

Probability Turns Material: The Boltzmann Equation

Original

Probability Turns Material: The Boltzmann Equation / Rondoni, Lamberto; Di Florio, Vincenzo. - In: ENTROPY. - ISSN 1099-4300. - ELETTRONICO. - 26:2(2024), pp. 1-19. [10.3390/e26020171]

Availability:

This version is available at: 11583/2987617 since: 2024-04-07T17:16:50Z

Publisher:

MDPI

Published

DOI:10.3390/e26020171

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

Probability Turns Material: The Boltzmann Equation

Lamberto Rondoni ^{1,2,*}  and Vincenzo Di Florio ^{1,2,3,*} ¹ Dipartimento di Scienze Matematiche, Politecnico di Torino, Corso Duca Degli Abruzzi 24, 10129 Turin, Italy² INFN, Sezione di Torino, Via Pietro Giuria 1, 10125 Turin, Italy³ Istituto Italiano di Tecnologia (IIT), Via Melen 83, B Block, 16163 Genoa, Italy

* Correspondence: lamberto.rondoni@polito.it (L.R.); vincenzo.diflorio@polito.it (V.D.F.)

Abstract: We review, under a modern light, the conditions that render the Boltzmann equation applicable. These are conditions that permit probability to behave like mass, thereby possessing clear and concrete content, whereas generally, this is not the case. Because science and technology are increasingly interested in small systems that violate the conditions of the Boltzmann equation, probability appears to be the only mathematical tool suitable for treating them. Therefore, Boltzmann's teachings remain relevant, and the present analysis provides a critical perspective useful for accurately interpreting the results of current applications of statistical mechanics.

Keywords: dynamical systems; probability; observables

1. Introduction

Statistical mechanics arose to interpret thermodynamic results from an atomistic perspective. That is an objectively difficult task because thermodynamics is deterministic. It views matter as a continuum characterized by a reduced number of variables, and it is irreversible in time. Partial differential equations are the suitable mathematical tool for such a description of physical objects. On the contrary, the microscopic dynamics of atoms are considered deterministic when dealing with fluids in non-extreme conditions. However, they are time-reversal invariant, with a few notable exceptions, such as Prigogine [1]. Being discrete rather than continuous, these dynamics are represented by systems of many ordinary differential equations. The difficulty, then, does not lie in doubts about the atomistic hypothesis, which have been dispelled long ago [2], when the Brownian motion was explained [3]. The difficulty concerns the above very different mathematical descriptions that have been developed for macroscopic and microscopic dynamics. Indeed, they are so different that they could even look incompatible.

With statistical physics, the gap has been filled adding statistical notions to the purely mechanical ones, and profiting from the mind-boggling number of degrees of freedom that must be integrated out to go from atoms to continuum [4]. This second aspect was initially overlooked by Boltzmann's contemporaries, who criticized him on the grounds of the laws of mechanics [5]. As a matter of fact, certain sentences in Boltzmann's fundamental paper [6] prompted criticism because they lack reference to probabilistic arguments and seem to imply a direct derivation of irreversible relaxation to equilibrium from the purely reversible dynamics of atoms:

"It has thus been rigorously proved that, whatever may be the initial distribution of kinetic energy, in the course of a very long time it must always necessarily approach the one found by Maxwell. The procedure used so far is of course nothing more than a mathematical artifice employed in order to give a rigorous proof of a theorem whose exact proof has not previously been found. It gains meaning by its applicability to the theory of polyatomic gas molecules".

Whether this interpretation of Boltzmann's intention is correct or not, Loschmidt and Zermelo legitimately, and correctly, pointed out that mechanics by itself does not justify



Citation: Rondoni, L.; Di Florio, V. Probability Turns Material: The Boltzmann Equation. *Entropy* **2024**, *26*, 171. <https://doi.org/10.3390/e26020171>

Academic Editor: Eun-jin Kim

Received: 5 January 2024

Revised: 4 February 2024

Accepted: 15 February 2024

Published: 17 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

the result. Boltzmann's reply, then, set things straight once and for all. Although, to date, the argument can be considered quite subtle. In a sense, Boltzmann agreed with the criticism, without admitting any error in his original point of view, but simply elaborating on it. He invoked a probabilistic argument, which does not belong to mechanics, but is powerfully unavoidable: given the number of molecules that make a macroscopic object, it is unreasonable not to take seriously the probabilistic predictions. In his own words, this is Boltzmann's answer to Zermelo:

"The applicability of probability theory to a particular case cannot of course be proved rigorously. If, out of 100,000 objects of a certain kind, about 100 are annually destroyed by fire, then we cannot be sure that this will happen next year. On the contrary, if the same conditions could be maintained for $10^{10^{10}}$ years, then during this time it would often happen that all 100,000 objects would burn up on the same day; and likewise there will be entire years during which not a single object is damaged. Despite this, every insurance company relies on probability theory. It is even more valid, on account of the huge number of molecules in a cubic millimetre, to adopt the assumption (which cannot be proved mathematically for any particular case) that when two gases of different kinds or at different temperatures are brought in contact, each molecule will have all the possible different states corresponding to the laws of probability and determined by the average values at the place in question, during a long period of time. These probability arguments cannot replace a direct analysis of the motion of each molecule; yet if one starts with a variety of initial conditions, all corresponding to the same average values (and therefore equivalent from the viewpoint of observation), one is entitled to expect that the results of both methods will agree, aside from some individual exceptions which will be even rarer than in the above example of 100,000 objects all burning on the same day. The assumption that these rare cases are not observed in nature is not strictly provable (nor is the entire mechanical picture itself) but in view of what has been said it is so natural and obvious, and so much in agreement with all experience with probabilities, from the method of least squares to the dice game, that any doubt on this point certainly cannot put in question the validity of the theory when it is otherwise so useful. It is completely incomprehensible to me how anyone can see a refutation of the applicability of probability theory in the fact that some other argument shows that exceptions must occur now and then over a period of eons of time; for probability theory itself teaches just the same thing".

(We would like to stress one aspect of this reply that is also quite important, although not directly related to the debate between Boltzmann and Zermelo, hence usually not considered. Boltzmann clearly indicates that a mathematical explanation of a physical phenomenon cannot be taken as applicable without restrictions. Indeed, any mathematical model is just that: it cannot replace reality in all its aspects. It suffices to keep this in mind to realize paradoxes may simply lie in the incorrect application of a theory to situations exiting its range of validity. The problem may be not lack of mathematical rigour, but of physical relevance.) One should note that, strictly speaking, Boltzmann reasonings apply to a special class of macroscopic objects, the rarefied gases. Nevertheless, over the years, the Boltzmann equation has been extended to an incredibly wide range of systems and phenomena. First, within the realm of physics, it has been extended to classical systems subjected to external forces, as well as to quantum and relativistic mechanics, with applications, e.g., in solid state physics and astrophysics, cf., e.g., [7,8]. Successively, applications have been produced in fields such as social sciences, economics, and traffic studies [9–13], just to mention a few examples. In mathematics, the study of the Boltzmann equation has also produced fundamental results, of general interest. For instance, the proof of the existence of solutions [14]. More recently, Cercignani's conjecture [15], that has been influential in understanding dissipation from a microscopic point of view, was proven to be substantially correct [16].

Even through all that, one comes to realize that probability, the primary tool of statistical physics, is crucial for accurately handling quantities related to discrete lumps of matter composed of numerous microscopic constituents, especially on the scale where

they seem to exhibit continuous behavior. But, thanks to this tool, statistical physics may in principle be used even beyond of the realm of thermodynamics if it is true, as it is, that even insurance companies rely on it. In fact, statistical physics has turned its attention more and more to the study of small systems, which are not guaranteed to be in Local Thermodynamic Equilibrium (LTE), and require thermodynamic quantities to be interpreted in a different or wider sense than in thermodynamics. Indeed, LTE requires, at least, that systems should possess a microscopic, mesoscopic and macroscopic scale in space and time w.r.t the observer. This condition cannot be ensured for all systems to which statistical physics is applied. This can lead, for instance, to finite size effects and fluctuations entailing blurred boundaries between different states of aggregation matter. This gives probabilities an even wider role in the description of small systems than in the statistical mechanics of macroscopic objects. Often, probability is the only sensible thing to compute. The trouble is that probability is immaterial.

In this paper, we analyze various notions of probability, highlighting this fact, and then demonstrate how it may be seen turning material within the framework of the Boltzmann equation. This may be useful in correctly interpreting a large fraction of present day statistical mechanics results.

2. Probability: Obscure and Practical

How can probability be useful? Karl Popper maintained that probabilistic results cannot be falsified, thus contradicting his concept of science: *For although probability statements play such a vitally important role in empirical science, they turn out to be in principle impervious to strict falsification* [17]. Further:

“In whatever way we may define the concept of probability, or whatever axiomatic formulations we choose: so long as the binomial formula is derivable within the system, probability statements will not be falsifiable. Probability hypotheses do not rule out anything observable; probability estimates cannot contradict, or be contradicted by, a basic statement; nor can they be contradicted by a conjunction of any finite number of basic statements; and accordingly not by any finite number of observations either” [17].

Of course, Popper’s view of science can be criticized, but there is a fact: suppose meteorologists have computed that the probability of rain tomorrow is 0.999. We go out carrying an umbrella, and we find the Sun is shining. Have we refuted the meteorologists’ prediction? Of course not. Consequently, if it rains it also does not prove anything about the mentioned calculations. So, why is probability interesting?

In the mathematical sciences, probability is a rather recent subject. In the XVI century, it was conceived to improve chances of winning in gambling. However, it rapidly came to constitute the basis of statistical and quantum mechanics, as well as of chaos theory and of the sciences of “complexity”. It is dominant in finance, in the interpretation of medical tests, in epidemiology, weather forecasts, etc. On the other hand, can we grasp the meaning of a statement like: “there is a probability of 10^{-40} that a nuclear power plant blows up”? When and how will we experience a number like that, so that we develop an opinion about its significance? What do scientists say?

Joh Archibald Wheeler, who would constantly make use of probabilities in his scientific activity, stated that *“Probability, like time, is a concept invented by humans, and humans have to bear the responsibility for the obscurities that attend it. Obscurities there are whether we consider probability defined as frequency or defined a la Bayes”* [18].

The great probabilist, Bruno de Finetti, in the preface of his treatise on probability theory said,

My thesis, paradoxically, and a little provocatively, but nonetheless genuinely, is simply this:

PROBABILITY DOES NOT EXIST

The abandonment of superstitious beliefs about the existence of Phlogiston, the Cosmic Ether, Absolute Space and Time, . . . , or Fairies and Witches, was an essential step along the road to scientific thinking. Probability, too, if regarded as something endowed with

some kind of objective existence, is no less a misleading misconception, an illusory attempt to exteriorize or materialize our true probabilistic beliefs [19].

Clearly, profound familiarity with the subject does not eliminate obscurities, and probability appears *immaterial*. Simple considerations may help us understand these statements.

The classical notion of probability gives an objective rule to attribute a probability to some expected event E . For **finitely many** possible events, N say, the classical probability is objectively defined by counting as

$$\text{Prob}(E) = \frac{N(E)}{N} \quad (1)$$

where $N(E)$ is number of possible outcomes yielding E . This means that the single events are declared equally probable. Coin tossing is the usual example of this kind: the single events are *heads* and *tails*; therefore, each of them is attributed a probability of $1/2$. It is a simple and appealing notion of probability, that looks particularly useful to deal with situations that totally defeat our understanding. However, it is obviously not suitable e.g., for tossing loaded coins or dice; or for YES or NO answers to “Is it going to rain?” Moreover, paradoxes arise if the set of possible events cannot be counted, particularly for continuous sets of events. Think, for instance, of the Bertrand Paradox: there are three different, apparently equally legitimate, probabilities for the event L , which represents the probability that a chord of a circle of unit radius is longer than $\sqrt{3}$, the length of the side of the inscribed equilateral triangle:

- given two points on the circumference, consider the chord joining them, and take one of the two points as the vertex of one inscribed equilateral triangle. If the chord lies outside the triangle, it is shorter than $\sqrt{3}$; if it intersects the triangle, it is longer than or equal to $\sqrt{3}$. Now, the vertices of the triangle delimit three arcs of equal length, and only one of them corresponds to a chord longer than $\sqrt{3}$. Therefore, the probability of L is $1/3$.
- consider a radius of the circle, and choose a point on it. There is a chord perpendicular to this radius, and its length is larger than $\sqrt{3}$ if the point is closer to the center of the circle than to the circumference. Therefore, the probability of L is $1/2$.
- take a point anywhere within the circle and construct a chord with this point as its midpoint. The chord is longer than $\sqrt{3}$ if the point lies within a concentric disk of radius $1/2$ the radius of the larger circle. The area of the smaller circle is $1/4$ of the full disk; therefore, the probability of L is $1/4$.

Despite the numerous *solutions* that have been devised to solve this conundrum, see, e.g., Ref. [20], it remains that they all refer to some kind of superior principle that can only be subjective. Hence, they cannot be taken as objectively valid, at least not in the sense of the mass of a stone. One may then adopt a subjective notion of probability at the start. This amounts to saying that probability should be assigned by the one interested in it, defining

$$\text{Prob}(E) = \text{maximum someone consistently puts at stake to win 1 in an unbiased game}$$

There is an immediate advantage with respect to the classical notion of probability: it is always possible to unambiguously associate a probability with an event, even in the question of whether or not I am a robot. But there are difficulties as well. First of all, different people may have different views on the event E , and that may make the scientific activity cumbersome.

In some cases, one may think that probabilities are objectively imposed by the sequence of events. One then may rely on a different kinds of counting events, thinking that at each time only a finite number of events has occurred, but assuming that extrapolation to

infinitely many events is possible. This constitutes the frequentist approach. Let N_t be the number of observations, or experiments repeated in time up to time t . Then,

$$\text{Prob}(E) = \lim_{t \rightarrow \infty} \frac{N(E; t)}{N_t}$$

where $N(E; t)$ is the number of outcomes regarding E in the time t . This notion of probability is very suggestive e.g., in dealing with situations that are often repeated in time or space, such as weather forecasts or car accidents. However, many events happen only once or a few times in a lifetime, and can be dramatically positive or negative, like nuclear power plants accidents or becoming Captain Regent of the San Marino Republic. Even harder problems are posed by events that happen frequently, but not enough for us to extract a reliable predictive rule, e.g., earthquakes or pandemics.

From these considerations, one might ask why bother at all to use probabilities. The fact is that the above has not stopped people from referring to probabilities. This is understandable if one realizes that probability, like any other mathematical tool, yields a model, one description of one aspect of reality, and does not intend to represent the whole reality. However, as such, it can change its meaning depending on the problem at hand; and it can do so without causing any inconvenience, enabling us to better understand the problems at hand. Indeed, James Clerk Maxwell argued that

They say that Understanding ought to work by the rules of right reason. These rules are, or ought to be, contained in Logic; but the actual science of Logic is conversant at present only with things either certain, impossible, or entirely doubtful, none of which (fortunately) we have to reason on. Therefore the true Logic for this world is the Calculus of Probabilities, which takes account of the magnitude of the probability (which is, or which ought to be in a reasonable man's mind). This branch of Math., which is generally thought to favour gambling, dicing, and wagering, and therefore highly immoral, is the only 'Mathematics for Practical Men', as we ought to be [21].

How can one be practical? The approach that has proven successful is refraining from adopting a specific interpretation of probability, in general, and to develop a mathematical theory within which unambiguous conclusions can be reached, by means of suitable calculations. In doing so, one must also bear in mind that results will refer to a rational framework, not to reality, and connection with reality will have to be assessed case by case. The mathematical framework suitable to this task is that of axiomatic probability, briefly summarized as follows:

- There is a set Ω called sample space, which contains all the “elementary” events ω ;
- there is a collection of subsets of Ω , $\mathcal{F}(\Omega)$ that has the structure of a σ -algebra of subsets $E \subset \Omega$, representing all (elementary and combined) events to which a probability is to be assigned;
- for all $E \in \mathcal{F}$, $P(E) \geq 0$, $P(\emptyset) = 0$, and $P(\Omega) = 1$, where \emptyset is the empty set;
- for $E_1 \cap E_2 = \emptyset$ one has $P(E_1 \cup E_2) = P(E_1) + P(E_2)$;
- given an infinite sequence of disjoint events, $\{E_i\}_{i=1}^{\infty}$ with $E_i \cap E_j = \emptyset$ if $i \neq j$, one has

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i) \quad (2)$$

Given these axioms, the rest follows, provided one has a suitable method of assigning probabilities to the elementary events. However, this method does not necessarily need to be a general rule endowed with a specific interpretation.

When do we need probabilities? When events (apparently) follow no rule, like the outcomes of coin tossing, or the trajectories of pollen grains in water, that are called random phenomena. Singularly, such phenomena follow no rule, but it may happen that their collective large-scale behaviour is described by (possibly evolving) probabilities that follow precise rules, which are called stochastic processes. Deterministic processes, on the other

hand, are singularly subjected to strict rules, such as those of classical mechanics, in which the future is totally determined by the initial conditions. However, because such evolutions require exact knowledge of the initial conditions, something that is never achievable in practice, probabilities play a fundamental role also in the case of deterministic evolutions, especially those that are known as chaotic. As a matter of fact, determinism and randomness are but points of view, more or less convenient, depending on many factors. One cannot empirically decide the “true” nature of a given phenomenon. One can only try to make our descriptions effective: i.e., satisfactory for our purposes, and we are satisfied when we can make reliable predictions.

Let us investigate how such an elusive and immaterial notion as probability became concrete and, in fact, “material” thanks to Boltzmann’s work and the Boltzmann equation. This may then serve as a guide to the use of probabilities in realms in which they often are the only suitable mathematical tool, but still bear the consequences of being “obscure” and “immaterial”.

3. Mass vs. Probability

Consider N particles, each with d degrees of freedom, which obey the equations of motion $\dot{x} = G(x)$, on their phase space \mathcal{M} , which is n -dimensional, with $n = 2dN$. Let Φ^t represent the operator, which yields the solution of the equation of motion at time t , $\Phi^t x$, if x is the initial condition. The functions of phase, $\mathcal{O} : \mathcal{M} \rightarrow \mathbb{R}$, are considered microscopic properties of such particles. Then, time averages

$$\overline{\mathcal{O}}(x) = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \mathcal{O}(\Phi^s x) ds \quad (3)$$

are postulated to represent macroscopically observable quantities. The idea is that a measurement tool and the macroscopic object of interest interact for a macroscopic time, which is very long compared to microscopic times, and the result is represented by the time average of the quantity of interest. Assuming that one is treating objects amenable to thermodynamic analysis, which are described by a handful of observables, the $t \rightarrow \infty$ limit results in a very good mathematical idealization of the measurement time, since that may last 10^{12} or more microscopic units. Moreover, the dependence on the initial condition x of a trajectory, $\gamma(x) = \{\Phi^t x\}_{t \in \mathbb{R}}$, representing the microscopic history of a given (single) object of interest, is irrelevant. (This is not necessarily true in small systems, in which local thermodynamic equilibrium is not verified; hence, thermodynamic relations may be inapplicable.) This suggests that the time average (3) may be obtained by replacing the impossible-to-compute phase space trajectory with a phase space probability distribution μ that weighs the different regions of \mathcal{M} proportionally to the frequency with which $\gamma(x)$ visits them:

$$\overline{\mathcal{O}}(x) = \int_{\mathcal{M}} \mathcal{O} d\mu = \mathbb{E}_{\mu}[\mathcal{O}]. \quad (4)$$

Here, the right hand side is called the phase space average. Because the left hand side of Equation (4) depends on x , while the right hand side does not, one must specify that the equality is not expected to hold for all $x \in \mathcal{M}$. At most, one can ask that it holds for μ -almost every $x \in \mathcal{M}$, i.e., as long as its possible violations are confined within a set $E \in \mathcal{M}$ of vanishing probability:

$$\mu(E) = \int_E d\mu = 0,$$

which by definition does not affect the integral in Equation (4).

The investigation of the validity of Equation (4) has led to the development of ergodic theory, concerning the properties of dynamical systems and their invariant measures. In its standard formulation, this mathematical theory states that there are four equivalent properties called *ergodicity*, including the validity of Equation (4). Letting (\mathcal{M}, μ, Φ) be a dynamical system with μ , an invariant measure under the dynamics Φ defined on the phase space \mathcal{M} , the following properties are equivalent:

- (E1) For every integrable function $\mathcal{O} : \mathcal{M} \mapsto \mathbb{R}$, time average and phase space average coincide for μ -almost every x , which means for all $x \in \mathcal{M}$ apart from a set E of vanishing measure, i.e., such that $\mu(E) = 0$:

$$\overline{\mathcal{O}}(x) = \mathbb{E}_\mu(\mathcal{O}), \quad \mu - \text{a.e. } x \in \mathcal{M} \tag{5}$$

- (E2) For every measurable $A \subset \mathcal{M}$ and for μ -a.e. $x \in \mathcal{M}$, the fraction of time a trajectory spends in A is given by:

$$\tau_A(x) = \mu(A), \quad \mu - \text{a.e. } x \in \mathcal{M} \tag{6}$$

where the fraction of time is defined by

$$\tau_A(x) = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \chi_A(\Phi^s x) ds \tag{7}$$

- (E3) There are no non-trivial integrals of motion. In other words, let \mathcal{O} be a function of phase, such that $\mathcal{O}(\Phi^t x) = \mathcal{O}(x)$ for all t and for μ -a.e. $x \in \mathcal{M}$, and let \mathcal{O} be integrable. Then,

$$\mathcal{O}(x) = \text{constant}, \quad \mu - \text{a.e. } x \in \mathcal{M} \tag{8}$$

- (E4) The dynamical system is metrically indecomposable. In other words, let A be an invariant measurable set, i.e., $\Phi^t A = A$ for all t . Then, either $\mu(A) = 1$ or $\mu(A) = 0$. When that is the case, the expression $\mathcal{M} = A \cup (\mathcal{M} \setminus A)$ is called metrically trivial decomposition of \mathcal{M} .

It is easy to see that ergodicity may be realized in simple dynamical systems whose phase space trajectories $\gamma(x)$ move parallel to each other and with the same phase space "velocity", i.e., same $G(x)$. Take for instance a two-dimensional torus as \mathcal{M} , and let trajectories be oriented so that they do not close on a periodic orbit. (Cutting the torus to obtain a square, with periodic boundary conditions, it suffices that the components of G be not rational with respect to each other.) This is but an example of an ergodic system whose trajectories remain close to each other in time, while they explore the whole phase space, thus satisfying E2. Now, imagine that probability belongs to phase space sets like mass belongs to chunks of matter moving in space. In other words, let us mimic the conservation of mass in real space with the conservation of probability in phase space, although the difference between the two is huge, as is shown even by the dimensions of real space, which is 3, and of \mathcal{M} , which is $O(2 \times 3 \times 10^{23})$ for spherical atoms. That means that probabilities may evolve, obeying

$$\mu_t(E) = \mu_0(\Phi^{-t}E) \tag{9}$$

and, if the evolving probability has a density, it satisfies the (generalized) Liouville equation

$$\frac{\partial f}{\partial t} = -\text{div}_\Gamma(fG) = -G \cdot \nabla_\Gamma f - f \text{div}_\Gamma G \tag{10}$$

If the equilibrium distribution is the microcanonical ensemble, which, for an ideal gas, is the uniform distribution of microstates with specific energy, and if the initial distribution is not uniform, the ergodic property does not lead in time to the equilibrium distribution. In our example of parallel trajectories, a lump of probability continues to go round and round in \mathcal{M} , never spreading to become uniform. This will be revealed by the phase space averages of observables that will correspondingly fluctuate in time, as the support of this initial probability moves within \mathcal{M} . If one pushes further the analogy between probability and mass, taking the spread of probability in phase space as representing the spread of the molecules of a gas in their container, ergodicity does not suffice to describe the *irreversible* relaxation of macroscopic states to their equilibrium. Therefore, a stronger property, that implies ergodicity, has been considered to describe the most common irreversible phenomenon, see, e.g., Ref. [22]. This condition is called *mixing*, and is defined by two equivalent properties:

(M1) For every pair of measurable sets in the phase space, $A, B \subset \mathcal{M}$, the following holds:

$$\lim_{t \rightarrow \infty} \mu(\Phi^{-t}A \cap B) = \mu(A)\mu(B), \quad \text{or} \quad \lim_{t \rightarrow \infty} \mu(A \cap \Phi^t B) = \mu(A)\mu(B) \quad (11)$$

(M2) for every pair of observables, e.g., $\mathcal{O}, \mathcal{P} \in L_2(\mathcal{M}, \mu)$, the following holds:

$$\lim_{t \rightarrow \infty} \int_{\mathcal{M}} (\mathcal{O} \circ \Phi^t) \mathcal{P} d\mu = \int_{\mathcal{M}} \mathcal{O} d\mu \int_{\mathcal{M}} \mathcal{P} d\mu, \quad (12)$$

$$\text{i.e., } \lim_{t \rightarrow \infty} \mathbb{E}_{\mu}[(\mathcal{O} \circ \Phi^t)\mathcal{P}] = \mathbb{E}_{\mu}(\mathcal{O})\mathbb{E}_{\mu}(\mathcal{P})$$

The meaning of M1 is that, in time Φ^t , B will occupy a fraction of A equal to its measure in the full \mathcal{M} . As A can be taken in any region of phase space, in particular within the support of an invariant measure under consideration, this means that an initial ensemble supported in B will reach in time and will stay in any corner of the support of μ . When the microcanonical ensemble is the relevant invariant measure, this means that the probability “contained” in B will eventually be uniformly distributed over the whole \mathcal{M} . This can be seen observing that the invariance of μ implies $\mu(A \cap \Phi^t B) = \mu(\Phi^{-t}(A \cap \Phi^t B)) = \mu(\Phi^{-t}A \cap B)$, and $\mu(A)\mu(\Phi^t B) = \mu(A)\mu(B)$. Then, for sufficiently long times, one may approximately write

$$\frac{\mu(A \cap \Phi^t B)}{\mu(A)} \approx \mu(B), \quad \text{provided } \mu(A) \neq 0 \quad (13)$$

which becomes exact in the $t \rightarrow \infty$ limit.

Property M2, in particular, expresses the decay of correlations of microscopic events in phase space. When the correlation function of the functions \mathcal{O} and \mathcal{P} , defined by

$$G(t) = \mathbb{E}_{\mu}[(\mathcal{O} \circ \Phi^t)\mathcal{P}] - \mathbb{E}_{\mu}(\mathcal{O})\mathbb{E}_{\mu}(\mathcal{P}) \quad (14)$$

does not vanish, the measurements of \mathcal{O} and \mathcal{P} are, in a sense, not independent. The decay in time of such correlation indicates a sort of irreversible loss of memory about the initial state. Indeed, a simple calculation appears to confirm that relaxation to an equilibrium state, for a Hamiltonian particle system, is achieved under mixing [23]:

$$\begin{aligned} \mathbb{E}_{\mu_t}(\mathcal{O}) &= \int_{\mathcal{M}} \mathcal{O} \rho_t dx = \int_{\mathcal{M}} \mathcal{O}(\rho_0 \circ \Phi^{-t}) dx \\ &= \int_{\mathcal{M}} (\mathcal{O} \circ \Phi^t) \rho_0 dx \xrightarrow{t \rightarrow \infty} \int_{\mathcal{M}} \mathcal{O} dx \int_{\mathcal{M}} \rho_0 dx = \mathbb{E}_{dx}(\mathcal{O}). \end{aligned} \quad (15)$$

Here, the dynamics is assumed to be phase space volume-preserving, so that the Jacobian determinant denoted by J^{-t} equals 1. Then, $\rho_0 = \text{constant}$ is the invariant probability density (its integral equals 1), and mixing has been applied, considering the probability density as a single observable. The result is that the initial value of the observable irreversibly converges to its microcanonical value. A mixing counterpart to the uniform translation on the two-dimensional torus is the so-called Arnold’s CAT map, defined by

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \text{ mod } 1. \quad (16)$$

Despite the apparent formal consistency, numerous difficulties arise with these phase space calculations. Although the same rules applying to mass have been applied to probability, the second obviously remains an *immaterial* abstract concept. For instance, mixing also holds for systems made of a single particle confined within convex walls [24], which surely cannot expand to fill its container, while the probability does. One conceptual difficulty with Equation (15) is that the probability density ρ_t is treated like a microscopic physical property of a given object, which it is not. It represents a probability, that is a completely different object. One microscopic quantity, the kinetic energy say, is an objective

quantity defined for a single concrete object of interest, which is represented by a single geometric point in the phase space. On the other hand, a probability may either refer to collections of systems, or sequences of experiments, each represented by a single phase in the phase space, or it can be subjectively introduced. In all cases, a distribution of phases in phase space is obtained, but the degree of crowding of phase points in different regions has no bearing on the evolution of a given single system. For example, beyond the cases of systems with a single particle, numerous large systems (high probability) could exist within a small region of phase space, each corresponding to uniform mass density. In such instances, the probability would evolve while the mass of the systems of interest remains unchanged. The evolution of phase space probabilities is not directly related to what pertains to thermodynamic quantities. That is solely determined by the equations of motion and the corresponding initial condition. Being associated with a higher or lower amount of companion phases does not modify its fate.

The question arises: under which conditions does probability turn *material*? Indeed, acknowledging, as Maxwell implied, that we have no better tool than probability does not imply that it can be used uncritically. We get a glimpse of that in Poincaré's considerations on the laws that govern the large number of molecules constituting a gas:

"It seems at first that the orderless collisions of this innumerable dust can only engender an inextricable chaos before which the analyst must retire. But the law of great numbers, that supreme law of chance, comes to our assistance. In face of a semi-disorder we should be forced to despair, but in extreme disorder this statistical law re-establishes a kind of average or mean order in which the mind can find itself again" [25].

Understanding these facts definitely helps in uncharted territories. Great numbers and extreme disorder seem to be essential, but often they are not present.

4. Hamiltonian Particle Systems and Ensembles

Because a priori it is not possible to decide which microscopic dynamics are better suited to the task of describing macroscopic systems, one almost universally begins with classical Hamiltonian mechanics, proven extremely successful in describing an incredible variety of macroscopic phenomena. This was Laplace's opinion:

"We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes".

(Note, however, that Laplace made that statement in an essay on probabilities. Indeed, he considered such an intellect out of our reach. This was not the manifesto of reductionism, as it is often portrayed, but the vindication of probabilities [26]). This was a very bold statement, considering the fact that experience was limited to large, observable objects, which typically come in small numbers, like planets and comets. In principle, there is no relation with the exceedingly large assemblies of invisibly tiny molecules that constitute a macroscopic body. This may be a cause of concern; hence, it must be subjected to testing whenever new phenomena are considered. The outcome is that statistical mechanics has greatly advanced by relying on Hamiltonian dynamics.

Consider a system of N particles, each having d degrees of freedom, characterized by a Hamiltonian $H = H(\Gamma)$, where $\Gamma = (\mathbf{q}, \mathbf{p}) \in \mathcal{M}$, and by the following equations of motion:

$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}, \quad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}} \quad (17)$$

As already mentioned, the connection with physical measurements begins from the observation that they take a positive time, and from the assumption that they yield the time

average of one phase function. If the measurement takes a time τ , and it starts when the microstate is Γ , the result of the measurement is postulated to be

$$\overline{\mathcal{O}}_{0,\tau}(\Gamma) = \frac{1}{\tau} \mathcal{O}_{0,\tau} = \frac{1}{\tau} \int_0^\tau \mathcal{O}(\Phi^t \Gamma) dt \quad (18)$$

where by the symbol $\mathcal{O}_{s,t}$ we mean the time integral of \mathcal{O} from time s to time t , and the bar denotes time averaging.

If one accepts this picture, it follows that measurements depend on τ and are therefore subjective because different observers may choose different initial times and different τ 's. They also depend on Γ , hence their result is a random variable because Γ is unknown. However, this contradicts Thermodynamics, which is an objective and deterministic theory, universally confirmed by experimental tests. One possibility to accord Equation (18) with experience is to assume that τ is very large, virtually infinitely larger than the microscopic scales concerning the evolution of Γ , and that over such a long time, \mathcal{O} has explored many times—hence with proper frequencies—its range of values so that the initial condition Γ is irrelevant. The same would hold if, in the time τ , generic phase space trajectories, with negligible exceptions, thoroughly explore the phase space, which is property E2 of Section 3. However, while E2 is sufficient for the result, it is not necessary, and typically impossible. One way or another, we obtain

$$\frac{1}{\tau} \int_0^\tau \mathcal{O}(\Phi^t \Gamma) dt \approx \overline{\mathcal{O}}(\Gamma) = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau \mathcal{O}(\Phi^t \Gamma) dt \approx o \in \mathbf{R} \quad (19)$$

where mathematically the first approximate equality is due to the fact that τ should be very large, although it cannot be infinite, and the second to the fact that the range of \mathcal{O} may not be perfectly explored. As long as the approximations fall below the scale of thermodynamic interest, equality can thus be used, and o represents the result of a measurement. This picture is justified by Fermi as follows:

“Studying the thermodynamical state of a homogeneous fluid of given volume at given temperature [...] we observe that there is an infinite number of states of molecular motion that correspond to it. With increasing time, the system exists successively in all the dynamical states that correspond to the given thermodynamical state. From this point of view we may say that a thermodynamical state is the ensemble of all the dynamical states through which, as a result of the molecular motion, the system is rapidly passing” [27].

Callen adds that, if the system passes sufficiently rapidly through representative atomic states, a measurement effectively averages over all them, and the system is in equilibrium [28]. If, on the other hand, the transitions from atomic state to atomic state is ineffective (e.g., the system is trapped in a small subset of states, or the transition rate is too slow) a macroscopic measurement does not yield a proper average over all possible atomic states, and the system is not in equilibrium. A different description must be adopted in such situations, which are in fact well known; think, e.g., of metal alloys subjected to thermal treatments. However, when the microscopic values of \mathcal{O} are indeed sufficiently rapidly explored, compared to macroscopic observation times, a single system of interest reveals itself through \mathcal{O} as the average over the *ensemble* of such possibilities. To the observer, the system revealed by \mathcal{O} *appears* like the average over the ensemble of its microscopic phases, and the result of the measurement equals an average with respect to a phase space probability distribution, called *ensemble*. However, this implicitly requires the macroscopic state to be stationary; if it shifts during the measurement, the observable cannot explore its range with the frequencies corresponding to that state, and different initial microstates may lead to different observable values. In fact, this is the case of, e.g., ageing systems.

Given a dynamical system like Equation (17), it is not obvious that Equation (19) holds, and even if it does, the statement is so complex that it may be impossible to prove. Therefore, the validity of Equation (19) is commonly postulated together with the assumption that there is an invariant probability distribution μ on \mathcal{M} , such that a measurement yields

Equation (4) for μ -almost every $x \in \mathcal{M}$, and for every phase function \mathcal{O} . What is obvious is that dense exploration of the hugely dimensional phase space cannot be the reasons why experience taught us that Equation (4) can be used, at least in the case of equilibrium systems, with the classical ensembles as probability distributions.

This fact can be explained without invoking the ergodic properties E1–E4. Given a single variable \mathcal{O} , its time average may appear like a full phase space average, even if its phase space trajectory $\Phi^t\Gamma$ explores a limited region of \mathcal{M} . It suffices that the range of values of \mathcal{O} be visited with proper frequencies; then the result is the same. Does this ever happen? Khinchin proved that this does indeed happen for the relevant observables of rarefied gases [29]. He noted that the subtleties of microscopic dynamics count little compared to the following:

- (a) macroscopic systems are made of very many particles: $N \gg 1$;
- (b) only several and special phase functions are physically relevant;
- (c) it does not matter if ensemble averages disagree with time averages on limited sets of trajectories.

For rarefied gases, the relevant phase functions are sums of molecular contributions, $\mathcal{O}(\Gamma) = \sum_{n=1}^N o(\mathbf{q}_n, \mathbf{p}_n)$, where $(\mathbf{q}_n, \mathbf{p}_n)$ is the vector of configurations and momentum of the n -th particle. These functions are appropriate for the pressure, temperature and density of rarefied gases, whose energy can also be expressed as $H = \sum_{n=1}^N h(\mathbf{q}_n, \mathbf{p}_n)$, because interactions among particles are energetically negligible. (Interactions are, however, essential for the condition of LTE, hence for the existence of thermodynamic properties, to be established.) Then, in the microcanonical case, which here means a uniform probability distribution in \mathcal{M} , Khinchin proved the validity of the following relation:

$$\text{Prob} \left(\frac{|\overline{\mathcal{O}} - \mathbb{E}_{\text{mc}}(\mathcal{O})|}{|\mathbb{E}_{\text{mc}}(\mathcal{O})|} \geq K_1 N^{-1/4} \right) \leq K_2 N^{-1/4}, \quad (20)$$

where K_1 and K_2 are positive constants, $\overline{\mathcal{O}}$ is the time average of \mathcal{O} , and $\mathbb{E}_{\text{mc}}(\mathcal{O})$ its phase space average in the microcanonical ensemble. In other words, the probability (in the microcanonical sense) that time averages differ by a small amount from the phase space averages is small if N is large: the larger N the smaller the probability of even smaller differences.

In this framework, the physically relevant ergodicity follows by and large from the $N \gg 1$ condition, combined with the validity of the law of large numbers. That makes the range of sum variables of interest quite narrow, close to a single constant, when the single molecular contributions o can be considered independent, identically distributed, random variables. Then, exploration of the range of all the observables of interest does not take a long time. (See, e.g., Ref. [30] for an interesting analysis of the observables that relax to an equilibrium state, while others do not relax, because the underlying Toda dynamics are integrable, hence, not ergodic.) The details of the microscopic dynamics become irrelevant, while ensembles, i.e., probabilities in phase space, turn considerably useful. Even though probability, per se, is an “immaterial” and abstract mathematical notion, it becomes quite useful in the description of equilibrium states, under the above conditions. With the derivation of the Boltzmann equation a further step is taken forward, allowing time evolution.

5. Boltzmann Equation from the Probabilistic BBGKY Hierarchy

Let us specify the above general treatment of probabilities in phase space to the case of N identical spherical hard particles of mass 1. These particles are subjected to no external forces, and only interact elastically when they collide. We may think of their interaction potential as being zero when the particle centers are at a distance larger than their diameter, and infinitely high when this distance equals their diameter, so that particles are neither deformed, nor do they penetrate each other, when they collide. Denote a probability density in the phase space \mathcal{M} by

$$f_N^{(N)}(\Gamma; t) = f_N^{(N)}(\mathbf{q}_1, \mathbf{p}_1; \dots; \mathbf{q}_N, \mathbf{p}_N; t) \tag{21}$$

with notation indicating the joint probability of N particles out of N . In absence of external forces, there are no accelerations between collisions; hence, $\dot{\mathbf{p}}_i = 0$ for every particle, and $\dot{\Gamma} = (\dot{\mathbf{q}}_1, 0; \dots; \dot{\mathbf{q}}_N, 0)$. The Liouville equation can be written as

$$\frac{\partial f_N^{(N)}}{\partial t} + \left(\text{div}_{\Gamma} \dot{\Gamma} + \dot{\Gamma} \cdot \nabla_{\Gamma} \right) f_N^{(N)} = \frac{\partial f_N^{(N)}}{\partial t} + \sum_{i=1}^N \mathbf{p}_i \cdot \nabla_{\mathbf{q}_i} f_N^{(N)} = 0 \tag{22}$$

where we have used the fact that the dynamics is Hamiltonian; hence, the first term in brackets is null. If we introduce the s -particle distribution function

$$f_N^{(s)}(\mathbf{q}_1, \mathbf{p}_1; \dots; \mathbf{q}_s, \mathbf{p}_s; t) = \int d\mathbf{q}_{s+1} d\mathbf{p}_{s+1} \dots d\mathbf{q}_N d\mathbf{p}_N f_N^{(N)}(\mathbf{q}_1, \mathbf{p}_1; \dots; \mathbf{q}_N, \mathbf{p}_N; t) \tag{23}$$

and we integrate the Liouville equation over the variables $\mathbf{q}_{s+1}, \mathbf{p}_{s+1} \dots \mathbf{q}_N, \mathbf{p}_N$, we obtain [31]:

$$\frac{\partial f_N^{(s)}}{\partial t} + \sum_{i=1}^s \mathbf{p}_i \cdot \nabla_{\mathbf{q}_i} f_N^{(s)} = F^{(s)}(f_N^{(s+1)}, f_N^{(s+1)'}) \tag{24}$$

where $F^{(s)}$ is a function of the joint $(s + 1)$ particles distribution before collision $f_N^{(s+1)'}$, and of the corresponding distribution after collision $f_N^{(s+1)}$. Therefore, Equation (24) is not closed: computing $f_N^{(s)}$ requires knowledge of $f_N^{(s+1)}$. The set of distributions $f_N^{(s)}$ and their evolution equations, Equation (24), constitute the *BBGKY hierarchy*. In particular, taking $s = 1$ and making $F^{(1)}$ explicit by solving the elastic collision dynamics, one obtains

$$\frac{\partial f_N^{(1)}}{\partial t} + \mathbf{p} \cdot \nabla_{\mathbf{q}} f_N^{(1)} = (N - 1)\sigma^2 \int_{\mathbb{R}} \int_{S_-} [f_N^{(2)'} - f_N^{(2)}] |(\mathbf{p} - \mathbf{p}_*) \cdot \mathbf{n}| d\mathbf{p}_* d\mathbf{n} \tag{25}$$

for the one-particle probability distribution function, obtained integrating out the coordinates and momenta of all particles but one. Here, σ is the diameter of one particle; hence, σ^2 is the collision cross section; \mathbf{n} is the unit vector in the direction going from the center of the test particle of position \mathbf{q} and momentum \mathbf{p} , labelled by 1, to that of another particle of momentum \mathbf{p}_* ; the range of integration S_- is the hemisphere within which $(\mathbf{p} - \mathbf{p}_*) \cdot \mathbf{n} < 0$, i.e., particles directed toward each other before collision; and $(N - 1)$ is the number of particles with which the test particle, can collide, assuming they all have the same probability of doing it, having integrated over all their momenta \mathbf{p}_* . Moreover, having denoted by $f_N^{(2)}$ the two-particle distribution function, whose expression immediately before and after the collision respectively takes the form

$$f_N^{(2)'} = f_N^{(2)}(\mathbf{q}, \mathbf{p}'; \mathbf{q} + \sigma\mathbf{n}, \mathbf{p}_*'; t) \quad \text{and} \quad f_N^{(2)} = f_N^{(2)}(\mathbf{q}, \mathbf{p}; \mathbf{q} + \sigma\mathbf{n}, \mathbf{p}_*; t) \tag{26}$$

where $\mathbf{q}_* = \mathbf{q} + \sigma\mathbf{n}$ gives the position of the second particle at the moment of collision, and

$$\mathbf{p}' = \mathbf{p} - [\mathbf{n} \cdot (\mathbf{p} - \mathbf{p}_*)]\mathbf{n} \quad \text{and} \quad \mathbf{p}_*' = \mathbf{p}_* + [\mathbf{n} \cdot (\mathbf{p} - \mathbf{p}_*)]\mathbf{n} \tag{27}$$

where \mathbf{p}' and $\mathbf{p}_*' are, respectively, the momenta of the test particle and of the other colliding particle before collision, while \mathbf{p} and \mathbf{p}_* denote the corresponding momenta after collision.$

To solve this equation, one needs a closure assumption on $f_N^{(2)}$. The equation

$$f_N^{(2)}(\mathbf{q}, \mathbf{p}; \mathbf{q}_*, \mathbf{p}_*; t) = f_N^{(1)}(\mathbf{q}, \mathbf{p}; t) f_N^{(1)}(\mathbf{q}_*, \mathbf{p}_*; t) \tag{28}$$

is known in the kinetic theory of gases as the *stosszahlansatz*, or hypothesis of molecular chaos. It amounts to assuming that when two particles collide, they are independent. As the dynamics of particles is deterministic, this statement may only have a statistical meaning,

justified by some kind of dynamical *randomness*, that can be legitimately called *chaos*. (In the theory of dynamical systems, the term chaos is often used to indicate systems that have at least one positive Lyapunov exponent. This notion was not available to Boltzmann and, indeed, he did not need this kind of chaos.) Then, writing f in place of $f_N^{(1)}$ to simplify the notation, and recalling that $\mathbf{v} = \mathbf{p}$ because we assumed unit masses, Equation (25) takes the form

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{q}} f = (N - 1)\sigma^2 \int_{\mathbb{R}} \int_{S_-} [f' f'_* - f f_*] (\mathbf{v} - \mathbf{v}_*) \cdot \mathbf{n} |d\mathbf{v}_* d\mathbf{n} \tag{29}$$

where primes mean before collision and $*$ second particle, and small particles, small σ , imply $\mathbf{q}_* \approx \mathbf{q}$. This is the celebrated *Boltzmann equation*. Given its probabilistic derivation, which begins with the distribution $f_N^{(N)}$ in phase space, $f(\mathbf{q}, \mathbf{p}, t)d\mathbf{q}d\mathbf{p}$ represents the probability that particle 1 be found in the volume element $d\mathbf{q}$ around \mathbf{q} , with momentum within the set $d\mathbf{p}$ around \mathbf{p} , at time t . The classical interpretation of probability we have so far adopted, then states that, given a very large collection of independent macroscopically equal systems, each made of N hard spheres, the fraction of them having particle 1 within $d\mathbf{q}d\mathbf{p}$ equals $f d\mathbf{q}d\mathbf{p}$. If one assumes, in addition, that particles are indistinguishable, a simple combinatorial factor turns $f d\mathbf{q}d\mathbf{p}$ into the fraction of ensemble members having one particle, whichever it is, in $d\mathbf{q}d\mathbf{p}$ at time t . It remains to connect this probability to material properties, like the density of the gas in $d\mathbf{q}d\mathbf{p}$.

6. Boltzmann Equation from Mass Balance in Real Space

The probability distribution f appearing in Equation (29) is obtained projecting the one for N particles defined on the corresponding $2dN$ dimensional phase space, down to the $2d$ dimensional one-particle space. Given a single system of interest and the ensemble interpretation of the probability in phase space, the condition $f_N^{(1)}(\mathbf{q}, \mathbf{p}; t) > 0$ does not imply that particle 1 actually lies within the volume $d\mathbf{q}d\mathbf{p}$, since only a fraction of the systems of the ensemble has particle 1 in $d\mathbf{q}d\mathbf{p}$: particle 1 of the complementary fraction of systems does not lie in $d\mathbf{q}d\mathbf{p}$. (If particles are indistinguishable and the 1-particle probability distribution has been correspondingly normalized, the same can be stated for any single particle of the systems of the ensemble.) In particular, $d\mathbf{q}d\mathbf{p}$ could be empty because the BBGKY hierarchy can be developed for any $N \geq 1$, and a few particles only occupy a limited region of space. Therefore, one could conclude that even the Boltzmann equation is *immaterial*, and there is no compelling reason to take it as a description of a given material system.

However, a physical derivation of the Boltzmann equation is also available. It does not start from a probability distribution on a phase space \mathcal{M} , which is a continuum of geometric points, each of which represents a whole N -particles system. This derivation refers to a single system made of a very large number N of small, but finite size, particles of mass 1, with positions and velocities in a given volume $\Delta\mathbf{q}\Delta\mathbf{p}$. The particles in such a volume constitute a discrete rather than a continuous set, and can be represented by a continuous density of mass only when many of them populate $\Delta\mathbf{q}\Delta\mathbf{p}$. Therefore, given the finite size of the particles, such a volume has to be small, to appear like a point on the macroscopic scale. However, unlike the volumes $d\mathbf{q}d\mathbf{p}$ of the previous section, which can be arbitrarily small, it cannot be infinitesimal. Indeed, it has to be much larger than the volume of a single particle, so that many of them can be lodged in it [32,33]. This highlights another essential difference between probability and mass densities. Let us denote by C_{ij} of size $\Delta\mathbf{q}\Delta\mathbf{p}$ one cell centered around a discrete set of lattice points $(\mathbf{q}_i, \mathbf{p}_j)$. The mass density within C_{ij} is defined as the mass of one particle times the number density $\rho(\mathbf{q}_i, \mathbf{p}_j; t)$, the number of particles per unit volume around $(\mathbf{q}_i, \mathbf{p}_j)$. Letting n_{ij} be the number of particles in cell C_{ij} , at time t , one may thus write

$$n_{ij}(t) = \rho(\mathbf{q}_i, \mathbf{p}_j; t)\Delta\mathbf{q}\Delta\mathbf{p}; \quad \sum_{i,j} n_{ij}(t) = N \tag{30}$$

and may try to approximate this set of discrete mass densities using a continuous function ρ , so that

$$N = \int_{V_\mu} \rho(\mathbf{q}, \mathbf{p}; t) d\mathbf{q} d\mathbf{p} \tag{31}$$

where V_μ is the set of positions and momenta available to particles of the system. This makes no sense unless n_{ij} is very large in each cell C_{ij} ; hence, such cells must be relatively large. Furthermore, particles have to collide: their motion has to be random, so that a uniform distribution can be rapidly achieved within each C_{ij} . Large n_{ij} , together with dynamical randomness, makes the particles entering or leaving a cell just a negligible number, compared to those staying inside. At the same time the cells must be sufficiently small compared to the observation scale, that they appear like a point. A wide separation of scales, possible only if N is very large, is required.

Let us now follow the motion of one set of particles lying in one cell $\Delta\mathbf{q}\Delta\mathbf{p}$ at time t , as they stroll in V_μ . Let us assume that ρ accurately describes the mass density as a continuum. In the absence of external forces, collisions are the only mechanism that allows velocities to change. If, in a time dt , the particles in $\Delta\mathbf{q}\Delta\mathbf{p}$ undergo no collisions, the position \mathbf{q} of one of them turns $\mathbf{q} + \mathbf{v}dt$, while its momentum \mathbf{p} is unchanged, and all the particles in the cell $\Delta\mathbf{q}\Delta\mathbf{p}$, centered about (\mathbf{q}, \mathbf{p}) , will be found in the cell $\Delta\hat{\mathbf{q}}\Delta\hat{\mathbf{p}}$ centered about $(\mathbf{q} + \mathbf{v}dt, \mathbf{p})$. This is achieved by evolving in time all points of the original cell with the same dynamics of the particles in it. Therefore, in the absence of collisions, one has

$$\rho(\mathbf{q} + \mathbf{v}dt, \mathbf{p}, t + dt) \Delta\hat{\mathbf{q}}\Delta\hat{\mathbf{p}} = \rho(\mathbf{q}, \mathbf{p}, t) \Delta\mathbf{q}\Delta\mathbf{p} \tag{32}$$

As \mathbf{q} and \mathbf{p} are canonically conjugate variables, and between two consecutive collisions each particle can be considered by itself as a whole Hamiltonian system, we obtain $\Delta\hat{\mathbf{q}}\Delta\hat{\mathbf{p}} = \Delta\mathbf{q}\Delta\mathbf{p}$, and

$$\rho(\mathbf{q} + \mathbf{v}dt, \mathbf{p}, t + dt) = \rho(\mathbf{q}, \mathbf{p}, t) \tag{33}$$

On the other hand, if particles collide, one additional term arises:

$$\rho(\mathbf{q} + \mathbf{v}dt, \mathbf{p}, t + dt) = \rho(\mathbf{q}, \mathbf{p}, t) + \left(\frac{d\rho}{dt} \right)_{coll} dt \tag{34}$$

which defines the collision term. Finally, expanding the left-hand side of Equation (34) to first order in dt , one obtains

$$\rho(\mathbf{q}, \mathbf{p}, t) + \frac{\partial\rho}{\partial t} dt + \mathbf{v} \cdot \nabla_{\mathbf{q}} \rho dt = \rho(\mathbf{q}, \mathbf{p}, t) + \left(\frac{d\rho}{dt} \right)_{coll} dt \tag{35}$$

because ρ depends on t both explicitly and implicitly through the position coordinates. Dividing by dt yields:

$$\frac{\partial\rho}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{q}} \rho = \left(\frac{d\rho}{dt} \right)_{coll} \tag{36}$$

where $\mathbf{v} \cdot \nabla_{\mathbf{q}} \rho$ represents streaming in position space. There is no streaming in momentum space, because no field acts on the particles, but particles may enter or exit $\Delta\mathbf{p}$ thanks to collisions.

Relation (36) expresses the conservation of mass, and does not require any interpretation of a probabilistic nature: it is based on the objective counting of particles (or measurement of mass density) and on their deterministic Hamiltonian dynamics. Assuming that instantaneous collisions of three or more particles are negligible in a gas of hard spheres, the collision term should be expressed as a function of the number of pairs of particles located inside $\Delta\mathbf{q}$ that may either enter (gain term) or leave (loss term) $\Delta\mathbf{p}$ because of collisions. Let $\rho^{(2)}$ be the density of such pairs. As the initial condition of the N particles is not known, going beyond Equation (36) requires some assumption on the collision term, analogously to Equation (25). One may then rely on the large value of N , on the smallness

of the particles, and on the disorder produced in configuration and velocity space by the collisions, which make sensible the factorization of $\rho^{(2)}$:

$$\rho^{(2)}(\mathbf{q}, \mathbf{p}; \mathbf{q} + \sigma \mathbf{n}, \mathbf{p}_*; t) = \rho^{(1)}(\mathbf{q}, \mathbf{p}; t) \rho^{(1)}(\mathbf{q} + \sigma \mathbf{n}, \mathbf{p}_*; t) \tag{37}$$

That, indeed, means that each of the $n_{\Delta \mathbf{q}}$ particles in an elementary volume $\Delta \mathbf{q}$ may collide with any of the remaining $n_{\Delta \mathbf{q}} - 1 \approx n_{\Delta \mathbf{q}}$ either for a gain or a loss of particles in the elementary volume $\Delta \mathbf{p}$. This is a rarefied gas condition, because it means that no particle can hide behind another one. Then, integrating over all possible velocities yields an expression like Equation (29), which makes sense only if particles are many and small. The equation for the mass distribution turns identical to Equation (29), introducing $\rho = \rho^{(1)} / N$. Therefore, the Boltzmann equation is obtained again. Its solution ρ enjoys the same properties that f enjoys.

At the same time, the conceptual difference between f of Equation (25) and ρ of Equation (36) could not be greater. In the first case one considers the fraction of identical abstract systems that enjoy certain properties out of the total of an ideal hypothetical continuum of identical systems; in the second case one considers the number of particles of a single concrete, experimentally observable system. Probabilistically, all calculations can be performed for a large collection of identical systems made of any number $N \geq 1$ of particles; physically, a single systems is considered and N must be huge. At this level of description, there is no stringent reason for one quite sophisticated abstract object like f to behave similarly to a material tangible object, such as ρ . Nevertheless, this is precisely what happens if probability is required to evolve like a fluid made of geometric points obeying the equations of motion of the system that each of them separately represents, and the *stosszahlansatz* holds. That may well be mathematically too hard to prove within a general Hamiltonian framework, also because it is not verified in a series of systems, hard to fully classify, cf. integrable systems. However, experience has demonstrated that it does hold in very many situations; situations in which we may legitimately state that *probability turns material*: probability can be formally identified with mass.

What is the cost of such an identification? When can the *stosszahlansatz* be justified? In the first place, one would think that large N makes particles independent before colliding, because they may have hardly met before, but this is not enough. Particles are not independent after having collided with each other. Moreover, the factorization condition

$$f_N^{(s)}(q_1, p_1, \dots, q_s, p_s; t) = \prod_{j=1}^s f_N^{(1)}(q_j, p_j; t) \tag{38}$$

is hard to obtain because particles do not overlap, and s of them occupy a volume of order $O(s\sigma^3)$. Thus, large s increases the region to be excluded from the free motion of particles, resulting in correlations among them that may persist over time. This suggests that $s = 2$ is the best candidate for the postulated independence. Also, large N helps in making rare the encounters between the same pair of particles only if the density is small; otherwise packed particles may collide forever with the same small group of companions. Then, σ should be small. On the other hand, collisions of hard spheres are defocussing, and have a randomizing effect on positions and velocities, which is beneficial for the decay of correlations. Therefore, N and σ should be tuned so that the total cross section is not negligible. Grad identified the scaling regime in which all the above is realized, now known as Boltzmann–Grad limit [34]. It consists in keeping $N\sigma^2$ positive and finite, while N grows:

$$N\sigma^2 = \text{constant} > 0 \quad \text{which implies} \quad \sigma^2 \sim \frac{1}{N}, \quad N\sigma^3 \sim N^{-\frac{1}{2}} \tag{39}$$

In this limit, the volume occupied by particles is negligible, while the total cross section is finite. More precisely, let λ denote the mean free path and τ the mean free time as, respectively, the typical microscopic length scale of the system, which is the distance

traveled on average by a particle between collisions, and the corresponding mean time. The first is defined by the ratio between the volume and the total cross section of the particles, the second by the ratio between this distance and the typical speed. Then, denoting by $|\Omega|$ the available volume, and letting $\beta = 1/k_B T$, so that the typical speed for particles of mass 1 is $\sqrt{2/\beta}$, we have

$$\lambda = \frac{|\Omega|}{N\pi\sigma^2}, \quad \tau = \sqrt{\frac{\beta}{2}} \frac{|\Omega|}{N\pi\sigma^2} \quad (40)$$

which remain finite in the Boltzmann–Grad limit. This implies a net effect produced by particle collisions, which is necessary to randomize their motion. Together with the vanishing of the excluded volume, the limit of Equation (40) is also necessary to make correlations rapidly decay, when N is large. Physically, this picture corresponds to a rarefied gas, in which particles collide, but the time average of their interaction energy is negligible compared to the time average of their kinetic energy. Whether the assumptions apply or not, only experience can tell. An important result implied by the Boltzmann–Grad limit is that $f_N^{(2)}$ remains factorized in time, if the solution of the s -th equation of the BBGKY hierarchy, $f_N^{(s)}$, exists is unique, and starts factorized as s factors $f_N^{(1)}$ for all s [31].

Arguably, the most relevant conceptual result stemming from the Boltzmann equation is the H -theorem. It can be concisely illustrated as follows: let the H -functional be defined by

$$H(f) = \int_{V \times \mathbb{R}^3} f \log(f) d\mathbf{p}d\mathbf{q} \quad (41)$$

where $V \subset \mathbb{R}^3$ is the volume in which a gas is contained. For f a solution of the Boltzmann equation within the volume V , the H -theorem states

$$\frac{d}{dt} H(f) \leq 0 \quad (42)$$

where the equality holds if and only if f is constant for $\mathbf{q} \in V$ and the \mathbf{p} are distributed as a Maxwellian probability distribution.

A point that may at times be overlooked, is that the solution of the Boltzmann equation requires a specification of initial and boundary conditions, like any other integro-differential equation. The form of the boundary of V is crucial, particularly for the validity of the H -theorem and, in general, for the overall solution. For instance, reflecting or periodic boundaries are commonly used, and the boundary is typically assumed to be piece-wise smooth, although works on irregular boundaries have been performed, cf. [35]. Moreover, V must be compact for physical properties such as pressure to have a definite equilibrium positive value.

The H -theorem shows that the Boltzmann equation exhibits irreversible behaviour, which appears in contrast to the time-reversal invariant laws of classical mechanics governing the motion of particles. In reality, various approximations and hypotheses, which are extraneous to the dynamics of an N particle system, have been introduced to pass from the evolution of probabilities in phase space to the Boltzmann equation. The N particle system has been approximated by a system consisting of infinitely many particles, verifying the molecular chaos hypothesis, through the Boltzmann–Grad limit, thus breaking time reversibility. The interesting fact is that such approximations and hypotheses perfectly represent the physics of a rarefied gas.

We conclude this section by noting that there exists quantum mechanical and relativistic generalizations of the Boltzmann equation, cf., e.g., Refs. [7,36]. These generalizations have proven successful in an exceedingly wide range of applications, including transport of electrons in solid conductors and in astrophysical plasmas. The most direct extensions are the classical ones, specifically concerning particles that are accelerated by external fields, such as electric fields. In this case, streaming in velocity space is introduced in addition to the streaming in position space. Denoting by \mathbf{F} the external force and by m the mass of the

particles, its effect on \mathbf{v} appears as the acceleration \mathbf{F}/m . This acceleration smoothly guides a particle inside or outside an element of velocity space, analogous to how velocity guides a particle inside or outside an element of position space. Consequently, the Boltzmann equation may be re-written as

$$\frac{\partial \rho}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{q}} \rho + \frac{\mathbf{F}}{m} \cdot \nabla_{\mathbf{v}} \rho = \left(\frac{d\rho}{dt} \right)_{coll} \quad (43)$$

In sufficiently hot plasmas made of charged particles of charge c , collisions can be neglected. Thus, denoting the electric and magnetic fields acting on these particles as \mathbf{E} and \mathbf{B} , respectively, Equation (43) can be re-written as

$$\frac{\partial \rho}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{q}} \rho + \frac{c}{m} (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \nabla_{\mathbf{v}} \rho = 0, \quad (44)$$

which is known as the Vlasov equation. In reality, variations of the Boltzmann equation are numerous. The kinetic theory of wave turbulence [37] finds applications from quantum to astrophysical phenomena, including anomalous transport. There are Boltzmann equations for thermostatted rather than Hamiltonian dynamics, and for a wide spectrum of variously interacting systems, such as the components of blood, or the individuals in a social network etc. It is impossible to list all of them. Of course, there is a difference between these variations and the original Boltzmann equation: the interaction rules are often rather loosely known, which makes the fundamental hypothesis of molecular chaos hard to justify. Nevertheless, these variations provide valuable insights into non-standard and even non-physical phenomena. The interested readers may easily find the relevant literature, which is not within our scope.

7. Concluding Remarks

In this paper we have reviewed some of the reasons that, in principle, could make probability an abstract physically inapplicable mathematical tool, while it is clear that it is arguably the most effective tool physicists can use. The point is that such effectiveness does not arise for free; it is the result of informed, repeatedly tested, and verified choices, as revealed by the derivation of the Boltzmann equation. Uncritical use of probabilities can easily lead to errors. Keeping in mind the teachings of statistical mechanics, which have unfolded since Boltzmann's glorious work of 180 years ago, is thus beneficial for future developments, which will venture in largely uncharted territory, often at variance with our established knowledge of macroscopic objects. Indeed, the microscopic fluctuations are not observable in the behaviour of macroscopic objects in thermodynamic conditions, although present day technology allows us to measure them in microscopic observations [38,39]. On the contrary, fluctuations may even dominate the behaviour of small systems, and require suitable approaches, whose results must be properly interpreted, especially if one insists on using the thermodynamic language. For instance, when dealing with situations in which there is no dissipation, one may find that the derivations of thermodynamic relations are only formal, and that the relevant quantities only have a statistical meaning. However, even more problematic may be the extension to nonequilibrium (dissipative or non-dissipative) states, where even the formal thermodynamic analogies may fail [40]. At present, one must proceed by examples, as a comprehensive theory of small and nonequilibrium systems is still missing, although notable works are being performed, see, e.g., Ref. [41] for one review.

Author Contributions: Writing—original draft preparation, L.R.; Writing—review and editing, L.R. and V.D.F. All authors have read and agreed to the published version of the manuscript.

Funding: Research funded by the Italian Ministry of University and Research (MUR) through the grant PRIN2022-PNRR project (No. P2022Z7ZAJ) "A Unitary Mathematical Framework for Modelling Muscular Dystrophies" (CUP: E53D23018070001). This work has been performed under the auspices of Italian National Group of Mathematical Physics (GNFM) of the Istituto Nazionale di Alta Matematica (INdAM).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: L.R. gratefully acknowledges the generous hospitality of the Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology, India, where part of this work has been performed.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Prigogine, I. *From Being to Becoming: Time and Complexity in the Physical Sciences*; Freeman & Co.: Dallas, TX, USA, 1980.
2. Feynman, R.P.; Leighton, R.B.; Sand, M. *The Feynman Lectures on Physics*; Addison Wesley: Reading, MA, USA, 1963; Volume 1.
3. Einstein, A. *Investigations on the Theory of the Brownian Movement*; Dover Publications: New York, NY, USA, 1956.
4. Chibbaro, S.; Rondoni, L.; Vulpiani, A. *Reductionism, Emergence and Levels of Reality. The Importance of Being Borderline*; Springer: Berlin/Heidelberg, Germany, 2014.
5. Cercignani, C. *Ludwig Boltzmann: The Man Who Trusted Atoms*; Oxford University Press: Oxford, UK, 1998.
6. Boltzmann, L. *Weitere Studien Liber das Wärmegleichgewicht unter Gasmolekülen*; Sitzungsberichte der Akademie der Wissenschaften; k. k. Hof- und Staatsdruckerei: Vienna, Austria, 1972; Volume 66, pp. 275–370; English translation in: S.G. Brush *Kinetic Theory; Irreversible Processes*; Pergamon Press: Oxford, UK, 1966; Volume 2, pp. 88–175.
7. Cercignani, C.; Kremer, G.M. *The Relativistic Boltzmann Equation: Theory and Applications*; Birkhäuser: Basel, Switzerland, 2002.
8. Haug, H.; Jauho, A.-P. *Quantum Kinetics in Transport and Optics of Semiconductors*; Springer: Berlin/Heidelberg, Germany, 2007.
9. Preziosi, L.; Rondoni, L. Conservative energy discretization of Boltzmann collision operator. *Q. Appl. Math.* **1999**, *57*, 699–721. [[CrossRef](#)]
10. Pareschi, L.; Toscani, G. *Interacting Multiagent Systems: Kinetic Equations and Monte Carlo Methods*; Oxford University Press: Oxford, UK, 2013.
11. Tosin, A.; Zanella, M. Boltzmann-type description with cutoff of Follow-the-Leader traffic models. In *Trails in Kinetic Theory: Foundational Aspects and Numerical Methods*; Springer: Cham, Switzerland, 2021; pp. 227–251.
12. Toscani, G.; Brugna, C. Wealth redistribution in Boltzmann-like models of conservative economies. In *Econophysics and Economics of Games, Social Choices and Quantitative Techniques*; Springer: Milan, Italy, 2010; pp. 71–82.
13. Albi, G.; Cristiani, E.; Pareschi, L.; Peri, D. Mathematical models and methods for crowd dynamics control. In *Crowd Dynamics, Volume 2: Theory, Models, and Applications*; Birkhäuser: Cham, Switzerland, 2020; pp. 159–197.
14. DiPerna, R.J.; Lions, P.L. On the Cauchy Problem for Boltzmann Equations: Global Existence and Weak Stability. *Ann. Math.* **1989**, *130*, 321–366. [[CrossRef](#)]
15. Cercignani, C. H-theorem and trend to equilibrium in the kinetic theory of gases. *Arch. Mech. Stosow.* **1982**, *34*, 231–241.
16. Villani, C. Cercignani’s conjecture is sometimes true and always almost true. *Commun. Math. Phys.* **2003**, *234*, 455–490. [[CrossRef](#)]
17. Popper, K. *The Logic of Scientific Discovery*; Routledge: London, UK, 1959.
18. Wheeler, J.A. Information, Physics, Quantum: The Search for Links. In Proceedings of the 3rd International Symposium on Foundations of Quantum Mechanics, Tokyo, Japan, 28–31 August 1989; pp. 54–368.
19. de Finetti, B. *Theory of Probability*; John Wiley & Sons: New York, NY, USA, 1990; Volume 1.
20. Jaynes, E.T. The Well-Posed Problem. *Found. Phys.* **1973**, *3*, 477. [[CrossRef](#)]
21. Maxwell, J.C. Chapter V—Letter to L Campbell Esq. In *The Life of James Clerk Maxwell*; Campbell, L., Garnett, W., Eds.; Mac Millan and Co.: London, UK, 1882.
22. Reichl, L.E. *A Modern Course in Statistical Physics*; Wiley-VCH: Weinheim, Germany, 1998.
23. Sinai, Y.G. *Introduction to Ergodic Theory*; Princeton University Press: Princeton, NJ, USA, 1976.
24. Bunimovich, L.A.; Sinai, Y.G.; Chernov, N.I. Statistical Properties of Two-Dimensional HYPERBOLIC billiards. *Russ. Math. Surv.* **1991**, *46*, 47. [[CrossRef](#)]
25. Poincaré, H. *Science and Method*; Thomas Nelson and Sons: London, UK, 1914.
26. Laplace, P. *Essai Philosophique sur les Probabilités*, 5th ed.; Courcier: Paris, France, 1825; English translation in 1952 as *Philosophical Essay on Probabilities*; Dover Publications: New York, NY, USA, 1814.
27. Fermi, E. *Thermodynamics*; Dover Publications: New York, NY, USA, 1956.
28. Callen, H.B. *Thermodynamics and an Introduction to Thermostatistics*; Wiley: Hoboken, NJ, USA, 1985.
29. Khinchin, A.I. *Mathematical Foundations of Statistical Mechanics*; Dover Publications: New York, NY, USA, 2013.
30. Baldovin, M.; Vulpiani, A.; Gradenigo, G. Statistical Mechanics of an Integrable System. *J. Stat. Phys.* **2021**, *183*, 41. [[CrossRef](#)]
31. Cercignani, C.; Illner, R.; Pulvirenti, M. *The Mathematical Theory of Dilute Gases*; Springer: Berlin/Heidelberg, Germany, 1994.
32. Kreuzer, H.J. *Nonequilibrium Thermodynamics and Its Statistical Foundations*; Clarendon Press: Oxford, UK, 1981.
33. Falcioni, M.; Palatella, L.; Pigolotti, S.; Rondoni, L.; Vulpiani, A. Initial growth of Boltzmann entropy and chaos in a large assembly of weakly interacting systems. *Phys. A* **2007**, *385*, 170. [[CrossRef](#)]
34. Grad, H. On the kinetic theory of rarefied gases. *Commun. Pure Appl. Math.* **1949**, *2*, 331. [[CrossRef](#)]

35. Arkeryd, L.; Heintz, A. On the solvability and asymptotics of the Boltzmann equation in irregular domains. *Commun. Partial Differ. Equ.* **1997**, *22*, 2129. [[CrossRef](#)]
36. Ashcroft, N.W.; Mermin, N.D. *Solid State Physics*; Harcourt College Publishers: Orlando, FL, USA, 1976.
37. Nazarenko, S. *Wave Turbulence*; Springer: Berlin/Heidelberg, Germany, 2011.
38. Bonaldi, M.; Conti, L.; Gregorio, P.D.; Rondoni, L.; Vedovato, G.; Vinante, A.; Bignotto, M.; Cerdonio, M.; Falferi, P.; Liguori, N.; et al. Nonequilibrium Steady-State Fluctuations in Actively Cooled Resonators. *Phys. Rev. Lett.* **2009**, *103*, 010601. [[CrossRef](#)]
39. Conti, L.; Gregorio, P.D.; Karapetyan, G.; Lazzaro, C.; Pegoraro, M.; Bonaldi, M.; Rondoni, L. Effects of breaking vibrational energy equipartition on measurements of temperature in macroscopic oscillators subject to heat flux. *J. Stat. Mech.* **2013**, *2013*, P12003. [[CrossRef](#)]
40. Cohen, E.G.D.; Rondoni, L. Particles, Maps and Irreversible Thermodynamics I. *Phys. A* **2002**, *306*, 117. [[CrossRef](#)]
41. Rondoni, L. Introduction to Nonequilibrium Statistical Physics and Its Foundations. In *Frontiers and Progress of Current Soft Matter Research*; Liu, X.-Y., Ed.; Soft and Biological Matter Series; Springer: Singapore, 2021.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.