out of a collection of realizations of a given process shares certain properties is often referred to as typicality.

1 Introduction

Statistical mechanics originates in the study of the properties of macroscopic bodies, *i.e.* of objects made of very large numbers of microscopic particles (atoms or molecules) whose dynamics follows mechanical laws, that are classical or quantum, depending on the case [LL80, Ma85]. One may formally write these equations for all the particles in the system, and may in principle solve them. However, the number of degrees of freedom, hence of equations to solve, is huge and, in addition, the initial conditions are not known, therefore the solution of this problem is impossible in practice. At the same time, knowledge of positions

and velocities of all the particles, which amounts to endless tables of numbers, is not particularly informative, if one is interested in temperature, pressure, elasticity, magnetization etc. This impossibility is hardly any concern.

On the other hand, whether the microscopic motions are very complicated or not, whether they are known or not, we observe that the macroscopic behaviours are relatively simple and understandable in terms of a reduced number of "observables" following relatively simple laws. In SM, the observed macroscopic simplicity, in spite of the expected complexity of the microscopic motions, is attributed to the statistical nature of the macroscopic laws, which reduces the complexity by averaging over the many degrees of freedom. The result is not purely mechanical [LL80, Ma85, CFLV08], and *qualitatively* differs from merely mechanical laws. That is why probabilities constitute the fundamental tool of SM. Their usage in physics has been pioneered by the founding fathers of SM, Maxwell, Boltzmann and Gibbs, who changed the very idea of the term *prediction* in physics and, as a consequence, in philosophy as well.

Given that this subject is one and a half century old, why should we discuss it today? In our opinion there are at least three good reasons for that:

(a) The subject is of interest by itself, both for scientists and philosophers, since it it exemplifies how a new phenomenon may *emerge* from the typical behaviour of a lower level one. In particular, the relation between microscopic and macroscopic laws is paradigmatic of how reductionist approaches to complex phenomena in many branches of science are prone to failure [CRV14a, CRV14b, B94, B02].

(b) The subject is pedagogically relevant: with respect to other appealing but rather speculative frameworks, like ecology or cosmology, it allows a concrete discussion of the main conceptual issues concerning the link between different levels of description of a given reality.

(c) The subject is important in the development of current technology: for instance, challenging frontiers for the applications of statistical physics are provided by systems with a small number of degrees of freedom, far from the thermodynamic limit, such as those of interesting in bio- and nano-technologies. Another frontier is given by non Hamiltonian models, which are considered appropriate in the description of granular materials, active matter, epidemics, etc. In these cases, one or both of the original assumptions of SM, namely the very many degrees of freedom and the Hamiltonian dynamics, are absent. Therefore the foundations and applicability of the theory have to be scrutinised [Ma85], in the light of a presumably even higher relevance of probabilities than in the original framework of SM [Z05].

The relevance of probability theory for SM stems from the original idea of Boltzmann, who associated macroscopic (thermodynamic) quantities to averages of mechanical observables of the microscopic constituents of matter. In particular, he adopted frequency of events as the basic notion of probability [VCCPV, G01].

In Boltzmann's SM, probability has no relation to measures of ignorance or uncertainty,[1] and it does not make any use of collections of identical objects.

This is part, instead, of Gibbs approach to SM, in which averages are computed with respect to probabilities that represent how the microscopic phases of large ensembles of identical objects are distributed in their phase spaces. This corresponds to the classical notion of probability, which differs from the frequentist notion, but it is commonly expected be equivalent to that. As computations of time-averages are much harder than ensemble calculations, one commonly accepts ergodic hypothesis, which amounts to such equivalence, and proceeds as prescribed by Gibbs. Therefore, a question arises about the link between the probabilistic computations of SM, and the results of laboratory experiments, which are conducted on a single realization of the macroscopic object under investigation.

In our opinion, the main theoretical issue to be addressed, in order to answer this question is the justification of *typicality*, *i.e.* of the fact that time averages of macroscopic quantities in the evolution of a single system are very close to averages of that quantity over ensembles of microscopically distinct but otherwise identical replicas of that system [G12]. This fundamental property can be seen as *emergent* in the proper limits.

To convince ourselves that this is not a hopeless project, we may refer to one of the best propositions used to link probability and physics, the Cournot's principle:

An event with very small probability will not happen.

Actually, this statement may be associated with the one in Jakob Bernoulli's celebrated book *Ars Conjectandi* (1713), which reads:

Something is morally certain if its probability is so close to certainty that shortfall is imperceptible.

We do not enter the debate about the validity of such a principle, see [SV01] for a nice analysis of it. However, we recall that eminent mathematicians, such as P. Levy, J. Hadamard, and A.N. Kolmogorov, considered the Cournot's principle as the only sensible connection between probability and the empirical world. That connection granted, Levy stressed the concrete character of probability, arguing that, at the ontological level:

Probability is a physical property just like length and weight.

In this chapter, we shall explain how macroscopic laws emerge as statistical laws from the microscopic ones: in passing from the microscopic realm to the macroscopic one, novel properties arise, which are alien to the microscopic realm. In summary: (i) we first illustrate the main ideas of Boltzmann, and the entailing ergodic hypothesis for systems made of very many degrees of freedom. Then, we will analyse some examples with regard to irreversibility and typicality: (ii) to this purpose, the Ehrenfest model will be used. This is a

[1] This inspired a whole branch of mathematics, known as ergodic theory, which represents one way of introducing probabilities in the analysis of the otherwise rigidly deterministic dynamical systems.

stochastic process concerning N non-interacting particles, and it can be rigorously analysed, showing that, in the $N \to \infty$ limit, irreversible behaviours characterize almost all realizations of the process; this will be followed by (iii) numerical simulations of single systems made of many particles, showing their irreversibility.

We note that conceptually a stochastic process is fundamentally different from the deterministic reversible dynamics of Hamiltonian particle systems. Nevertheless, it can be rigorously proven that particular deterministic systems can be mapped into stochastic processes, and when that is not the case, the presence of chaos in interacting particle systems effectively amounts to a certain degree of dynamical randmness. Assuming that this is the case for systems of interest, as it has been repeatedly demonstrated in the literature, and as we will also show, the stochastic process we consider turns useful because it allows a pregnant quantitative analysis of the onset of typicality in the large N limit.

2 Probability and real world

Discussing the foundations of SM necessarily starts with the two seminal contributions given by Boltzmann [C98, G01]:

I. the use of probabilities in the calculation of physical quantities;

II. the link between microscopic dynamics (mechanical laws) and macroscopic properties (thermodynamics).

The second is formalised by Boltzmann's celebrated relation

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

where S is the entropy of a given state, and W is number of possible microscopic configurations corresponding that macroscopic state. In the Hamiltonian dynamics picture, this number is then identified with the phase space volume occupied by the relevant microscopic phases. For a system of N particles each with d degrees of freedom, a microscopic phase is a $2dN$-dimensional vector, $\Gamma = (Q_1, P_1; Q_2, P_2; ...; Q_N, P_N)$, whose components are the d-dimensional coordinates Q_i and momenta P_i, $i = 1, ..., N$ of all particles. The volume in the phase space is thus defined for a fixed energy as

$$W(E, V, N) = \frac{1}{N! h^{3N}} \int \delta(H(\mathbf{Q}, \mathbf{P}) - E) d^{3N}\mathbf{Q} d^{3N}\mathbf{P} \,, \tag{2}$$

where h is the $2d$-dimensional volume of a small cell, that we may think refers to a single particle [LL80].[2]

From a philosophical standpoint, Eq.(1) plays the role of a *bridge law* [CRV14a], connecting the atomic level to the macroscopic one, and constitutes the fundamental ingredient of SM, that justifies all its applications to condensed matter physics and chemistry.

[2]The use of the symbol h should not lead to believe that quantum effects are taken into consideration. In the present picture, quantum mechanics plays no role.

Linked to point I is the ergodic hypothesis, which connects dynamics and probability. This is done as follows. Consider a macroscopic object of N interacting particles each with 3 degrees of freedom, and let the microscopic state be described by $\mathbf{X} \in R^{6N}$. A measurement of some macroscopic quantity, for instance the pressure, is supposed to last much longer than the molecular time scales, and the result of the measurement is taken to be the time average over the measurement time $\mathcal{T}$, of some mechanical property that is function of $\mathbf{X}$. The measurement tool is therefore said to effectively compute the following quantity:

$$\bar{A}^{\mathcal{T}} = \frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} A(\mathbf{X}(t))dt \ . \tag{3}$$

In principle, the computation of $\bar{A}^{\mathcal{T}}$ requires the initial condition $\mathbf{X}(0)$, and the determination of the time evolution following from that initial phase, $\mathbf{X}(t)$. Given that $\mathbf{X}(t)$ represents the complete microscopic motion, this is surely beyond any human capability.

Boltzmann's ingenious idea to overcome this difficulty, *i.e.* the ergodic hypothesis, is to replace the time average with a suitable average on the the phase space. He assumed that

$$\lim_{\mathcal{T}\to\infty} \frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} A(\mathbf{X}(t))dt = \int A(\mathbf{X})\rho(\mathbf{X})d\mathbf{X} \ , \tag{4}$$

where $\rho(\mathbf{X})$ is the suitable probability density. The sense of this hypothesis is that the physically relevant but impossible computation of the time-average can actually be turned into a (generally exceedingly simpler) probabilistic computation. In particular, if the ergodic hypothesis is assumed to be valid, it is easy to derive also the canonical Boltzmann-Gibbs distribution for a system which exchanges energy with an external environment, and then deduce the corresponding thermodynamics. If successful, this process achieves the goal of SM.

The issue is now whether the ergodic hypothesis is valid or not in the cases of physical interest. Unfortunately, many numerical investigations, starting from the FPUT (Fermi, Pasta, Ulam and Tsingou) work [G07] on chains of non linear oscillators, as well as rigorous mathematical results, notably the KAM (Kolmogorov, Arnold and Moser) theorem on non integrable systems [D14], show that the ergodic hypothesis does not hold rigorously in the form given above, if generic functions of phase A are considered. One could, thus, naively conclude that ergodicity cannot be taken as central in the foundations of SM, and that it could even be misleading. As a matter of fact, one finds that the ergodic hypothesis cannot be so lightly discarded. Indeed, it turns out that:

(a) the Boltzmann-Gibbs probability distributions (the classical ensembles) are valid;

(b) molecular dynamics gives correct results, being generally based on the ergodic hypothesis.

To understand these facts, one may recall the original Boltzmann's reasoning, that has since been considered the standard explanation of the success of the ergodic hypothesis [LL80, Ma85]. even though it has been variously challenged in time [B96, G05]. That reasoning has been made mathematically rigorous by Khinchin [K49]. In a nutshell, Khinchin's argues that the ergodic hypothesis is "practically" true, as far as physical phenomena are considered, if

- $N \gg 1$;
- suitable observables are selected;
- one allows for failure of (4) in a small region of the phase space.

Khinchin proved that, for the class of separable sum functions defined by

$$f(\mathbf{X}) = \sum_{n=1}^{N} f_n(\mathbf{q}_n, \mathbf{p}_n), \tag{5}$$

where each f_n represents a single particle contribution, the following holds

$$Prob\Big(\frac{|\delta f(\mathbf{X})|}{N} \geq \frac{c_1}{N^{1/4}}\Big) \leq \frac{c_2}{N^{1/4}} \tag{6}$$

where c_1, c_2 are constants, and $\delta f(\mathbf{X})$ is the difference between the time average starting at $\mathbf{X}$ and the average value computed in the microcanonical ensemble. It is worth noting that many interesting microscopic functions of phase are sum functions, like kinetic energy, the momentum etc. Then, Khinchin considered non interacting systems, whose energy (Hamiltonian) is expressed as

$$H = \sum_{n} h_1(\mathbf{q}_n, \mathbf{p}_n). \tag{7}$$

where each h_1 term is the energy of one particle. That is a serious limitation of the approach, but Mazur and van der Linden generalised Khinchin's result to the physically more interesting case of (weakly) interacting particles[ML63], whose hamiltonian can be written as

$$H = \sum_{n} h_1(\mathbf{q}_n, \mathbf{p}_n) + \sum_{n,n'} V(|\mathbf{q}_n - \mathbf{q}_{n'}|) \ . \tag{8}$$

In brief, it has been proven that, although the ergodic hypothesis as formulated above is not generally rigorously true for physically interesting systems, it remains valid for physically relevant observables of a wide class of systems made of very many particles. Indeed, in this case, violations of the hypothesis are restricted to negligibly small regions of phase space, in which the system may fall with a probability of order $O(N^{-1/4})$, that vanishes in the $N \to \infty$ limit, but is definitely irrelevant "already" for 10^{24} particles.

The main ingredients of this reasoning are the large value of N, together with the fact that one only needs to consider a special class of phase functions.

This makes by and large marginal the role of the details of the microscopic dynamics, apart from the fact that it must preserve phase space volumes, like Hamiltonian dynamics does.

One consequence of having a restricted set of observables and and very large N, is that one may separate different space and time *scales*. The fact that $N \gg 1$ implies that particles are much smaller than the macroscopic body they constitute. Moreover, when these particles are allowed to move almost freely in space, thanks to their weak interactions, their mean-free path λ has also to be much smaller than the characteristic macroscopic length L of the object they belong to. Such distances can then be associated with the corresponding time-scales, *i.e.* the times needed to cross them with the particles average velocity.

The separation of scales is fundamental for the emergence of novel phenomena, when passing from one level of description to another [B94, K13, D15]. Indeed, it is required for spatial correlations to be negligible over distances that are very small on the macroscopic scale, which is the basis for quantities such as the internal energy, the entropy etc. to be extensive, as observed in thermodynamics. Moreover, the separation of time scales allows the realization of local thermodynamic equilibrium in sufficiently short times, that the average performed by a measurement appears to account for all the possible values the observable of interest can take. Consequently, the initial condition is irrelevant, and the ergodic hypthesis is vindicated.

That is the content of the condition known as typicality, which states that extensive observables will stay close to their mean value; and we now see that such a condition is better established if the number of particles is larger. Therefore, in this framework, the "atypical" behaviours can be considered of vanishing probability when dealing with macroscopic objects, in agreement with thermodynamics, that is deterministic and excludes them.

We conclude this section noting that while the approach illustrated above provides a convincing basis for the applicability of SM to the description of macroscopic objects, it does not cover the whole spectrum of relevant problems. In particular, given a generic initial condition, and an observable $\mathcal{O}$, estimates of the minimum value of the measurement time $\mathcal{T}$, such that $\bar{\mathcal{O}}^{\mathcal{T}} \simeq \langle \mathcal{O} \rangle$, are hardly available. This problem has been widely investigated since the FPUT numerical experiment, in which the 1-dimensional nature of the system hinders the decay of various kinds of correlations, making the local thermodynamic equilibrium hard to establish [G07]. While this does not allow a direct connection with thermodynamics, it does not prevent the use of SM, which in this respect generalizes the macroscopic theories to small systems, like 1-dimensional systems must be.

2.1 Statistical mechanics as statistical inference?

As argued above, we believe that Boltzmann's justification of SM, based on the large number of degrees of freedom and on typicality, is conceptually satisfactory when dealing with the mergence of macroscopic phenomena from microscopic dynamics. Nevertheless, there exists a radical anti-dynamical point of view which takes SM as a mere form of statistical inference, and not like a theory of

objective physical reality.

This view is philosophically pragmatic and anti-realistic, and it implies that probabilities measure the degree of truth of a logical proposition, rather than describing the state of a given material object. This approach has become quite fashionable in the framework of "complex systems", and can be traced back to the work of Jaynes [J57], who expressed this idea through the maximum entropy principle (MaxEnt): a general rule for finding the probability of a given event when only partial information is available. In a nutshell, the principle proceeds as follows: given the expected values c_i of m independent functions f_i, defined on a space of coordinates x, a probability distribution ρ is constructed, in such a way that

$$c_i = \int f_i(x)\rho(x)dx \equiv \langle f_i \rangle \quad i = 1, ..., m \ . \tag{9}$$

As the name anticipates, the construction proceeds by maximisation of a formal entropy function H, under the constraints (9), which is thought to generalize the Gibbs entropy [G06, CFLV08]. In practice, using the Lagrange multipliers procedure, the probability density ρ is obtained maximising

$$H = -\int \rho(x) \ln \rho(x)dx, \tag{10}$$

under the constraints $c_i = \langle f_i \rangle$. One then obtains:

$$\rho(x) = \frac{1}{Z} \exp \sum_{i=1}^{m} \lambda_i f_i(x) \ , \tag{11}$$

where the parameters $\lambda_1, .., \lambda_m$ depend on the values $c_1, ..., c_m$. This approach may indeed be applied to the statistical mechanics of systems with a fixed number of particles; for instance, fizing the value of the mean energy leads to the usual canonical Gibbs distribution in a very simple and elegant fashion [P11, U95]. Therefore, the MaxEnt appears as a cornucopia, out of which one can extract in a straightforward way the main results of SM.

This conceptual issue deserves a critical discussion. Indeed, the interest of such an approach comes from the fact that most phenomena of scientific interest, notably the biological phenomena, lack a reliable theory, while there is good amount of data concerning them. Two difficulties immediately arise:

(a) the ancient saying *"ex nihilo nihil"* continues to be appealing;

(b) *unperformed experiments have no results.*

In this respect, a caustic, but insightful example was conceived by Shang-Keng Ma [Ma85]:

"How many days a year does it rain in Hsinchu?" [3] *One might reply "As there are two possibilities, to rain or not to rain, and I am completely ignorant about Hsinchu, therefore it rains six months in a year."*

[3] Hsinchu is a chinese city on the Pacific ocean.

The important point Ma wants to make is that it is not possible to infer something about a real phenomenon, thanks to our ignorance. As recalled in the previous section, probability in SM is used in relation to objective frequencies, it is the ratio of numbers extracted from concrete observations. In the MaxEnt framework, it is instead related to the degree of uncertainty or of our ignorance about an event: lack of knowledge is used to produce kowledge.

Apart from these very general considerations, the weakest technical aspect of the MaxEnt approach is its dependence on the choice of the variables needed to represent a given phenomenon. Tis fact can be understood as follows. Given a property X, whose values x are distributed according to the probability density ρ_X, one realizes that the "entropy" $H_X = -\int \rho_X(x) \ln \rho_X(x) dx$ is not an intrinsic quantity of the phenomenon X, hence it is unclear how H can characterize X. For instance, changing parametrisation, *i.e.* using the coordinates $y = f(x)$ in place of x, where f is an invertible function, the entropy of the same phenomenon turns:

$$H_Y = -\int \rho_Y(y) \ln \rho_Y(y) dy = H_X + \int \rho_X(x) \ln |f'(x)| dx \ . \tag{12}$$

Therefore the MaxEnt gives different solutions if different variables are adopted to describe the very same phenomenon. In order to avoid this unacceptable condition, Jaynes later proposed a more sophisticated version of the MaxEnt, in terms of the relative "entropy":

$$H^* = -\int \rho(x) \ln \Big[\frac{\rho(x)}{q(x)}\Big] dx \ , \tag{13}$$

where q is a known probability density. H^* at variance with the entropy H, does not depend on the choice of variables. Nonetheless, H^* now depends on the distribution q, hence the problem is merely shifted toward the selection of such a probability density, which is analogous to the problem of choosing the proper variables.

For instance, knowledge of the mean energy and taking a uniform distribution, say $q = \text{const}$, leads to the correct Gibbs distribution, but this q is, in principle, either a totally arbitrary choice, or it amounts to an *a priori* knowledge of the correct result. Analogoulsy, while the correct variables for the description of equilibrium thermodynamic systems are well known, because they concern comparatively very simple phenomena, which have been investigated for very long, the same cannot be stated about generic systems, such as the complex ones for which the MaxEnt principle is supposed to provide a theoretical framework.

In conclusion, even the second, more elaborate method, is not truly predictive. Therefore, although the MaxEnt principle can be considered a neat and elegant way of deriving Gibbs-like probability distributions, when they are known to apply, we see no reason to found SM on it. Presumably, it may be useful to gain insight on a given phenomenon, in the absence of informed guiding principles, to be tested together with other alternatives, but one should keep in mind that it may lead into error, since Gibbs-like probability distributions are not generic, not even in relatively simply physical phenomena [ARV17].

3 The old debated problem of irreversibility

Typicality, which we have first discussed in the case of equilibrium systems, plays an important role also in the case of irreversible non-equilibrium phenomena. To illustrate this fact, let us begin with two simple observations:

(a) microscopic mechanical laws are invariant under time reversal:

$$t \to -t \ , \ \mathbf{q} \to \mathbf{q} \ , \ \mathbf{p} \to -\mathbf{p} \, . \tag{14}$$

(b) the macroscopic world is described by irreversible laws, *e.g.* the Fick equation for the diffusion of a scalar concentration C

$$\partial_t C = D \Delta C \, , \tag{15}$$

where D is the diffusivity of the scalar.

The question thus arises: is it possible to derive macroscopic (irreversible) equations starting from a microscopic (reversible) description [L93]?

This fundamental question constitutes the core of the objections raised by Loschmidt and Zermelo about Boltzmann's celebrated H-theorem, which describes an irreversible evolution from non-equilibrium toward equilibrium states [H09]. Loschmidt tackled directly the issue of reversibility, while Zermelo applied Poincaré's recurrence theorem, that shortly earlier had been demonstrated. The theorem states that, given a conservative system, like the newtonian ones we consider, and an initial condition in its phase space, the entailing evolution will sooner or later come back arbitrarily close to the sarting point. In other words, there is "recurrence". Therefore, if a function of phase increases for a while,[4] sooner or later it has to decrease; which apparently means that the second law of thermodynamics cannot be derived from the newtonian dynamics of a system made of N particles.

Beside technical points, Boltzmann rapidly understood and refuted Zermelo's mathematically correctly formulated paradox, explaining the physical content of his theory. First of all, one must realize that physics, like all measurements one can perform, is about specific space and time scales. Then, Boltzmann's point of view was masterly summarised by Smoluchowski as follows : *A process appears irreversible when the initial state has a recurrence time which is long compared to the time of observation* In fact, Zermelo's paradox is physically irrelevant because, as rigorously proven by Kac [K57] for ergodic systems with N degrees of freedom, the recurrence-time goes like

$$\langle T_R \rangle = \tau_o C^N \ , \tag{16}$$

where τ_o is a typical time, and $C > 1$ depends on the desired precision of recurrence. Therefore, the mathematcally correct Zermelo's argument is physically

[4] Most notably the opposite of the H-functional taken by Boltzmann to mirror the entropy of an isolated system.

irrelevant because, given N for a macroscopic system, the corresponding recurrence time is unphysically and ridiculously huge: well beyond many ages of our universe for just a cubic centimetre of air.

Loschmidt raised a subtler criticism, that requires more elaborate analysis.

3.1 Use and abuse of probabilities (ensembles) and entropies

Gibbs ensembles are one of the cornerstones of SM, yet we believe that they are often introduced in very unfortunate fashions. For instance, in standard textbooks[] one can finds rather obscure statements, such as

"an ensemble is an infinite collection of identical systems"

Gibbs' goal, who acknowledged Maxwell and Boltzmann for introducing ensembles [G06], was to use them in order to reformulate the Boltzmann's probabilistic approach based on the ergodic hypothesis. He then defined an *ensemble* as an infinite (imaginary) collection of macroscopically identical systems, that differ in their microscopic phases. Mathematically, such a collection could be intuitively and efficiently represented by a distribution of points in the phase space. The physical reasons behind the applicability of this idea have been outlined *e.g.* by Fermi[F56].

He explained that an ensemble represents the microscopic states explored by the dynamics of a single system, in the time taken by a measurement, but only under some conditions. In particular, the transitions from microscopic state to microscopic state must be much faster than the measurement.

Therefore, from the thermodynamic perspective, which is a deterministic description of single systems, taking too seriously statistical ensembles may be misleading and, in fact, a source of errors. Different is the case of systems that are not of thermodynamic interest, for which probabilities may be the only sensible information,[5] which however we do not treat here.

Let us consider, for example, the entropy of a given system, and let $\rho(\mathbf{X},t)$ be a probability distribution of its microscopic states in the phase space. The so called Gibbs-entropy is then defined as:

$$S_G(t) = -k_B \int \rho(\mathbf{X},t) \ln \rho(\mathbf{X},t)\, d\mathbf{X} = S_G(0)\,. \tag{17}$$

One may be tempted to think of the dynamics of the collection of ensemble members described by $\rho(\mathbf{X},t)$ as of the molecuels of a certain system, and their evolution in phase space as the diffusion of molecules in real-space. However, one should note that phase space is an abstract, exceedingly high-dimensional space, that is totally different fro the 3-dimensional real space. Phase points

[5] When dealing with non-macroscopic systems, thermodynamics does not strictly apply. This is the case, for instance, of Brownian particles immersed in a liquid. In this case, only a probabilistic, *ensemble*, description appears interesting and feasible.

are not molecules,[6] hence the evolution of their density has in general no physical content at all. In fact, if the N particles of the system obey the Hamilton equations of motion, the celebrated Liouville Theorem states that volumes in phase space are conserved by the time evolution. An immeadiate consequence of which is that the Gibbs entropy S_G is constant in time[H09]. In other words, while the Gibbs entropy correctly yields the equilibrium thermodynamic entropy of the system, it does not represent the growing entropy of an isolated nonequilibrium system: the Gibbs entropy is not a suitable SM counterpart of the thermodynamic entropy.

To overcome this difficulty, many auhtors have introduced a coarse-graining of the phase space, *i.e.* a partition of phase space made of cells of given small size, say ϵ, and a corersponding coarse-grained version of the probability density and of the Gibbs entropy. The probability for the microscopic phase to lie in the $i-$the cell at time t is expressed by:

$$P_\epsilon(i,t) = \int_{\Lambda_\epsilon(i)} \rho(\mathbf{X},t)\, d\mathbf{X} \ , \tag{18}$$

and the corresponding coarse grained Gibbs entropy is defined by:

$$S_{G,\epsilon}(t) = -k_B \sum_i P_\epsilon(i,t) \ln P_\epsilon(i,t) \ . \tag{19}$$

Then, unlike S_G, the quantity $S_{G,\epsilon}$ is not constant in time: it grows monotonically, if ρ is not invariant, till a maximum is reached. But this success in describig an evolving "entropy" is only apparent, and not real. One of the main, far from unique, difficulties that this method faces is that the evolution of $S_{G,\epsilon}$ is not intrinsic, but depends on ϵ. Also, it has been proven that $S_{G,\epsilon}$ does not grow for a while: it remains constant up to a crossover time $t_* \sim \ln(1/\epsilon)$, which grows without bounds, when ϵ decreases [CFLV08, FPPRV]. Physically this makes no sense; it is analogous to state that the heat generated by burning one litre of gasoline depends on how accurately we observe the phenomenon, and if we observe it very accurately, no heat is generated...

3.2 The H theorem

A physical framework in which macroscopic irreversibility emerges out of microscopic reversible dynamics is afforded by the celebrated H-Theorem, which Boltzmann derived within the kinetic theory of gases. Here, one starts from the one particle distribution function $f(\mathbf{q},\mathbf{p},t)$, which represents the mass density in the so- called μ- space, *i.e.* the space of a single particle coordinates and momenta, which 6-dimensional for particles with 3 degrees of freedom. In the limit of dilute monoatomic gas, with rather subtle assumptions, Boltzmann derived

[6] For instance, they have no extension and do not interact, while molecules occupy a certain volume and interect with each other. Moreover, one phase point represents a whole N-particles system, which something totally different from one of the N particles.

the time evolution equation of f, which takes the form:

$$\frac{\partial}{\partial t} f(\mathbf{q}, \mathbf{p}, t) + \sum_j \frac{p_j}{m} \frac{\partial}{\partial q_j} f(\mathbf{q}, \mathbf{p}, t) + \sum_j \dot{p_j} \frac{\partial}{\partial p_j} f(\mathbf{q}, \mathbf{p}, t) = C(f, f) \ , \qquad (20)$$

where $C(f, f)$ is a bilinear integral term which accounts for the (weak) interactions among the particles. This equation implies that the quantity:

$$S_B(t) = -H(t) = -k_B \int f(\mathbf{q}, \mathbf{p}, t) \ln f(\mathbf{q}, \mathbf{p}, t) \, d\mathbf{q} \, d\mathbf{p} \qquad (21)$$

constanty increases, until an equilibrium state is reached [H09]:

$$\frac{dS_B(t)}{dt} \geq 0 \ , \quad \text{where} \ \text{“} = \text{”} \ \text{holds only at equilibrium} \qquad (22)$$

Boltzmann could then identify his entropy S_B with the thermodynamic entropy of an isolated dilute gas. Equation (22) is called “H-theorem”. The physical content is that the second law of thermodynamics is obtained via S_B, (21), casting the laws of classical mechanics, which are reversible, into a suitable probabilistic framework.

In addition to the recurrence paradox formulated by Zermelo, which we have observed to be physically irrelevant, another paradox has been devised to contradict Boltzmann's approach and his H-theroem: the reversibility paradox, usually attributed to Loschmidt [7]. In fact, this paradox is mathematically justified, like Zermelo's paradox, and can be equally dismissed, as irrelevant for the physics of macroscopic systems.

To understand that, let us recall the assumptions underlying the Boltzmann theory. First of all, the number of particles N is very large. Then, the one particle distribution function $f(\mathbf{q}, \mathbf{p}, t)$ that is the main theoretical object in the theory, can be seen as an empirical distribution function, concerning the positions and velocities of the N particles, formally expressed by:

$$f(\mathbf{q}, \mathbf{p}, t) = \frac{1}{N} \sum_{n=1}^{N} \delta[\mathbf{q} - \mathbf{q}_n(t)] \delta[\mathbf{p} - \mathbf{p}_n(t)] \, . \qquad (23)$$

Therefore,

(i) $f(\mathbf{q}, \mathbf{p}, t)$ is a well defined macroscopic observable: the number density of particles;

(ii) $f(\mathbf{q}, \mathbf{p}, t)$ can be measured (*e.g.* in numerical simulations) and such a measurement concerns the single system under investigation, made of the N particles: there is no need to refer to the statistical ensembles.

[7] The content of the paradox is the following. Given that microscopic dynamics is reversible in time, if we were able to reverse time, the dynamics should trace back its trajectory, and therefore also S_B should decrease.

It has then been proven by Lanford that the microscopic Hamiltonian, reversible, dynamics is not incompatible with the H-theorem[L81]. Indeed, given a hard-sphere system; considering the Boltzmann-Grad limit, *i.e.* $N \to \infty$, $\sigma \to 0$ so that $N\sigma^2 \to constant$, where σ is the diameter of the particles; and starting from an initial condition in a *"good set"*, one obtains that $f(\mathbf{q}, \mathbf{p}, t)$ evolves as prescribed by the Boltzmann equation, hence the H-theorem holds. In other words, Lanford proved that

$$f(\mathbf{q}, \mathbf{p}, t) \simeq f_B(\mathbf{q}, \mathbf{p}, t) \ , \tag{24}$$

where $f_B(\mathbf{q}, \mathbf{p}, t)$ is the solution of the Boltzmann equation. This result holds for a short time, which is a fraction of the mean-collision time, for $N \gg 1$ and a typical $\mathbf{X}(0)$; but it is enough to prove rigorously that Hamiltonian dynamics, in the proper limit, does not violate the Boltzmann equation, and one obtains an irreversible behaviour from a microscopic reversible dynamics [L93, CRV14a, CFLV08, CCCV16].

It is further worth noting that the validity of the H-theorem does not rest on the details of the particles interactions, as long as they exert a short range repulsion. This is important, from a physical point of view, since it means that the result is quite general.

3.3 Again about entropies and probability

Although at first glance the Gibbs and the Boltzmann entropies look similar, their dynamical conceptual and physical meaning, hence their behaviours are totally different. Both entropies correctly describe equilibrium states, but the Gibbs entropy is defined in terms of the very abstract notion of phase space probability density or of *ensemble*, while the Boltzmann entropy is derived from the very material property which is the number of particles of one concrete system occupying a given spacial volume, with velocity in a given cube of the velocity space. Therefore, some understanding of the connection between such diverse entropies is desirable.

Roughly speaking, two main points of view are generally adopted: the subjective and the objective interpretation of probability. According to the subjective interpretation, probability is a degree of belief in something. One of the most influential followers of such a view is Jaynes, who claimed that the entropy of a physical system depends on the observers' knowledge of it, or on their (informed) belief concerning the phenomenon of interest. In the objective interpretation, the probability of an event is instead determined by the the physics of the system of interest and not by the available or missing information.

This difference allows us to distinguish between thermodynamic irreversibility, and the relaxation of a phase-space probability distribution $\rho(\mathbf{X}, t)$ to an invariant (constant in time) distribution, further clarifying that an abstract ensemble must not be confused with a given macroscopic system. Indeed, for dynamical system exhibiting a good degree of chaos, one commonly observes that $\rho(\mathbf{X}, t)$ converges in time to an invariant probability distribution,

$\rho(\mathbf{X}, t) \to \rho_{inv}(\mathbf{X})$. In other words, the ensemble averages of all phase functions "irreversibly" converge to given values.

This is not the irreversibility the second law of thermodynamics speaks about! In the thermodynamic case, the observables of interest of systems prepared in the same way, evolve in the same fashion. If an enseble average converges, the different elements of the ensemble, hence their observables, may evolve in totally different and inconsistent ways. Again, this is a consequence of the fact that phase-space points are not particles, and their probability density is not a mass distribution. In the phase space, a single system is represented by just one point and an actual experiment follows a single trajectory, not a cloud of points from which a collection of different trajectories arises.

The physically relevant issue is that a single macroscopic system behaves irreversibly and in a unique fashion, starting from a generic initial microscopic state.

In summary, contrarily to some perhaps fashionable claims [PS79], there is no direct link between the convergence process of probabilities in phase space, and the thermodynamic irreversibility. For this reason, the only way to pursue the program of SM for macroscipic objects is to take an objective approach to probability, which is the Boltzmann framework. That does not diminish in any way the importance of dynamical system theory in other problems [CFLV08], even when the number of degrees of freedom is large[BJPV].

4 Typicality and irreversibility

The above discussion introduces the question of *typicality*, which is related to the one raised by various philosophers of science, regarding the role of the microcanonical distribution (ensemble) in the description of constant energy (isolated) systems. We argue that there are very good reasons to assign a privileged role to the microcanonical ensemble, compared to other probability distributions, that are equally invariant under the Hamiltonian evolution. This rather technical subject concerning deterministic dynamics, can be cast in a suggestive framework, once it has been shown that the dynamics of interacting particles does commonly and effectively result in a certain kind of randomness. We thus illustrate the notion of typicality, and the connection between deterministic and stochastic evolution with some examples.

4.1 Typicality in stochastic models

A popular model whose simplicity allows a neat discussion of typicality is the well known Ehrenfest flea model [BCV19], that is jokingly referred to the fleas that jump back and forth between two dogs. The model consists of a Markov chain [G18] representing N "particles", each of which can either be in a box called A, or in another box called B. The state of the Markov chain at time t is identified by the number n_t of particles in A, and the evolution is ruled by a

stochastic law, with given transition probabilities:

$$P_{n\to n-1} = \frac{n}{N} \ , \ P_{n\to n+1} = 1 - \frac{n}{N} \, .$$

dictating how the state $n_t = n$ changes in one time step to become $n_{t+1} = n \pm 1$. In our SM language, the state n_t can be seen as the "macroscopic" state of the system of interest, while the the corresponding "microscopic" configuration is defined by the list of the particles lying in box A, considered as distinguishable particles. The equilibrium (macroscopic) state is expressed by $n_{eq} = N/2$.

The evolution of an ensemble of initial conditions starting from a given state n_0 can be described computing not only the mean population $\langle n_t \rangle$ but also its variance $\sigma_t^2 = \langle n_t^2 \rangle - \langle n_t \rangle^2$. One obtains:

$$\langle n_t \rangle = \frac{N}{2} + \Big(1 - \frac{2}{N}\Big)^t \Delta_0 \ , \ \sigma_t^2 = \frac{N}{4} + \Big(1 - \frac{4}{N}\Big)^t \Big(\Delta_0^2 - \frac{N}{4}\Big) + \Big(1 - \frac{2}{N}\Big)^{2t} \Delta_0^2 \, , \quad (25)$$

where $\Delta_0 = n_0 - N/2$. The main result is that $\langle n_t \rangle \to n_{eq} = N/2$, exponentially fast with a characteristic time $\tau_c = -1/\ln(1 - 2/N) \simeq N/2$ and a standard deviation σ_t that goes to its equilibrium value $\sqrt{N}/2$ with a characteristic time $O(N)$.

These results for $\langle n_t \rangle$ and σ_t are obtained at the level of the ensemble, *i.e.* as averages over the behaviour of all possible single N-particles (macroscopic) systems obeying Eq.(25). What about a single macroscopic object, *i.e.* a single realization of the process?

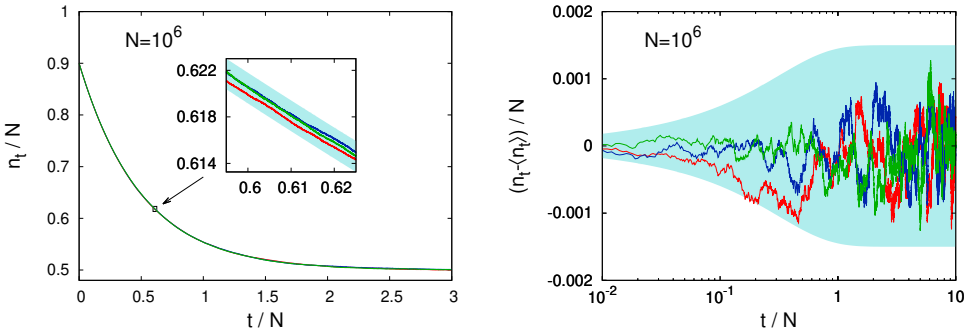

Figure 1: Several realisations of the time evolution of the state of the Ehrenfest flea model, n_t, for $N = 10^6$. The coloured region corresponds to three standard deviations from the running mean: $\langle n_t \rangle - 3\sigma_t < n_t < \langle n_t \rangle + 3\sigma_t$, from [BCV19].

Figure 1 illustrates the result of numerical simulations, showing that for large N, the single object behaves "typically". In more mathematical terms, one has:

$$\text{Prob}\Big(n_t \simeq < n_t > \ \text{ for any } t \in [0, T]\Big) \simeq 1 \ \text{ where } \ T = O(N) \, , \quad (26)$$

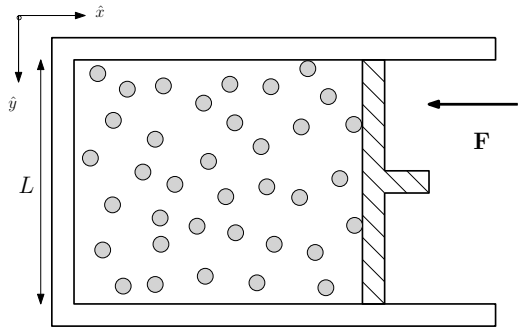

Figure 2: Sketch of a determinisitc (Hamiltonian) model of a gas confined by a piston.

which means that n_t practically behaves as the average $\langle n_t \rangle$ in almost all cases.

Consider now a far from equilibrium initial condition, *e.g.* $n_0 \simeq N$. It is possible to show that, for $N \gg 1$, up to a time $O(N/2)$, *i.e.* as long as n_t remains far from n_{eq}, each single realization of n_t stays "close" to the time dependent average $\langle n_t \rangle$. Indeed using tools of probability theory, one can show [BCV19]:

$$\text{Prob}\Big(\frac{|n_t - \langle n_t \rangle|}{N} < \epsilon_N \text{ for any } t \in [0, T] \Big) \geq 1 - a_N \,, \tag{27}$$

with the quantities $\epsilon_N \to 0$, and $a_N \to 0$ as $N \to \infty$. Taking $\epsilon_N \sim N^{-B}$ with $0 < B < 1/3$, one has $a_N \sim N^{-A}$ with $A > 0$. For instance, $B = 0.2$ implies $A \geq 0.2$

That means that the overwhelming majority of realisations of the stochastic process n_t remains close to $\langle n_t \rangle$ for a long time, if N is sufficiently large. In other words, every macroscopic measurement on the systems has a very low probability of resulting sensibly different from the expected value. This is the conceptual meaning of "tipicality" in SM.

4.2 Typicality in large deterministic systems

The above, exactly solvable stochastic model, neatly quantifies the notion of typicality, but it may appear inappropriate in one investigation concerning the statistical properties of particle systems obeying deterministic equations of motion. The gap between the stochastic and the deterministic realm is however bridged by standard deterministic particle systems, whose properties evolve as erratically as they do in random processes.

Consider, for instance, a channel containing N particles of mass m, closed by a fixed vertical wall on the left, and by a frinctionless piston of mass M on the right[CCCV16]. The piston motion is determined by a constant force F and by its collisions with particles inside the channel, see Fig.2 . The Hamiltonian

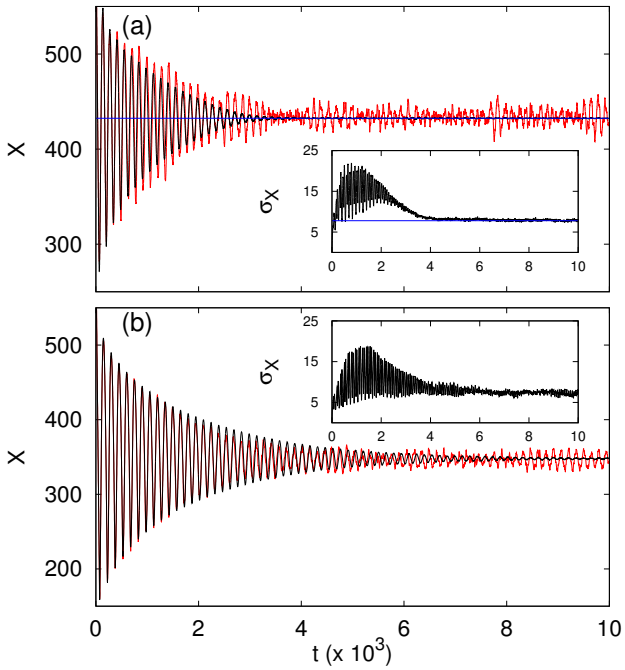

Figure 3: $X(t)$ vs. t for $N = 1024, M = 50, m = 1, F = 150$ and $X(0) = X_{eq} + 10\sigma_{eq}$ in a chaotic piston (a), and in a non chaotic piston (b). Red lines represent $X(t)$ for a single realization; black lines refer to the ensemble average $\langle X(t)\rangle$,[CCCV16].

of this system reads:

$$H = \frac{P^2}{2M} + \sum_i \frac{p_i^2}{2m} + \sum_{i<j} U(|\mathbf{q}_i - \mathbf{q}_j|) + U_w(\mathbf{q}_1, ..., \mathbf{q}_N, X) + FX\,,$$

where U is the interacting potential among the particles, and U_w denotes the interaction of the particles with the wall. In the case of point particles, $U = 0$ and U_w yields elastic collisions. Then, the dynamics is not chaotic, and it is easy to find the "equilibrium" position of the piston, $\langle X\rangle$, and its variance σ_X^2. In presence of interactions, *e.g.* for interaction potentials like:

$$U(r) = \frac{U_o}{r^{12}} \quad , \quad U_w = U_o \sum_i \frac{1}{|x_i - X|^{12}}$$

the equations of motion can be solved numerically, and reveal one positive Lyapunov exponent, *i.e.* chaos.

Figure 3 illustrates the irreversible behavior of the states $X(t)$ of one chaotic and of one non-chaotic instances of the piston model. Their initial conditions $X(0)$ are typical assuming a fixed $X(0)$ which is far from equilibrium, which

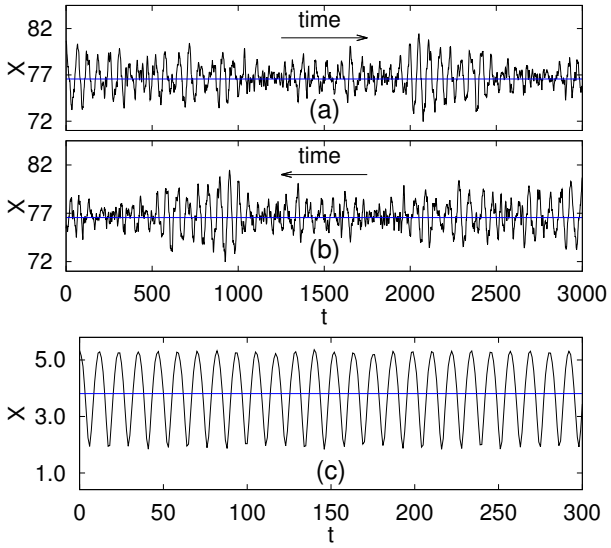

Figure 4: $X(t)$ vs t, for initial condition $X(0) = X_{eq} + 3\sigma_{eq}$; $N = 1024, M = 50, m = 1, F = 150$ $X(t)$ vs t in (a) and (b); $N = 4, M = 50, m = 1, F = 150$ in (c), from [CCCV16].

are determined by the models parameters. In particular, the positions of the particles are initially distributed uniformly in the interval $[0, X(0)]$, while the velocities initially follow a Maxwell-Boltzmann distribution at a temperature T different from the equilibrium temperature T_{eq}, so that $|X(0) - X_{eq}| \gg \sigma_{eq}$, where the subscript eq refers to the equilibrium state.

The result is that the single trajectories are typical: although far from equilibrium, fluctuations about the corresponding ensemble averages are small compared to such averages, as in the case of the stochastic Ehrenfest model. This supports the anticipated analogy between stochastic and deterministic systems, both in the presence and in the absence of chaos, demonstrating that positive Lyapunov exponents are not required for the randomness associated with a many particles system irreversible behaviour. As a matter of fact, the numerical results for our deterministic reversible dynamics look rather similar to those for the stochastic Ehrenfest model, explicitly showing why irreversibility can be understood as an emergent property of a single system under proper initial conditions, when N turns sufficiently large[CCCV16].

These results should be contrasted with those of Fig.4, which reports the behaviour of a small N system starting from an initial condition close to equilibrium. It is well evident the absence of irreversible behaviour.

5 Summary and conclusion

In this contribution we discussed the notion of irreversibility and its relation to typicality, in the framework of the SM of macroscopic systems made of classical particles. We pointed out the dangers associated with an uninformed use of probabilities in phase space, *i.e.* of statistical ensemble. The fact is that the thermodynamic laws ruling the behaviour of macroscopic objects are deterministic and, for a given initial marcoscopic state, always lead to the same evolution. Therefore, there is no need to average over ensembles of differently evolving objects: in the case one really wants to perform such an avarage, the results will be equal to that given by one of those objects. We then formalized this fact with the notion of *typicality* which, for a macroscopic observable M, may be expressed as:

$$\mathrm{Prob}\Big(M(t) \simeq \langle M(t) \rangle\Big) \simeq 1 \;\; when \;\; N \gg 1 \tag{28}$$

where Prob represents the fraction of phase space trajectories enjoying a given property, and the property considered is that almost all trajectories separately behave as the their average.

In our investigation, two examples have been analysed, in order to illustrate how typicality and irreversibility arise in the evolution of systems made of many particles. The first is a stochastic process, known as the Ehrenfest flea model, which is exactly solvable hence allows us to obtain analytical expressions of the quantities of interest. While this is not a particle system like the Hamiltonian systems considered by Boltzmann, it gives a clear picture of the emergence of irreversibility in its large N limit. We have then investigated two versions of a gas confined by a moving piston, which are in fact Hamiltonian particles systems, and we have then shown that the stochastic and the deterministic systems have totally analogous behaviours, confirming the relevance of the stochastic description, in the case of large N. Finally, we have shown that for small N irreversibility in a single system, in the sense of Eq.(28), does not hold. Therefore, this is indeed and emergent property of the macroscipic realm.

Our results have been obtained under the following conditions:

1. the system is made of a very large number of particles, *i.e.* $N \gg 1$;
2. the initial condition is very far from equilibrium, meaning
   - $|n_0 - N/2| \gg \sqrt{N}$ in the Ehrenfest model
   - $|X(0) - X_{eq}| \gg \sigma_{eq}$ in piston model.

As a final remark, let us observe that the notion of typicality we have associated with the thermodynamic laws, pertains also to the Boltzmann approach. It consists of an *objective* operation: counting the cases of interest, and realizing that, with no exception, they behave in the same prescribed way. However, counting requires a finite number of objects. Ensembles, on the other hand, refers to the probability of a continuum. This is one of the technicalities that requires further investigation.

References

[ARV17] G. Auletta, L. Rondoni, and A. Vulpiani, *On the relevance of the maximum entropy principle in non-equilibrium statistical mechanics* The European Physical Journal Special Topics, **226**, 2327 (2017)

[BCV19] M. Baldovin, L. Caprini and A. Vulpiani, *Irreversibility and typicality: a simple analytical result for the Ehrenfest model* Physica A, **524**, 422 (2019).

[B02] R.W. Batterman, *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence* (Oxford University Press, 2002)

[B94] M.V. Berry *Asymptotics, singularities and the reduction of theories* Logic, Methodology and Philosophy of Science **IX**, 597 (1994)

[BJPV] T. Bohr, H.M. Jensen, G. Paladin, and A. Vulpiani *Dynamical systems approach to turbulence* (Cambridge University Press, 1998)

[B96] J. Bricmont *Science of chaos or chaos in science?* Ann. N.Y. Acad. Sci. **775**, 131 (1996)

[CFLV08] P. Castiglione, M. Falcioni, A. Lesne and A. Vulpiani, *Chaos and coarse graining in statistical mechanics* (Cambridge University Press, 2008).

[C98] C. Cercignani, *Ludwig Boltzmann: the man who trusted atoms* (Oxford Univ. Press, 1998).

[CCCV16] L. Cerino, F. Cecconi, M. Cencini, and A. Vulpiani, *The role of the number of degrees of freedom and chaos in macroscopic irreversibility* Physica A **442**, 486 (2016).

[CRV14a] S. Chibbaro, L. Rondoni, and A. Vulpiani, *Reductionism, Emergence and Levels of Reality* (Stringer-Verlag, 2014).

[CRV14b] S. Chibbaro, L. Rondoni, and A. Vulpiani, *On the Foundations of Statistical Mechanics: Ergodicity, Many Degrees of Freedom and Inference* Communications in Theoretical Physics **62**, 469 (2014).

[D15] B. Drossel *On the relation between the second law of thermodynamics and classical and quantum mechanics* in B. Falkenburg and M. Morrison (Ed.s) *Why more is different*, Springer, 2015, pp 41

[D14] H.S. Dumas, *The KAM Story* (World Scientific, 2014)

[FPPRV] M Falcioni, L Palatella, S Pigolotti, L Rondoni, A Vulpiani, *Initial growth of Boltzmann entropy and chaos in a large assembly of weakly interacting systems*, Physica A **385** 170 (2007)

[F56] E Fermi, *Thermodynamics*, (Dover Publications, 1956).

[G07] G. Gallavotti (Ed.) *The Fermi-Pasta-Ulam problem: a status report* (Springer-Verlag, 2007).

[G05] P. Gaspard *Chaos, scattering and statistical mechanics* (Cambridge University Press, 2005)

[G06] J.W. Gibbs it The scientific papers of J. Willard Gibbs, Vol 1 (Longmans, Green and Company. 1906)

[G18] B,V. Gnedenko *Theory of probability* (Routledge, 2018)

[G12] S. Goldstein, *Typicality and notions of probability in physics*, in Y. Ben- Menahem and M. Hemmo (Eds.), *Probability in physics*, Springer-Verlag, 2012, pp. 59.

[G01] S. Goldstein, *Boltzmann's approach to statistical mechanics*, in J. Bricmont et al. (Eds.) *Chance in physics*, Springer-Verlag, 2001, pp. 39.

[H09] K. Huang *Introduction to statistical physics* (CRC press, 2009)

[J57] E.T. Jaynes *Information theory and statistical mechanics* Physical Review, **106**, 620 (1957)

[K57] M. Kac, *Probability and related topics in physical sciences*, (American Mathematical Soc., 1957).

[K13] L.P. Kadanoff *Relating theories via renormalization.* Studies in History and Philosophy of Science Part B, **44**, 22 (2013).

[K49] A.I. Khinchin, *Mathematical Foundations of Statistical Mechanics*, (Dover, 1949).

[LL80] L D Landau and E M Lifshitz, *Statistical Physics*, (Butterworth-Heinemann, 1980)

[L81] O.E. Lanford, *The hard sphere gas in the Boltzmann- Grad limit*, Physica A **106**, 70 (1981).

[L93] J. L. Lebowitz, *Boltzmann's entropy and time's arrow*, Physics Today **46**, 32 (1993).

[Ma85] S. K. Ma, *Statistical Mechanics*, (World Scientific, Singapore, 1985)

[ML63] P. Mazur and J. van der Linden, *Asymptotic form of the structure function for real systems*, J. Math. Phys. **4**, 271 (1963).

[P11] L. Peliti *Statistical mechanics in a nutshell* (Princeton University Press, 2011)

[PS79] I.Prigogine and I. Stengers *La nouvelle alliance: métamorphose de la science* (Gallimard, 1979)

[SV01] G. Shafer and V. Vovk, *Probability and Finance* (J. Wiley and Sons Inc., 2001)

[U95] J. Uffink *Can the maximum entropy principle be explained as a consistency requirement?* Studies in History and Philosophy of Science Part B **26**, 223 (1995)

[VCCPV] A. Vulpiani, F. Cecconi, M. Cencini, A. Puglisi, and D. Vergni (Ed.s) *Large deviations in physics. The Legacy of the Law of Large Numbers* (Springer, 2014).

[Z05] N. Zanghì, *I fondamenti concettuali dell'approccio statistico in fisica*, in V. Allori, M. Dorato, F. Laudisa e N. Zanghì (Eds.), *La natura delle cose: Introduzione ai fondamenti e alla filosofia della fisica*, Carocci, 2005, pp. 202.

[Z18] W.H. Zurek, *Eliminating ensembles from equilibrium statistical physics: Maxwell's demon, Szilard's engine, and thermodynamics via entanglement*, Phys. Rep. **755**, 1 (2018)