Compendium of the foundations of classical statistical physics

Jos Uffink

Institute for History and Foundations of Science, Universiteit Utrecht

PO Box 80 000, 3508 TA Utrecht, The Netherlands

February 2, 2006

ABSTRACT

Roughly speaking, classical statistical physics is the branch of theoretical physics that aims to account for the thermal behaviour of macroscopic bodies in terms of a classical mechanical model of their microscopic constituents, with the help of probabilistic assumptions. In the last century and a half, a fair number of approaches have been developed to meet this aim. This study of their foundations assesses their coherence and analyzes the motivations for their basic assumptions, and the interpretations of their central concepts. The most outstanding foundational problems are the explanation of time-asymmetry in thermal behaviour, the relative autonomy of thermal phenomena from their microscopic underpinning, and the meaning of probability.

A more or less historic survey is given of the work of Maxwell, Boltzmann and Gibbs in statistical physics, and the problems and objections to which their work gave rise. Next, we review some modern approaches to (i) equilibrium statistical mechanics, such as ergodic theory and the theory of the thermodynamic limit; and to (ii) non-equilibrium statistical mechanics as provided by Lanford's work on the Boltzmann equation, the so-called Bogolyubov-Born-Green-Kirkwood-Yvon approach, and stochastic approaches such as 'coarse-graining' and the 'open systems' approach. In all cases, we focus on the subtle interplay between probabilistic assumptions, dynamical assumptions, initial conditions and other ingredients used in these approaches.

Keywords: Statistical mechanics, kinetic gas theory, thermodynamics, entropy, probability, irreversibility

[0]email: uffink@phys.uu.nl

Contents

1 Introduction

It has been said that an advantage of having a mature, formalized version of a theory is that one may forget its preceding history. This saying is certainly true for the purpose of studying the conceptual structure of a physical theory. In a discussion of the foundations of classical mechanics, for example, one need not consider the work of the Parisian scholastics. In the foundations of quantum mechanics, one may start from the von Neumann axioms, and disregard the preceding "old" quantum theory. Statistical physics, however, has not yet developed a set of generally accepted formal axioms, and consequently we have no choice but to dwell on its history.

This is not because attempts to chart the foundations of statistical physics have been absent, or scarce (e.g. Ehrenfest & Ehrenfest-Afanassjewa 1912, ter Haar 1955, Penrose 1979, Sklar 1993, Emch & Liu 2001). Rather, the picture that emerges from such studies is that statistical physics has developed into a number of different schools, each with its own programme and technical apparatus. Unlike quantum theory or relativity, this field lacks a common set of assumptions that is accepted by most of the participants; although there is, of course, overlap. But one common denominator seems to be that nearly all schools claim the founding fathers, Maxwell, Boltzmann and Gibbs as their champions.

Broadly understood, statistical physics may be characterized as a branch of physics intended to describe the thermal behaviour and properties of matter in bulk, i.e. of macroscopic dimensions in relation to its microscopic corpuscular constituents and their dynamics.[1] In this review, we shall only deal with approaches that assume a finite number of microscopic constituents, governed by classical dynamics. (See Emch (2006) for a discussion of quantum statistical physics that also addresses infinite systems.)

The above description is deliberately vague; it does not yet specify what thermal behaviour is, and being a characterization in terms of intentions, leaves open by what methods the goals may be achieved. Let us expand a bit. There are two basic ingredients in statistical physics. The first is a mechanical model of a macroscopic material system. For example, a gas may be modeled as a system of point particles, or as hard spheres, or as composite objects, etc. Similarly, one may employ lattice models for solids, and so forth. In general, the particulars of the mechanical model, and its dynamics, will depend on the system of interest.

The second ingredient of the theory on which all approaches agree is the introduction of probability and statistical considerations. Sometimes, textbooks explain the need for this ingredient by pointing to the fact that an exact solution of the equations of motion for mechanical models with a large number of degrees of freedom is unfeasible. But this motivation from deficiency surely under-

[1] The terms "in bulk" and the distinction "micro/macroscopic" should be understood in a relative sense. Thus, statistical physics may apply to a galaxy or nebula, in which the constituent stars are considered as 'microscopic constituents'.

estimates the constructive and explanatory role that probability plays in statistical physics. A slightly better motivation, also found in many textbooks, is that even if the dynamical equations could be solved in detail, most of these details would turn out to be irrelevant for the purpose of characterizing the thermal behaviour. There is some truth in this observation, yet it can hardly be satisfactory as it stands. Certainly, not all details about the microdynamics are irrelevant, e.g. in phase transitions, and one naturally wishes for more concrete information about exactly which details are irrelevant and which are not.

One of the foremost foundational problems in statistical physics is thus to specify and to clarify the status of probabilistic assumptions in the theory. As we shall see, this task already leads to a rough distinction between approaches in which probability arises as a notion explicitly defined in mechanical terms (kinetic theory), and approaches in which it is a conceptually independent ingredient (statistical mechanics).

Next, there are ingredients on which much less consensus can be found. Here is a (partial) list:

- Assumptions about the overwhelmingly large number of microscopic constituents (typically of the order of 10^{23} or more).
- An assumption about the erratic nature of the dynamics (e.g. ergodicity).
- The choice of special initial conditions.
- The role of external influences on the system, i.e., assumptions about whether the system is open to the exchange of energy/momentum with its environment, in combination with an assumed sensitivity of the dynamics under such external disturbances.
- Symmetry of macroscopic quantities under permutation of the microscopic constituents.
- Limits in the resolution or experimental accuracy of macroscopic observers.
- Appeal to a time-asymmetric principle of causality.

The role of each of these ingredients in the recipe of statistical physics is controversial. What many "chefs" regard as absolutely essential and indispensable, is argued to be insufficient or superfluous by many others. A major goal in the foundations of statistical physics should therefore lie in an attempt to sort out which subset of the above ideas can be formulated in a precise and coherent manner to obtain a unified and sufficiently general framework for a theory of statistical physics.

Another issue in which the preceding discussion has been vague is what is meant by the thermal behaviour and properties of macroscopic matter. There are two sources on which one may draw in order to delineate this topic. The first is by comparison to other (older) traditions in theoretical physics that have the same goal as statistical physics but do not rely on the two main ingredients above viz. a mechanical model and probabilistic arguments. There are two main examples: thermodynamics

and hydrodynamics. The other source, of course, is observation. This provides a rich supply of phenomena, some of which have thus far withstood full theoretical explanation (e.g. turbulence).

Obviously, a measure of success for statistical physics can be found in the question to what extent this approach succeeds in reproducing the results of earlier, non-statistical theories, where they are empirically adequate, and in improving upon them where they are not. Thus, the foundations of statistical physics also provides a testing ground for philosophical ideas about inter-theory relations, like reduction (cf. Brush 1977, Sklar 1993, Batterman 2002). However I will not go into this issue. The remainder of this introduction will be devoted to a rough sketch of the four theories mentioned, i.e. thermodynamics, hydrodynamics, kinetic theory and statistical physics.

1.1 Thermodynamics.

Orthodox thermodynamics is an approach associated with the names of Clausius, Kelvin, and Planck. Here, one aims to describe the thermal properties of macroscopic bodies while deliberately avoiding commitment to any hypothesis about the microscopic entities that might constitute the bodies in question. Instead, the approach aims to derive certain general laws, valid for all such bodies, from a restricted set of empirical principles.

In this approach the macroscopic body (or thermodynamic system) is conceived of as a sort of black box, which may interact with its environment by means of work and heat exchange. The most basic empirical principle is that macroscopic bodies when left to themselves, i.e. when isolated from an environment, eventually settle down in an equilibrium state in which no further observable changes occur. Moreover, for simple, homogeneous bodies, this equilibrium state is fully characterized by the values of a small number of macroscopic variables.

Other empirical principles state which types of processes are regarded as impossible. By ingenious arguments one can then derive from these principles the existence of certain quantities (in particular: absolute temperature, energy and entropy) as 'state functions', i.e. functions defined on a space of thermodynamical equilibrium states for all such systems.

While the theory focuses on processes, the description it can afford of such processes is extremely limited. In general, a process will take a system through a sequence of non-equilibrium states, for which the thermodynamic state functions are not defined, and thus cannot be characterized in detail with the tools afforded by the theory. Therefore one limits oneself to the consideration of special types of processes, namely those that begin and end in an equilibrium state. Even more special are those processes that proceed so delicately and slowly that up to an arbitrarily small error one may assume that the system remains in equilibrium throughout the entire process. The latter processes are called *quasistatic*, or sometimes *reversible*.[2]

[2]The reader may be warned, however, that there are many different meanings to the term 'reversible' in thermodynamics.

Of course, since equilibrium states are by definition assumed to remain in equilibrium if unperturbed, all such processes are triggered by an external intervention such as pushing a piston or removing a partition. For the first type of process, orthodox thermodynamics can only relate the initial and final state. The second type of process can be (approximately) represented as a curve in the equilibrium state space.

The advantage of the approach is its generality. Though developed originally for the study of gases and liquids, by the late nineteenth century, it could be extended to the behaviour of magnets and other systems. Indeed, the independence of hypotheses about its micro-constituents means that the methods of orthodox thermodynamics can also be –and have been– applied to essentially quantum-mechanical systems (like photon gases) or to more exotic objects like black holes (see (Rovelli 2006).

With regard to the foundations of statistical physics, two aspects of thermodynamics are of outstanding importance. First, the challenge is to provide a counterpart for the very concept of equilibrium states and to provide a counterpart for the thermodynamic law that all isolated systems not in equilibrium evolve towards an equilibrium state. Secondly, statistical physics should give an account of the Second Law of thermodynamics, i.e. the statement that entropy cannot decrease in an adiabatically isolated system. Obviously, such counterparts will be statistical; i.e. they will hold on average or with high probability, but will not coincide with the unexceptionally general statements of thermodynamics.

1.2 Hydrodynamics

It would be a mistake to believe that the goals of statistical physics are exhausted by reproducing the laws of thermodynamics. There are many other traditions in theoretical physics that provide a much more detailed, yet less general, characterization of thermal behaviour. A concrete example is hydrodynamics or fluid dynamics. In contrast to thermodynamics, hydrodynamics does rely on an assumption about microscopic constitution. It models a fluid as a continuous medium or plenum. It is, in modern parlance, a field theory. Moreover it aims to describe the evolution of certain macroscopic quantities in the course of time, i.e. during non-equilibrium processes. As such it is an example of a theory which is much more informative and detailed than thermodynamics, at the price, of course, that its empirical scope is restricted to fluids.

Without going in detail (for a more comprehensive account, see e.g. Landau & Lifshitz 1987, de Groot & Mazur 1961), hydrodynamics assumes there are three fundamental fields: the mass density $\rho(\vec{x}, t)$, a velocity field $\vec{v}(\vec{x}, t)$ and a temperature field $T(\vec{x}, t)$. There are also three fundamental field equations, which express, in a differential form, the conservation of mass, momentum and energy.

See Uffink (2001) for a discussion.

Unfortunately, these equations introduce further quantities: the pressure $P(\vec{x}, t)$, the stress tensor $\pi(\vec{x}, t)$, the energy density $u(\vec{x}, t)$, the shear and bulk viscosities η and ζ and thermal conductivity κ, each of which has to be related to the fundamental fields by means of various constitutive relations and equations of state (dependent on the fluid concerned), in order to close the field equations, i.e. to make them susceptible to solution.

The resulting equations are explicitly asymmetric under time reversal. Yet another remarkable feature of hydrodynamics is the fact that the equations can be closed at all. That is: the specification of only a handful of macroscopic quantities is needed to predict the evolution of those quantities. Their behaviour is in other words *autonomous*. This same autonomy also holds for other theories or equations used to describe processes in systems out of equilibrium: for example the theories of diffusion, electrical conduction in metals, the Fourier heat equation etc. In spite of a huge number of microscopic degrees of freedom, the evolution of a few macroscopic quantities generally seems to depend only on the instantaneous values of these macroscopic quantities. Apart from accounting for the asymmetry under time reversal displayed by such theories, statistical physics should also ideally explain this remarkable autonomy of their evolution equations.

1.3 Kinetic theory

I turn to the second group of theories we need to consider: those that do rely on hypotheses or modeling assumptions about the internal microscopic constitution or dynamics of the systems considered. As mentioned, they can be divided into two rough subgroups: kinetic theory and statistical mechanics.

Kinetic theory, also called the kinetic theory of gases, the dynamical theory of gases, the molecular-kinetic theory of heat etc., takes as its main starting point the assumption that systems (gases in particular) consist of molecules. The thermal properties and behaviour are then related in particular to the motion of these molecules.

The earliest modern version of a kinetic theory is Daniel Bernoulli's (1731). Bernoulli's work was not followed by further developments along the same line for almost a century. But it regained new interest in the mid-nineteenth century. The theory developed into a more general and elaborate framework in the hands of Clausius, Maxwell and Boltzmann. Clausius extended Bernoulli's model by taking into account the collisions between the particles, in order to show that the formidable molecular speeds (in the order of 10^3 m/s) were compatible with relatively slow rates of diffusion. However, he did not develop a systematic treatment of collisions and their effects. It was Maxwell who was the first to realize that collisions would tend to produce particles moving at a variety of speeds, rather than a single common speed, and proceeded to ask how probable the various values of the velocity would be in a state of equilibrium. Maxwell thus introduced the concept of probability

and statistical considerations into kinetic theory.

From 1868 onwards, Boltzmann took Maxwell's investigations further. In his famous memoir of 1872 he obtained an equation for the evolution of the distribution function, the Boltzmann equation, and claimed that every non-stationary distribution function for an isolated gas would evolve towards the Maxwellian form, i.e. towards the equilibrium state. However, along the way, Boltzmann had made various assumptions and idealizations, e.g. neglecting the effect of multi-particle collisions, which restrict his derivations' validity to dilute gases, as well as the *Stoßzahlansatz*, developed by Maxwell in 1867, (or 'hypothesis of molecular disorder' as he later called it).

The Boltzmann equation, or variations of this equation, is the physicists' work-horse in gas theory. The hydrodynamical equations can be derived from it, as well as other transport equations. However, it is well known that it is only an approximation, and commonly regarded as a first step in a hierarchy of more detailed equations. But the foremost conceptual problem is its time-asymmetric nature, which highlights the fact that the Boltzmann equation itself could not be derived from mechanics alone. During Boltzmann's lifetime, this led to two famous objections, the reversibility objection (*Umkehreinwand*) by Loschmidt and the recurrence objection (*Wiederkehreinwand*) by Zermelo. A third important challenge, only put forward much more recently by Lanford (1975), concerns the consistency of the Boltzmann equation with the assumption that the gas system is a mechanical system governed by Hamiltonian dynamics.

1.4 Statistical mechanics

There is only a vague borderline between kinetic theory and statistical mechanics. The main distinctive criterion, as drawn by the Ehrenfests (1912) is this. Kinetic theory is what the Ehrenfests call "the older formulation of statistico-mechanical investigations" or "kineto-statistics of the molecule". Here, molecular states, in particular their velocities, are regarded as stochastic variables, and probabilities are attached to such molecular states of motion. These probabilities themselves are determined by the state of the total gas system. They are conceived of either as the relative *number* of molecules with a particular state, or the relative *time* during which a molecule has that state. (Maxwell employed the first option, Boltzmann wavered between the two.) It is important to stress that in both options the "probabilities" in question are determined by the *mechanical* properties of the gas. Hence there is really no clear separation between mechanical and statistical concepts in this approach.

Gradually, a transition was made to what the Ehrenfests called a "modern formulation of statistico-mechanical investigations" or "kineto-statistics of the gas model", or what is nowadays known as statistical mechanics. In this latter approach, probabilities are not attached to the state of a molecule but to the state of the entire gas system. Thus, the state of the gas, instead of determining the probability

distribution, now itself becomes a stochastic variable.

A merit of this latter approach is that interactions between molecules can be taken into account. Indeed, the approach is not necessarily restricted to gases, but might in principle also be applied to liquids or solids. (This is why the name 'gas theory' is abandoned.) The price to be paid however, is that the probabilities themselves become more abstract. Since probabilities are attributed to the mechanical states of the total system, they are no longer determined by such mechanical states. Instead, in statistical mechanics, the probabilities are usually conceived of as being determined by means of an 'ensemble', i.e. a fictitious collection of replicas of the system in question. But whatever role one may wish to assign to this construction, the main point is that probability is now an independent concept, no longer reducible to mechanical properties of the system.

It is not easy to pinpoint this transition in the course of the history, except to say that Maxwell's work in the 1860s definitely belong to the first category, and Gibbs' book of 1902 to the second. Boltzmann's own works fall somewhere in the middle ground. His earlier contributions clearly belong to the kinetic theory of gases (although his 1868 paper already applies probability to an entire gas system); while his work after 1877 is usually seen as elements in the theory of statistical mechanics. However, Boltzmann himself never indicated a clear distinction between these two different theories, and any attempt to draw a demarcation at an exact location in his work seems somewhat arbitrary.

From a conceptual point of view, the transition from kinetic gas theory to statistical mechanics poses two main foundational questions. First: on what grounds do we choose a particular ensemble, or the probability distribution characterizing the ensemble? Gibbs did not enter into a systematic discussion of this problem, but only discussed special cases of equilibrium ensembles (i.e. canonical, micro-canonical etc.) for which the probability distribution was stipulated by some special simple form. A second problem is to relate the ensemble-based probabilities to the probabilities obtained in the earlier kinetic approach for a single gas model.

The Ehrenfests (1912) paper was the first to recognize these questions, and to provide a partial answer. Namely: Assuming a certain hypothesis of Boltzmann's, which they dubbed the *ergodic hypothesis*, they pointed out that for an isolated system the micro-canonical distribution is the unique stationary probability distribution. Hence, if one demands that an ensemble of isolated systems describing thermal equilibrium must be represented by a stationary distribution, the only choice for this purpose is the micro-canonical one. Similarly, they pointed out that under the ergodic hypothesis, infinite time averages and ensemble averages were identical. This, then, would provide a desired link between the probabilities of the older kinetic gas theory and those of statistical mechanics, at least in equilibrium and in the infinite time limit. Yet the Ehrenfests simultaneously expressed strong doubts about the validity of the ergodic hypothesis. These doubts were soon substantiated when in 1913

Rosenthal and Plancherel proved that the hypothesis was untenable for realistic gas models.

The Ehrenfests' reconstruction of Boltzmann's work thus gave a prominent role to the ergodic hypothesis, suggesting that it played a fundamental and lasting role in his thinking. Although this view indeed produces a more coherent view of his multi-faceted work, it is certainly not historically correct. Boltzmann himself also had grave doubts about this hypothesis, and expressly avoided it whenever he could, in particular in his two great papers of 1872 and 1877b. Since the Ehrenfests, many authors have presented accounts of Boltzmann's work. Particularly important are Klein (1973) and Brush (1976).

Nevertheless, the analysis of the Ehrenfests did thus lead to a somewhat clearly delineated programme for or view about the foundations of statistical physics, in which ergodicity was a crucial feature. The demise of the original ergodic hypothesis did not halt the programme; the hypothesis was replaced by an alternative (weaker) hypothesis, i.e. that the system is '*metrically transitive*' (nowadays, the name 'ergodic' is often used as synonym). What is more, certain mathematical results of Birkhoff and von Neumann (the ergodic theorem) showed that for ergodic systems in this new sense, the desired results could indeed be proven, modulo a few mathematical provisos that at first did not attract much attention.

Thus there arose the ergodic or "standard" view on the foundations of statistical mechanics; (see, e.g. Khinchin 1949, p. 44). On that view, the formalism of statistical mechanics emerges as follows: A concrete system, say a container with gas, is represented as a mechanical system with a very large number of degrees of freedom. All physical quantities are functions of the dynamical variables of the system, or, what amounts to the same thing, are functions on its phase space. However, experiments or observation of such physical quantities do not record the instantaneous values of these physical quantities. Instead, every observation must last a duration which may be extremely short by human standards, but will be extremely long on the microscopic level, i.e. one in which the microstate has experienced many changes, e.g. because of the incessant molecular collisions. Hence, all we can register are *time averages* of the physical quantities over a very long periods of time. These averages are thus empirically meaningful. Unfortunately they are theoretically and analytically obstreperous. Time averages depend on the trajectory and can only be computed by integration of the equations of motion. The expectation value of the phase function over a given ensemble, the *phase average* has the opposite qualities, i.e. it is easy to compute, but not immediately empirically relevant. However, ergodicity ensures that the two averages are equal (almost everywhere). Thus, one can combine the best of both worlds, and identify the theoretically convenient with the empirically meaningful.

While statistical mechanics is clearly a more powerful theory than kinetic theory, it is, like thermodynamics, particularly successful in explaining and modeling gases and other systems in equilibrium. Non-equilibrium statistical mechanics remains a field where extra problems appear.

1.5 Prospectus

The structure of this chapter is as follows. In Section 2, I will provide a brief exposition of orthodox thermodynamics, and in subsection 2.2 an even briefer review of some less-than-orthodox approaches to thermodynamics. Section 3 looks at the kinetic theory of gases, focusing in particular on Maxwell's ground-breaking papers of 1860 and 1867, and investigates the meaning and status of Maxwell's probabilistic arguments.

Section 3.3 is devoted to (a selection of) Boltzmann's works, which, as mentioned above, may be characterized as in between kinetic theory and statistical mechanics. The focus will be on his 1868 paper and his most celebrated papers of 1872 and 1877. Also, the objections from Loschmidt (1877) and Zermelo (1897) are discussed, together with Boltzmann's responses. Our discussion emphasizes the variety of assumptions and methods used by Boltzmann over the years, and the open-endedness of his results: the ergodic hypothesis, the *Stoßzahlansatz*, the combinatorial argument of 1877, and a statistical reading of the H-theorem that he advocated in the 1890s.

Next, Section 5 presents an account of Gibbs' (1902) version of statistical mechanics and emphasizes the essential differences between his and Boltzmann's approach. Sections 6 and 7 give an overview of some more recent developments in statistical mechanics, In particular, we review some results in modern ergodic theory, as well as approaches that aim to develop a more systematic account of non-equilibrium theory, such as the BBGKY approach (named after Bogolyubov, Born, Green, Kirkwood and Yvon) and the approach of Lanford. Section 7 extends this discussion for a combination of approaches, here united under the name *stochastic dynamics* that includes those known as 'coarse-graining' and 'interventionism' or 'open systems'. In all cases we shall look at the question whether or how such approaches succeed in a satisfactory treatment of non-equilibrium.

As this prospectus makes clear, the choice of topics is highly selective. There are many important topics and developments in the foundations of statistical physics that I will not touch. I list the most conspicuous of those here together with some references for readers that wish to learn more about them.

- Maxwell's demon and Landauer's principle: (Klein 1970, Earman & Norton 1998, 1999, Leff& Rex 2003, Bennett 2003, Norton 1005, Maroney 2005, Ladyman et al. 2006).
- Boltzmann's work in the 1880s (e.g. on monocyclic systems) (Klein 1972, 1974, Bierhalter 1992, Gallavotti 1999, Uffink 2005).
- Gibbs' paradox (van Kampen 1984, Jaynes 1992, Huggett 1999, Saunders 2006).
- Branch systems (Schrödinger 1950, Reichenbach 1956, Kroes 1985, Winsberg 2004).
- Subjective interpretation of probability in statistical mechanics (Tolman 1938, Jaynes 1983, von Plato 1991, van Lith 2001a, Balian 2005).

- Prigogine and the Brussels-Austin school (Obcemea & Brändas 1983, Batterman 1991, Karakostas 1996, Edens 2001, Bishop 2004).

2 Orthodox thermodynamics

2.1 The Clausius-Kelvin-Planck approach

Thermodynamics is a theory that aims to characterize macroscopic physical bodies in terms of macroscopically observable quantities (typically: temperature, pressure, volume, etc.,) and to describe their changes under certain types of interactions (typically exchange of heat or work with an environment).

The classical version of the theory, which evolved around 1850, adopted as a methodological starting point that the fundamental laws of the theory should be independent of any particular hypothesis about the microscopic constitution of the bodies concerned. Rather, they should be based on empirical principles, i.e. boldly generalized statements of experimental facts, not on hypothetical and hence untestable assumptions such as the atomic hypothesis.

The reasons for this methodology were twofold. First, the dominant view on the goal of science was the positivist-empirical philosophy which greatly valued directly testable empirical statements above speculative hypotheses. But the sway of the positivist view was never so complete that physicists avoided speculation altogether. In fact many of the main founders of thermodynamics eagerly indulged in embracing particular hypotheses of their own about the microphysical constitution of matter.

The second reason is more pragmatic. The multitude of microphysical hypotheses and conjectures was already so great in the mid-nineteenth century, and the prospect of deciding between them so dim, that it was a clear advantage to obtain and present results that did not depend on such assumptions. Thus, when Clausius stated in 1857 that he firmly believed in the molecular-kinetic view on the nature of gases, he also mentioned that he had not previously revealed this opinion in order not to mix this conviction with his work on thermodynamics proper (Clausius 1857, p. 353).[3]

Proceeding somewhat ahistorically,[4] one might say that the first central concept in thermodynamics is that of *equilibrium*. It is taken as a fact of experience that macroscopic bodies in a finite volume, when left to themselves, i.e. isolated from an environment eventually settle down in a stationary state in which no further observable changes occur (the 'Minus First Law', cf. page 20). This stationary state is called a (thermal) *equilibrium state*. Moreover, for simple, homogeneous bodies, this state is

[3]The wisdom of this choice becomes clear if we compare his fame to that of Rankine. Rankine actually predated Clausius in finding the entropy function (which he called 'thermodynamic potential'). However, this result was largely ignored due to the fact that it was imbedded in Rankine's rather complicated theory of atomic vortices.

[4]I refer to Uffink (2001) for more details.

fully characterized by the values of a small number of macroscopic variables. In particular, for fluids (i.e. gases or liquids), two independent variables suffice to determine the equilibrium state.

For fluids, the three variables pressure p, temperature θ and volume V, are thus related by a so-called equation of state, where, following Euler, it has become customary to express pressure as a function of the two remaining variables:

$$p = p(\theta, V) \tag{1}$$

The form of this function differs for different fluids; for n moles of an ideal gas it is given by:

$$p(\theta, V) = nR\theta/V \tag{2}$$

where R is the gas constant and θ is measured on the gas thermometer scale.

The content of thermodynamics developed out of three ingredients. The first is the science of calorimetry, which was already developed to theoretical perfection in the eighteenth century, in particular by Joseph Black (Fox 1971, Truesdell 1980, Chang 2003,2004). It involved the study of the thermal changes in a body under the addition of or withdrawal of heat to the system. Of course, the (silent) presupposition here is that this process of heat exchange proceeds so delicately and slowly that the system may always be regarded as remaining in equilibrium. In modern terms, it proceeds 'quasi-statically'. Thus, the equation of state remains valid during the process.

The tools of calorimetry are those of differential calculus. For an infinitesimal increment $đQ$ of heat added to a fluid, one puts

$$đQ = c_V d\theta + \Lambda_\theta dV, \tag{3}$$

where c_V is called the heat capacity at constant volume and Λ_θ the latent heat at constant temperature. Both c_V and Λ_θ are assumed to be functions of θ and V. The notation $đ$ is used to indicate that the heat increment $đQ$ is not necessarily an exact differential, i.e. Q is not assumed to be a function of state.

The total heat Q added to a fluid during a process can thus be expressed as a line integral along a path $\mathcal{P}$ in the (θ, V) plane

$$Q(\mathcal{P}) = \int_{\mathcal{P}} đQ = \int_{\mathcal{P}} (c_V d\theta + \Lambda_\theta dV) \tag{4}$$

A treatment similar to the above can be given for the quasistatic heat exchange of more general thermal bodies than fluids. Indeed, calorimetry was sufficiently general to describe phase transitions, say from water to ice, by assuming a discontinuity in Λ_θ.

All this is independent of the question whether heat itself is a substance or not. Indeed, Black

himself wished to remain neutral on this issue. Even so, much of the terminology of calorimetry somehow invites the supposition that heat is a substance, usually called caloric, and many eighteenth and early nineteenth century authors adopted this view (Fox 1971). In such a view it makes sense to speak of the amount of heat contained in a body, and this would entail that $đQ$ must be an exact differential (or in other words: $Q(\mathcal{P})$ must be the same for all paths $\mathcal{P}$ with the same initial and final points). But this turned out to be empirically false, when the effects of the performance of work were taken into account.

Investigations in the 1840s (by Joule and Mayer among others) led to the conviction that heat and work are "equivalent"; or somewhat more precisely, that in every cyclic process $\mathcal{C}$, the amount of heat $Q(\mathcal{C})$ absorbed by the system is proportional to the amount of work performed by the system. Or, taking $W(\mathcal{C})$ as positive when performed *on* the system :

$$JQ(\mathcal{C}) + W(\mathcal{C}) = 0 \tag{5}$$

where $J \approx 4.2\mathrm{Nm/Cal}$ is Joule's constant, which modern convention takes equal to 1. This is the so-called *First Law of thermodynamics*.

For quasistatic processes this can again be expressed as a line integral in a state space Ω_{eq} of thermodynamic equilibrium states

$$\oint_{\mathcal{C}} (đQ + đW) = 0 \tag{6}$$

where

$$đW = -pdV. \tag{7}$$

Assuming the validity of (6) for all cyclic paths in the equilibrium state space implies the existence of a function U on Ω_{eq} such that

$$dU = đQ + đW. \tag{8}$$

The third ingredient of thermodynamics evolved from the study of the relations between heat and work, in particular the efficiency of heat engines. In 1824, Carnot obtained the following theorem.

> CARNOT'S THEOREM: Consider any system that performs a cyclic process $\mathcal{C}$ during which (a) an amount of heat $Q_+(\mathcal{C})$ is absorbed from a heat reservoir at temperature θ_+, (b) an amount of heat $Q_-(\mathcal{C})$ is given off to a reservoir at a temperature θ_-, with $\theta_- < \theta_+$, (c) there is no heat exchange at other stages of the cycles, and (d) some work $W(\mathcal{C})$ is done on a third body. Let $\eta(\mathcal{C}) := \frac{W(\mathcal{C})}{Q_+(\mathcal{C})}$ be the efficiency of the cycle. Then:
>
> (1) All quasistatic cycles have the same efficiency. This efficiency is a univer-

sal function of the two temperatures, i.e.,

$$\eta(\mathcal{C}) = \eta(\theta_+, \theta_-). \tag{9}$$

(2) All other cycles have a efficiency which is less or equal to that of the quasi-static cycle.

Carnot arrived at this result by assuming that heat was a conserved substance (and thus: $Q_+(\mathcal{C}) = Q_-(\mathcal{C})$ for all $\mathcal{C}$), as well as a principle that excluded the construction of a perpetuum mobile (of the first kind).

In actual fact, Carnot did not use the quasistatic/non-quasistatic dichotomy to characterize the two parts of his theorem. [5]

In fact, he used two different characterizations of the cycles that would produce maximum efficiency. (a): In his proof that Carnot cycles belong to class (1), the crucial assumption is that they "might have been performed in an inverse direction and order"(Carnot 1824, p. 11). But a little later (p. 13), he proposed a necessary and sufficient condition for a cycle to produce maximum efficiency, namely (b): In all stages which involve heat exchange, only bodies of equal temperature are put in thermal contact, or rather: their temperatures differ by a vanishingly small amount.

Carnot's theorem is remarkable since it did not need any assumption about the nature of the thermal system on which the cycle was carried out. Thus, when his work first became known to the physics community (Thomson, later known as Lord Kelvin, 1848) it was recognized as an important clue towards a general theory dealing with both heat and work exchange, for which Kelvin coined the name 'thermodynamics'. Indeed, Kelvin already showed in his first paper (1848) on the subject that Carnot's universal function η could be used to devise an absolute scale for temperature, i.e. one that did not depend on properties of a particular substance.

Unfortunately, around the very same period it became clear that Carnot's assumption of the conservation of heat violated the First Law. In a series of papers Clausius and Kelvin re-established Carnot's theorem on a different footing (i.e. on the first law (5) or, in this case $Q_+(\mathcal{C}) = Q_-(\mathcal{C}) + W(\mathcal{C})$, and a principle that excluded perpetual motion of the second kind) and transformed his results into general propositions that characterize general thermodynamical systems and their changes under the influence of heat and work. For the most part, these investigations were concerned with the first part of Carnot's theorem only. They led to what is nowadays called the first part of the Second Law; as follows.

First, Kelvin reformulated his 1848 absolute temperature scale into a new one, $T(\theta)$, in which

[5] Indeed, Truesdell (1980) argues that this characterization of his theorem is incorrect. See Uffink (2001) for further discussions.

the universal efficiency could be expressed explicitly as:

$$\eta(T_+, T_-) = 1 - \frac{T_-}{T_+}, \tag{10}$$

where $T_i = T(\theta_i)$. Since the efficiency η is also expressed by $W/Q_+ = 1 - (Q_-/Q_+)$, this is equivalent to

$$\frac{Q_-}{T_-} = \frac{Q_+}{T_+}. \tag{11}$$

Next, changing the sign convention to one in which Q is positive if absorbed and negative if given off by the system, and generalizing for cycles in which an arbitrary number of heat reservoirs are involved, one gets:

$$\sum_i \frac{Q_i}{T_i} = 0. \tag{12}$$

In the case where the system is taken through a quasistatic cycle in which the heat reservoirs have a continuously varying temperature during this cycle, this generalizes to

$$\oint_C \frac{đQ}{T} = 0. \tag{13}$$

Here, T still refers to the temperature of the heat reservoirs with which the system interacts, not to its own temperature. Yet Carnot's necessary and sufficient criterion of reversibility itself requires that during all stages of the process that involve heat exchange, the temperatures of the heat reservoir and system should be equal. Hence, in this case one may equate T with the temperature of the system itself.

The virtue of this result is that the integral (13) can now be entirely expressed in terms of quantities of the system. By a well-known theorem, applied by Clausius in 1865, it follows that there exists a function, called entropy S, defined on the equilibrium states of the system such that

$$S(s_1) - S(s_2) = \int_{s_1}^{s_2} \frac{đQ}{T} \tag{14}$$

or, as it more usually known:

$$\frac{đQ}{T} = dS. \tag{15}$$

This result is frequently expressed as follows: $đQ$ has an *integrating divisor* (namely T): division by T turns the inexact (incomplete, non-integrable) differential $đQ$ into an exact (complete, integrable) differential. For one mole of ideal gas (i.e. a fluid for which c_V is constant, Λ_θ vanishes and the ideal gas law (2) applies), one finds, for example:

$$S(T, V) = c_V \ln T + R \ln V + const. \tag{16}$$

The existence of this entropy function also allows for a convenient reformulation of the First Law for quasistatic processes (8) as

$$dU = TdS - pdV, \tag{17}$$

now too expressed in terms of properties of the system of interest.

However important this first part of the Second Law is by itself, it never led to much dispute or controversy. By contrast, the extension of the above results to cover the second part of Carnot's theorem gave rise to considerably more thought, and depends also intimately on what is understood by '(ir)reversible processes'.

The second part of Carnot's theorem was at first treated in a much more step-motherly fashion. Clausius' (1854) only devoted a single paragraph to it, obtaining the result that for "irreversible" cycles

$$\oint \frac{dQ}{T} \leq 0. \tag{18}$$

But this result is much less easy to apply, since the temperature T here refers to that of the heat reservoir with which the system is in contact, not (necessarily) that of the system itself.

Clausius put the irreversible processes in a more prominent role in his 1865 paper. If an irreversible cyclic process consists of a general, i.e. possibly non-quasistatic stage, from s_i to s_f, and a quasistatic stage, from s_f back to s_i, one may write (18) as

$$\int_{s_i \ \text{non-qs}}^{s_f} \frac{đQ}{T} + \int_{s_f \ \text{qs}}^{s_i} \frac{đQ}{T} \leq 0. \tag{19}$$

Applying (14) to the second term in the left hand side, one obtains

$$\int_{s_i \ \text{non-qs}}^{s_f} \frac{đQ}{T} \leq S(s_f) - S(s_i) \tag{20}$$

If we assume moreover that the generally non-quasistatic process is adiabatic, i.e. $đQ = 0$, the result is

$$S(s_i) \leq S(s_f). \tag{21}$$

In other words, in any adiabatic process the entropy of the final state cannot be less than that of the initial state.

Remarks: 1. The notation $\oint$ for cyclic integrals, and $đ$ for inexact differentials is modern. Clausius, and Boltzmann after him, would simply write $\int \frac{dQ}{T}$ for the left-hand side of (13) and (18).

2. An important point to note is that Clausius' formulation of the Second Law, strictly speaking, does not require a general monotonic increase of entropy for any adiabatically isolated system in the

course of time. Indeed, in orthodox thermodynamics, entropy is defined only for equilibrium states. Therefore it is meaningless within this theory to ask how the entropy of a system changes during a non-quasistatic process. All one can say in general is that when a system starts out in an equilibrium state, and ends, after an adiabatic process, again in an equilibrium state, the entropy of the latter state is not less than that of the former.

Still, the Second Law has often been understood as demanding continuous monotonic increase of entropy in the course of time, and often expressed, for adiabatically isolated systems, in a more stringent form

$$\frac{dS}{dt} \geq 0. \tag{22}$$

There is, however, no basis for this demand in orthodox thermodynamics.

3. Another common misunderstanding of the Second Law is that it would only require the non-decrease of entropy for processes in *isolated* systems. It should be noted that this is only part of the result Clausius derived: the Second Law holds more generally for *adiabatic* processes, i.e., processes during which the system remains adiabatically insulated. In other words, the system may be subject to arbitrary interactions with the environment, except those that involve heat exchange. (For example: stirring a liquid in a thermos flask, as in Joule's 'paddle wheel' experiment.)

4. Another point to be noted is that Clausius' result that the entropy in an adiabatically isolated system can never decrease is derived from the *assumption* that one can find a quasistatic process that connects the final to the initial state, in order to complete a cycle. Indeed, if such a process did not exist, the entropy difference of these two states would not be defined. The existence of such quasistatic processes is not problematic in many intended applications (e.g. if s_f and s_i are equilibrium states of a fluid); but it may be far from obvious in more general settings (for instance if one considers processes far from equilibrium in a complex system, such as a living cell). This warning that the increase of entropy is thus conditional on the existence of quasistatic transitions has been pointed out already by Kirchhoff (1894, p. 69).

5. Apart from the well-known First and Second Laws of thermodynamics, later authors have identified some more basic assumptions or empirical principles in the theory that are often assumed silently in traditional presentations—or sometimes explicitly but unnamed—which may claim a similar fundamental status.

The most familiar of these is the so-called *Zeroth Law*, a term coined by Fowler & Guggenheim (1939). To introduce this, consider the *relation of thermal equilibrium*. This is the relationship holding between the equilibrium states of two systems, whenever it is the case that the composite system, consisting of these two systems, would be found in an equilibrium state if the two systems are placed in direct thermal contact—i.e., an interaction by which they are only allowed to exchange heat. The zeroth law is now that the assumption that this is a *transitive* relationship, i.e. if it holds

for the states of two bodies A and B, and also for the states of bodies B and C, it likewise holds for bodies A and C.[6] The idea of elevating this to a fundamental 'Law', is that this assumption, which underlies the concept of temperature, can only be motivated on empirical grounds.

Another such assumption, again often stated but rarely named, is that any system contained in a finite volume, if left to itself, tends to evolve towards an equilibrium state. This has also sometimes been called a 'zeroth law' (cf. Uhlenbeck & Ford 1963, p.5; Lebowitz 1994, p. 135) in unfortunate competition with Fowler & Guggenheim's nomenclature. The name *Minus First Law* has therefore been proposed by Brown & Uffink (2001). Note that this assumption already introduces an explicitly time-asymmetric element, which is deeper than—and does not follow from—the Second Law. However, most nineteenth (and many twentieth) century authors did not appreciate this distinction, and as we shall see below, this Minus First Law is often subsumed under the Second Law.

2.2 Less orthodox versions of thermodynamics

Even within the framework of orthodox thermodynamics, there are approaches that differ from the Clausius-Kelvin-Planck approach. The foremost of those is undoubtedly the approach developed by Gibbs in 1873–1878 (Gibbs 1906). Gibbs' approach differs much in spirit from his European colleagues. No effort is devoted to relate the existence or uniqueness of the thermodynamic state variables U T or S to empirical principles. There existence is simply assumed. Also, Gibbs focused on the description of equilibrium states, rather than processes.

Previous authors usually regarded the choice of variables in order to represent a thermodynamic quantity as a matter of convention, like the choice of a coordinate system on the thermodynamic (equilibrium) state space. For a fluid, one could equally well choose the variables (p, V), (V, T), etc., as long as they are independent and characterize a unique thermodynamic equilibrium state.[7] Hence one could equally well express the quantities U, S, etc. in terms of any such set of variables. However, Gibbs had the deep insight that some choices are 'better' than others, in the sense that if, e.g., the entropy is presented as a function of energy and volume, $S(U, V)$, (or energy as a function of entropy and volume, $U(S, V)$) all other thermodynamic quantities could be determined from it, while this is generally not true for other choices. For example, if one knows only that for one mole of gas $S(T, V)$ is given by (2), one cannot deduce the equations of state $p = RT/V$ and $U = c_V T$. In contrast, if the function $S(U, V) = c_V \ln U + R \ln V +$ const.' is given, one obtains these equations from its partial derivatives: $\frac{p}{T} = (\frac{\partial S}{\partial V})_U$ and $\frac{1}{T} = (\frac{\partial S}{\partial U})_V$.

[6] Actually, transitivity alone is not enough. The assumption actually needed is that thermal equilibrium is an *equivalence* relation, i.e., it is transitive, reflective and symmetric (cf. Boyling 1972, p. 45)

[7] The latter condition may well fail: A fluid like water can exist at different equilibrium states with the same p, V, but different T (Thomsen& Hartka 1962)

For this reason, Gibbs called

$$dU = TdS - pdV \quad \text{or} \quad dS = \frac{1}{T}dU + \frac{p}{T}dV \tag{23}$$

the *fundamental equation*.[8] Of course this does not mean that other choices of variables are inferior. Instead, one can find equivalent fundamental equations for such pairs of variables too, in terms of the Legendre transforms of U. (Namely: the Helmholtz free energy $F = U - TS$ for the pair (T, V); the enthalpy $U + pV$ for (p, S), and the Gibbs free energy $U + pV - TS$ for (p, T).) Further, Gibbs extended these considerations from homogeneous fluids to heterogeneous bodies, consisting of several chemical components and physical phases.

Another major novelty is that Gibbs proposed a variational principle to distinguish stable from neutral and unstable equilibria. (Roughly, this principle entails that for stable equilibrium the function $S(U, V)$ should be concave.) This criterium serves to be of great value in characterizing phase transitions in thermodynamic systems, e.g. the Van der Waals gas (Maxwell used it to obtain his famous "Maxwell construction" or equal area rule (Klein 1978)). Gibbs work also proved important in the development of chemical thermodynamics, and physical chemistry.

Another group of approaches in orthodox thermodynamics is concerned particularly with creating a more rigorous formal framework for the theory. This is often called *axiomatic thermodynamics*. Of course, choosing to pursue a physical theory in an axiomatic framework does not by itself imply any preference for a choice in its physical assumptions or philosophical outlook. Yet the Clausius-Kelvin-Planck approach relies on empirical principles and intuitive concepts that may seem clear enough in their relation to experience—but are often surprisingly hard to define. Hence, axiomatic approaches tend to replace these empirical principles by statements that are conceptually more precise, but also more abstract, and thus arguably further removed from experience. The first example of this work is Carathéodory (1909). Later axiomatic approaches were pursued, among others, by Giles (1964), Boyling (1972), Jauch (1972, 1975), and by Lieb & Yngvason (1999). All these approaches differ in their choice of primitive concepts, in the formulation of their axioms, and hence also in the results obtained and goals achieved. However, in a rough sense, one might say they all focus particularly on demonstrating under what conditions one might guarantee the mathematical existence and uniqueness of entropy and other state functions within an appropriate structure.

Since the 1940s a great deal of work has been done on what is known as "*non-equilibrium thermodynamics*" or "thermodynamics of irreversible processes" (see e.g. de Groot 1951, Prigogine 1955, de

[8]Note how Gibbs' outlook differs here from the Clausius-Kelvin-Planck view: These authors would look upon (23) as a statement of the first law of thermodynamics, interpreting the differentials as infinitesimal increments during a quasistatic *process*, cf. (17). For Gibbs, on the other hand, (23) does not represent a process but a differential equation on the thermodynamic state space whose solution $U(S, V)$ or $S(V, U)$ contains *all* information about the equilibrium properties of the system, including the equations of state, the specific and latent heat, the compressibility, etc.— much more than just First Law.

Groot & Mazur 1961, Yourgrau et al. 1966, Truesdell 1969, Müller 2003). This type of work aims to extend orthodox thermodynamics into the direction of a description of systems in non-equilibrium states. Typically, one postulates that thermodynamic quantities are represented as continuously variable fields in space and time, with equilibrium conditions holding approximately within each infinitesimal region within the thermodynamic system. Again, it may be noted that workers in the field seem to be divided into different schools (using names such as "extended thermodynamics","generalized thermodynamics", "rational thermodynamics", etc.) that do not at all agree with each other (see Hutter & Wang 2003).

This type of work has produced many successful applications. But it seems fair to say that until now almost all attention has gone to towards practical application. For example, questions of the type that axiomatic thermodynamics attempts to answer, (e.g.: Under what conditions can we show the existence and uniqueness of the non-equilibrium quantities used in the formalism?) are largely unanswered, and indeed have given rise to some scepticism (cf. Meixner 1969, Meixner 1970). Another inherent restriction of this theory is that by relying on the assumption that non-equilibrium states can, at least in an infinitesimal local region, be well approximated by an equilibrium state, the approach is incapable of encompassing systems that are very far from equilibrium, such as in turbulence or living cells.)

The final type of approach that ought to be mentioned is that of *statistical thermodynamics*. The basic idea here is that while still refraining from introducing hypotheses about the microscopic constituents of thermodynamic systems, one rejects a key assumption of orthodox thermodynamics, namely, that a state of equilibrium is one in which all quantities attain constant values, in order to accommodate fluctuation phenomena such as Brownian motion or thermal noise. Thus the idea becomes to represent at least some of the thermodynamic quantities as random quantities, that in the course of time attain various values with various probabilities. Work in this direction has been done by Szilard (1925), Mandelbrot (1956, 1962, 1964), and Tisza & Quay (1963).

Of course the crucial question is then how to choose the appropriate probability distributions. One approach, elaborated in particular by Tisza (1966), taking its inspiration from Einstein (1910), relies on a inversion of Boltzmann's principle: whereas Boltzmann argued (within statistical mechanics) that the thermodynamic notion of entropy could be identified with the logarithm of a probability; Einstein argued that in thermodynamics, where the concept of entropy is already given, one may define the relative probability of two equilibrium states by the exponent of their entropy difference. Other approaches have borrowed more sophisticated results from mathematical statistics. For example, Mandelbrot used the Pitman-Koopman-Darmois theorem, which states that sufficient estimators exist only for the "exponential family" of probability distributions to derive the canonical probability distribution from the postulate that energy be a sufficient estimator of the system's temperature (see

also Uffink & van Lith 1999).

3 Kinetic theory from Bernoulli to Maxwell

3.1 Probability in the mid-nineteenth century

Probability theory has a history dating back at least two centuries before the appearance of statistical physics. Usually, one places the birth of this theory in the correspondence of Pascal and Fermat around 1650. It was refined into a mature mathematical discipline in the work of Jacob Bernoulli (1713), Abraham de Moivre (1738) and Pierre-Simon de Laplace (1813) (cf. Hacking 1975).

In this tradition, often called 'classical probability', the notion of probability is conceived of as a measure of the degree of certainty of our beliefs. Two points are important to note here. First, in this particular view, probability resides in the mind. There is nothing like uncertainty or chance in Nature. In fact, all authors in the classical tradition emphasize their adherence to strict determinism, either by appeal to divine omniscience (Bernoulli, de Moivre) or by appeal to the laws of mechanics and the initial conditions (Laplace). A probability hence represents a judgment about some state of affairs, and not an intrinsic property of this state of affairs. Hence, the classical authors never conceived that probability has any role to play in a description of nature or physical processes as such.[9] Secondly, although Bernoulli himself used the term "subjective" to emphasize the fact that probability refers to us, and the knowledge we possess, the classical interpretation does not go so far as modern adherents to a subjective interpretation of probability who conceive of probability as the degrees of belief of an arbitrary (although coherent) person, who may base his beliefs on personal whims, prejudice and private opinion.

This classical conception of probability would, of course, remain a view without any bite, if it were not accompanied by some rule for assigning values to probabilities in specific cases. The only such available rule is the so-called 'principle of insufficient reason': whenever we have no reason to believe that one case rather than another is realized, we should assign them equal probabilities (cf. Uffink 1995). A closely related version is the rule that two or more variables should be independent whenever we have no reason to believe that they influence each other.

While the classical view was the dominant, indeed the only existent, view on probability for the whole period from 1650 to 1813, it began to erode around 1830. There are several reasons for this, but perhaps the most important is, paradoxically, the huge success with which the theory was being applied to the most varied subjects. In the wake of Laplace's influential *Essai philosophique*

[9] Daniel Bernoulli might serve as an example. He was well acquainted with the work on probability of his uncle Jacob and, indeed, himself one of the foremost probabilists of the eighteenth century. Yet, in his work on kinetic gas theory (to be discussed in section 3.2), he did not find any occasion to draw a connection between these two fields of his own expertise.

sûr les Probabilités, scientists found applications of probability theory in jurisdiction, demography, social science, hereditary research, etc. In fact, one may say: almost everywhere except physics (cf. Hacking 1990). The striking regularity found in the frequencies of mass phenomena, and observations that (say) the number of raindrops per second on a tile follows the same pattern as the number of soldiers in the Prussian army killed each year by a kick from their horse, led to the alternative view that probability was not so much a representation of subjective (un)certainty, but rather the expression of a particular regularity in nature (Poisson, Quetelet). From these days onward we find mention of the idea of *laws of probability*, i.e. the idea that theorems of probability theory reflect lawlike behaviour to which Nature adheres. In this alternative, frequentist view of probability, there is no obvious place for the principle of insufficient reason. Instead, the obvious way to determine the values of probabilities is to collect empirical data on the frequencies on occurrences of events. However, a well-articulated alternative to the classical concept of probability did not emerge before the end of the century, and (arguably) not before 1919— and then within in a few years there were no less than three alternatives: a logical interpretation by Keynes, a frequentist interpretation by von Mises and a subjective interpretation by Ramsey and De Finetti. See Fine (1973), Galavotti (2004) or Emch (2005) for a more detailed exposition.

Summing up roughly, one may say that around 1850 the field of probability was in a state of flux and confusion. Two competing viewpoints, the classical and the frequency interpretation, were available, and often mixed together in a confusing hodgepodge. The result was well-characterized in a famous remark of Poincaré (1896) that all mathematicians seem to believe that the laws of probability refer to statements learned from experience, while all natural scientists seem to think they are theorems derived by pure mathematics.

The work of Maxwell and Boltzmann in the 1860s emerged just in the middle of this confusing era. It is only natural that their work should reflect the ambiguity that the probability concept had acquired in the first half of the nineteenth century. Nevertheless, it seems that they mainly thought of probability in terms of frequencies, as an objective quantity, which characterizes a many-particle system, and that could be explicitly defined in terms of its mechanical state. This, however, is less clear for Maxwell than for Boltzmann.

Gradually, probability was emancipated from this mechanical background. Some isolated papers of Boltzmann (1871b) and Maxwell (1879) already pursued the idea that probabilities characterize an *ensemble* of many many-particle systems rather than a single system. Gibbs's 1902 book adopted this as a uniform coherent viewpoint. However, this ensemble interpretation is still sufficiently vague to be susceptible to different readings. A subjective view of ensembles, closely related to the classical interpretation of Bernoulli and Laplace, has emerged in the 1950s in the work of Jaynes. This paper, will omit a further discussion of this approach. I refer to (Jaynes 1983, Uffink 1995,1996, Balian

2005) for more details.

3.2 From Bernoulli to Maxwell (1860)

The kinetic theory of gases (sometimes called: dynamical theory of gases) is commonly traced back to a passage in Daniel Bernoulli's *Hydrodynamica* of 1738. Previous authors were, of course, quite familiar with the view that gases are composed of a large but finite number of microscopic particles. Yet they usually explained the phenomenon of gas pressure by a static model, assuming repulsive forces between these particles.

Bernoulli's discussion is the first to explain pressure as being due to their motion. He considered a gas as consisting of a great number of particles, moving hither and thither through empty space, and exerting pressure by their incessant collisions on the walls. With this model, Bernoulli was able to obtain the ideal gas law $pV = const.$ at constant temperature, predicted corrections to this law at high densities, and argued that the temperature could be taken as proportional to the square of the velocity of the particles. Despite this initial success, no further results were obtained in kinetic gas theory during the next century. By contrast, the view that modeled a gas as a continuum proved much more fertile, since it allowed the use of powerful tools of calculus. Indeed, the few works in the kinetic theory in the early nineteenth century e.g. by Waterston and Herapath were almost entirely ignored by their contemporaries (cf. Brush 1976).

Nevertheless, the kinetic view was revived in the 1850s, in works by Kronig and Clausius. The main stimulus for this revival was the Joule-Mayer principle of the equivalence of heat and work, which led to the First Law of thermodynamics, and made it seem more plausible that heat itself was just a form of motion of gas particles. (A point well-captured in the title of Clausius' 1857 paper: "The kind of motion we call heat", subsequently adopted by Stephen Brush (1976) for his work on the history of this period.)

Clausius also recognized the importance of mutual collisions between the particles of the gas, in order to explain the relative slowness of diffusion when compared with the enormous speed of the particles (estimated at values of 400 m/s or more at ordinary room temperature). Indeed, he argued that in spite of their great speed, the mean free path, i.e. the distance a particle typically travels between two collision, could be quite small (of the order of micrometers) so that the mean displacement per second of particles is accordingly much smaller.

However, Clausius did not pay much attention to the consideration that such collisions would also change the magnitude of the velocities. Indeed, although his work sometimes mentions phrases like "mean speed" or "laws of probability" he does not specify a precise averaging procedure or probability assumption, and his calculations often proceed by crude simplifications (e.g. assuming that all but one of the particles are at rest).

3.2.1 Maxwell (1860)

It was Maxwell's paper of 1860 that really marks the re-birth of kinetic theory. Maxwell realized that if a gas consists of a great number N of moving particles, their velocities will suffer incessant change due to mutual collisions, and that a gas in a stationary state should therefore consist of a mixture of slower and faster particles. More importantly, for Maxwell this was not just an annoying complication to be replaced by simplifying assumptions, but the very feature that deserved further study.

He thus posed the question

> Prop. IV. To find the average number of particles whose velocities lie between given limits, after a great number of collisions among a great number of equal particles. (Maxwell 1860, p. 380).

Denoting this desired average number as $Nf(\vec{v})d^3\vec{v}$, he found a solution to this problem by imposing two assumptions: the distribution function $f(\vec{v})$ should (i) factorize into functions of the orthogonal components of velocity, i.e. there exists some function g such that:

$$f(\vec{v}) = g(v_x)g(v_y)g(v_z), \tag{24}$$

and (ii) be spherically symmetric, i.e.,

$$f(\vec{v}) \text{ depends only on } v = \|\vec{v}\|. \tag{25}$$

He observed that these functional equations can only be satisfied if

$$f(\vec{v}) = Ae^{-v^2/B}, \tag{26}$$

where the constant A is determined by normalization: $A = (B\pi)^{-3/2}$; and constant B is determined by relating the mean squared velocity to the absolute temperature—i.e., adopting modern notation: $\frac{3}{2}kT = \frac{m}{2}\langle v^2\rangle$—to obtain:

$$f(\vec{v}) = \left(\frac{m}{2\pi kT}\right)^{3/2} e^{-mv^2/2kT}. \tag{27}$$

Maxwell's result led to some novel and unexpected predictions, the most striking being that the viscosity of a gas should be independent of its density, which was, nevertheless, subsequently experimentally verified. Another famous prediction of Maxwell was that in this model the ratio of the specific heats $\gamma = \frac{c_V}{c_p}$ must take the value of $\frac{4}{3}$. This did not agree with the experimentally obtained value of $\gamma = 1.408$.[10]

[10]More generally, $c_V/c_p = (f+2)/f$ where f is the number of degrees of freedom of a molecule. This so-called c_V/c_p

Maxwell's paper is the first to characterize the state of a gas by a distribution function f. It is also the first to call $f(\vec{v})d^3\vec{v}$ a *probability*. Clearly, Maxwell adopted a frequency interpretation of probability. The probability for the velocity to lie within a certain range $d^3\vec{v}$ is nothing but the relative number of particles in the gas with a velocity in that range. It refers to an objective, mechanical property of the gas system, and not to our opinions.

Now an obvious problem with this view is that if the gas contains a finite number of particles, the distribution of velocities must necessarily be discrete, i.e., in Dirac delta notation:

$$f(\vec{v}) = \frac{1}{N}\sum_{i=1}^{N}\delta(\vec{v} - \vec{v}_i), \tag{28}$$

and if the energy of the gas is finite and fixed, the distribution should have a bounded support. The function (26) has neither of these properties.

It is not clear how Maxwell would have responded to such problems. It seems plausible that he would have seen the function (26) as representing only a good enough approximation,[11] in some sense, to the actual state of the gas but not to be taken too literally, just like actual frequencies in a chance experiment never match completely with their expected values. This is suggested by Maxwell's own illustration of the continuous distribution function as a discrete cloud of points, each of which representing the endpoint of a velocity vector (cf. Fig. 1 from (Maxwell 1875)). This suggests he thought of an actual distribution more along the lines of (28) than (26). But this leaves the question open in what sense the Maxwell distribution approximates the actual distribution of velocities.

One option here would be to put more emphasis on the phrase "average" in the above quote from Maxwell. That is, maybe f is not intended to represent an actual distribution of velocities but an averaged one. But then, what kind of average? Since an average over the particles has already been performed, the only reasonable options could be an average over time or averaging over an ensemble of similar gas systems. But I can find no evidence that Maxwell conceived of such procedures in this paper. Perhaps one might argue that the distribution (26) is intended as an expectation, i.e. that it represents a reasonable mind's guess about the number of particles with a certain velocity. But in that case, Maxwell's interpretation of probability ultimately becomes classical.

However this may be, it is remarkable that the kinetic theory was thus able to make progress beyond Bernoulli's work by importing mathematical methods (functional equations) involving the

anomaly haunted gas theory for another half century. The experimental value around 1.4 is partly due to the circumstance that most ordinary gases have diatomic molecules for which, classically, $f = 6$. Quantum theory is needed to explain that one of these degrees is "frozen" at room temperature. Experimental agreement with Maxwell's prediction was first obtained by Kundt and Warburg in 1875 for mercury vapour (For more details, see Brush 1976, p. 353–356).

[11] This view was also expressed by Boltzmann (1896b). He wrote, for example: "For a finite number of molecules the Maxwell distribution can never be realized exactly, but only as a good approximation" (Abh., III, p. 569).

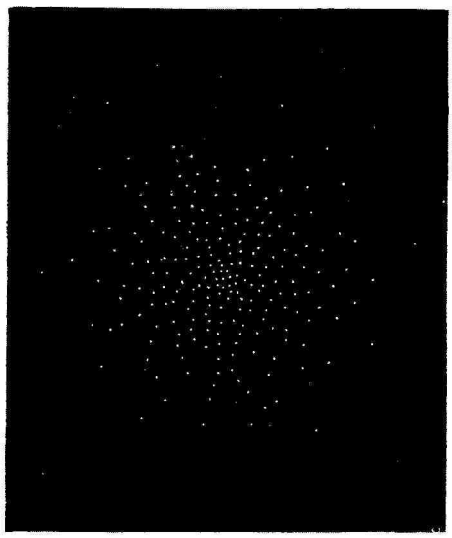

Diagram of Velocities.

Figure 1: *An illustration of the Maxwell distribution from (Maxwell 1875). Every dot represents the end-point of a velocity vector.*

representation of a state by continuous functions; though at the price of making this state concept more abstractly connected to physical reality.

A more pressing problem is that the assumptions (24, 25) Maxwell used to derive the form of his distribution do not sit well with its intended frequency interpretation. They seem to reflect a priori desiderata of symmetry, and are perhaps motivated by an appeal to some form of the principle of insufficient reason, in the sense that if there is, in our knowledge, no reason to expect a dependence between the various orthogonal components of velocity, we are entitled to assume they are independent.

This reading of Maxwell's motivations is suggested by the fact that in 1867 he described his 1860 assumption (24) as "the assumption that the probability of a molecule having a velocity resolved parallel to x lying between given limits is not in any way affected by the *knowledge* that the molecule has a given velocity resolved parallel to y" (Maxwell 1867, emphasis added).

It has been pointed out (see e.g. Brush 1976, Vol. II, pp. 183–188) that Maxwell's 1860 argument seems to have been heavily inspired by Herschel's (1850) review of Quetelet's work on probability. This review essay contained a strikingly similar argument, applied to a marksman shooting at a target, in order to determine the probability that a bullet will land at some distance from the target. What is more, Herschel's essay is firmly committed to the classical interpretation of probability and gives the principle of insufficient reason a central role. Indeed, he explains the (analogue of) condition (25)

as "nothing more than the expression of our state of *complete* ignorance of the causes of the errors [i.e. the deviation from the target] and their mode of action" (Herschel 1850, p. 398, emphasis in the original). If Maxwell indeed borrowed so much from Herschel, it seems likely that he would also have approved of, or at least be inspired by, this motivation of condition (25).[12]

Whatever may have been Maxwell's original motivation for these assumptions, their dubious nature is also clear from the fact that, in spite of his formulation of the problem (i.e. to determine the form of the function f "after a great number of collisions"), they do not refer to collisions at all. Indeed, it would seem that any motivation for their validity would just as well apply to a gas model consisting of non-colliding (e.g. perfectly transparent) particles as well. As Maxwell himself later remarked about certain derivations in the works of others, one might say that the condition "after a great number of collisions" is intended "rather for the sake of enabling the reader to form a mental image of the material system than as a condition for the demonstration" (Maxwell (1879) Garber, Brush & Everitt 1995, p. 359).

3.3 Maxwell (1867)

Whatever the merits and problems of his first paper, Maxwell's next paper on gas theory of 1867 rejected his previous attempt to derive the distribution function from the assumptions (24, 25) as "precarious" and proposed a completely different argument. This time, he considered a model of point particles with equal masses interacting by means of a repulsive central force, proportional to the fifth power of their mutual distance. What is more important, this time the collisions are used in the argument.

Maxwell considers an elastic collision between a pair of particles such that the initial velocities are $\vec{v}_1, \vec{v}_2$ and final velocities $\vec{v}_1{}', \vec{v}_2{}'$).[13] These quantities are related by the conservation laws of momentum and energy, yielding four equations, and two parameters depending on the geometrical factors of the collision process.

It is convenient to consider a coordinate frame such that particle 1 is at rest in the origin, and the relative velocity $\vec{v}_2 - \vec{v}_1$ is directed along the negative z axis, and to use cylindrical coordinates. If (b, ϕ, z) denote the coordinates of the trajectory of the centre of particle 2, we then have $b = const.$, $\phi = const$, $z(t) = z_0 - \|\vec{v}_2 - \vec{v}_1\|t$ before the collision. In the case where the particles are elastic hard spheres, a collision will take place only if the impact parameter b is less than the diameter d of

[12] It is interesting to note that Herschel's review prompted an early and biting criticism of the principle of insufficient reason as applied to frequencies of events by Leslie Ellis, containing the famous observation: "Mere ignorance is no ground for any inference whatsoever. *Ex nihilo nihil*. It cannot be that because we are ignorant of the matter, we know something about it" (Ellis 1850). It is not certain, however, whether Maxwell knew of this critique.

[13] In view of the infinite range of the interaction, 'initial' and 'final' are to be understood in an asymptotic sense, i.e. in the limits $t \longrightarrow \pm\infty$. An alternative followed in the text is to replace Maxwell's (1867) model with the hard spheres he had considered in 1860.

the spheres. The velocities after the collision are then determined by $\|\vec{v}_1 - \vec{v}_2\|, b$ and ϕ. Transformed back to the laboratory frame, the final velocities $\vec{v}_1{}', \vec{v}_2{}'$ can then be expressed as functions of $\vec{v}_1$, $\vec{v}_2$, b and ϕ.

Maxwell now assumes what the Ehrenfests later called the *Stoßzahlansatz*: the number of collisions during a time dt, say $N(\vec{v}_1, \vec{v}_2)$, in which the initial velocities $\vec{v}_1, \vec{v}_2$ within an element $d^3\vec{v}_1 d^3\vec{v}_2$ are changed into final velocities $\vec{v}_1{}', \vec{v}_2{}'$ in an element $d^3\vec{v}_1{}' d^3\vec{v}_2{}'$ within a spatial volume element $dV = bdbd\phi dz = \|\vec{v}_1 - \vec{v}_2\| \, bdbd\phi dt$ is proportional to the product of the number of particles with velocity $\vec{v}_1$ within $d^3\vec{v}_1$ (i.e. $Nf(\vec{v}_1)d\vec{v}_1$), and those with velocity $\vec{v}_2$ within $d^3\vec{v}_2$ (i.e. $Nf(\vec{v}_2)d^3\vec{v}_2$), and that spatial volume element. Thus:

$$N(\vec{v}_1, \vec{v}_2) = N^2 f(\vec{v}_1) f(\vec{v}_2) \|\vec{v}_2 - \vec{v}_1\| d^3\vec{v}_1 d^3\vec{v}_2 bdbd\phi dt. \tag{29}$$

Due to the time reversal invariance of the collision laws, a similar consideration applies to the so-called inverse collisions, in which initial velocities $\vec{v}_1{}', \vec{v}_2{}'$ and final velocities $\vec{v}_1$ and $\vec{v}_2$ are interchanged. Their number is proportional to

$$N(\vec{v}_1{}', \vec{v}_2{}') = N^2 f(\vec{v}_1{}') f(\vec{v}_1{}') \|\vec{v}_2{}' - \vec{v}_1{}'\| d^3\vec{v}_1{}' d\vec{v}_2{}' bdbd\phi dt \tag{30}$$

Maxwell argues that the distribution of velocities will remain stationary, i.e. unaltered in the course of time, if the number of collisions of these two kinds are equal, i.e. if

$$N(\vec{v}_1{}', \vec{v}_2{}') = N(\vec{v}_1, \vec{v}_2). \tag{31}$$

Moreover, the collision laws entail that $\|\vec{v}_2{}' - \vec{v}_1{}'\| = \|\vec{v}_2 - \vec{v}_1\|$ and $d^3\vec{v}_1{}' d^3\vec{v}_2{}' = d^3\vec{v}_1 d^3\vec{v}_2$. Hence, the condition (31) may be simplified to

$$f(\vec{v}_1) f(\vec{v}_2) = f(\vec{v}_1{}') f(\vec{v}_2{}'), \quad \text{for all } \vec{v}_1, \vec{v}_2. \tag{32}$$

This is the case for the Maxwellian distribution (26). Therefore, Maxwell says, the distribution (26) is a "possible" form.

He goes on to claim that it is also the *only* possible form for a stationary distribution.This claim, i.e. that stationarity of the distribution f can only arise under (32) is nowadays also called the principle of *detailed balancing* (cf. Tolman 1938, p. 165).[14] Although his argument is rather brief, the idea seems to be that for a distribution violating (32), there must (because of the *Stoßzahlansatz*) be two

[14]The reader might be warned, however, that the name 'detailed balancing' is also used to cover somewhat different ideas than expressed here (Tolman 1938, p. 521).

velocity pairs[15] $\vec{v}_1, \vec{v}_2$ and $\vec{u}_1, \vec{u}_2$, satisfying $\vec{v}_1 + \vec{v}_2 = \vec{u}_1 + \vec{u}_2$ and $v_1^2 + v_2^2 = u_1^2 + u_2^2$, such that the collisions would predominantly transform $(\vec{v}_1, \vec{v}_2) \longrightarrow (\vec{u}_1, \vec{u}_2)$ rather than $(\vec{u}_1, \vec{u_2}) \longrightarrow (\vec{v}_1, \vec{v}_2)$. But then, since the distribution is stationary, there must be a third pair of velocities, $(\vec{w}_1, \vec{w}_2)$, satisfying similar relations, for which the collisions predominantly produce transitions $(\vec{u}_1, \vec{u_2}) \longrightarrow (\vec{w}_1, \vec{w}_2)$, etc. Now, the distribution can only remain stationary if any such sequence closes into a cycle. Hence there would be cycles of velocity pairs $(\vec{v}_1, \vec{v}_2) \longrightarrow (\vec{u}_1, \vec{u}_2) \longrightarrow \ldots \longrightarrow (\vec{v}_1, \vec{v}_2)$ which the colliding particles go through, eventually returning to their original velocities.

Maxwell then argues: "Now it is impossible to assign a reason why the successive velocities of a molecule should be arranged in this cycle rather than in the reverse order" (Maxwell 1867, p.45). Therefore, he argues, these two cycles should be equally probable, and, hence, a collision cycle of the type $(\vec{v}_1, \vec{v}_2) \longrightarrow (\vec{v}_1{}', \vec{v}_2{}')$ is already equally probable as a collision cycle of the type $(\vec{v}_1{}', \vec{v}_2{}') \longrightarrow (\vec{v}_1, \vec{v}_2)$, i.e. condition (32) holds.

Comments. First, a clear advantage of Maxwell's 1867 derivation of the distribution function (26) is that the collisions play a crucial role. The argument would not apply if there were no collisions between molecules. A second point to note is that the distribution (26) is singled out because of its *stationarity*, instead of its spherical symmetry and factorization properties. This is also a major improvement upon his previous paper, since stationarity is essential to thermal equilibrium.

A crucial element in the argument is still an assumption about independence. But now, in the *Stoßzahlansatz*, the initial velocities of colliding particles are assumed independent, instead of the orthogonal velocity components of a single particle. Maxwell does not expand on why we should assume this *ansatz*; he clearly regarded it as obvious. Yet it seems plausible to argue that he must have had in the back of his mind some version of the principle of insufficient reason, i.e., that we are entitled to treat the initial velocities of two colliding particles as independent because we have no reason to assume otherwise. Although still an argument from insufficient reason, this is at least a much more plausible application than in the 1860 paper.

A main defect of the paper is his sketchy claim that the Maxwell distribution (26) would be the unique stationary distribution. This claim may be broken in two parts: (a) the cycle argument just discussed, leading Maxwell to argue for detailed balancing; and (b) the claim that the Maxwell distribution is uniquely compatible with this condition.

A demonstration for part (b) was not provided by Maxwell at all; but this gap was soon bridged by Boltzmann (1868)—and Maxwell gave Boltzmann due credit for this proof. But part (a) is more interesting. We have seen that Maxwell here explicitly relied on reasoning from insufficient reason.

[15] Actually, Maxwell, discusses only velocities of a single molecule. For clarity, I have transposed his argument to a discussion of pairs.

He was criticized on this point by Boltzmann (1872) and also by Guthrie (1874).

Boltzmann argued that Maxwell was guilty of begging the question. If we suppose that the two cycles did not occur equally often, then this supposition by itself would provide a reason for assigning unequal probabilities to the two types of collisions.[16] This argument by Boltzmann indicates, at least in my opinion that he was much less prepared than Maxwell to argue in terms of insufficient reason. Indeed, as we shall see in Section 3.3, his view on probability seems much more thoroughly frequentist than Maxwell.

In fact Boltzmann later repeatedly mentioned the counterexample of a gas in which all particles are lined up so that they only collide centrally, and move perpendicularly between parallel walls (Boltzmann 1872, Abh. I p. 358, Boltzmann 1878 Abh. II p. 285). In this case, the velocity distribution

$$\frac{1}{2}\left(\delta(v - v_0) + \delta(v + v_0)\right) \tag{33}$$

is stationary too.

Some final remarks on Maxwell's work: As we have seen, it is not easy to pinpoint Maxwell's interpretation of probability. In his (1860), he identifies the probability of a particular molecular state with the relative number of particles that possess this state.[17] Yet, we have also seen that he relates probability to a state of knowledge. Thus, his position may be characterized as somewhere between the classical and the frequentist view.

Note that Maxwell never made any attempt to reproduce the second law. Rather he seems to have been content with the statistical description of thermal equilibrium in gases.[18] All his writings after 1867 indicate that he was convinced that a derivation of the Second Law from mechanical principles was impossible. Indeed, his remarks on the Second Law generally point to the view that the Second Law "has only statistical certainty" (letter to Tait, undated; Garber, Brush & Everitt 1995, p. 180), and that statistical considerations were foreign to the principles of mechanics. Indeed, Maxwell was quite amused to see Boltzmann and Clausius engage in a dispute about who had been the first to reduce the Second Law of thermodynamics to mechanics:

[16] More precisely, Boltzmann argued as follows: "In order to prove the impossibility [of the hypothesis] that the velocity of [a pair of] molecule[s] changes more often from $[(\vec{v}_1, \vec{v}_2)$ to $(\vec{v}_1{}', \vec{v}_2{}')]$ than the converse, Maxwell says that there should then exist a closed series of velocities that would be traversed rather in one order than the other. This, however, could not be, he claims, because one could not indicate a reason, why molecules would rather traverse the cycle in one order than the other. But it appears to me that this last claim already presupposes as proven what is to be proved. Indeed, if we assume as already proven that the velocities change as often from $(\vec{v}_1, \vec{v}_2)$ to $(\vec{v}_1{}', \vec{v}_2{}')$ as conversely, then of course there is no reason why the cycle should rather be run through in one order than the other. But if we assume that the statement to be proven is not yet proved, then the very fact that the velocities of the molecules prefer to change rather from $(\vec{v}_1, \vec{v}_2)$ to $(\vec{u}_1, \vec{u}_2)$ than conversely, rather from $(\vec{u}_1, \vec{u}_2)$ to $(\vec{w}_1, \vec{w}_2)$ than conversely, etc. would provide the reason why the cycle is traversed rather one way than the other" (Abh. I, p. 319).

[17] Curiously, this terminology is completely absent in his 1867 paper.

[18] Apart from a rather lame argument in (Maxwell 1860) analyzed by (Brush 1976, p.344).

> It is rare sport to see those learned Germans contending the priority of the discovery that the 2nd law of $\theta\Delta$cs is the 'Hamiltonsche Prinzip', [...] The Hamiltonsche Prinzip, the while, soars along in a region unvexed by statistical considerations, while the German Icari flap their waxen wings in nephelococcygia[19] amid those cloudy forms which the ignorance and finitude of human science have invested with the incommunable attributes of the invisible Queen of Heaven (letter to Tait, 1873; Garber, Brush & Everitt 1995, p. 225)

Clearly, Maxwell saw a derivation of the Second Law from pure mechanics, "unvexed by statistical considerations", as an illusion. This point appears even more vividly in his thought experiment of the "Maxwell demon", by which he showed how the laws of mechanics could be exploited to produce a violation of the Second Law. For an entry in the extensive literature on Maxwell's demon, I refer to (Earman & Norton 1998,1999, Leff & Rex 2003, Bennett 2003, Norton 2005).

But neither did Maxwell make any effort to reproduce the Second Law on a *unified* statistical/mechanical basis. Indeed, the scanty comments he made on the topic (e.g. in Maxwell 1873, Maxwell 1878b) rather seem to point in another direction. He distinguishes between what he calls the 'statistical method' and the 'historical' or 'dynamical' (or sometimes 'kinetic') method. These are two modes of description for the same system. But rather than unifying them, Maxwell suggests they are competing, or even incompatible—one is tempted to say "complementary"– methods, and that it depends on our own knowledge, abilities, and interests which of the two is appropriate. For example:

> In dealing with masses of matter, while we do not perceive the individual molecules, we are *compelled* to adopt what I have described as the statistical method, and to abandon the strict dynamical method, in which we follow every motion by the calculus (Maxwell 1872, p. 309, emphasis added).

In this respect, his position stands in sharp contrast to that of Boltzmann, who made the project of finding this unified basis his lifework.

4 Boltzmann[20]

4.1 Early work: Stoßzahlansatz *and ergodic hypothesis*

Boltzmann had already been considering the problem of finding a mechanical derivation of the Second Law in a paper of 1866. At that time, he did not know of Maxwell's work. But in 1868, he had

[19] 'Cloudcuckooland", an illusionary place in Aristophanes' *The Birds*.
[20] Parts of this section were also published in (Uffink 2004).

read both Maxwell's papers of 1860 and 1867. Like Maxwell, he focuses on the study of gases in thermal equilibrium, instead of the Second Law. He also adopts Maxwell's idea of characterizing thermal equilibrium by a probability distribution, and the *Stoßzahlansatz* as the central dynamical assumption. But along the way in this extensive paper, Boltzmann comes to introduce an entirely different alternative approach, relying on what we now call the ergodic hypothesis.

As we saw in section 3.3, Maxwell had derived his equilibrium distribution for two special gas models (i.e. a hard sphere gas in 1860 and a model of point particles with a central r^5 repulsive force acting between them in 1867). He had noticed that the distribution, once attained, will remain stationary in time (when the gas remains isolated), and also argued (but not very convincingly) that it was the *only* such stationary distribution.

In the first section of his (1868a), Boltzmann aims to reproduce and improve these results for a system of an infinite number of hard discs moving in a plane. He regards it as obvious that the equilibrium distribution should be independent of the position of the discs, and that every direction of their velocities is equally probable. It is therefore sufficient to consider the probability distribution over the various values of the velocity $v = \|\vec{v}\|$. However, Boltzmann started out with a somewhat different interpretation of probability in mind than Maxwell. He introduced the probability distribution as follows:

> Let $\phi(v)dv$ be the sum of all the instants of time during which the velocity of a disc in the course of a very long time lies between v and $v + dv$, and let N be the number of discs which on average are located in a unit surface area, then
>
> $$N\phi(v)dv \tag{34}$$
>
> is the number of discs per unit surface whose velocities lie between v and $v + dv$ (Abh. I, p. 50).[21]

Thus, $\phi(v)dv$ is introduced as the relative *time* during which a (given) disc has a particular velocity. But, in the same breath, this is identified with the relative *number* of discs with this velocity.

This remarkable quote shows how he identified two different meanings for the same function. We shall see that this equivocation returned in different guises again and again in Boltzmann's writings.[22] Indeed, it is, I believe, the very heart of the ergodic problem, put forward so prominently by the Ehrenfests (cf. paragraph 6.1). Either way, of course, whether we average over time or particles,

[21] Here and below, "Abh." refers to the three volumes of Boltzmann's collected scientific papers (Boltzmann 1909).

[22] This is not to say that he always conflated these two interpretations of probability. Some papers employ a clear and consistent choice for one interpretation only. But then that choice differs between papers, or even in different sections of a single paper. In fact, in (Boltzmann 1871c) he even multiplied probabilities with different interpretations into one equation to obtain a joint probability. But then in (1872) he conflates them again. Even in his last paper (Boltzmann & Nabl 1904), Boltzmann identifies two meanings of probability with a simple-minded argument.

probabilities are defined here in strictly mechanical terms, and are therefore objective properties of the gas.

Next he goes into a detailed mechanical description of a two-disc collision process. If the variables which specify the initial velocities of two discs before the collision lie within a given infinitesimal range, Boltzmann determines how the collision will transform the initial values of these variables $(\vec{v}_i, \vec{v}_j)$ into the final values $(\vec{v}'_i, \vec{v}'_j)$ in another range. At this point a two-dimensional analogy of the *Stoßzahlansatz* is introduced to obtain the number of collisions per unit of time. As in Maxwell's treatment, this amounts to assuming that the number of such collisions is proportional to the product $\phi(v_1)\phi(v_2)$. In fact:

$$N(\vec{v}_1, \vec{v}_2) \propto N^2 \frac{\phi(v_1)}{v_1} \frac{\phi(v_2)}{v_2} \|\vec{v}_2 - \vec{v}_1\| dv_1 dv_2 dt \tag{35}$$

where the proportionality constant depends on the geometry of the collision.

He observes that if, for all velocities v_i, v_j and all pairs of discs i, j, the collisions that transform the values of the velocities (v_i, v_j) from a first range $dv_i dv_j$ into values v'_i, v'_j within the range $dv'_i dv'_j$ occur equally often as conversely (i.e., equally often as those collisions that transform initial velocities v'_i, v'_j within $dv'_i dv'_j$ into final values v_i, v_j within $dv_i dv_j$), the distribution ϕ will remain stationary. He states "This distribution is therefore the desired one" (Abh. I p. 55). Actually, this is the first occasion in the paper at which the desideratum of stationarity of the probability distribution is mentioned.

Using the two-dimensional version of the *Stoßzahlansatz* this desideratum leads to

$$\frac{\phi(v_i)}{v_i} \frac{\phi(v_j)}{v_j} = \frac{\phi(v'_i)}{v'_i} \frac{\phi(v'_j)}{v'_j} \tag{36}$$

He shows (Abh. I, p. 57) that the only function obeying condition (36) for all choices of $v_1, v_2, v'_1 v'_2$, compatible with the energy equation $v_1^2 + v_2^2 = v_1'^2 + v_2'^2$, is of the form

$$\phi(v) = 2hve^{-hv^2}, \tag{37}$$

for some constant h. Putting $f(v) := v\phi(v)$ we thus obtain the two-dimensional version of the Maxwell distribution (26). Boltzmann does not address the issue of whether the condition (36) is necessary for the stationarity of ϕ.

In the next subsections of (1868a), Boltzmann repeats the derivation, each time in slightly different settings. First, he goes over to the three-dimensional version of the problem, assuming a system of hard spheres, and supposes that one special sphere is accelerated by an external potential $V(\vec{x})$. He shows that if the velocities of all other spheres are distributed according to the Maxwellian distribution (26), the probability distribution of finding the special sphere at place $\vec{x}$ and velocity $\vec{v}$ is

$f(\vec{v},\vec{x}) \propto e^{-h(\frac{1}{2}mv^2+V(\vec{x}))}$ (Abh. I, p.63). In a subsequent subsection, he replaces the spheres by material points with a short-range interaction potential and reaches a similar result.

At this point, (the end of Section I of the (1868a) paper), the argument suddenly switches course. Instead of continuing in this fashion, Boltzmann announces (Abh. I p. 80) that all the cases treated, and others yet untreated, follow from a much more general theorem. This theorem, which, as we shall see relies on the ergodic hypothesis, is the subject of the second and third Section of the paper. I will limit the discussion to the third section and rely partly on Maxwell's (1879) exposition, which is somewhat simpler and clearer than Boltzmann's own.

4.1.1 The ergodic hypothesis

Consider a general mechanical system of N material points, each with mass m, subject to an arbitrary time-independent potential.[23] In modern notation, let $x = (\vec{q}_1, \vec{p}_1; \ldots; \vec{q}_N, \vec{p}_N)$ denote the canonical position coordinates and momenta of the system. Its Hamiltonian is then[24]

$$H(x) = \frac{1}{2m}\sum_i^N \vec{p}_i^2 + U(\vec{q}_1, \ldots, \vec{q}_N). \tag{38}$$

The state x may be represented as a phase point in the mechanical phase space Γ. Under the Hamiltonian equations of motion, this phase point evolves in time, and thus describes a trajectory x_t ($t \in \mathbb{R}$). This trajectory is constrained to lie on a given energy hypersurface $\Gamma_E = \{x \in \Gamma \,:\, H(x) = E\}$. Boltzmann asks for the probability (i.e. the fraction of time during a very long period) that the phase point lies in a region $dx = d^3\vec{q}_1 \cdots d^3\vec{p}_N$, which we may write as:

$$\rho(x)dx = \chi(x)\delta(H(x) - E)dx. \tag{39}$$

for some function χ. Boltzmann seems to assume implicitly that this distribution is stationary. This property would of course be guaranteed if the "very long period" were understood as an infinite time. He argues, by Liouville's theorem, that χ is a constant for all points on the energy hypersurface that are "possible", i.e. that are actually traversed by the trajectory. For all other points χ vanishes. If we neglect those latter points, the function χ must be constant over the entire energy hypersurface, and the probability density ρ takes the form

$$\rho_{\mathrm{mc}}(x) = \frac{1}{\omega(E)}\delta(H(x) - E), \tag{40}$$

[23] Although Boltzmann does not mention it at this stage, his previous section added the stipulation that the particles are enclosed in a finite space, surrounded by perfectly elastic walls.

[24] Actually Boltzmann allows the N masses to be different but restricts the potential as being due to external and mutual two-particle forces only, i.e. $H(x) = \sum_i \frac{\vec{p}_i^2}{2m_i} + \sum_{i\leq j} U_{ij}(\|\vec{q}_i - \vec{q}_j\|) + \sum_i U_i(\vec{q}_i)$.

the micro-canonical distribution, where

$$\omega(E) = \int \delta H(x) = E dx \tag{41}$$

is the so-called structure function.

In particular, one can now form the marginal probability density for the positions $\vec{q}_1, \ldots, \vec{q}_N$ by integrating over the momenta:

$$\rho_{\rm mc}(\vec{q}_1, \ldots, \vec{q}_N) := \int \rho_{\rm mc}(x)\, d^3\vec{p}_1 \cdots d^3\vec{p}_N = \frac{2m}{\omega(E)} \int \delta\left(\sum \vec{p}_i^{\,2} - 2m(E - U(q))\right) d\vec{p}_1 \cdots d\vec{p}_N. \tag{42}$$

The integral over the momenta can be evaluated explicitly (it is $2R^{-1}$ times the surface area of a hypersphere with radius $R = \sqrt{2m(E-U)}$ in $n = 3N$ dimensions), to obtain

$$\rho_{\rm mc}(\vec{q}_1, \ldots, \vec{q}_N) = \frac{2m(\pi)^{n/2}}{\omega(E)\Gamma(\frac{n}{2})} (2m(E - U(q))^{(n-2)/2}, \tag{43}$$

where Γ denotes Euler's gamma function: $\Gamma(x) := \int_0^\infty t^{x-1} e^{-t} dt$.

Similarly, the marginal probability density for finding the first particle with a given momentum component p_{1x} as well as finding the positions of all particles at $\vec{q}_1, \ldots, \vec{q}_N$ is

$$\begin{aligned} \rho_{\rm mc}(p_{1x}, \vec{q}_1 \ldots, \vec{q}_N) &= \int \rho_{\rm mc}(x)\, dp_{1y} dp_{1z} d^3\vec{p}_2 \cdots d^3\vec{p}_N \\ &= \frac{2m\pi^{(n-1)/2}}{\omega(E)\Gamma(\frac{n-1}{2})} \left(2m(E - U(q)) - p_{1x}^2\right)^{(n-3)/2}. \end{aligned} \tag{44}$$

These two results can be conveniently presented in the form of the conditional probability that the x-component of momentum of the first particle has a value between p and $p + dp$, given that the positions have the values $\vec{q}_1 \ldots, \vec{q}_N$, by taking the ratio of (44) and (43):

$$\rho_{\rm mc}(p \,|\, \vec{q}_1, \ldots, \vec{q}_N) dp = \frac{1}{\sqrt{2m\pi}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \frac{(E - U - \frac{p^2}{2m})^{(n-2)/2}}{(E-U)^{(n-3)/2}} dp. \tag{45}$$

This, in essence, is the general theorem Boltzmann had announced. Further, he shows that in the limit where $n \longrightarrow \infty$, and the kinetic energy per degree of freedom $\kappa := (E - U)/n$ remains constant, the expression (45) approaches

$$\frac{1}{\sqrt{4\pi m\kappa}} \exp\left(-\frac{p^2}{4m\kappa}\right) dp. \tag{46}$$

This probability thus takes the same form as the Maxwell distribution (26), if one equates $\kappa = \frac{1}{2}kT$. Presumably, it is this result that Boltzmann had in mind when he claimed that all the special cases he

has discussed in section 1 of his paper, would follow from the general theorem. One ought to note however, that since U, and therefore κ, depends on the coordinates, the condition $\kappa =$ constant is different for different values of $(\vec{q}_1, \ldots, \vec{q}_n)$.

Some comments on this result.

1. The difference between this approach and that relying on the *Stoßzahlansatz* is rather striking. Instead of concentrating on a gas model in which particles are assumed to move freely except for their occasional collisions, Boltzmann here assumes a much more general Hamiltonian model with an arbitrary interaction potential $U(\vec{q}_1, \ldots \vec{q}_N)$. Moreover, the probability density ρ is defined over phase space, instead of the space of molecular velocities. This is the first occasion where probability considerations are applied to the state of the mechanical system as whole, instead of its individual particles. If the transition between kinetic gas theory and statistical mechanics may be identified with this caesura, (as argued by the Ehrenfests (1912) and by Klein (1973)) it would seem that the transition has already been made right here.

But of course, for Boltzmann the transition did not involve a major conceptual move, thanks to his conception of probability as a relative time. Thus, the probability of a particular state of the total system is still identified with the fraction of time in which that state is occupied by the system. In other words, he had no need for ensembles or non-mechanical probabilistic assumptions.

However, one should note that the equivocation between relative time and relative number of particles, which was comparatively harmless in the first section of the 1868 paper, is now no longer possible in the interpretation of ρ. Consequently, the conditional probability $\rho(p|\vec{q}_1, \ldots \vec{q}_N)$ gives us the relative time that the total system is in a state for which particle 1 has a momentum with x-component between p and $p+dp$, for given values of all the positions. There is no immediate route to conclude that this has anything to do with the relative number of particles with the momentum p. In fact, there is no guarantee that the probability (45) for particle 1 will be the same for other particles too, unless we use the assumption that U is invariant under permutation of the particles. Thus, in spite of their identical form, the distribution (46) has a very different meaning than (26).

2. The transition from (45) to (46), by letting the number of particles become infinite, also seems to be the first instance of a thermodynamical limit. Since the Maxwell distribution is thus recovered only in this limit, Boltzmann's procedure resolves some questions raised above concerning Maxwell's distribution. For a finite number of particles, the distribution (45) always has a finite support, i.e. $\rho_{mc} = 0$ for those values of $p_i^2 \geq 2m(E - U)$. Thus, we do not run into trouble with the finite amount of energy in the gas.

3. Most importantly, the results (45,46) open up a perspective of great generality. It suggests that the probability of the molecular velocities for an isolated system in a stationary state will always assume the Maxwellian form if the number of particles tends to infinity. Notably, this proof seems

to completely dispense with any particular assumption about collisions, or other details of the mechanical model involved, apart from the assumption that it is Hamiltonian. Indeed it need not even represent a gas.

4. The main weakness of the present result is its assumption that the trajectory actually visits all points on the energy hypersurface. This is nowadays called the *ergodic hypothesis*.[25]

Boltzmann returned to this issue on the final page of the paper (Abh. I, p. 96). He notes there that there might be exceptions to his theorem, for example, when the trajectory is periodic. However, Boltzmann observed, such cases would be sensitive to the slightest disturbance from outside. They would be destroyed, e.g. by the interaction of a single free atom that happened to be passing by. He argued that these exceptions would thus only provide cases of unstable equilibrium.

Still, Boltzmann must have felt unsatisfied with his own argument. According to an editorial footnote in his collected works (Abh. I p.96), Boltzmann's personal copy of the paper contains a hand-written remark in the margin stating that the point was still dubious and that it had not been proven that, even in the presence of interaction with a single external atom, the system would traverse all possible values compatible with the energy equation.

4.1.2 Doubts about the ergodic hypothesis

Boltzmann's next paper (1868b) was devoted to checking the validity of the ergodic hypothesis in a relatively simple solvable mechanical model. This paper also gives a nice metaphoric formulation of the ergodic hypothesis: if the phase point were a light source, and its motion exceedingly swift, the entire energy surface would appear to us as homogeneously illuminated (Abh. I, p. 103). However,

[25]The literature contains some surprising confusion about how the hypothesis got its name. The Ehrenfests borrowed the name from Boltzmann's concept of an "*Ergode*", which he introduced in (Boltzmann 1884) and also discussed in his Lectures on Gas Theory (Boltzmann 1898). But what did Boltzmann actually understood by an *Ergode*? Brush points out in his translation of (Boltzmann 1898, p. 297), and similarly in (Brush 1976, p. 364), that Boltzmann used the name to denote a stationary ensemble, characterized by the microcanonical distribution in phase space. In other words, in in the context of Boltzmann's (1898) an *Ergode* is just an microcanonical ensemble, and seems to have nothing to do to do with the so-called ergodic hypothesis. Brush criticized the Ehrenfests for causing confusion by their terminology.

However, in his original (1884) introduction of the phrase, the name *Ergode* is used for a stationary ensemble with only a *single* integral of motion , i.e. its total energy. As a consequence, the ensemble is indeed micro-canonical, but, what is more, every member of the ensemble satisfies the hypothesis of traversing every phase point with the given total energy. Indeed, in this context, being an element of an *Ergode* implies satisfaction of this hypothesis. Thus, the Ehrenfests were actually justified in baptizing the hypothesis "ergodic".

Another dispute has emerged concerning the etymology of the term. The common opinion, going back at least to the Ehrenfests has been that the word derived from *ergos* (work) and *hodos* (path). Gallavotti (1994) has argued however that "undoubtedly" it derives from *ergos* and *eidos* (similar). Now one must grant Gallavotti that it ought to expected that the etymology of the suffix "-ode" of ergode is identical to that of other words Boltzmann coined in this paper, like *Holode, Monode, Orthode* and *Planode*; and that a reference to path would be somewhat unnatural in these last four cases. However, I don't believe a reference to "eidos" would be more natural. Moreover, it seems to me that if Boltzmann intended this etymology, he would have written "Ergoide" in analogy to planetoid, ellipsoid etc. That he was familiar with this common usage is substantiated by him coining the term "*Momentoide*" for momentum-like degrees of freedom (i.e. those that contribute a quadratic term to the Hamiltonian) in (Boltzmann 1892). The argument mentioned by Cercignani (1998, p. 141) (that Gallavotti's father is a classicist) fails to convince me in this matter.

his doubts were still not laid to rest. His next paper on gas theory (1871a) returns to the study of a detailed mechanical gas model, this time consisting of polyatomic molecules, and avoids any reliance on the ergodic hypothesis. And when he did return to the ergodic hypothesis in (1871b), it was with much more caution. Indeed, it is here that he actually first described the worrying assumption as an *hypothesis*, formulated as follows:

> The great irregularity of the thermal motion and the multitude of forces that act on a body make it probable that its atoms, due to the motion we call heat, traverse all positions and velocities which are compatible with the principle of [conservation of] energy (Abh. I p. 284).[26]

Note that Boltzmann formulates this hypothesis for an arbitrary body, i.e. it is not restricted to gases. He also remarks, at the end of the paper, that "the proof that this hypothesis is fulfilled for thermal bodies, or even is fullfillable, has not been provided" (Abh. I p. 287).

There is a major confusion among modern commentators about the role and status of the ergodic hypothesis in Boltzmann's thinking. Indeed, the question has often been raised how Boltzmann could ever have believed that a trajectory traverses *all* points on the energy hypersurface, since, as the Ehrenfests conjectured in 1911, and was shown almost immediately in 1913 by Plancherel and Rosenthal, this is mathematically impossible when the energy hypersurface has a dimension larger than 1 (cf. paragraph 6.1).

It is a fact that both his (1868a, Abh. I, p.96) and (1871b, Abh. I, p.284) mention external disturbances as an ingredient in the motivation for the ergodic hypothesis. This might be taken as evidence for 'interventionism', i.e. the viewpoint that such external influences are crucial in the explanation of thermal phenomena (cf: Blatt 1959, Ridderbos & Redhead 1998). Yet even though Boltzmann clearly expressed the thought that these disturbances might help to motivate the ergodic hypothesis, he never took the idea very seriously. The marginal note in the (1868a) paper mentioned above indicated that, even if the system is disturbed, there is still no easy proof of the ergodic hypothesis, and all his further investigations concerning this hypothesis assume a system that is either completely isolated from its environment, or at most acted upon by a static external force. Thus, interventionism did not play a significant role in his thinking.[27]

It has also been suggested, in view of Boltzmann's later habit of discretizing continuous variables, that he somehow thought of the energy hypersurface as a discrete manifold containing only finitely many discrete cells (Gallavotti 1994). On this reading, obviously, the mathematical no-go theorems

[26] An equivalent formulation of the ergodic hypothesis is that the Hamiltonian is the only independent integral of the Hamiltonian equations of motion. This version is given in the same paper (Boltzmann 1909, p. 281-2)

[27] Indeed, on the rare occasions on which he later mentioned external disturbances, it was only to say that they are "not necessary" (Boltzmann 1895b). See also Boltzmann (1896, §91).

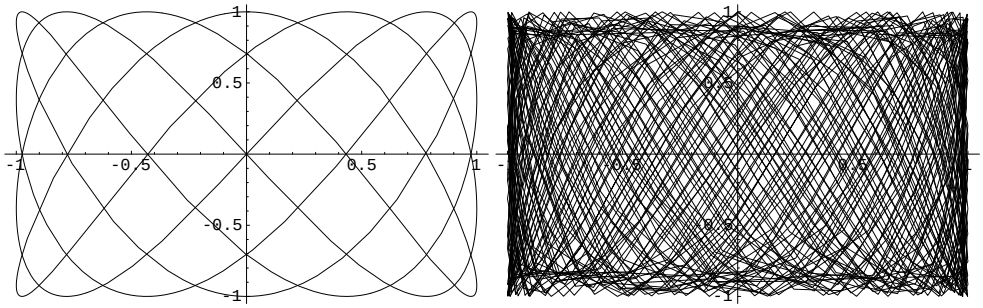

Figure 2: *Trajectories in configuration space for a two-dimensional harmonic oscillator with potential* $U(x, y) = ax^2 + by^2$. *Illustrating the distinction between (i) the case where* $\sqrt{a/b}$ *is rational (here 4/7) and (ii) irrational* $(1/e)$. *Only a fragment of the latter trajectory has been drawn.*

of Rosenthal and Plancherel no longer apply. Now it is definitely true that Boltzmann developed a preference towards discretizing continuous variables, and would later apply this procedure more and more (although usually adding that this was fictitious and purely for purposes of illustration and more easy understanding, cf. paragraph 4.2.2). However, there is no evidence in the (1868) and (1871b) papers that Boltzmann implicitly assumed a discrete structure of the mechanical phase space or the energy hypersurface.

Instead, the context of his (1871b) makes clear enough how he intended the hypothesis, as has already been argued by Brush (1976). Immediately preceding the section in which the hypothesis is introduced, Boltzmann discusses trajectories for a simple example: a two-dimensional harmonic oscillator with potential $U(x, y) = ax^2 + by^2$. For this system, the configuration point (x, y) moves through the surface of a rectangle. (Cf. Fig. 2. See also Cercignani (1998, p. 148).) He then notes that if a/b is rational, (actually: if $\sqrt{a/b}$ is rational) this motion is periodic. However, if this value is irrational, the trajectory will, in the course of time, traverse "*almählich die ganze Fläche*" (Abh. I, p. 271) of the rectangle. He says that in this case x and y are *independent*, since for each values of x an infinity of values for y in any interval in its range are possible. The very fact that Boltzmann considers intervals for the values of x and y of arbitrary small sizes, and stressed the distinction between rational and irrational values of the ratio a/b, indicates that he did *not* silently presuppose that phase space was essentially discrete, where those distinctions would make no sense.

Now clearly, in modern language, one should say that if $\sqrt{a/b}$ is irrational the trajectory is *dense* in the rectangle, but not that it traverses all points. Boltzmann did not possess this language. In fact, he could not have been aware of Cantor's insight that the continuum contains more than a countable infinity of points. Thus, the correct statement that, in the case that $\sqrt{a/b}$ is irrational, the trajectory will traverse, for each value of x, an infinity of values of y within any interval however small, could easily have led him to believe (incorrectly) that *all* values of x and y are traversed in the course of

time.

It thus seems eminently plausible, in view of the fact that this discussion immediately precedes the formulation of the ergodic hypothesis, that Boltzmann's understanding of the ergodic hypothesis is really what Ehrenfests dubbed the *quasi-ergodic hypothesis*: the assumption that the trajectory is dense (i.e. passes arbitrarily close to every point) on the energy hypersurface.[28] The quasi-ergodic hypothesis is not mathematically impossible in higher-dimensional phase spaces. However, the quasi-ergodic hypothesis does not entail the desired conclusion that the only stationary probability distribution over the energy surface is micro-canonical.

Nevertheless, Boltzmann remained sceptical about the validity of his hypothesis, and attempted to explore different routes to his goal of characterizing thermal equilibrium in mechanics. Indeed, both the preceding (1871a) and his next paper (1871c) present alternative arguments, with the explicit recommendation that they avoid hypotheses. In fact, he did not return to the ergodic hypothesis at all until the 1880s (stimulated by Maxwell's 1879 review of the last section of Boltzmann's 1868 paper). At that time, perhaps feeling fortified by Maxwell's authority, he was to express much more confidence in the ergodic hypothesis. However, after 1885, this confidence disappears again, and although he mentions the hypothesis occasionally in later papers, he never assumes its validity. Most notably, the ergodic hypothesis is not even mentioned in his *Lectures on Gas Theory* (1896, 1898).

To sum up, what role did the ergodic hypothesis play for Boltzmann? It seems that Boltzmann regarded the ergodic hypothesis as a special dynamical assumption that may or may not be true, depending on the nature of the system, and perhaps also on its initial state and the disturbances from its environment. Its role was simply to help derive a result of great generality: for any system for which the hypothesis is true, its equilibrium state is characterized by (45), from which an analogy to the Maxwell distribution may be recovered in the limit $N \longrightarrow \infty$, regardless of any details of the inter-particle interactions, or indeed whether the system represented is a gas, fluid, solid or any other thermal body.

As we discussed in paragraph 1.4, the Ehrenfests (1912) have suggested that the ergodic hypothesis played a much more fundamental role. In particular, if the hypothesis is true, averaging over an (infinitely) long time would be identical to phase averaging with the microcanonical distribution. Thus, they suggested that Boltzmann relied on the ergodic hypothesis in order to equate time averages and phase averages, or in other words, to equate two meanings of probability (relative time and relative volume in phase space.) There is however *no* evidence that Boltzmann ever followed this line of reasoning neither in the 1870s, nor later. He simply never gave any justification for equivocating time and particle averages, or phase averages, at all. Presumably, he thought nothing much depended

[28] Or some hypothesis compatible with the quasi-ergodic hypothesis. As it happens, Boltzmann's example is also compatible with the measure-theoretical hypothesis of 'metric transitivity' (cf. paragraph 6.1).

on this issue and that it was a matter of taste.

4.2 The Boltzmann equation and H-theorem (1872)

In 1872 Boltzmann published one of his most important papers. It contained two celebrated results nowadays known as the Boltzmann equation and the H-theorem. The latter result was the basis of Boltzmann's renewed claim to have obtained a general theorem corresponding to the Second Law. This paper has been studied and commented upon by numerous authors, and an entire translation of the text has been provided by (Brush 1966). Thus, for the present purposes, a succinct summary of the main points might have been sufficient. However, there is still dispute among modern commentators about its actual content.

The issue at stake in this dispute is the question whether the results obtained in this paper are presented as necessary consequences of the mechanical equations of motion, or whether Boltzmann explicitly acknowledged that they would allow for exceptions. Klein has written:

> I can find no indication in his 1872 memoir that Boltzmann conceived of possible exceptions to the H-theorem, as he later called it (Klein 1973, p. 73).

Klein argues that Boltzmann only came to acknowledge the existence of such exceptions thanks to Loschmidt's critique in 1877. An opposite opinion is expressed by von Plato (1994). Calling Klein's view a "popular image", he argues that, already in 1872, Boltzmann was well aware that his H-theorem had exceptions, and thus "already had a full hand against his future critics". Indeed, von Plato states that

> Contrary to a widely held opinion, Boltzmann is not in 1872 claiming that the Second Law and the Maxwellian distribution are *necessary* consequences of kinetic theory (von Plato 1994, p. 81).

So it might be of some interest to try and settle this dispute.

Boltzmann (1872) starts with an appraisal of the role of probability theory in the context of gas theory. The number of particles in a gas is so enormous, and their movements are so swift that we can observe nothing but average values. The determination of averages is the province of probability calculus. Therefore, "the problems of the mechanical theory of heat are really problems in probability calculus" (Abh. I, p. 317). But, Boltzmann says, it would be a mistake to believe that the theory of heat would therefore contain uncertainties.

He emphasizes that one should not confuse incompletely proven assertions with rigorously derived theorems of probability theory. The latter are necessary consequences of their premisses, just like in any other theory. They will be confirmed by experience as soon as one has observed a sufficiently large number of cases. This last condition, however, should be no significant problem in

the theory of heat because of the enormous number of molecules in macroscopic bodies. Yet, in this context, one has to make doubly sure that we proceed with the utmost rigour.

Thus, the message expressed in the opening pages of this paper seems clear enough: the results Boltzmann is about to derive are advertised as doubly checked and utterly rigorous. Still, they are theoretical. Their relationship with experience might be less secure, since any probability statement is only reproduced in observations by sufficiently large numbers of independent data. Thus, Boltzmann would have allowed for exceptions in the relationship between theory and observation, but not in the relation between premisses and conclusion.

He continues by saying what he means by probability, and repeats its equivocation as a fraction of time and the relative number of particles that we have seen earlier in 1868:

> If one wants [...] to build up an exact theory [...] it is before all necessary to determine the probabilities of the various states that one and the same molecule assumes in the course of a very long time, and that occur simultaneously for different molecules. That is, one must calculate how the number of those molecules whose states lie between certain limits relates to the total number of molecules (Abh. I p. 317).

However, this equivocation is not vicious. For most of the paper the intended meaning of probability is always the relative number of molecules with a particular molecular state. Only at the final stages of his paper (Abh. I, p. 400) does the time-average interpretation of probability (suddenly) recur.

Boltzmann says that both Maxwell and he had attempted the determination of these probabilities for a gas system but without reaching a complete solution. Yet, on a closer inspection, "it seems not so unlikely that these probabilities can be derived on the basis of the equations of motion alone..." (Abh. I, p. 317). Indeed, he announces, he has solved this problem for gases whose molecules consist of an arbitrary number of atoms. His aim is to prove that whatever the initial distribution of state in such a system of gas molecules, it must inevitably approach the distribution characterized by the Maxwellian form (ibid. p. 320).

The next section specializes to the simplest case of monatomic gases and also provides a more complete specification of the problem he aims to solve. The gas molecules are contained in a fixed vessel with perfectly elastic walls. They interact with each other only when they approach each other at very small distances. These interactions can be mimicked as collisions between elastic bodies. Indeed, these bodies are modeled as hard spheres (Abh I, p. 320). Boltzmann represents the state of the gas by a time-dependent distribution function $f_t(\vec{v})$, called the "distribution of state", which gives us, at each time t, the relative number of molecules with velocity between $\vec{v}$ and $\vec{v} + d^3\vec{v}$.[29]

He also states two more special assumptions:

[29] Actually Boltzmann formulated the discussion in terms of a distribution function over kinetic energy rather than velocity. I have transposed this into the latter, nowadays more common formulation.

1. Already in the initial state of the gas, each direction of velocity is equally probable. That is:

$$f_0(\vec{v}) = f_0(v). \tag{47}$$

   It is assumed as obvious that this will also hold for any later time.

2. The gas is spatially uniform within the container. That is, the relative number of molecules with their velocities in any given interval, and their positions in a particular spatial region R does not depend on the location of R in the available volume.

The next and crucial assumption used by Boltzmann to calculate the change in the number of particles with a velocity $\vec{v}_1$ per unit time, is the *Stoßzahlansatz*, (29) and (30).

For modern readers, there are also a few unstated assumptions that go into the construction of this equation. First, the number of molecules must be large enough so that the (discrete) distribution of their velocities can be well approximated by a continuous and differentiable function f. Secondly, f changes under the effect of binary collisions only. This means that the density of the gas should be low (so that three-particle collisions can be ignored) but not too low (which would make collisions too infrequent to change f at all). These two requirements are already hard enough to put in a mathematically precise form. The modern explicitation is that of taking the so-called Boltzmann-Grad limit (cf. paragraph 6.4). The final (unstated) assumption is that all the above assumptions remain valid in the course of time.

He addresses his aim by constructing a differentio-integral evolution equation for f_t, by taking the difference of (29) and (30) and integrating over all variables except $\vec{v}_1$ and t. The result (in a modern notation) is the *Boltzmann equation*:

$$\frac{\partial f_t(\vec{v}_1)}{\partial t} = N \int_0^d b db \int_0^{2\pi} d\phi \int_{\mathbb{R}^3} d^3\vec{v}_2 \, \|\vec{v}_2 - \vec{v}_1\| \left(f_t(\vec{v}_1{}') f_t(\vec{v}_2{}') - f_t(\vec{v}_1) f_t(\vec{v}_2) \right) \tag{48}$$

which describes the change of f in the course of time, when this function at some initial time is given. (Recall from paragraph 3.3 that the primed velocities are to be thought of as functions of the unprimed velocities and the geometrical parameters of the collision: $\vec{v}_i' = \vec{v}_i{}'(\vec{v}_1, \vec{v}_2, b, \phi)$, and d denotes the diameter of the hard spheres.)

4.2.1 The H-theorem

Assuming that the Boltzmann equation (48) is valid for all times, one can prove, after a few well-known manipulations, that the following quantity

$$H[f_t] := \int f_t(\vec{v}) \ln f_t(\vec{v}) d^3\vec{v} \tag{49}$$

decreases monotonically in time, i.e.

$$\frac{dH[f_t]}{dt} \leq 0; \tag{50}$$

as well as its stationarity for the Maxwell distribution, i.e.:

$$\frac{dH[f_t]}{dt} = 0 \ \ (\forall t) \ \ \text{iff} \ \ f_t(v) = Ae^{-Bv^2}. \tag{51}$$

Boltzmann concludes Section I of the paper as follows:

> It has thus been rigorously proved that whatever may have been the initial distribution of kinetic energy, in the course of time it must necessarily approach the form found by Maxwell. [...] This [proof] actually gains much in significance because of its applicability to the theory of multi-atomic gas molecules. There too, one can prove for a certain quantity $[H]$ that, because of the molecular motion, this quantity can only decrease or in the limiting case remain constant. Thus, one may prove that because of the atomic movement in systems consisting of arbitrarily many material points, there always exists a quantity which, due to these atomic movements, cannot increase, and this quantity agrees, up to a constant factor, exactly with the value that I found in [(Boltzmann 1871c)] for the well-known integral $\int dQ/T$.
>
> This provides an analytical proof of the Second Law in a way completely different from those attempted so far. Up till now, one has attempted to proof that $\int dQ/T = 0$ for a reversible (*umkehrbaren*) cyclic[30] process, which however does not prove that for an irreversible cyclic process, which is the only one that occurs in nature, it is always negative; the reversible process being merely an idealization, which can be approached more or less but never perfectly. Here, however, we immediately reach the result that $\int dQ/T$ is in general negative and zero only in a limit case... (Abh. I, p. 345)

Thus, as in his 1866 paper, Boltzmann claims to have a rigorous, analytical and general proof of the Second Law. From our study of the paper until now, (i.e. section I) it appears that Klein's interpretation is more plausible than von Plato's. I postpone a further discussion of this dispute to paragraph 4.2.3, after a brief look at the other sections of the paper.

4.2.2 Further sections of Boltzmann (1872)

Section II is entitled "Replacement of integrals by sums" and devoted to a repetition of the earlier arguments, now assuming that the kinetic energies of the molecules can only take values in a discrete

[30]The term "cyclic" is missing in Brush's translation, although the original text does speak of "Kreisprozeß". The special notation $\oint$ for cyclic integrals was not introduced until much later.

set $\{0, \epsilon, 2\epsilon, \ldots, p\epsilon\}$. Boltzmann shows that in the limit $\epsilon \longrightarrow 0$, $p\epsilon \longrightarrow \infty$ the same results are recovered.

Many readers have been surprised by this exercise, which seems rather superfluous both from a didactic and a logical point of view. (However, some have felt that it foreshadowed the advent of quantum theory.) Boltzmann offers as motivation for the detour that the discrete approach is clearer than the previous one. He argues that integrals only have a symbolic meaning, as a sum of infinitely many infinitesimal elements, and that a discrete calculation yields more understanding. He does not argue, however, that it is closer to physical reality. Be that as it may, the section does eventually take the limit, and recovers the same results as before.

The third section treats the case where the gas is non-uniform, i.e., when condition 2 above is dropped. For this case, Boltzmann introduces a generalized distribution function $f_t(\vec{r}, \vec{v})$, such that $f_t d^3\vec{r} d^3\vec{v}$ represents the relative number of particles with a position in a volume element $d^3\vec{r}$ around $\vec{r}$ and a velocity in an element $d^3\vec{v}$ around $\vec{v}$.

He obtains a corresponding generalized Boltzmann equation:

$$\frac{\partial f_t(\vec{r}, \vec{v})}{\partial t} + \vec{v} \cdot \nabla_x f_t + \frac{\vec{F}}{m} \cdot \nabla_v f_t = \\ N \int b db d\phi d^3\vec{v}_2 \, \|\vec{v}_2 - \vec{v}_1\| \left(f_t(\vec{r}, \vec{v}_1')) f_t(\vec{r}, \vec{v}_2') - f_t(\vec{r}, \vec{v}_1)) f_t(\vec{r}, \vec{v}_2) \right) \tag{52}$$

where $\vec{F}$ denotes an external force field on the gas. The quantity H now takes the form $H[f_t] := \int f_t(\vec{r}, \vec{v}) d^3\vec{r} d^3\vec{v}$; and a generalization of the H-theorem $dH/dt \leq 0$ is obtained.

The last three sections are devoted to polyatomic molecules, and aim to obtain generalized results for this case too. The key ingredient for doing so is, of course, an appropriately generalized *Stoßzahlansatz*. The formulation of this assumption is essentially the same as the one given in his paper on poly-atomic molecules (1871a), which was later shown wrong and corrected by Lorentz. I will not go into this issue (cf. Lorentz 1887, Boltzmann 1887b, Tolman 1938).

An interesting passage occurs at the very end of the paper, where he expands on the relationship between H and entropy. He considers a monatomic gas in equilibrium. The stationary distribution of state is given as:

$$f^*(\vec{r}, \vec{v}) = V^{-1} \left(\frac{3m}{4\pi T} \right)^{3/2} \exp(\frac{-3mv^2}{4T}) \tag{53}$$

where V is the volume of the container. (Note that in comparison with (27), Boltzmann adopts units for temperature that make $k = 2/3$.) He shows that

$$H[f^*] := \int f^* \log f^* dx dv = -N \log V \left(\frac{4\pi T}{3m} \right)^{3/2} - \frac{3}{2} N; \tag{54}$$

which agrees (assuming $S = -kNH[f^*]$) with the thermodynamical expression for the ideal gas

(16) up to an additive constant. A similar result holds for the polyatomic gas.

4.2.3 Remarks and problems

1. The role of probability. As we have seen, the H-theorem formed the basis of a renewed claim by Boltzmann to have obtained a theorem corresponding to the full Second Law (i.e. including both parts) at least for gases. A main difference from his 1866 claim, is that he now strongly emphasizes the role of probability calculus in his derivation. It is clear that the conception of probability expounded here is thoroughly frequentist and that he takes 'the laws of probability' as empirical statements. Furthermore, probabilities can be fully expressed in mechanical terms: the probability distribution f is nothing but the relative number of particles whose molecular states lie within certain limits. Thus, there is no conflict between his claims that on the one hand, "the problems of the mechanical theory of heat are really problems in probability calculus" and that the probabilities themselves are derived on the basis of the equations of motion alone, on the other hand. Indeed, it seems to me that Boltzmann's emphasis on the crucial role of probability in this paper is only intended to convey that probability theory provides a particularly useful and appropriate language for discussing mechanical problems in gas theory. There is no indication in this paper yet that probability theory could play a role by furnishing assumptions of a non-mechanical nature, i.e., independent of the equations of motion (cf. Boltzmann & Nabl 1904, p. 520).

2. The role of the Stoßzahlansatz. Note that Boltzmann stresses the generality, rigour and "analyticity" of his proof. He puts no emphasis on the special assumptions that go into the argument. Indeed, the *Stoßzahlansatz*, later identified as the key assumption that is responsible for the time-asymmetry of the H-theorem, is announced as follows

> The determination [of the number of collisions] can only be obtained in a truly tedious manner, by consideration of the relative velocities of both particles. But since this consideration has, apart from its tediousness, not the slightest difficulty, nor any special interest, and because the result is so simple that one might almost say it is self-evident I will only state this result." (Abh. I, p. 323)

It thus seems natural that Boltzmann's contemporaries must have understood him as claiming that the H-theorem followed necessarily from the dynamics of the mechanical gas model.[31] I can find no evidence in the paper that he intended this claim to be read with a pinch of salt, as (von Plato 1991, p.. 81) has argued.

[31] Indeed this is *exactly* how Boltzmann's claims were understood. For example, the recommendation written in 1888 for his membership of the Prussian Academy of Sciences mentions as his main feat that Boltzmann had proven that, whatever its initial state, a gas must necessarily approach the Maxwellian distribution (Kirsten & Körber 1975, p.109).

Is there then no evidence at all for von Plato's reading of the paper? Von Plato refers to a passage from Section II, where Boltzmann repeats the previous analysis by assuming that energy can take on only discrete values, and replacing all integrals by sums. He recovers, of course, the same conclusion, but now adds a side remark, which touches upon the case of non-uniform gases:

> Whatever may have been the initial distribution of states, there is one and only one distribution which will be approached in the course of time. [...] This statement has been proved for the case where the distribution of states was already initially uniform. It must also be valid when this is not the case, i.e. when the molecules are initially distributed in such a way that in the course of time they mix among themselves more and more, so that after a very long time the distribution of states becomes uniform. This will always be the case, with the exception of very special cases, e.g. when all molecules were initially situated along a straight line, and were reflected by the walls onto this line (Abh. I, p. 358).

It is this last remark that, apparently, led to the view that after all Boltzmann did already conceive of exceptions to his claims. However, I should say that this passage does not convince me. True enough, Boltzmann in the above quote indicates that there are exceptions. But he mentions them only in connection with an *extension* of his results to the case when the gas is not initially uniform, i.e. when condition (2) above is dropped. There can be no doubt that under the assumption of the conditions (1) and (2), Boltzmann claimed the rigorous validity of the H-theorem. (Curiously, his more systematic treatment of the non-uniform gas (Section III of (1872)) does not mention any exception to the claim that "H can only decrease" (Abh. I p. 362).

As a matter of fact, when Loschmidt formulated the objection, it happened to be by means of an example of a non-uniform gas (although nothing essential depended on this). Thus, if Boltzmann had in 1872 a "full hand against his future critics", as von Plato claims, one would expect his reply to Loschmidt's objection to point out that Loschmidt was correct but that he had already anticipated the objection. Instead, he accused Loschmidt of a fallacy (see paragraph 4.3 below).

But apart from the historical issue of whether Boltzmann did or did not envisage exceptions to his H-theorem, it seems more important to ask what kind of justification Boltzmann might have adduced for the *Stoßzahlansatz*. An attempt to answer this question must be somewhat speculative, since, as we have seen, Boltzmann presented the assumption as "almost self-evident" and "having no special interest", and hence presumably as not in need of further explanation. Still the following remarks may be made with some confidence.

First, we have seen that Maxwell's earlier usage of the assumption was never far away from an argument from insufficient reason. Thus, in his approach, one could think of the *Stoßzahlansatz* as expressing that we have no reason to expect any influence or correlation between any pair of particles

that are about to collide. The assumption would then appear as a probabilistic assumption, reflecting a 'reasonable judgment', independent from mechanics.

In contrast, Boltzmann's critique of Maxwell's approach (cf. footnote 16) suggests that he did not buy this arguments from insufficient reason. But since the *Stoßzahlansatz* clearly cannot be conceived of as an assumption about dynamics —like the ergodic hypothesis—, this leaves only the option that it must be due to a special assumption about the mechanical state of the gas. Indeed, in the years 1895-6, when Boltzmann acknowledged the need for the *ansatz* in the proof of his H-theorem more explicitly —referring to it as "Assumption A" (Boltzmann 1895) or "the hypothesis of molecular disorder" (Boltzmann 1896)—, he formulated it as an assumption *about* the state of the gas.

Yet, even in those years, he would also formulate the hypothesis as expressing that "haphazard governs freely" (Boltzmann 1895, Abh. III, p. 546) or "that the laws of probability are applicable for finding the number of collisions" (Boltzmann 1895b). Similarly, he describes states for which the hypothesis fails as contrived "so as to intentionally violate the laws of probability"(Boltzmann 1896, §3). However, I think these quotations should not be read as claims that the *Stoßzahlansatz* was a consequence of probability theory itself. Rather, given Boltzmann's empirical understanding of "the laws of probability", they suggest that Boltzmann thought that, as a matter of empirical fact, the assumption would 'almost always' hold, even if the gas was initially very far from equilibrium.

3. The H-theorem and the Second Law. Note that Boltzmann misconstrues, or perhaps understates, the significance of his results. Both the Boltzmann equation and the H-theorem refer to a body of gas in a fixed container that evolves in isolation from its environment. There is no question of heat being exchanged by the gas during a process, let alone in an irreversible cyclic process. His comparison in the quotation on page 46 with Clausius' integral $\int dQ/T$ (i.e. $\oint đQ/T$ in equation (18) above) is therefore really completely out of place.

The true import of Boltzmann's results is rather that they provide (i) a generalization of the entropy concept to non-equilibrium states,[32] and (ii)a claim that this non-equilibrium entropy $-kH$ increases monotonically as the isolated gas evolves for non-equilibrium towards an equilibrium state. The relationship with the Second Law is, therefore, somewhat indirect: On the one hand, Boltzmann proves much more than was required, since the second law does not speak of non-equilibrium entropy, nor of monotonic increase; on the other hand it proves also less, since Boltzmann does not consider the increase of entropy in general adiabatic processes.

[32]Boltzmann emphasized that his expression for entropy should be seen as an *extension* of thermodynamic entropy to non-equilibrium states in (1877b, Abh. II, p. 218; 1896, §5) . Of course there is no guarantee that this generalization is the *unique* candidate for a non-equilibrium entropy.

4.3 Boltzmann (1877a): the reversibility objection

According to Klein (1973), Boltzmann seemed to have been satisfied with his treatments of 1871 and 1872 and turned his attention to other matters for a couple of years. He did come back to gas theory in 1875 to discuss an extension of the Boltzmann equation to gases subjected to external forces. But this paper does not present any fundamental changes of thought. (However, it does contain some further elucidation, for example, it mentions for the first time that the derivation of the Boltzmann equation requires that the gas is so dilute that collisions between three or more particles simultaneously can be ignored).

However, the 1875 paper did contain a result which, two years later, led to a debate with Loschmidt. Boltzmann showed that (52) implied that a gas in equilibrium in an external force field (such as the earth's gravity) should have the same average kinetic energy at all heights and therefore, a uniform temperature; while its pressure and density would of course vary with height. This conclusion conflicted with the intuition that when molecules travel upwards, they must do work against the gravitational field, and pay for this by having a lower kinetic energy at greater heights.

Now Boltzmann (1875) was not the first to reach the contrary result, and Loschmidt was not the first to challenge it. Maxwell and Guthrie entered into a debate on the very same topic in 1873. But actually their main point of contention need not concern us very much. The discussion between Loschmidt and Boltzmann is particularly important for quite another issue, which Loschmidt only introduced as an side remark. Considering a gas container in a homogeneous gravitational field, Loschmidt discussed a situation where initially all atoms except one lie at rest at the bottom of the container. The single moving atom could then, by collisions, stir the others and send them into motion until a "stationary state", characterized by the Maxwell distribution, is obtained. He continues

> By the way, one should be careful about the claim that in a system in which the so-called stationary state has been achieved, starting from an arbitrary initial state, this average state can remain intact for all times. I believe, rather, that one can make this prediction only for a short while with full confidence.
>
> Indeed, if in the above case, after a time τ which is long enough to obtain the stationary state, one suddenly assumes that the velocities of all atoms are reversed, we would obtain an initial state that would appear to have the same character as the stationary state. For a fairly long time this would be appropriate, but gradually the stationary state would deteriorate, and after passage of the time τ we would inevitable return to our initial state: only one atom has absorbed all kinetic energy of the system [...], while all other molecules lie still on the bottom of the container.
>
> Obviously, in every arbitrary system the course of events must be become retrograde

when the velocities of all its elements are reversed (Loschmidt 1876, p. 139).

4.3.1 Boltzmann's response (1877a)

Boltzmann's response to Loschmidt is somewhat confusing. On the one hand, he acknowledges that Loschmidt's objection is "quite ingenious and of great significance for the correct understanding of the Second Law." However, he also brands the objection as a "fallacy" and a "sophism".[33] But then, two pages later again, the argument is "of the greatest importance since it shows how intimately connected are the Second Law and probability theory."

The gist of the response is this. First, Boltzmann captures the essential core of the problem in an admirably clear fashion:

> "Every attempt to prove, from the nature of bodies and the laws of interaction for the forces they exert among each other, without any assumption about initial conditions, that
>
> $$\int \frac{dQ}{T} \leq 0 \tag{55}$$
>
> must be in vain" (Abh. II. p.119–121).

The point raised here is usually known as the *reversibility objection*. And since the H-theorem (which only received this name in the 1890s) was presented in 1872 as a general proof that $\int \frac{dQ}{T} \leq 0$ (cf. the long quotation on page 46), it would imply that this theorem was invalid. Boltzmann aims to show, however, that this objection is a fallacy. His argument might be dissected into 5 central points.

1. Conceding that the proof cannot be given. Boltzmann says that a proof that every distribution must with absolute necessity evolve towards a uniform distribution cannot be given, claiming that this fact "is already taught by probability theory". Indeed, he argues, even a very non-uniform distribution of state is, although improbable to the highest degree, not impossible. Thus, he admits that there are initial states for which H increases, just as well as those for which H decreases. This admission, of course, is hard to rhyme with his professed purpose of showing that it is fallacious to conclude that some assumption about the initial state would be needed.

Note that this passage announces a major conceptual shift. Whereas the 1872 paper treated the distribution of state f_t as if it *defines* probability (i.e. of molecular velocities), this time the distribution of states is itself something which can be to a higher or lesser degree "probable". That is: probabilities are *attributed* to distributions of state, i.e. the distribution of state itself is treated as

[33] The very fact that Boltzmann called this conclusion —which by all means and standards is *correct*— a fallacy shows, in my opinion, that he had not anticipated the objection. In fact, how much Boltzmann had yet to learn from Loschmidt's objection is evident when we compare this judgment to a quotation from his *Lectures on Gas Theory* (1898, p. 442): "this one-sidedness [of the H-theorem] lies uniquely and solely in the initial conditions."

a random variable. This shift in viewpoint became more explicit in his (1877b); as we will discuss in section 4.4 below.

2. Rethinking the meaning of "probability". Boltzmann argues that every distribution of state, whether uniform or non-uniform, is equally improbable. But there are "infinitely many" more uniform distributions of state than non-uniform distributions. Here we witness another conceptual shift. In (1872), the term "distribution of state" referred to the function $f(\vec{v})$ or $f(\vec{r}, \vec{v})$, representing the relative numbers of molecules with various molecular states. In that sense, there would, of course, only be a *single* uniform distribution of state: the Maxwellian distribution function (53). But since Boltzmann now claims there are many, he apparently uses the term "distribution of state" to denote a much more detailed description, that includes the velocity and position of every individual molecule, so that permutations of the molecules yield a different distribution of state. That is, he uses the term in the sense of what we would nowadays call a microstate, and what he himself would call a "*Komplexion*" a few months later in his (1877b)—on which occasion he would reserve the name 'distribution of state' for the macrostate.

Note that Boltzmann assumes every Komplexion to be equally probable (or improbable) so that the probability of a particular distribution of state is determined by the relative numbers. Indeed he remarks that it might be interesting to calculate the probabilities of state distributions by determining the ratio of their numbers; this suggestion is also worked out in his subsequent paper of 1877b.

This, indeed, marks another conceptual change. Not only are probabilities attributed to distributions of state instead of being defined by them; they are determined by an equiprobability assumption. Boltzmann does not explicitly motivate the assumption. In view of the discussion in paragraph 3.1, one might conjecture that he must have had something like Laplace's principle of insufficient reason in mind, which makes any two cases which, according to our information are equally possible, also equally probable. But this would indicate an even larger conceptual change; and not just because Boltzmann is broadly a frequentist concerning probability. Also, the principle of insufficient reason, or any similar assumption, makes sense only from the view point that probability is a non-mechanical notion: it reflects our belief or information about a system. I cannot find any evidence that he accepted this idea. Of course it is also possible to conjecture that he silently fell back upon the ergodic hypothesis. But this conjecture also seems unlikely, given his avoidance of the hypothesis since 1871.

3. A claim about evolutions. Boltzmann says: "Only from the fact that there are so many more uniform than non-uniform distributions of state [i.e.: microstates] follows the larger probability that the distribution will become uniform in the course of time" (p. 120). More explicitly, he continues:

> [...] one can prove that infinitely many more initial states evolve after a long time towards a more uniform distribution of states than to a less uniform one, and that even

in that latter case, these states will become uniform after an even longer time (Abh. II, p. 120)[34]

Note that this is a claim about evolutions of microstates. In fact, it is the first case of what the Ehrenfests later called a *statistical H-theorem*, but what is perhaps better called a *statistical reading* of the H-theorem, since in spite of Boltzmann's assertion, no proof is offered.

4. The (im)probability of Loschmidt's initial state. Boltzmann maintains that the initial conditions considered by Loschmidt only have a minute probability. This is because it is obtained by a time evolution and velocity reversal of a non-uniform microstate. Since both time evolution and velocity reversal are one-to-one mappings (or more to the point: they preserve the Liouville measure), these operations should not affect the number or probability of states. Hence, the probability of Loschmidt's state is equal to that of the special non-uniform state from which it is constructed. But by point 2 above, there are infinitely many more uniform states than non-uniform states, so the probability of Loschmidt's state is extraordinarily small.

5. From (im)probability to (im)possibility. The final ingredient of Boltzmann's response is the claim that whatever has an extraordinarily small probability is practically impossible.

The conclusion of Boltzmann's argument, based on these five points, is that the state selected by Loschmidt may be considered as practically impossible. Note that this is a completely static argument; i.e., its logic relies merely on the points 1,2,4 and 5, and makes no assumption about evolutions, apart from the general feature that the dynamical evolution conserves states (or measure). Indeed, point 3, i.e. the statistical reading of the H-theorem, is not used in the argument.

As a consequence, the argument, although perfectly consistent, shows more than Boltzmann can possibly have wanted. The same reasoning that implies Loschmidt's initial state can be ignored, also excludes other non-uniform states. In particular, the same probability should be assigned to Loschmidt's initial state *without* the reversal of velocities. But that state *can* be produced in the laboratory, and, presumably, should not be considered as practically impossible. Indeed, if we adopt the rule that all non-uniform states are to be ignored on account of their low probability, we end up with a consideration of uniform states only, i.e. the theory would be reduced to a description of equilibrium, and the H-theorem reduced to $dH/dt = 0$, and any time-asymmetry is lost.

This, surely, is too cheap a victory over Loschmidt's objection. What one would like to see in Boltzmann's argument is a greater role for assumptions about the time evolution in order to substantiate his statistical reading of the H-theorem.

[34]The clause about 'the latter case' is absent in the translation by (Brush 2003, p. 366).

Summing up: From this point on, we shall see that Boltzmann emphasizes even more strongly the close relations between the Second Law and probability theory. Even so, it is not always clear what these relations are exactly. Further, one may question whether his considerations of the probability of the initial state hit the nail on the head. Probability theory is equally neutral to the direction of time as is mechanics.

The true source of the reversibility problem was only identified by Burbury (1894a) and Bryan (1894) after Boltzmann's lecture in Oxford, which created a intense debate in the columns of *Nature*. They pointed out that the *Stoßzahlansatz* already contained a time-asymmetric assumption.

Indeed, this assumption requires that the number of collisions of the kind $(\vec{v}_1, \vec{v}_2) \longrightarrow (\vec{v}_1{}', \vec{v}_2{}')$ is proportional to the product $f(\vec{v}_1)f(\vec{v}_2)$ where, $\vec{v}_1, \vec{v}_2$ are the velocities *before* the collisions. If we would replace this by the requirement that the number of collisions is proportional to the product for the velocities $\vec{v}_1{}', \vec{v}_2{}'$ *after* the collision, we would obtain, by a similar reasoning, $dH/dt \geq 0$. The question is now, of course, why we should prefer one assumption above the other, without falling into some kind of double standard. (I refer to Price (1996) for a detailed discussion of this danger.) One thing is certain, and that is that any such preference cannot be obtained from mechanics and probability theory alone.

4.4 Boltzmann (1877b): the combinatorial argument

Boltzmann's next paper (1877b) is often seen as a major departure from the conceptual basis employed in his previous work. Indeed, the conceptual shifts already indicated implicitly in his reply to Loschmidt become in this article explicit. Indeed, according to (ter Haar 1955, p. 296) and (Klein 1973, p. 83), it is this paper that marks the transition from kinetic theory to statistical mechanics. Further, the paper presents the famous link between entropy and 'probability' that later became known as "Boltzmann's principle", and was engraved on his tombstone as "$S = k \log W$".

Boltzmann's begins the paper by stating that his goal is to elucidate the relationship between the Second Law and probability calculus. He notes he has repeatedly emphasized that the Second Law is related to probability calculus. In particular he points out that the 1872 paper confirmed this relationship by showing that a certain quantity [i.e. H] can only decrease, and must therefore obtain its minimum value in the state of thermal equilibrium. Yet, this connection of the Second Law with probability theory became even more apparent in his previous paper (1877a). Boltzmann states that he will now solve the problem mentioned in that paper, of calculating the probabilities of various distributions of state by determining the ratio of their numbers.

He also announces that, when a system starts in an improbable state, it will always evolve towards more probable states, until it reaches the most probable state, i.e. that of thermal equilibrium. When this is applied to the Second Law, he says, "we can identify that quantity which is usually called

entropy, with the probability of the state in question." And: "According to the present interpretation, [the Second Law] states nothing else but that the probability of the total state of a composite system always increases" [Abh. II, pp. 165-6]. Exactly how all this is meant, he says, will become clear later in the article.

4.4.1 The combinatorial argument

Succinctly, and rephrased in the Ehrenfests' terminology, the argument is as follows. Apart from Γ, the mechanical phase space containing the possible states x for the total gas system, we consider the so-called μ-space, i.e. the state space of a single molecule. For monatomic gases, this space is just a six-dimensional Euclidean space with $(\vec{r}, \vec{v})$ as coordinates. With each mechanical state x we can associate a collection of N points in μ-space; one for each molecule.

Now, partition μ-space into m disjoint cells: $\mu = \omega_1 \cup \ldots \cup \omega_m$. These cells are taken to be rectangular in the coordinates and of equal size. Further, it is assumed that the energy of each molecule in cell ω_i in has a value ϵ_i, depending only on i. For each x, henceforth also called the *microstate* (Boltzmann's term was the *Komplexion*), we define the *macrostate* or 'distribution of state' as $Z := (n_1, \ldots, n_m)$, with n_i the number of particles whose molecular state is in cell ω_i. The relation between macro- and microstate is obviously non-unique since many different microstates, e.g. obtained by permuting the molecules, lead to the same macrostate. One may associate with every given macrostate Z_0 the corresponding set of microstates:

$$\Gamma_{Z_0} := \{x \in \Gamma \ : \ Z(x) = Z_0\}. \tag{56}$$

The phase space volume $|\Gamma_{Z_0}|$ of this set is proportional to the number of permutations of the particles that do not change the macrostate Z_0. Indeed, when the six-dimensional volume of the cells ω_i is $\delta\omega$, i.e., the same for each cell, the phase space volume of the set Γ_Z is

$$|\Gamma_Z| = \frac{N!}{n_1! \cdots n_m!} (\delta\omega)^N. \tag{57}$$

Moreover, assuming that $n_i \gg 1$ for all i and using the Stirling approximation for the factorials, one finds

$$\ln \Gamma_Z \approx N \ln N - \sum_i n_i \ln n_i + N \ln \delta\omega. \tag{58}$$

This expression is in fact proportional to a discrete approximation of the H-function. Indeed, putting

$$n_i = N f(\vec{r}_i, \vec{v}_i) \delta\omega \tag{59}$$

where $(\vec{r}_i, \vec{v}_i)$ are the coordinates of a representative point in ω_i, we find

$$\begin{aligned}\sum_i n_i \ln n_i &= \sum_i N f(\vec{r}_i, \vec{v}_i)\ \ln\Big(N f(\vec{r}_i, \vec{v}_i)\delta\omega\Big)\,\delta\omega \\ &\approx N\int f(\vec{r}, \vec{v})\,\Big(\ln f(\vec{r}, \vec{v}) + \ln N + \ln\delta\omega\Big)\,d^3\vec{r}d^3\vec{v} \\ &= NH + N\ln N + N\ln\delta\omega; \end{aligned} \tag{60}$$

and therefore, in view of (58):

$$-NH \approx \ln|\Gamma_Z|. \tag{61}$$

And since Boltzmann had already identified $-kNH$ with the entropy of a macrostate, one can also take entropy as proportional to the logarithm of the volume of the corresponding region in phase space. Today, $\ln|\Gamma_Z|$ is often called the *Boltzmann entropy*.

Boltzmann next considers the question for which choice of Z does the region Γ_Z have maximal size, under the constraints of a given total number of particles N, and a total energy E:

$$N = \sum_{i=1}^{m} n_i, \qquad\qquad E = \sum_{i=1}^{m} n_i\epsilon_i. \tag{62}$$

This problem can easily be solved with the Lagrange multiplier technique. Under the Stirling approximation (58) one finds

$$n_i = \mu e^{\lambda\epsilon_i}, \tag{63}$$

which is a discrete version of the Maxwell distribution. (Here, μ an λ are determined in terms of N and E by the constraints (62).)

Boltzmann proposes to take the macrostate with the largest volume as representing equilibrium. More generally, he also refers to these volumes as the "probability" or "permutability" of the macrostate. He therefore now expresses the Second Law as a tendency for the system to evolve towards ever more probable macrostates, until, in equilibrium, it has reached the most probable state.

4.4.2 Remarks and problems

1. the role of dynamics. In the present argument, no dynamical assumption has been made. In particular, it is not relevant to the argument whether the ergodic hypothesis holds, or how the particles collide. At first sight, it might seem that this makes the present argument more general than the previous one. Indeed, Boltzmann suggests at the end of the paper (Abh. II p. 223) that the same argument might be applicable also to dense gases and even to solids.

However, it should be noticed that the assumption that the total energy can be expressed in the form $E = \sum_i n_i \epsilon_i$ where the energy of each particle depends only on the cell in which it is located, and not on the state of other particles is very strong. This can only be maintained, independently of the number N, if there is no interaction at all between the particles. The validity of the argument is thus really restricted to ideal gases (cf. Uhlenbeck and Ford 1963).

2. The choice of cells. One might perhaps hope, at first sight, that the procedure of partitioning μ-space into cells is only a technical or didactic device and can be eliminated by finally taking a limit in which $\delta\omega \longrightarrow 0$; similar to the procedure of his 1872 paper. This hope is dashed because the expression (58) diverges. Indeed, the whole prospect of using combinatorics would disappear if we did not adopt a finite partition. But also the special choice to give all cells equal volume in position and velocity variables is not quite self-evident, as Boltzmann himself shows. In fact, before he develops the argument given here, his paper presents a discussion in which the particles are characterized by their energy instead of position and velocity. This leads him to carve up μ-space into cells of equal size $\delta\epsilon$ in energy. He then shows that the combinatorial argument *fails* to reproduce the desired Maxwell distribution for particles moving in 3 spatial dimensions.[35] This failure is then remedied (Abh. II, p. 190) by switching to a choice of equally sized cells in $\delta\omega$ in position and velocity. The latter choice is apparently 'right', in the sense that leads to the desired result. However, since the choice clearly cannot be relegated to a matter of convention, it leaves open the question of justification.

Modern commentators are utterly divided in the search for a direction in which a motivation for the choice of the size of these cells can be found. Some argue that the choice should be made in accordance with the actual finite resolution of measuring instruments or human observation capabilities. The question whether these do in fact favour a partition into cells of equal phase space volume has hardly been touched upon. Others (Popper 1982, Redhead 1995) reject an appeal to observation capacities on the grounds that these would introduce a 'subjective' or 'anthropocentric' element into the explanation of irreversibility (see also Jaynes 1965, Grünbaum 1973, Denbigh & Denbigh 1985, Ridderbos 2002).

3. Micro versus macro. The essential step in the argument is the distinction between micro- and macrostates. This is indeed the decisive new element, that allowed Boltzmann a complete reinterpretation of the notion and role of probability.

In 1872 and before, the distribution of state f was *identified* with a probability (namely of a

[35]The problem is that for an ideal gas, where all energy is kinetic, $\delta\epsilon \propto v\delta v$. On the other hand, for three-dimensional particles, $\delta\omega \propto v^2\delta v$. The function f derived from (59) and (63) thus has a different dependence on v in the two cases. As Boltzmann notes, the two choices are compatible for particles in two dimensions (i.e. discs moving in a plane).

molecular state, cf. Remark 1 of paragraph 4.2.3). On the other hand, in the present work it, or its discrete analogue Z, is a description of the macrostate of the gas, to which a probability is *assigned*. Essentially, the role of the distribution of state has been shifted from defining a probability measure to being a stochastic variable. Its previous role is taken over by a new idea: Probabilities are not assigned to the particles, but to the macrostate of the gas as a whole, and measured by the corresponding volume in phase space.

Another novelty is that Boltzmann has changed his concept of equilibrium. Whereas previously the defining characteristic of equilibrium was its stationarity, in Boltzmann's new view it is conceived as the macrostate (i.e. a region in phase space) that takes up the largest volume. As a result, a system in a state of equilibrium need not remain there: in the course of time, the microstate of the system may fluctuate in and out of this equilibrium region. Boltzmann briefly investigated the probability of such fluctuations in his (Boltzmann 1878). Almost thirty years later, the experimental predictions for fluctuation phenomena by Einstein and Smoluchowski provided striking empirical successes for statistical mechanics.

4. But what about evolutions? Perhaps the most important issue is this. What exactly is the relation of the 1877b paper to Loschmidt's objection and Boltzmann's primary reply to it (1877a)? The primary reply (cf. paragraph 4.3) can be read as an announcement of two subjects of further investigation:

> From the relative numbers of the various distributions of state, one might even be able to calculate their probabilities. This could lead to an interesting method of determining thermal equilibrium (Abh. II, p. 121)

This is a problem about equilibrium. The second announcement was that Boltzmann said "The case is completely analogous for the Second Law" (Abh. II, p. 121). Because there are so very many more uniform than non-uniform distributions, it should be extraordinarily improbable that a system should evolve from a uniform distribution of states to a non-uniform distribution of states. This is a problem about evolution (cf. point 3 of section 4.3). In other words, one would like to see that something like the statistical H-theorem actually holds.

Boltzmann's (1877b) is widely read as a follow-up to these announcements. Indeed, Boltzmann repeats the first quote above in the introduction of the paper (Abh. II, p. 165), indicating that he will address this problem. And so he does, extensively. Yet he also states:

> Our main goal is not to linger on a discussion of thermal equilibrium, but to investigate the relations of probability with the Second Law of thermodynamics (Abh. II, p. 166).

Thus, the main goal of 1877b is apparently to address the problem concerning evolutions and to show how they relate to the Second Law. Indeed, this is what one would naturally expect since the

reversibility objection is, after all, a problem concerned with evolutions. Even so, a remarkable fact is that the 1877b paper hardly ever touches its self-professed "main goal" at all. As a matter of fact, I can find only one passage in the remainder of the paper where a connection with the Second Law is mentioned.

It occurs in Section V (Abh. II, p. 216-7). After showing that in equilibrium states for monatomic gases the 'permutability measure' $\ln|\Gamma_Z|$ (for which Boltzmann's notation is Ω) is proportional to the thermodynamical entropy, up to an arbitrary additive constant, he concludes that, by choosing the constant appropriately:[36]

$$\int \frac{dQ}{T} = \frac{2}{3}\Omega \left[= \frac{2}{3}\ln|\Gamma_Z|\right] \tag{64}$$

and adds:

> It is known that when a system of bodies goes through reversible changes, the total sum of the entropies of all these bodies remains constant; but when there are among these processes also irreversible (nicht umkehrbar) changes, then the total entropy must necessarily increase. This follows from the familiar circumstance that $\int dQ/T$ is negative for an irreversible cyclic process. In view of (64), the sum of all permutability measures of all bodies $\sum \Omega$, or their total permutability measure, must also increase. Hence, permutability is a quantity which is, up to a multiplicative and additive constant, identical to entropy, but which retains a meaning also during the passage of an irreversible body [sic– read: "process"], in the course of which it continually increases (Abh. II p.217)

How does this settle the problem about evolutions, and does it provide a satisfactory refutation of the reversibility objection? In the literature, there are at least four views about what Boltzmann's response actually intended or accomplished.

4α. *Relying on the separation between micro- and macroscales:* A view that has been voiced recently, e.g. by Goldstein (2001), is that Boltzmann had, by his own argument, adequately and straightforwardly explained why entropy should tend to increase. In particular, this view argues, the fact of the overwhelmingly large phase space volume of the set $\Gamma_{\rm eq}$ of all equilibrium phase points, compared to the set of non-equilibrium points already provides a sufficient argument.

> For a non-equilibrium phase point x of energy E, the Hamiltonian dynamics governing the motion x_t arising from x would have to be ridiculously special to avoid reasonably quickly carrying x_t into $\Gamma_{\rm eq}$ and keeping it there for an extremely long time —unless, of course x itself were ridiculously special (Goldstein 2001, p. 6).

[36] Actually, equation (64) is the closest he got to the famous formula on his tombstone, since $\Omega = \ln W$, and Boltzmann adopts a temperature scale that makes $k = 2/3$.

In fact, this view may lay some claim to being historically faithful. As we have seen, Boltzmann's (1877a) did claim that the large probability for an evolution towards equilibrium did follow from the large differences in number of states.

The main difficulty with this view is that, from a modern perspective, it is hard to maintain that it is adequate. States don't evolve into other states just because there are more of the latter, or because they make up a set of larger measure. The evolution of a system depends only on its initial state and its Hamiltonian. Questions about evolution can only be answered by means of an appeal to dynamics, not by the measure of sets alone. To take an extreme example, the trajectory covered by x_t, i.e. the set $\{x_t \ : \ t \in \mathbb{R}\}$ is a set of measure zero anyway; and hence very special. By contrast, its complement, i.e. the set of states *not* visited by a given trajectory is huge: it has measure one. Certainly, we cannot argue that the system cannot avoid wandering into the set of states that it does not visit. Another example is that of a system of non-interacting particles, e.g., the ideal gas. In this case, all the energies of the individual particles are conserved, and because of these conserved quantities, the phase point can only visit a very restricted region of phase space.[37]

The lesson is, of course, that in order to obtain any satisfactory argument why the system should tend to evolve from non-equilibrium states to the equilibrium state, we should make some assumptions about its dynamics. In any case, judgments like "reasonable" or "ridiculous" remain partly a matter of taste. The reversibility objection is a request for mathematical proof (which, as the saying goes, is something that even convinces an unreasonable person).

4β. *Relying on the ergodic hypothesis:* A second, and perhaps the most well-known, view to this problem is the one supplied by the Ehrenfests. In essence, they suggest that Boltzmann somehow relied on the ergodic hypothesis in his argument.

It is indeed evident that if the ergodic hypothesis holds, a state will spend time in the various regions of the energy hypersurface in phase space in proportion to their volume. That is to say, during the evolution of the system along its trajectory, regions with a small volume, corresponding to highly non-uniform distributions of state are visited only sporadically, and regions with larger volume, corresponding to more uniform distributions of state more often.

This should also make it plausible that if a system starts out from a very small region (an improbable state) it will display a tendency to evolve towards the overwhelmingly larger equilibrium state. Of course, this 'tendency' would have to be interpreted in a qualified sense: the same ergodic hypothesis would imply that the system cannot stay inside the equilibrium state forever and thus there would necessarily be fluctuations in and out of equilibrium. Indeed, one would have to state that the

[37] It is somewhat ironic to note, in view of remark 1 above, that this is the only case compatible with Boltzmann's argument. This gives rise to Khinchin's "methodological paradox" (cf. 101).

tendency to evolve from improbable to probable states is itself a probabilistic affair: as something that holds true for most of the initial states, or for most of the time, or as some or other form of average behaviour. In short, we would then hopefully obtain some statistical version of the H-theorem. What exactly the statistical H-theorem should say remains an open problem in the Ehrenfests' point of view. Indeed they distinguish between several interpretations (the so-called 'concentration curve' and the 'bundle of H-curves' Ehrenfest & Ehrenfest-Afanassjewa (1912, p. 31–35)).

Now, it is undeniable that the Ehrenfests' reading of Boltzmann's intentions has some clear advantages. In particular, even though nobody has yet succeeded in proving a statistical H-theorem on the basis of the ergodic hypothesis, or on the basis of the assumption of metric transitivity (cf. paragraph 6.1, one might hope that some statistical version of the H-theorem is true.

One problem here is that the assumptions Boltzmann used in his paper are restricted to non-interacting molecules, for which the ergodic hypothesis is demonstrably false. But even more importantly, it is clear that Boltzmann did not follow this line of argument in 1877b at all. Indeed, he nowhere mentions the ergodic hypothesis. In fact he later commented on the relation between the 1877b paper and the ergodic hypothesis of 1868, saying:

> On that occasion [i.e. in (1877b)] ... I did not wish to touch upon the question whether a system is capable of traversing all possible states compatible with the equation of energy (Boltzmann 1881a, Abh. II p. 572).

4γ. *Relying on the H-theorem:* A third point of view, one to which this author adhered until recently, is that, in (1877b) Boltzmann simply relied on the validity of the H-theorem of 1872. After all, it was the 1872 paper that proposed to interpret $-NH$ as entropy (modulo multiplicative and additive constants), on the basis of the alleged theorem that it could never decrease. The 1877b paper presents a new proposal, to link the entropy of a macrostate with $\ln|\Gamma_Z|$. But this proposal is motivated, if not derived, by showing that $\ln|\Gamma_Z|$ is (approximately) equal to $-NH$, as in (61), whose interpretation as entropy was established in (1872). It thus seems plausible to conjecture that Boltzmann's thinking relied on the results of that paper, and that the claim that states will evolve from improbable to probable states, i.e. that $\ln|\Gamma_Z|$ shows a tendency to increase in time, likewise relied on the H-theorem he had proved there.[38] The drawback of this reading is that it makes Boltzmann's response to the reversibility objection quite untenable. Since the objection as formulated in his (1877a) calls the validity of the H-theorem into question, a response that *presupposes* the validity of this theorem is of no help at all.

[38]The conjecture is supported by the fact Boltzmann's later exposition in (1896) is presented along this line.

4δ. *Bypassing the H-theorem:* Janssen (2002) has a different reading. He notes: "In Boltzmann's 1877 paper the statement that systems never evolve from more probable to less probable states is presented only as a new way of phrasing the Second Law, not as a consequence of the H-theorem" (p. 13). Indeed, any explicit reference to the H-theorem is absent in the 1877b paper. However, what we are to make of this is not quite certain. The earlier paper (1877a) did not mention the theorem either, but only discussed "any attempt to prove that $\int \frac{dQ}{T} \leq 0$". Still, this is commonly seen as an implicit reference to what is now known as the H-theorem, but which did not yet have a particular name at that time. Indeed, the H-theorem itself was characterized in 1872 only as a new proof that $\int \frac{dQ}{T} \leq 0$ (cf. the quotation on page 46). So, the fact that the H-theorem is not explicitly mentioned in (1877b) is not by itself a decisive argument that he did not intend to refer to it.

Even so, the fact that he presented the increase of entropy as something which was well-known and did not refer to the 1872 paper at all, does make Janssen's reading plausible. So, perhaps Boltzmann merely relied on the empirical validity of the Second Law as a ground for this statement, and not at all on any proposition from kinetic theory of gases.[39] This, of course, would undermine even more strongly the point of view that Boltzmann had a statistical version of the H-theorem, or indeed any theorem at all, about the probability of time evolution.

The reversibility objection was not about a relationship between the phenomenological Second Law and the H-theorem, but about the relationship between the H-theorem and the mechanical equations of motion. So even though Janssen's reading makes Boltzmann's views consistent, it does not make the 1877b paper provide a valid answer to Loschmidt's objection.

4ϵ. *The urn analogy—victory by definition?* At the risk of perhaps overworking the issue, I also want to suggest a fifth reading. Boltzmann's (1877b) contains an elaborate discussion of repeated drawings from an urn. In modern terms, he considers a Bernoulli process, i.e., a sequence of independent identically distributed repetitions of an experiment with a finite number of possible outcomes. To be concrete, consider an urn filled with m differently labeled lots, and a sequence of N drawings, in which the lot i is drawn n_i times ($\sum_{i=1}^{m} n_i = N$). He represents this sequence by a "distribution of state" $Z = (n_1, \ldots, n_m)$. In this discussion, the probability of these distributions of state is at first identified with the (normalized) number of permutations by which Z can be produced. In other words

$$\mathrm{Prob}(Z) \propto \frac{N!}{n_1! \cdots n_m!}. \tag{65}$$

[39] Further support for this reading can be gathered from later passages. For example, Boltzmann (1897b) writes "Experience shows that a system of interacting bodies is always found 'initially' in an improbable state and will soon reach the most probable state (that of equilibrium). (Abh.III, p. 607). Here too, Boltzmann presents the tendency to evolve from improbable to more probable states as a fact of experience rather than the consequence of any theorem.

But halfway this discussion (Abh. II, p. 171), he argues that one can redefine probabilities in an alternative fashion, namely, as the relative frequency of occurrence during *later* drawings of a sequence of N lots. Thus, even when, on a particular trial, an improbable state Z occurred, we can still argue that on a later drawings, a more probable state will occur. Boltzmann speaks about the changes in Z during the consecutive repetitions as an *evolution*. He then says:

> The most probable distribution of state must therefore be defined as that one to which most [states] will evolve to (Abh. II, p. 172).

Although he does not make the point quite explicitly, the discussion of urn drawings is undoubtedly meant as an analogy for the evolution of the distribution of state in a gas. Hence, it is not implausible that, in the latter case too, Boltzmann might have thought that *by definition* the most probable distribution of state is the one that most states will evolve to. And this, in turn, would mean that he regarded the problem about evolutions not as something to be proved, and that might depend on the validity of specific dynamical assumptions like the ergodic hypothesis or the *Stoßzahlansatz*, but as something already settled from the outset. This would certainly explain why Boltzmann did not bother to address the issue further.

Even so, this reading too has serious objections. Apart from the fact that it is not a wise idea to redefine concepts in the middle of an argument, the analogy between the evolution of an isolated gas and a Bernoulli process is shaky. In the first case, the evolution is governed by deterministic laws of motion; in the latter one simply avoids any reference to underlying dynamics by the stipulation of the probabilistic independence of repeated drawings. However, see paragraph 6.2.4.

To sum up this discussion of Boltzmann's answer to the reversibility objection: it seems that on all above readings of his two 1877 papers, the lacuna between what Boltzmann had achieved and what he needed to do to answer Loschmidt satisfactorily — i.e. to address the issue of the evolution of distributions of state and to prove that non-uniform distributions tend, in some *statistical* sense, to uniform ones, or to prove any other reformulation of the H-theorem — remains striking.

4.5 The recurrence objection

4.5.1 Poincaré

In 1890, in his famous treatise on the three-body problem of celestial mechanics, Poincaré derived what is nowadays called the recurrence theorem . Roughly speaking, the theorem says that for every mechanical system with a bounded phase space, almost every initial state of the system will, after some finite time, return to a state arbitrarily closely to this initial state, and indeed repeat this infinitely often.

In modern terms, the theorem can be formulated as follows:

> RECURRENCE THEOREM: Consider a dynamical system[40] $\langle \Gamma, \mathcal{A}, \mu, T \rangle$ with $\mu(\Gamma) < \infty$. Let $A \in \mathcal{A}$ be any measurable subset of Γ,and define, for a given time τ, the set
>
> $$B = \{x \ : \ x \in A \ \& \ \forall t \geq \tau : T_t x \notin A\} \tag{66}$$
>
> Then
>
> $$\mu(B) = 0. \tag{67}$$

In particular, for a Hamiltonian system, if we choose Γ to be the energy hypersurface Γ_E, take A to be a 'tiny' region in Γ_E, say an open ball of diameter ϵ in canonical coordinates, the theorem says that the set of points in this region whose evolution is such that they will, after some time τ, never return to region A, has measure zero. In other words, almost every trajectory starting within A will after any finite time we choose, later return to A.

Poincaré had already expressed his objections against the tenability of a mechanical explanation of irreversible phenomena in thermodynamics earlier (e.g. Poincaré 1889). But armed with his new theorem, he could make the point even stronger. In his (1893), he argued that the mechanical conception of heat is in contradiction with our experience of irreversible processes. According to the English kinetic theories, says Poincaré:

> [t]he world tends at first towards a state where it remains for a long time without apparent change; and this is consistent with experience; but it does not remain that way forever, it the theorem cited above is not violated; it merely stays there for an enormously long time, a time which is longer the more numerous are the molecules. This state will not be the final death of the universe but a sort of slumber, from which it will awake after millions and millions of centuries.
>
> According to this theory, to see heat pass from a cold body into a warm one, it will not be necessary to have the acute vision, the intelligence and the dexterity of Maxwell's demon; it will suffice to have a little patience (Brush 2003, p.380).

He concludes that these consequences contradict experience and lead to a "definite condemnation of mechanism" (Brush 2003, p.381).

Of course, Poincaré's "little patience", even for "millions and millions of centuries" is a rather optimistic understatement. Boltzmann later estimated the time needed for a recurrence in 1 cc of air to be $10^{10^{19}}$ seconds (see below): utterly beyond the bounds of experience. Poincaré's claim that the

[40] See section 6.1 for a definition of dynamical systems. But in short: Γ is a phase space, $\mathcal{A}$ a family of measurable subsets of Γ and T is a one-parameter continuous group of time evolutions $T_t : \Gamma \times \mathbb{R} \longrightarrow \Gamma$.

results of kinetic theory are contradicted by experience is thus too hasty.

Poincaré's article does not seem to have been noticed in the contemporary German-language physics community —perhaps because he criticized English theories only. However, Boltzmann was alerted to the problem when a slightly different argument was put forward by Zermelo in 1896. The foremost difference is that in Zermelo's argument experience does not play a role.

4.5.2 Zermelo's argument

Zermelo (1896a) points out that for a Hamiltonian mechanical system with a bounded phase space, Poincaré's theorem implies that, apart from a set of singular states, every state must recur almost exactly to its initial state, and indeed repeat this recurrence arbitrarily often. As a consequence, for any continuous function F on phase space, $F(x_t)$ cannot be monotonically increasing in time, (except when the initial state is singular); whenever there is a finite increase, there must also be a corresponding decrease when the initial state recurs. (see (Olsen 1993) for a modern proof of this claim) Thus, it would be impossible to obtain 'irreversible' processes. Along the way, Zermelo points out a number of options to avoid the problem.

1. Either we assume that the gas system has no bounded phase space. This could be achieved by letting the particles reach infinite distances or infinite velocities. The first option is however excluded by the assumption that a gas is contained in a finite volume. The second option could be achieved when the gas consists point particles which attract each other at small distances, (e.g. an $F \propto r^{-2}$ inter-particle attractive force can accelerate them toward arbitrarily high velocities.) However, on physical grounds one ought to assume that there is always repulsion between particles at very small distances.

2. Another possibility is to assume that the particles act upon each other by velocity-dependent forces. This, however would lead either to a violation of the conservation of energy or the law of action and reaction, both of which are essential to atomic theory.

3. The H-theorem holds only for those special initial states which are the exception to the recurrence theorem, and we assume that only those states are realized in nature. This option would be unrefutable, says Zermelo. Indeed, the reversibility objection has already shown that not all initial states can correspond to the Second Law. However, here we would have to exclude the overwhelming majority of all imaginable initial states, since the exceptions to the Recurrence Theorem only make up a set of total extension (i.e. in modern language: measure) zero. Moreover, the smallest change in the state variables would transform a singular state into a recurring state, and thus suffice to destroy the assumption. Therefore, this assumption "would be quite unique in physics and I do not believe that anyone would be satisfied with it for very long."

This leaves only two major options:

4. The Carnot-Clausius principle must be altered.[41]

5. The kinetic theory must be formulated in an essentially different way, or even be given up altogether.

Zermelo does not express any preference between these last two options. He concludes that his aim has been to explain as clearly as possible what can be proved rigorously, and hopes that this will contribute to a renewed discussion and final solution of the problem.

I would like to emphasize that, in my opinion, Zermelo's argument is entirely correct. If he can be faulted for anything, it is only that he had not noticed that in his very recent papers, Boltzmann had already been putting a different gloss on the H-theorem.

4.5.3 Boltzmann's response

Boltzmann's (1896b) response opens by stating that he had repeatedly pointed out that the theorems of gas are statistical. In particular, he says, he had often emphasized as clearly as possible that the Maxwell distribution law is not a theorem from ordinary mechanics and cannot be proven from mechanical assumptions.[42] Similarly, from the molecular viewpoint, the Second Law appears merely as a probability statement. He continues with a sarcastic remark:

> Zermelo's paper shows that my writings have been misunderstood; nevertheless it pleases me for it appears to be the first indication that these works have been noticed in Germany.[43]

Boltzmann agrees that Poincaré's recurrence theorem is "obviously correct", but claims that Zermelo's application of the theorem to gas theory is not. His counter argument is very similar to his (1895) presentation in *Nature*, a paper that Zermelo had clearly missed.

In more detail, this argument runs as follows. Consider a gas in a vessel with perfectly smooth and elastic walls, in an arbitrary initial state and let it evolve in the course of time. At each time t we can calculate $H(t)$. Further, consider a graph of this function, which Boltzmann called: *the H-curve*. In his second reply to Zermelo (Boltzmann 1897a), he actually produced a diagram. A rough an modernized version of such an H-curve is sketched in Fig. 3.

[41] By this term, Zermelo obviously referred to the Second Law, presumably including the Zeroth Law.

[42] This is, as we have seen, a point Boltzmann had been making since 1877. However, one might note that just a few years earlier, Boltzmann (1892), after giving yet another derivation of the Maxwell distribution (this time generalized to a gas of hard bodies with an arbitrary number of degrees of freedom that contribute quadratic terms to the Hamiltonian), had concluded: "I believe therefore that its correctness [i.e. of the Maxwell distribution law] as a theorem of analytical mechanics can hardly be doubted" (Abh.III p.432). But as we have seen on other occasions, for Boltzmann, statements that some result depended essentially on probability theory, and the statement that it could be derived as a mechanical theorem, need not exclude each other.

[43] Eight years earlier, Boltzmann had been offered the prestigious chair in Berlin as successor of Kirchhoff, and membership of the Prussian Academy. The complaint that his works did not draw attention in Germany is thus hard to take seriously.

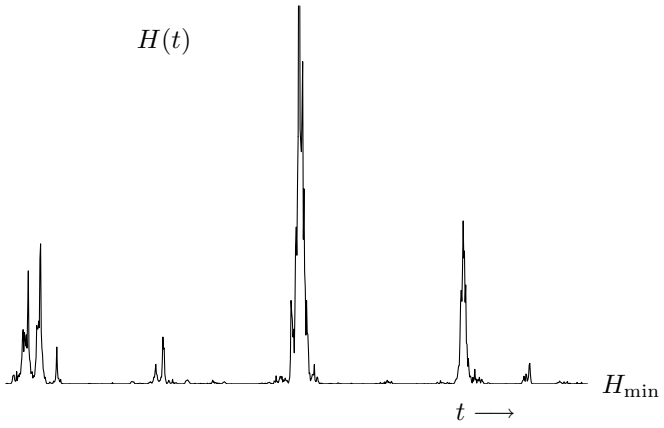

Figure 3: *A (stylized) example of an H-curve*

Barring all cases in which the motion is 'regular', e.g. when all the molecules move in one common plane, Boltzmann claims the following properties of the curve:

(i). For most of the time, $H(t)$ will be very close to its minimum value, say $H_{\min}$. Moreover, whenever the value of $H(t)$ is very close to $H_{\min}$, the distribution of molecular velocities deviates only very little from the Maxwell distribution.

(ii). The curve will occasionally, but very rarely, rise to a peak or summit, that may be well above $H_{\min}$.

(iii). The probability of a peak decreases extremely rapidly with its height.

Now suppose that, at some initial time $t = 0$, the function takes a very high value H_0, well above the minimum value. Then, Boltzmann says, it will be enormously probable that the state will, in the course of time, approach the Maxwell distribution, i.e., $H(t)$ will decrease towards $H_{\min}$; and subsequently remain there for an enormously long time, so that the state will deviate only very little from the Maxwell distribution during vanishingly short durations. Nevertheless, if one waits even longer, one will encounter a new peak, and indeed, the original state will eventually recur. In a mathematical sense, therefore, these evolutions are periodic, in full conformity with Poincaré's recurrence theorem.

What, then, is the failure of Zermelo's argument? Zermelo had claimed that only very special states have the property of continually approaching the Maxwell distribution, and that these special states taken together make up an infinitely small number compared to the totality of possible states. This is incorrect, Boltzmann says. For the overwhelming majority of states, the H-curve has the qualitative character sketched above.

Boltzmann also took issue with (what he claimed to be Zermelo's) conclusion that the mechanical

viewpoint must somehow be changed or given up. This conclusion would only be justified, he argues, if this viewpoint led to some consequence that contradicted experience. But, Boltzmann claims, the duration of the recurrence times is so large that no one will live to observe them.

To substantiate this claim about the length of the recurrence time, he presents, in an appendix an estimate of the recurrence time for 1 cc of air at room temperature and pressure. Assuming there are 10^9 molecules in this sample,[44] and choosing cells in the corresponding μ-space as six-dimensional cubes of width 10^{-9} m in (physical) space and 1 m/s in velocity space, Boltzmann calculates the number of different macrostates, i.e. the number of different ways in which the molecules can be distributed over these cells as (roughly) 10^{10^9}. He then assumes that, before a recurrence of a previous macrostate, the system has to pass through *all* other macrostates. Even if the molecules collide very often, so that the system changes its macrostate 10^{27} times per second, the total time it takes to go through this huge number of macrostates will still take $10^{10^9-27} \approx 10^{10^9}$ seconds. In fact, this time is so immensely large that its order of magnitude is not affected whether we express it in seconds, years, millennia, or what have you.

The upshot is, according to Boltzmann: if we adopt the view that heat is a form of motion of the molecules, obeying the general laws of mechanics, and assume that the initial state of a system is very unlikely, we arrive at a theorem which corresponds to the Second Law for all observed phenomena. He ends with another sarcasm:

> All the objections raised against the mechanical view of Nature are therefore empty and rest on errors. But whoever cannot overcome the difficulties, which a clear understanding of the theorems of gas theory poses, should indeed follow the advice of Mr Zermelo and decide to give up the theory completely. (Abh. III p. 576).

4.5.4 Zermelo's reply

Zermelo (1896b) notes that Boltzmann's response confirms his views by admitting that the Poincaré theorem is correct and applicable to a closed system of gas molecules. Hence, in such a system, "all [sic] motions are *periodic* and not *irreversible* in the strict sense". Thus, kinetic gas theory cannot assert that there is a strict monotonic increase of entropy as the Second Law would require. He adds: "I think this general clarification was not at all superfluous" (Brush 2003, p. 404).

Therefore, Zermelo argues, his main point had been conceded: there is indeed a conflict between thermodynamics and kinetic theory, and it remains a matter of taste which of the two is abandoned. Zermelo admits that observation of the Poincaré recurrences may well fall beyond the bounds of

[44] Actually, modern estimates put the number of molecules in 1cc of air closer to 10^{19}, which would make Boltzmann's estimate for recurrence time even larger still, i.e. $10^{10^{19}}$.

human experience. He points out (correctly) that Boltzmann's estimate of the recurrence time presupposes that the system visits *all* other cells in phase space before recurring to an initial state. This estimate is inconclusive, since the latter assumption is somewhat ad hoc. In general, these recurrence times need not "come out so 'comfortingly' large" (Brush 2003, p. 405). But, as I stressed before, the relation with experience simply was no issue in Zermelo's objection.

The main body of Zermelo's reply is taken by an analysis of the justification of and consequences drawn from Boltzmann's assumption that the initial state is very improbable, i.e., that H_0 is very high. Zermelo argues that even in order to obtain an approximate or empirical analogue of the Second Law, as Boltzmann envisaged, i.e. an approach to a long-lasting, but not permanent equilibrium state, it would not suffice to show this result for one particular initial state. Rather, one would have to show that evolutions *always* take place in the same sense, at least during observable time spans.

As Zermelo understands it, Boltzmann does not merely assume that the initial state has a very high value for H, but also that, as a rule, the initial state lies on a maximum, or has just passed a maximum. If this assumption is granted, then it is obvious that one can only observe a decreasing flank of the H-curve. However, Zermelo protests, one could have chosen any time as the initial time. In order to obtain a satisfactorily general result, the additional assumption would thus have to apply at all times. But then the H-curve would have to consist entirely of maxima. But this leads to nonsense, Zermelo argues, since the curve cannot be constant. Zermelo concludes that Boltzmann's assumptions about the initial state are thus in need of further *physical* explanation.

Further, Zermelo points out that probability theory, by itself, is neutral with respect to the direction of time, so that no preference for evolutions in a particular sense can be derived from it. He also points out that Boltzmann apparently equates the duration of a state and its extension (i.e. the relative time spent in a region and the relative volume of that region in phase space). "I cannot find that he has actually *proved* this property" (Brush 2003, p. 406).

4.5.5 Boltzmann's second reply

In his second reply (1897a), Boltzmann rebuts Zermelo's demand for a physical explanation of his assumptions about the initial state of the system with the claim that the question is not what will happen to an arbitrarily chosen initial state, but rather what will happen to a system in the present state of the universe.

He argues that one should depart from the (admittedly unprovable) assumption that the universe (or at least a very large part of the universe that surrounds us started in a very improbable state and still is in an improbable state. If one then considers a small system (e.g. a gas) that is suddenly isolated from the rest of the universe, there are the following possibilities: (i) The system may already be in equilibrium, i.e. H is close to its minimum value. This, Boltzmann says, is by far the most probable

case. But among the few cases in which the system is not in equilibrium, the most probable case is (ii) that H will be on a maximum of the H-curve, so that it will decrease in both directions of time. Even more rare is the case in which (iii) the initial value of H will fall on a decreasing flank of the H curve. But such cases are just as frequent as those in which (iv) H falls on an increasing flank.[45]

Thus, Boltzmann's explanation for the claim that H is initially on a maximum is that this would be the most likely case for a system not in equilibrium, which isolated from the rest of the universe in its present state.

This occasion is perhaps the first time that Boltzmann advanced an explanation of his claims as being due to an assumption about initial state of the system, ultimately tied to an assumption about the initial conditions of the universe. Today, this is often called the *past-hypothesis* (cf. Albert 2000, Winsberg 2004, Callender 2004, Earman 2006).

He ends his reply with the observation that while the mechanical conception of gas theory agrees with the Clausius-Carnot conception [i.e. thermodynamics] in all observable phenomena, a virtue of the mechanical view is that it might eventually predict new phenomena, in particular for the motion of small bodies suspended in fluids. These prophetic words were substantiated eight years later in Einstein's work on Brownian motion.

However, he does not respond to Zermelo's requests for more definite proofs of the claims (1) –(3), or of the equality of phase space volume and time averages in particular. He bluntly states that he has thirty years of priority in measuring probabilities by means of phase space volume (which is true) and adds that he has always had done so (which is false). Even so, one cannot interpret this claim of Boltzmann as a rejection of the time average conception of probability. A few lines below, he claims that the most probable states will also occur most frequently, except for a vanishingly small number of initial states. He does not enter into a proof of this. Once again, this provides an instance where the Ehrenfests conjectured that Boltzmann might have had the ergodic hypothesis in the back of his mind.

4.5.6 Remarks

Boltzmann's replies to Zermelo have been recommended as "superbly clear and right on the money" (Lebowitz 1999, p. S347). However, as will clear from the above and the following remarks, I do not share this view. See also (Klein 1973, Curd 1982, Batterman 1990, Cercignani 1998, Brush 1999, Earman 2006) for other commentaries on the Zermelo-Boltzmann dispute.

[45] the Ehrenfests (1912) later added a final possible case (v): H may initially be on a local minimum of the H-curve, so that it increases in both directions of time. But by a similar reasoning, that case is even less probable than the cases mentioned by Boltzmann.

1. The issues at stake It is clear that, in at least one main point of the dispute, Boltzmann and Zermelo had been talking past each other. When Zermelo argued that in the kinetic theory of gases there can be no continual approach towards a final stationary state, he obviously meant this in the sense of a limit $t \longrightarrow \infty$. But Boltzmann's reply indicates that he took the "approach" as something that is not certain but only probable, and as lasting for a very long, but finite time. His graph of the H-curve makes abundantly clear that $\lim_{t \longrightarrow \infty}$H(t) does not exist.

It is true that his statistical reading of the H-theorem, as laid down in the claims (1)–(3) above, was already explicit in (Boltzmann 1895), and thus Boltzmann could claim with some justification that his work had been overlooked. But in fairness, one must note that, even in this period, Boltzmann was sending mixed messages to his readers. Indeed, the first volume of Boltzmann's *Lectures on Gas Theory*, published in 1896, stressed, much like his original (1872) paper on the H-theorem, the necessity and exceptionless generality of the H-theorem, adding only that the theorem depended on the assumption of molecular disorder (as he then called the *Stoßzahlansatz*):[46] "

> [T]he quantity designated as H can only decrease; at most it can remain constant.[...] The only assumption we have made here is that the distribution of velocities was initially 'molecularly disordered' and remains disordered. Under this condition we have therefore proved that the quantity called H can only decrease and that the distribution of velocities must necessarily approach the Maxwell distribution ever more closely (Boltzmann 1896, § 5, p. 38).

Zermelo might not have been alone in presuming that Boltzmann had intended this last claim literally, and was at least equally justified in pointing out that Boltzmann's clarification "was not at all superfluous".

On the other hand, Boltzmann misrepresented Zermelo's argument as concluding that the mechanical view should be given up. As we have seen, Zermelo only argued for a *dilemma* between the strict validity of the kinetic theory and the strict validity of thermodynamics. Empirical matters were not relevant to Zermelo's analysis. Still, Boltzmann is obviously correct when he says that the objection does not yet unearth a conflict with experience. Thus, his response would have been more successful as a counter-argument to Poincaré than to Zermelo.

2. The statistical reading of the H-theorem. Another point concerns the set of claims (1)–(3) that Boltzmann lays down for the behaviour of the H-curve. Together, they form perhaps the

[46] in his reply to Zermelo, Boltzmann claimed that his discussion of the H-theorem in the *Lectures on Gas theory* was intended under the explicitly emphasized assumption that the number of molecules was infinite, so that the recurrence theorem did not apply. However, I can find no mention of such an assumption in this context. On the contrary, the first occasion on which this latter assumption appears is in §6 on page 46 where it is introduced as "an assumption we shall make later", suggesting that the previous discussion did *not* depend on in it.

most clearly stated and explicit form of the "statistical reading of the H-theorem" (cf remark 3 on page 53). Yet they only have a loose connection to the original theorem. It is unclear, for example, whether these claims still depend on the *Stoßzahlansatz*, the assumption that the gas is dilute, etc. It thus remains a reasonable question what argument we have for their validity. Boltzmann offers none. In his 1895 paper in *Nature*, he argued as if he had proved as much in his earlier papers, and added tersely: "I will not here repeat the proofs given in my papers"(Abh. III p. 541). But surely, Boltzmann never proved anything concerning the probability of the time evolution of H, and at this point there remains a gap in his theory. Of course, one might speculate on ways to bridge this gap; e.g. that Boltzmann implicitly and silently relied on the ergodic hypothesis, as the Ehrenfests suggested or in other ways, but I refrain from discussing this further. The most successful modern attempt so far to formulate and prove a statistical H-theorem has been provided by Lanford, see paragraph 6.4 below.

5 Gibbs' Statistical Mechanics

The birth of statistical mechanics in a strict sense, i.e. as a coherent and systematic theory, is marked by the appearance of J.W. Gibbs's book (1902) which carries this title: *Elementary Principles in Statistical Mechanics; developed with especial reference to the rational foundation of thermodynamics*. His point of departure is a general mechanical system governed by Hamiltonian equations of motion, whose (micro)states are represented by points in the mechanical phase space Γ.

Gibbs avoids specific hypotheses about the microscopic constitution of such a system. He refers to the well-known problem concerning the anomalous values of the specific heat for gases consisting of diatomic molecules (mentioned in footnote 10), and remarks:

> Difficulties of this kind have deterred the author from attempting to explain the mysteries of nature, and have forced him to be contented with the more modest aim of deducing some of the more obvious propositions relating to the statistical branch of mechanics (Gibbs 1902, p. viii).

It is clear from this quote that Gibbs' main concern was with logical coherence, and less with the molecular constitution. (Indeed, only the very last chapter of the book is devoted to systems composed of molecules.) This sets his approach apart from Maxwell and Boltzmann.[47]

The only two ingredients in Gibbs' logical scheme are mechanics and probability. Probability is introduced here as an ingredient not reducible to the mechanical state of an individual system, but by means of the now familiar "ensemble":

[47] It also sets him apart from the approach of Einstein who, in a series of papers (1902, 1903, 1904) independently developed a formalism closely related to that of Gibbs, but used it as a probe to obtain empirical tests for the molecular/atomic hypothesis (cf. Gearhart 1990, Navarro 1998, Uffink 2006).

> We may imagine a great number of systems of the same nature, but differing in the configurations and velocities which they have at a given instant, and differing not merely infinitesimally, but it may be so as to embrace every conceivable combination of configuration and velocities. And here we may set the problem, not to follow a particular system through its succession of configurations, but to determine how the whole number of systems will be distributed among the various conceivable configurations and velocities at any required time, when the distribution has been given for some one time (Gibbs 1902, p. v).

and

> What we know about a body can generally be described most accurately and most simply by saying that it is one taken at random from a great number (ensemble) of bodies which are completely described. (p. 163)

Note that Gibbs is somewhat non-committal about any particular interpretation of probability. (Of course, most of the presently distinguished interpretations of probability were only elaborated since the 1920s, and we cannot suppose Gibbs to have pre-knowledge of those distinctions.) A modern frequentist (for whom a probability of an event is the frequency with which that event occurs in a long sequence of similar cases) will have no difficulty with Gibbs' reference to an ensemble, and will presumably identify that notion with von Mises' notion of a *Kollektiv*. On the other hand, authors like Jaynes who favour a subjectivist interpretation of probability (in which the probability of an event is understood as a state of knowledge or belief about that event) have emphasized that in Gibbs' approach the ensemble is merely 'imagined' and a tool for representing our knowledge.

The ensemble is usually presented in the form of a probability density function ρ over Γ, such that $\int_A \rho(x)dx$ is the relative number of systems in the ensemble whose microstate $x = (\vec{q}_1, \vec{p_1}; \ldots; \vec{q}_N, \vec{p}_N)$ lies in the region A. The evolution of an ensemble density ρ_0 at time $t = 0$ is dictated by the Hamiltonian equations of motion. In terms of the (formal) time evolution operator T_t, we get

$$\rho_t(x) = \rho_0(T_{-t}x) \tag{68}$$

or, in differential form:

$$\frac{\partial \rho_t(x)}{\partial t} = \{H, \rho\} \tag{69}$$

where $\{\cdot, \cdot\}$ denotes the Poisson bracket:

$$\{H, \rho\} = \sum_{i=1}^{N} \frac{\partial H}{\partial \vec{q}_i} \frac{\partial \rho}{\partial \vec{p}_i} - \frac{\partial H}{\partial \vec{p}_i} \frac{\partial \rho}{\partial \vec{q}_i} \tag{70}$$

A case of special interest is that in which the density function is stationary, i.e.

$$\forall t: \quad \frac{\partial \rho_t(x)}{\partial t} = 0. \tag{71}$$

This is what Gibbs calls the condition of *statistical equilibrium*. Gibbs notes that any density which can be written as a function of the Hamiltonian is stationary, and proceeds to distinguish special cases, of which the most important are:

$$\rho_E(x) = \frac{1}{\omega(E)}\delta(H(x) - E) \quad \text{(microcanonical)} \tag{72}$$

$$\rho_\theta(x) = \frac{1}{Z(\theta)}\exp(-H(x)/\theta) \quad \text{(canonical)} \tag{73}$$

$$\rho_{\theta,\alpha}(x, N) = \frac{1}{N!Z(\theta,\alpha)}\exp(-H(x)/\theta + \alpha N) \quad \text{(grand-canonical)} \tag{74}$$

where $\omega(E)$, $Z(\theta)$ and $Z(\theta, \alpha)$ are normalization factors. In the following I will mainly discuss the canonical and microcanonical ensembles.

5.1 Thermodynamic analogies for statistical equilibrium

As indicated by the subtitle of the book, Gibbs' main goal was to provide a 'rational foundation' for thermodynamics. He approaches this issue quite cautiously, by pointing out certain analogies between relations holding for the canonical and microcanonical ensembles and results of thermodynamics. At no point does Gibbs claim to have reduced thermodynamics to statistical mechanics.

The very first analogy noticed by Gibbs is in the case of two systems, A and B put into thermal contact. This is modeled in statistical mechanics by taking the product phase space, $\Gamma_{AB} = \Gamma_A \times \Gamma_B$, and a Hamiltonian $H_{AB} = H_A + H_B + H_{\rm int}$. If both A and B are described by canonical ensembles and if $H_{\rm int}$ is 'infinitely small' compared to the system Hamiltonian, then the combined system will be in statistical equilibrium if $\theta_A = \theta_B$. This, he says, "is entirely analogous to ... the corresponding case in thermodynamics" where "the most simple test of the equality of temperature of two bodies is that they remain in thermal equilibrium when brought into thermal contact" (ibid. p. 37). Clearly, Gibbs invites us to think of statistical equilibrium as analogous to thermal equilibrium, and θ as the analogue of the temperature of the system.[48]

A second point of analogy is in reproducing the 'fundamental equation' (23) of thermodynamics:

$$dU = TdS + \sum_i F_i da_i \tag{75}$$

[48] A more elaborate discussion of the properties of the parameter θ and their analogies to temperature, is in Einstein (1902). That discussion also addresses the transitivity of thermal equilibrium, i.e. the Zeroth Law of thermodynamics (cf. paragraph 2).

where a_i are the so-called external parameters (e.g. volume) and F_i the associated generalized forces (e.g. minus the pressure). For the canonical ensemble, Gibbs derives a relation formally similar to the above fundamental equation:[49]

$$d\langle H\rangle = \theta d\sigma - \sum_i \langle A_i\rangle da_i. \tag{76}$$

Here, $\langle H\rangle$ is the expectation value of the Hamiltonian in the canonical ensemble, θ the modulus of the ensemble, σ the so-called Gibbs entropy of the canonical distribution:

$$\sigma[\rho_\theta] = -\int \rho_\theta(x)\ln\rho_\theta(x)dx, \tag{77}$$

a_i are parameters in the form of the Hamiltonian and the $\langle A_i\rangle = \langle\frac{\partial H}{\partial a_i}\rangle$ represent the 'generalized forces'.[50] The equation suggests that the canonical ensemble averages might serve as analogues of the corresponding thermodynamic quantities, and θ and σ as analogues of respectively temperature and entropy.[51]

Note the peculiarly different role of θ and σ in (76): these are not expectations of phase space functions, but a parameter and a functional of the ensemble density ρ_θ. This has a significant conceptual implication. The former quantities may be thought of as averages, taken over the ensemble of some property possessed by each individual system in the ensemble. But for temperature θ and entropy σ, this is not so. In the case of θ one can diminish this contrast— at least when H is the sum of a kinetic and a potential energy term and the kinetic part is quadratic in the momenta, i.e. $H = \sum_i \alpha_i p_i^2 + U(q_1,\ldots q_n)$—because of the well-known equipartition theorem. This theorem says that θ equals twice the expected kinetic energy for each degree of freedom:

$$\frac{\theta}{2} = \alpha_i\langle p_i^2\rangle_\theta. \tag{78}$$

Thus, in this case, one can find phase functions whose canonical expectation values are equal to θ, and regard the value of such a function as corresponding to the temperature of an individual system.[52]

[49] See (Uhlenbeck and Ford 1963, van Lith 2001b) for details.

[50] A more delicate argument is needed if one wishes to verify that $-\langle\frac{\partial H}{\partial V}\rangle$ can really be identified with pressure, i.e. the average force per unit area on the walls of the container. Such an argument is given by Martin-Löf (1979, p. 21–25)

[51] A crucial assumption in this derivation is that the differential expressions represent infinitesimal elements of quasistatic processes during which the probability density always retains its canonical shape. This assumption is in conflict with a dynamical evolution (van Lith 2001b, p. 141).

[52] For proposals of more generally defined phase functions that can serve as an analogy of temperature, see (Rugh 1997, Jepps et al. 2000).

But *no* function χ on phase space exists such that

$$\sigma[\rho_\theta] = \langle \chi \rangle_\theta \quad \text{for all } \theta. \tag{79}$$

Thus, the Gibbs entropy cannot be interpreted as an average of some property of the individual members of the ensemble.

The next question is whether a differential equation similar to (76) can be obtained also for the microcanonical ensemble. In this case, it is natural to consider the same expressions $\langle A_i \rangle$ and $\langle H \rangle$ as above, but now taken as expectations with respect to the microcanonical ensemble, so that obviously $\langle H \rangle_{\rm mc} = E$. The problem is then to find the microcanonical analogies to T and S. Gibbs (1902, p. 124–128, 169–171) proposes the following:

$$T \quad \longleftrightarrow \quad \left(\frac{d \ln \Omega(E)}{dE} \right)^{-1}, \tag{80}$$

$$S \quad \longleftrightarrow \quad \ln \Omega(E), \tag{81}$$

where

$$\Omega(E) \quad := \quad \int_{H(x) \leq E} dp_1 \ldots dq_n \tag{82}$$

is known as the integrated structure function.

Remarkably, in a later passage, Gibbs (1902, p. 172–178) also provides a second pair of analogies to temperature and entropy, namely:

$$T \quad \longleftrightarrow \quad \left(\frac{d \ln \omega(E)}{dE} \right)^{-1} \tag{83}$$

$$S \quad \longleftrightarrow \quad \ln \omega(E), \tag{84}$$

where ω is the structure function

$$\omega(E) = \frac{d\Omega(E)}{dE} = \int_{H(x)=E} dx.$$

For this choice, the relation (75) is again reproduced. Thus, there appears to be a variety of choice for statistical mechanical quantities that may serve as thermodynamic analogue. Although Gibbs discussed various pro's and con's of the two sets, —depending on such issues as whether we regard the energy or the temperature as an independent variable, and whether we prefer expected values of most probable values— he does not reach a clear preference for one of them. (As he put it, system (80,81) is the more natural, while system (83,84) is the simpler of the two.) Still, Gibbs argued (ibid.,

p. 183) that the two sets of analogies will approximately coincide for a very large number degrees of freedom. Nevertheless, this means there remains an underdetermination in his approach that one can hope to avoid only in the thermodynamic limit.

The expressions (81) and (84) are also known as the 'volume entropy' and the 'surface entropy'. In modern textbooks the latter choice has been by far the most popular, perhaps because it coincides with the Gibbs entropy for the microcanonical ensemble: $\sigma[\rho_E] = \ln \omega(E)$. However, it has been pointed out that there are also general theoretical reasons to prefer the volume entropy (81), in particular because it is, unlike the surface entropy, an adiabatic invariant (see Hertz 1910, Rugh 2001, Campisi 2005).

Of course, all of this is restricted to (statistical) equilibrium. In the case of non-equilibrium, one would obviously like to obtain further thermodynamical analogies that recover the approach to equilibrium (the 'Minus First Law', cf. p. 20) and an increase in entropy for adiabatic processes that start and end in equilibrium, or even to reproduce the kinetic equations on a full statistical mechanical basis. What Gibbs had to say on such issues will be the subject of the paragraphs 5.3 and 5.4.

But Gibbs also noted that a comparison of temperature and entropy with their analogies in statistical mechanics "would not be complete without a consideration of their differences with respect to units and zeros and the numbers used for their numerical specification" (Gibbs 1902, p.183). This will be taken up below in §5.2.

5.2 Units, zeros and the factor N!

The various expressions Gibbs proposed as analogies for entropy, i.e. (77,81,84), were presented without any discussion of 'units and zeros', i.e. of their physical dimension and the constants that may be added to these expressions. This was only natural because Gibbs singled out those expressions for their formal merit of reproducing the fundamental equation, in which only the combination TdS appears. He discussed the question of the physical dimension of entropy by noting that the fundamental equation remains invariant if we multiply the analogue for temperature —i.e. the parameter θ in the canonical case, or the functions (80 or (83) for the microcanonical case— by some constant K and the corresponding analogues for entropy — (77), (81) and (84)— by $1/K$. Applied to the simple case of the monatomic ideal gas of N molecules, he concluded that, in order to equate the analogues of temperature to the ideal gas temperature, $1/K$ should be set equal to

$$\frac{1}{K} = \frac{2}{3}\frac{c_V}{N}, \tag{85}$$

where c_V is the specific heat at constant volume. He notes that "this value had been recognized by physicists as a constant independent of the kind of monatomic gas considered" (Gibbs 1902, p. 185).

Indeed, in modern notation, $1/K = k$, i.e. Boltzmann's constant.

Concerning the question of 'zeros', Gibbs noted that all the expressions proposed as analogy of entropy had the dimension of the logarithm of phase space volume and are thus affected by the choice of our units for length mass and time in the form of some additional constant (cf. Gibbs 1902, p. 19,183). But even if some choice for such units is fixed, further constants could be added to the statistical analogs of entropy, i.e. arbitrary expressions that may depend on anything not varied in the fundamental equation. However, their values would disappear when differences of entropy are compared. And since only entropy differences have physical meaning, a question of determining these constants would thus appear to be immaterial. However, Gibbs went on to argue that "the principle that the entropy of any body has an arbitrary additive constant is subject to limitations when different quantities of the same substance are compared"(Gibbs 1902, p. 206). He formulated further conditions on how the additive constant may depend on the number N of particles in his final chapter.

Gibbs starts this investigation by raising the following problem. Consider the phase (i.e. microstate) $(\vec{q}_1, \vec{p_1}; \ldots; \vec{q}_N, \vec{p}_N)$ of an N-particle system where the particles are said to be "indistinguishable", "entirely similar" or "perfectly similar".[53] Now, if we perform a permutation on the particles of such a system, should we regard the result as a different phase or not? Gibbs first argues that it "seems in accordance with the spirit of the statistical method" to regard such phases as the same. It might be urged, he says, that for such particles no identity is possible except that of qualities, and when comparing the permuted and unpermuted system, "nothing remains on which to base the identification of any particular particle of the first system with any particular particle of the second" (Gibbs 1902, p. 187).

However, he immediately rejects this argument, stating that all this would be true for systems with "simultaneous objective existence", but hardly applies to the "creations of the imagination". On the contrary, Gibbs argues:

> "The perfect similarity of several particles of a system will not in the least interfere with the identification of a particular particle in one case and with a particular particle in another. The question is one to be decided in accordance with the requirements of practical convenience in the discussion of the problems with which we are engaged" (Gibbs 1902, p. 188)

He continues therefore by exploring both options, calling the viewpoint in which permuted phases are regarded as identical the *generic* phase, and that in which they are seen as distinct the *specific* phase. In modern terms the generic phase space is obtained as the quotient space of the specific phase space

[53] Presumably, these terms mean (at least) that the Hamiltonian is invariant under their permutation, i.e. they have equal mass and interact in exactly the same way.

obtained by identifying all phase points that differ by a permutation (see Leinaas & Myrheim 1977). In general, there are $N!$ different permutations on the phase of a system of N particles,[54] and there are thus $N!$ different specific phases corresponding to one generic phase. This reduces the generic phase space measure by an overall factor of $\frac{1}{N!}$ in comparison to the specific phase space. Since the analogies to entropy all have a dimension equal to the logarithm of phase space measure, this factor shows up as an further additive constant to the entropy, namely $-\ln N!$ in comparison to an entropy calculated from the specific phase. Gibbs concludes that when N is constant, "it is therefore immaterial whether we use [the generic entropy] or [the specific entropy], since this only affects the arbitrary constant of integration which is added to the entropy (Gibbs 1902, p. 206).[55]

However, Gibbs points out that this is *not* the case if we compare the entropies of systems with different number of particles. For example, consider two identical gases, each with the same energy U, volume V and number of particles N, in contiguous containers, and let the entropy of each gas be written as $S(U, V, N)$. Gibbs puts the entropy of the total system equal to the sum of the entropies:

$$S_{\mathrm{tot}} = 2S(U, V, N). \tag{86}$$

Now suppose a valve is opened, making a connection between the two containers. Gibbs says that "we do not regard this as making any change in the entropy, although the gases diffuse into one another, and this process would increase the entropy if the gases were different" (Gibbs 1902, p. 206-7). Therefore, the entropy in this new situation is

$$S'_{\mathrm{tot}} = S_{\mathrm{tot}}. \tag{87}$$

But the new system, is a gas with energy $2U$, volume $2V$, and particle number $2N$. Therefore, we obtain:

$$S'_{\mathrm{tot}} = S(2U, 2V, 2N) = 2S(U, V, N), \tag{88}$$

where the right-hand side equation expresses the *extensivity* of entropy. This condition is satisfied (at least for large N) by the generic entropy but not by the specific entropy. Gibbs concludes "it is evident therefore that it is equilibrium with respect to generic phases, and not that with respect to specific, with which we have to do in the evaluation of entropy, ... except in the thermodynamics of bodies in which the number of molecules of the various kinds is constant" (Gibbs 1902, p. 207).

The issue expressed in these final pages is perhaps the most controversial in Gibbs' book; at least it has generated much further discussion. Many later authors have argued that the insertion of a factor

[54] This assumes that the molecular states $\vec{p}_i, \vec{q}_i)$ of the particles do not coincide. However the points in specific phase space for which one or more molecular states do coincide constitute a set of Lebesgue measure zero anyway.

[55] The same conclusion also obtains for the Boltzmann entropy (61) (Huggett 1999).

$1/N!$ in the phase space measure is obligatory to obtain "correct" results and, ultimately due to a lack of any metaphysical identity or "haecceity" of the perfectly similar particles considered. Some have even gone on to argue that quantum mechanics is needed to explain this. For example, Huang (1987, p. 154) writes "It is not possible to understand classically why we must divide [...] by $N!$ to obtain the correct counting of states. The reason is inherently quantum mechanical ...". However, many others deny this (Becker 1967, van Kampen 1984, Ray 1984). It would take me too far afield to discuss the various views and widespread confusion on this issue.

Let it suffice to note that Gibbs rejected arguments from the metaphysics of identity for the creations of the imagination. (I presume this may be taken to express that the phases of an N-particles system are theoretical constructs, rather than material objects.) Further, Gibbs did not claim that the generic view was correct and the specific view of incorrect; he preferred to settle the question by "practical convenience". There are indeed several aspects of his argument that rely on assumptions that may be argued to be conventional. for example the 'additivity' demand (86) could be expanded to read more fully:

$$S_{\rm tot}(U_1, V_1, N_1; U_2, V_2, N_2) + K_{\rm tot} = S_1(U_1, V_1, N_1) + K_1 + S_2(U_2, V_2, N_2) + K_2, \qquad (89)$$

Applied to the special case where S_1 and S_2 are identical functions taken at the same values of their arguments. The point to note here is that this relation only leads to (86) if we also employ the conventions $K_{\rm tot} = K_1 + K_2$ and $K_1 = K_2$. Also, his cautious choice of words concerning (87) —"we do not regard this as making any change"— suggest that he wants to leave open whether this equation expresses a fact or a conventional choice on our part. But by and large, it seems fair to say that Gibbs' criterion for practical convenience is simply the recovery of the properties usually assumed to hold for thermodynamic entropy.

As a final remark, note that the contrast mentioned here in passing by Gibbs, i.e. that in thermodynamics the mixing of identical gases, by allowing them to diffuse into one another, does not change the entropy, whereas this process does increase entropy if the gases are different, implicitly refers to an earlier discussion of this issue in his 1875 paper (Gibbs 1906, pp. 165–167). The contrast between the entropy of mixing of identical fluids and that of different fluids noted on that occasion is now commonly known as the *Gibbs paradox*. (More precisely, this 'paradox' is that the entropy of mixing different fluids is a constant ($kT \ln 2$ in the above case) as long as the substances are different, and vanishes abruptly when they are perfectly similar; thus negating the intuitive expectation one might have had that the entropy of mixing should diminish gradually when the substances become more and more alike). Now note that in the the specific view, mixing different substances and mixing identical substances both lead to an entropy increase: in that view there is no Gibbs paradox, since there is no abrupt change when the substances become more and more alike. On the other hand,

the adoption of the generic view, i.e. the division of the phase space measure by $N!$, is used by Gibbs to recover the usual properties of thermodynamic entropy *including* the Gibbs paradox — the discontinuity between mixing of different and identical gases.

Still, many authors seem to believe that the division by $N!$ is a procedure that *solves* the Gibbs paradox. But this is clearly not the case; instead, it is the specific viewpoint that avoids the paradox, while the generic viewpoint recovers the Gibbs paradox for the statistical mechanical analogies to entropy. The irony of it all is that, in statistical mechanics, the term "Gibbs paradox" is sometimes used to mean or imply the *absence* of the original Gibbs paradox in the specific point of view, so that a resolution of *this* "Gibbs paradox" requires the return of the original paradox.

5.3 Gibbs on the increase of entropy

As we have seen, the Gibbs entropy may be defined as a functional on arbitrary probability density functions ρ on phase space Γ:[56]

$$\sigma[\rho] = -\int \rho(x) \ln \rho(x) dx \tag{90}$$

This expression has many well-known and useful properties. For example, under all probability densities restricted to the energy hypersurface $H(x) = E$, the microcanonical density (72) has the highest entropy. Similarly, one can show that of all distributions ρ with a given expectation value $\langle H \rangle_\rho$, the canonical distribution (73) has the highest entropy, and that of all distributions for which both $\langle H \rangle$ and $\langle N \rangle$ are given, the grand-canonical ensemble has the highest entropy.

But suppose that ρ is not stationary. It will therefore evolve in the course of time, as given by $\rho_t(x) = \rho(T_{-t}x)$. One might ask whether this entropy will increase in the course of time. However, Liouville's theorem implies immediately

$$\sigma[\rho_t] = \sigma[\rho_0]. \tag{91}$$

In spite of the superficial similarity to Boltzmann's H, the Gibbs entropy thus remains constant in time. The explanation of the Second Law, or an approach to equilibrium, cannot be so simple.

However, Gibbs warns us to proceed with great caution. Liouville's theorem can be interpreted as stating that the motion of ρ_t can be likened to motion in a (multidimensional) incompressible fluid. He thus compared the evolution of ρ to that of the stirring of a dye in a incompressible medium

[56] Gibbs actually does not use the term entropy for this expression. He calls the function $\ln \rho$ the "index of probability", and $-\sigma$ "the average index of probability". As we have seen, Gibbs proposed more than one candidate for entropy in the microcanonical ensemble, and was well aware that: " [t]here may be [...], and there are, other expressions that may be thought to have some claim to be regarded as the [...] entropy with respect to systems of a finite number of degrees of freedom" (Gibbs 1902, p. 169).

(Gibbs 1902, p. 143-151). In this case too, the average density of the dye, as well as the average of any function of its density, does not change. Still, it is a familiar fact of experience that by stirring tends to bring about a uniform mixture, or a state with uniform density, for which the expression $-\int \rho \ln \rho \, dx$ would have increased to attain its maximum value.

Gibbs saw the resolution of this contradiction in the definition of the notion of density. This, of course, is commonly taken as the limit of the quantity of dye in a spatial volume element, when the latter goes to zero. If we apply this definition, i.e. take this limit first, and then consider the stirring motion, we will arrive at the conclusion that $-\int \rho \ln \rho dx$ remains constant. But if we consider the density defined for a fixed finite (non-zero) volume element, and then stir for an indefinitely long time, the density may become 'sensibly' uniform, a result which is not affected if we subsequently let the volume elements become vanishingly small. The problem, as Gibbs saw it, is therefore one of the order in which we proceed to take two limits.

Gibbs was aware that not all motions in phase space produce this tendency toward statistical equilibrium, just as not every motion in an incompressible fluid stirs a dy to a sensibly homogeneous mixture. Nevertheless, as he concluded tentatively,: "We might perhaps fairly infer from such considerations as have been adduced that an approach to a limiting condition of statistical equilibrium is the general rule, when the initial condition is not of that character" (Gibbs 1902, p. 148).

5.4 Coarse graining

The most common modern elaboration of Gibbs' ideas is by taking recourse to a partitioning of phase space in cells, usually called "*coarse graining*. Instead of studying the original distribution function $\rho(x)$ we replace $\rho(x)dx$ by its phase average over each cell, by the mapping:

$$\mathcal{CG} : \rho(x) \mapsto \mathcal{CG}\rho(x) = \sum_i \hat{\rho}(i) \mathbb{1}_{\omega_i}(x), \tag{92}$$

where

$$\hat{\rho}(i) := \frac{\int_{\omega_i} \rho(x) dx}{\int_{\omega_i} dx}, \tag{93}$$

and $\mathbb{1}$ denotes the characteristic function:

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{elsewhere.} \end{cases} \tag{94}$$

The usual idea is that such a partition matches the finite precision of our observational capabilities, so that a "coarse grained" distribution might be taken as a sufficient description of what is observable. Obviously, the average value of any function on Γ that does not vary too much within cells is

approximately the same, whether we use the fine-grained or the coarse-grained distribution.

For any ρ one can also define the coarse grained entropy $\Sigma[\rho]$ as the composition of (92) and (90):

$$\Sigma[\rho] := \sigma[\mathcal{C}\mathcal{G}\rho]. \tag{95}$$

This coarse grained entropy need not be conserved in time. Indeed, it is easy to show (cf. Tolman 1938, p. 172) that:

$$\Sigma[\rho] \geq \sigma[\rho]. \tag{96}$$

Hence, if we assume that at some initial time that $\rho_0 = \mathcal{C}\mathcal{G}\rho_0$, e.g. if $\rho_0 \propto \frac{1}{V_i}\mathbb{1}_{\omega_i}$ for some cell i, then for all t:

$$\Sigma[\rho_t] \geq \sigma[\rho_t] = \sigma[\rho_0] = \Sigma[\rho_0]. \tag{97}$$

However, this does not imply that $\Sigma[\rho_t]$ is non-decreasing or that it approaches a limiting value as $t \longrightarrow \infty$.

If a property, similar to the stirring of a dye holds for the dynamical evolution of ρ_t, one may have

$$\lim_{t\longrightarrow\infty} \Sigma[\rho_t] = \Sigma[\rho_{\mathrm{mc}}] \tag{98}$$

and hence, an approach towards equilibrium could emerge on the coarse-grained level. This convergence will of course demand a non-trivial assumption about the dynamics. In modern work this assumption is that the system has the *mixing* property (see paragraph 6.1).

5.5 Comments

Gibbs' statistical mechanics has produced a formalism with clearly delineated concepts and methods, using only Hamiltonian mechanics and probability theory. It can and is routinely used to calculate equilibrium properties of gases and other systems by introducing a specific form of the Hamiltonian. The main problems that Gibbs has left open are, first, the motivation for the special choice of the equilibrium ensembles and, second, that the quantities serving as thermodynamic analogies are not uniquely defined. However, much careful work has been devoted to show that, under certain assumptions about tempered interaction of molecules, unique thermodynamic state functions, with their desired properties are obtained in the 'thermodynamic limit' (cf. §6.3.1).

1. Motivating the choice of ensemble. While Gibbs had not much more to offer in recommendation of these three ensembles than their simplicity as candidates for representation for equilibrium, modern views often provide an additional story. First, the microcanonical ensemble is particularly singled out for describing an ensemble of systems in thermal isolation with a fixed energy E.

Arguments for this purpose come in different kinds. As argued by Boltzmann (1868), and shown more clearly by Einstein (1902), the microcanonical ensemble is the *unique* stationary density for an isolated ensemble of systems with fixed energy, if one assumes the ergodic hypothesis. Unfortunately, for this argument, the ergodic hypothesis is false for any system that has a phase space of dimension 2 or higher (cf. paragraph 6.1).

A related but more promising argument relies on the theorem that the measure P_{mc} associated with the microcanonical ensemble via $P_{\mathrm{mc}}(A) = \int_A \rho_{\mathrm{mc}}(x)dx$ is the unique stationary measure under all measures that are absolutely continuous with respect to P_{mc}, if one assumes that the system is metrically transitive (again, see paragraph 6.1).

This argument is applicable for more general systems, but its conclusion is weaker. In particular, one would now have to argue that physically interesting systems are indeed metrically transitive, and why measures that are not absolutely continuous with respect to the microcanonical one are somehow to be disregarded. The first problem is still an open question, even for the hard-spheres model (as we shall see in paragraph 6.1.1). The second question can be answered in a variety of ways.

For example, Penrose (1979, p. 1941) adopts a principle that every ensemble should be representable by a (piecewise) continuous density function, in order to rule out "physically unreasonable cases". (This postulate implies absolute continuity of the ensemble measure with respect to the microcanonical measure by virtue of the Radon-Nikodym theorem.) See Kurth (1960, p. 78) for a similar postulate. Another argument, proposed by Malament & Zabell (1980), assumes that the measure P associated with a physically meaningful ensemble should have a property called 'translation continuity. Roughly, this notion means that the probability assigned to any measurable set should be a continuous function under small displacements of that set within the energy hypersurface. Malament & Zabell show that this property is equivalent to absolute continuity of P with respect to μ_{mc}, and thus singles out the microcanonical measure uniquely if the system is metrically transitive (see van Lith 2001b, for a more extensive discussion).

A third approach, due to Tolman and Jaynes, more or less postulates the microcanonical density, as a appropriate description of our knowledge about the microstate of a system with given energy (regardless of whether the system is metrically transitive or not).

Once the microcanonical ensemble is in place as a privileged description of an isolated system with a fixed energy, one can motivate the corresponding status for the other ensembles with relatively less effort. The canonical distribution is shown to provide the description of a small system S_1 in weak energetic contact with a larger system S_2, acting as a 'heat bath' (see Gibbs 1902, p. 180–183). Here, it is assumed that the total system is isolated and described by a microcanonical ensemble, where the total system has a Hamiltonian $H_{\mathrm{tot}} = H_1 + H_2 + H_{\mathrm{int}}$ with $H_2 \gg H_1 \gg H_{\mathrm{int}}$. More elaborate versions of such an argument are given by Einstein (1902) and Martin-Löf (1979).

Similarly, the grand-canonical ensemble can be derived for a small system that can exchange both energy and particles with a large system. (see van Kampen 1984).

2. The 'equivalence' of ensembles. It is often argued in physics textbooks that the choice between these different ensembles (say the canonical and microcanonical) is deprived of practical relevance by a claim that they are all "equivalent". (See (Lorentz 1916, p. 32) for perhaps the earliest version of this argument, or Thompson 1972, p. 72, Huang 1987, p. 161-2,) for recent statements.) What is meant by this claim is that if the number of constituents increases, $N \longrightarrow \infty$, and the total Hamiltonian is proportional to N, the thermodynamic relations derived from each of them will coincide in this limit.

However, these arguments should not be mistaken as settling the *empirical* equivalence of the various ensembles, even in this limit. For example, it can be shown that the microcanonical ensemble admits the description of certain metastable thermodynamic states, (e.g. with negative heat capacity) that are excluded in the canonical ensemble (see Touchette 2003, Touchette et al. 2004, and literature cited therein).

3. The coarse-grained entropy. The coarse-graining approach is reminiscent of Boltzmann's construction of cells in his (1877b); cf. the discussion in paragraph 4.4). The main difference is that here one assumes a partition on phase-space Γ, where Boltzmann adopted it in the μ-space. Nevertheless, the same issues about the origin or status of a privileged partition can be debated (cf. p. 58). If one assumes that the partition is intended to represent what we *know* about the system, i.e. if one argues that all we know is whether its state falls in a particular cell ω_i, it can be argued that the its status is subjective. If one argues that the partition is meant to represent limitations in the precision of human observational possibilities, perhaps enriched by instruments, i.e. that we cannot observe more about the system than that its state is in some cell ω_i, one might argue that its choice is objective, in the sense that there are objective facts about what a given epistemic community can observe or not. Of course, one can then still maintain that the status of the coarse-graining would then be anthropocentric (see also the discussion in §7.5). However, note that Gibbs himself did not argue for a preferential size of the cells in phase space, but for taking the limit in which their size goes to zero in a different order.

4. Statistical equilibrium. Finally, a remark about Gibbs' notion of equilibrium. This is fundamentally different from Boltzmann's 1877 notion of equilibrium as the macrostate corresponding to the region occupying the largest volume in phase space (cf. section 4.4). For Gibbs, statistical equilibrium can only apply to an ensemble. And since any given system can be regarded as belonging to an infinity of different ensembles, it makes no sense to say whether an individual system is

in statistical equilibrium or not. In contrast, in Boltzmann's case, equilibrium can be attributed to a single system (namely if the microstate of that system is an element of the set $\Gamma_{\mathrm{eq}} \subset \Gamma$). But it is not guaranteed to remain there for all times.

Thus, one might say that in comparison with the orthodox thermodynamical notion of equilibrium (which is both stationary and a property of an individual system) Boltzmann (1877b) and Gibbs each made an opposite choice about which aspect to preserve and which aspect to sacrifice. See (Uffink 1996b, Callender 1999, Lavis 2005) for further discussions.

6 Modern approaches to statistical mechanics

This section will leave the more or less historical account followed in the previous sections behind, and present a selective overview of some influential modern approaches to statistical physics. In particular, we focus on ergodic theory (§ 6.1–6.2), the theory of the thermodynamic limit §6.3, the work of Lanford on the Boltzmann equation (§6.4), and the BBGKY approach in §6.5.

6.1 Ergodic theory

When the Ehrenfests critically reviewed Boltzmann's and Gibbs' approach to statistical physics in their renowned Encyclopedia article (1912), they identified three issues related to the ergodic hypothesis.

1. The ambiguity in Boltzmann's usage of "probability" of a phase space region (as either the relative volume of the region or the relative time spent in the region by the trajectory of the system).
2. The privileged status of the microcanonical probability distribution or other probability distributions that depend only on the Hamiltonian.
3. Boltzmann's argument that the microstate of a system, initially prepared in a region of phase space corresponding to a non-equilibrium macrostate, should tend to evolve in such a way that its trajectory will spend an overwhelmingly large majority of its time inside the region of phase space corresponding to the equilibrium macrostate Γ_{eq}.

In all these three problems, a more or less definite solution is obtained by adopting the ergodic hypothesis. Thus, the Ehrenfests suggested that Boltzmann's answer to the above problems *depended* on the ergodic hypothesis. As we have seen, this is correct only for Boltzmann's treatment of issue (2) in his (1868a). The doubtful status of the ergodic hypothesis, of course, highlighted the unresolved status of these problems in the Ehrenfests' point of view.

In later works the "ergodic problem" has become more exclusively associated with the first issue on the list above, i.e., the problem of showing the equality of phase and time averages. This problem

can be formulated as follows. Consider a Hamiltonian system and some function f defined on its phase space Γ. The (infinite) time average of f, for a system with initial state x_0 may be defined as:

$$\overline{f(x_0)} = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(T_t x_0) dt \tag{99}$$

where T_t is the evolution operator. On the other hand, for an ensemble of systems with density $\rho_t(x)$, the ensemble average of f is

$$\langle f \rangle_t = \int f(x) \rho_t(x) dx. \tag{100}$$

The ergodic problem is the question whether, or under which circumstances, the time average and ensemble average are equal, i.e.: $\overline{f(x_0)} \stackrel{?}{=} \langle f \rangle_t$. Note that there are immediate differences between these averages. $\overline{f}$ depends on the initial state x_0, in contrast to $\langle f \rangle$. Indeed, each choice of an initial phase point gives rise to another trajectory in phase space, and thus gives, in general, another time average. Secondly, $\langle f \rangle$ will in general depend on time, whereas $\overline{f}$ is time-independent. Hence, a general affirmative answer to the problem cannot be expected.

However, in the case of a stationary ensemble (statistical equilibrium) the last disanalogy disappears. Choosing an even more special case, the microcanonical ensemble ρ_{mc}, the simplest version of the ergodic problem is the question:

$$\overline{f(x_0)} \stackrel{?}{=} \langle f \rangle_{\mathrm{mc}}. \tag{101}$$

Now it is obvious that if Boltzmann's ergodic hypothesis is true, i.e. if the trajectory of the system traverses all points on the energy hypersurface Γ_E, the desired equality holds. Indeed, take two arbitrary points x and y in Γ_E. The ergodic hypothesis implies that there is a time τ such that $y = T_\tau x$. Hence:

$$\begin{aligned}
\overline{f(y)} &= \lim_{T\to\infty} \frac{1}{T} \int_0^T f(T_{t+\tau} x) dt \\
&= \lim_{T\to\infty} \frac{1}{T} \left(\int_0^\tau f(T_t x) dt + \int_0^T f(T_t x) dt \right) \\
&= \lim_{T\to\infty} \frac{1}{T} \int_0^T f(T_t x) dt = \overline{f(x)}
\end{aligned}$$

In other words, $\overline{f}$ must be constant over Γ_E, and hence, also equal to the microcanonical expectation value.

For later reference we note another corollary: the ergodic hypothesis implies that ρ_{mc} is the only stationary density on Γ_E (cf. section 4.1).

The Ehrenfests doubted the validity of the ergodic hypothesis, as Boltzmann had himself, and therefore proposed an alternative, which they called the *quasi-ergodic hypothesis*. This states that the trajectory lies dense in Γ_E, i.e., x_t will pass through every open subset in Γ_E, and thus come arbitrarily close to every point in Γ_E. The system may be called quasi-ergodic if this holds for all its trajectories. As we have seen, this formulation seems actually closer to what Boltzmann may have intended, at least in 1871, than his own literal formulation of the hypothesis.

Not long after the Ehrenfests' review, the mathematical proof was delivered that the ergodic hypothesis cannot hold if Γ_E is a more than one-dimensional manifold (Rosenthal 1913, Plancherel 1913). The quasi-ergodic hypothesis, on the other hand, cannot be immediately dismissed. In fact, it may very well be satisfied for Hamiltonian systems of interest to statistical mechanics. Unfortunately, it has remained unclear how it may contribute to a solution to the ergodic problem. One might hope, at first sight, that for a quasi-ergodic system time averages and microcanonical averages coincide for continuous functions, and that the microcanonical density ρ_{mc} is the only continuous stationary density. But even this is unknown. It is known that quasi-ergodic systems may fail to have a unique stationary measure (Nemytskii and Stepanov 1960, p. 392). This is not to say that quasi-ergodicity has remained a completely infertile notion. In topological ergodic theory, the condition is known under the name of "minimality", and implies several interesting theorems (see Petersen 1983, p. 152ff).

While the Rosenthal-Plancherel result seemed to toll an early death knell over ergodic theory in 1913, a unexpected revival occurred in the early 1930s. These new results were made possible by the stormy developments in mathematics and triggered by Koopman's results, showing how Hamiltonian dynamics might be embedded in a Hilbert space formalism where the evolution operators T_t are represented as a unitary group. This made a whole array of mathematical techniques (e.g. spectral analysis) available for a new attack on the problem.

The first result was obtained by von Neumann in a paper under the promising (but misleading) title "Proof of the Quasi-Ergodic Hypothesis" (1932). His theorem was strengthened by G.D. Birkhoff in a paper entitled "Proof of the Ergodic Theorem" (1931), and published even before von Neumann's.

Since their work, and all later work in ergodic theory, involves more precise mathematical notions, it may be worthwhile first to introduce a more abstract setting of the problem. An abstract *dynamical system* is defined as a tuple $\langle \Gamma, \mathcal{A}, \mu, T \rangle$, where Γ as an arbitrary set, $\mathcal{A}$ is a σ-algebra of subsets of Γ, called the 'measurable' sets in Γ, and μ is a probability measure on Γ, and T denotes a one-parameter group of one-to-one transformations T_t on Γ (with $t \in \mathbb{R}$ or $t \in \mathbb{Z}$) that represent the evolution operators. The transformations T_t are assumed to be measure-preserving, i.e. $\mu(T_t A) = \mu(A)$ for all $A \in \mathcal{A}$. In the more concrete setting of statistical mechanics, one may take Γ

to be the energy hypersurface, $\mathcal{A}$ the collection of its Borel subsets, μ the microcanonical probability measure and T the evolution induced by the Hamiltonian equations.

The von Neumann-Birkhoff ergodic theorem can be formulated as follows:

> ERGODIC THEOREM: Let $\langle \Gamma, \mathcal{A}, \mu, T \rangle$ be any dynamical system and f be an integrable function on Γ. Then
>
> (i) $\overline{f(x)} = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(T_t x)dt$ exists for almost all x;
> i.e. the set of states $x \in \Gamma$ for which $\overline{f(x)}$ does not exist has μ-measure zero.
>
> (ii) $\overline{f(x)} = \langle f \rangle_\mu$ for almost all x iff the system is metrically transitive.

Here, metric transitivity means that it is impossible is to carve up Γ in two regions of positive measure such that any trajectory starting in one region never crosses into the other. More precisely:

> METRIC TRANSITIVITY: A dynamical system is called metrically transitive[57] iff the following holds: for any partition of Γ into disjoint sets A_1, A_2 such that $T_t A_1 = A_1$ and $T_t A_2 = A_2$, it holds that $\mu(A_1) = 0$ or $\mu(A_2) = 0$.

It is not difficult to see why this theorem may be thought of as a successful solution of the original ergodic problem under a slight reinterpretation. First, metric transitivity captures in a measure-theoretic sense the idea that trajectories wander wildly across the energy hypersurface, allowing only exceptions for a measure zero set. Secondly, the theorem ensures the equality of time and microcanonical ensemble average, although only for integrable functions and, again, with the exception of a measure zero set. But that seemed good enough for the taste of most physicists.

The ergodic theorem was therefore celebrated as a major victory. In the words of Reichenbach:

> Boltzmann introduced [...] under the name of *ergodic hypothesis* [...] the hypothesis that the phase point passes through every point of the energy hypersurface. This formulation is easily shown to be untenable. It was replaced by P. and T. Ehrenfest by the formulation that the path comes close to every point within any small distance ϵ which we select and which is greater than 0.
>
> There still remained the question whether the ergodic hypothesis must be regarded as an independent supposition or whether it is derivable from the canonical equations, as Liouville's theorem is.
>
> This problem[...] was finally solved through ingenious investigations by John von Neumann and George Birkhoff, who were able to show that the second alternative is true. [...] With von Neumann and Birkhoff's theorem, deterministic physics has reached its

[57] This name is somewhat unfortunate, since the condition has nothing to do with metric in the sense of distance, but is purely measure-theoretical. Metrically transitive systems are also called 'metrically indecomposable' or, especially in the later literature 'ergodic'. I will stick to the older name in order to avoid confusion with the ergodic hypothesis.

> highest degree of perfection: the strict determinism of elementary processes is shown to lead to statistical laws for macroscopic occurrences." (Reichenbach 1956, p. 78)

Unfortunately, nearly everything stated in this quotation is untrue.

6.1.1 Problems

1. Do metrically transitive systems exist? An immediate question is of course whether metrically transitive systems exist. In a mathematical sense of 'exist' the answer is affirmative. More interesting is the question of whether one can show metric transitivity for any model that is realistic enough to be relevant to statistical mechanics.

A few mechanical systems have been explicitly proven to be metrically transitive. For example: one hard sphere moving in a vessel with a convex scatterer, or a disc confined to move in a 'stadium' (two parallel line-segments connected by two half circles) or its three-dimensional analogue: one hard sphere moving in a cylinder, closed on both sides by half-spheres. But in statistical mechanics one is interested in systems with many particles.

In 1963, Sinai announced he had found a proof that a gas consisting of N hard spheres is metrically transitive. The ergodic theorem thus finally seemed to be relevant to physically interesting gas models. Of course, the hard-spheres-model is an idealization too, but the general expectation among physicists was that a transition to more sophisticated models of a gas system would only make the metric transitivity even more likely and plausible, even though admittedly harder to prove.

The problem proves to be extraordinarily tedious, and Sinai's proof was complicated and, actually, never completely published. But many partial results were. In fact, the development of ideas and techniques needed for the effort contributed much to the emergence of a vigorous mathematical theory, nowadays called 'ergodic theory'. And since Sinai's claim seemed so desirable, many books and articles presented the claim as a solid proven fact (e.g. Lebowitz & Penrose 1973, Sklar 1993).

But by the 1980s, the delay in the publication of a complete proof started to foster some doubts about the validity of the claim. Finally, Sinai and Chernov (1987, p. 185) wrote: "The announcement made in [(Sinai 1963)] for the general case must be regarded as immature." What has been shown rigorously is that a system of three hard spheres is metrically transitive. Recently, the problem has been taken further by Szász (1996) and Simányi and Szász (1999). They have ascertained that for a model of N hard spheres, the ergodic component, i.e. a subset of the energy hypersurface on which the assumption of metric transitivity holds has positive measure. The full problem, however, still awaits solution.

2. Infinite times. In the definition of the time average (99) the limit $T \rightarrow \infty$ is taken. This brings along a number of problems:

(i). The time average is interesting because it is experimentally accessible. The hope is that it represents the equilibrium value of f. But the limit $T \to \infty$ tells us nothing about what happens in a finite time. What is empirically accessible, at best, is the quantity $\frac{1}{T}\int_0^T f(T_t x_0)dt$ for a large but finite T. This expression can still deviate arbitrarily far from the limiting value.

(ii). The limit may even exist while the system is not in equilibrium. A time-averaged value need not be an equilibrium value, because in general

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(T_t x)\, dt \neq \lim_{t\to\infty} f(T_t x). \tag{102}$$

For periodical motions, for example, the left-hand side exists but the right-hand side does not.

(iii) Empirically, equilibrium often sets in quite rapidly. But the time T needed to make $\frac{1}{T}\int_0^T f(T_t x_0)dt$ even remotely close to $\langle f\rangle_{\mathrm{mc}}$ might be enormous, namely of the order of Boltzmann's estimate of the Poincaré-recurrence times! (See also Jaynes (1967, p. 94).)

3. The measure-zero problem. The result that the ergodic theorem provides is that for metrically transitive systems $\overline{f(x)} = \langle f\rangle_{\mathrm{mc}}$ except for a set of microstates with measure zero. So the suggestion here is that this set of exceptions is in some sense negligible. And, as judged from the probability measure μ_{mc}, that is obviously true. But a set of measure zero need not be negligible in any other sense. It is well-known that if one compares 'smallness in measure' with other natural criteria by which one can judge the 'size' of sets, e.g. by their cardinality, dimension or Baire category, the comparisons do not match. Sets of measure zero can be surprisingly large by many other standards Sklar (1993, pp. 181–188).

More importantly, one might choose another measure μ', such that μ-measure zero sets are no longer sets of μ'-measure zero and conversely. It is of course the choice of the measure that determines which sets have measure zero. Thus, if one decides to disregard or neglect sets with a microcanonical measure zero, a privileged status of the microcanonical measure is already presupposed. But this means the virtue of the ergodic theorem as a means of motivating a privileged role of the microcanonical measure is diminished to a self-congratulating one.

6.2 The mixing property, K systems and Bernoulli systems

Ergodic theory, the mathematical field that emerged from the theorems of Birkhoff and von Neumann, may be characterized as a rigorous exploration of the question to what extent a deterministic, time-reversal invariant dynamical system may give rise to random-like behaviour on a macroscopic scale, by assuming various special properties on its dynamics.

In its modern carnation, this theory distinguishes a hierarchy of such properties that consists of

various strengthenings of metric transitivity. Perhaps the most important are the mixing property, the property of being a 'K system' and the Bernoulli systems. The higher up one goes this ladder, the more 'random' behaviour is displayed. The evolution at the microlevel is in all cases provided by the deterministic evolution laws. In the (extensive) literature on the subject, many more steps in the hierarchy are distinguished (such as 'weak mixing', 'weak Bernoulli', 'very weak Bernoulli' etc.), and also some properties that do not fit into a strict linear hierarchy (like the 'Anosov' property, which relies on topological notions rather than on a purely measure-theoretical characterization of dynamical systems). It falls beyond the scope of this paper to discuss them.

6.2.1 Mixing

The idea of *mixing* is usually attributed to Gibbs, in his comparison of the evolution of ensembles with stirring of a dye into an incompressible fluid (cf. section 5.4). Even if initially the fluid and the dye particles occupy separate volumes, stirring will eventually distribute the dye particles homogeneously over the fluid. The formal definition is:

MIXING: A dynamical system $\langle \Gamma, \mathcal{A}, \mu, T \rangle$ is called mixing iff $\forall A, B \in \mathcal{A}$

$$\lim_{t \to \infty} \mu(T_t A \cap B) = \mu(A)\mu(B). \tag{103}$$

In an intuitive sense the mixing property expresses the idea that the dynamical evolution will thoroughly stir the phase points in such a way that points initially contained in A eventually become homogeneously distributed over all measurable subsets B of Γ. One can easily show that *mixing* is indeed a stronger property than metric transitivity, by applying the condition to an invariant set A and choosing $B = A$. The converse statement does not hold. (E.g.: the one-dimensional harmonic oscillator is metrically transitive but not mixing).

Again, there is an interesting corollary in terms of probability measures or densities. Consider a mixing system, and a time-dependent probability density ρ_t, such that ρ_t is *absolutely continuous* with respect to the microcanonical measure μ. (This means that all sets $A \in \mathcal{A}$ with $\mu(A) = 0$, also have $\int_A \rho_t(x)dx = 0$, or equivalently, that ρ_t is a proper density function that is integrable with respect to μ.) In this case, the probability measure associated with ρ_t converges, as $t \longrightarrow \infty$, to the microcanonical measure. Thus, an ensemble of mixing systems with an absolutely continuous density will asymptotically approach to statistical equilibrium. Note that the same result will also hold for $t \longrightarrow -\infty$, so that there is no conflict with the time reversal invariance. Is it in conflict with Poincaré's recurrence theorem? No, the recurrence theorem is concerned with microstates (phase points), and not probability densities. Even when almost all trajectories eventually return close by their original starting point, the recurrence time will differ for each phase point, so that the evolution

of an ensemble of such points can show a definite approach to statistical equilibrium.

Note also that if the result were used as an argument for the privileged status of the microcanonical measure (viz., as the unique measure that all absolutely continuous probability distributions evolve towards), the strategy would again be marred by the point that the condition of absolute continuity already refers to the microcanonical measure as a privileged choice.

Despite the elegance of the mixing property, we can more or less repeat the critical remarks made in the context of the ergodic theorem. In the first place, the condition considers the limit $t \to \infty$, which implies nothing about the rate at which convergence takes place. Secondly, the condition imposed is trivially true if we choose A or B to be sets of measure zero. Thus, the mixing property says nothing about the behaviour of such sets during time evolution. And thirdly, one is still faced with the question whether the mixing property holds for systems that are physically relevant for statistical mechanics. And since the property is strictly stronger than metric transitivity, this problem is at least as hard.

6.2.2 K systems

The next important concept is that of a K *system* ('K' after Kolmogorov). For simplicity, we assume that time is discrete, such that $T_t = T^t$, for $t \in \mathbb{Z}$. There is a perfectly analogously defined concept for continuous time, called K flows (cf. Emch, this volume, Definition 10.3.2).

> K SYSTEM:[58] A dynamical system $\langle \Gamma, \mathcal{A}, \mu, T \rangle$ is called a *K system* if there is a subalgebra $\mathcal{A}_0 \subset \mathcal{A}$, such that
>
> 1. $T^n \mathcal{A}_0 \subset T^m \mathcal{A}_0$ for times $m < n$; where $\subset$ denotes proper inclusion.
> 2. the smallest σ-algebra containing $\cup_{n=1}^{\infty} T^{-n} \mathcal{A}_0$ is $\mathcal{A}$.
> 3. $\cap_{n=1}^{\infty} T^n \mathcal{A}_0 = \mathcal{N}$, where $\mathcal{N}$ is the σ-algebra containing only sets of μ-measure zero or one.

At first sight, this definition may appear forbiddingly abstract. One may gain some intuition by means of the following example. Consider a finite partition $\alpha = \{A_1, \ldots A_m\}$ of Γ into disjoint cells and the so-called *coarse-grained history* of the state of the system with respect to that partition. That is, instead of the detailed trajectory x_t, we only keep a record of the labels i of the cell A_i in which the state is located at each instant of time, until time t=0:

$$\ldots i_{-k}, \ldots, i_{-3}, i_{-2}, i_{-1}, i_0 \qquad i_{-k} \in \{1, \ldots, m\}, \quad k \in \mathbb{N}. \tag{104}$$

[58] There is a considerable variation in the formulation of this definition (Cornfeld, Fomin & Sinai 1982, Batterman 1991, Berkovitz et al. 2006). The present formulation adds one more. It is identical to more common definitions if one replaces n and m in the exponents of T by $-n$ and $-m$ respectively.

This sequence is completely determined by the microstate x at $t = 0$:

$$i_{-k}(x) = \sum_{j=1}^{m} j \mathbb{1}_{A_j}(T^{-k}x) \tag{105}$$

where $\mathbb{1}$ denotes the characteristic function (94). Yet, as we shall see, for a K system, this sequence typically behaves in certain respects like a random sequence. Observe that

$$i_{-k}(x) = j \Longleftrightarrow T^{-k}x \in A_j \Longleftrightarrow x \in T^k A_j; \tag{106}$$

so we can alternatively express the coarse-grained history by means of evolutions applied to the cells in the partition. If $T\alpha := \{TA_1, \ldots, TA_m\}$, let $\alpha \vee T\alpha := \{A_i \cup TA_j \ : i, j = 1, \ldots m\}$ denote the common refinement of α and $T\alpha$. Saying that x belongs to $A_i \cup TA_j$ is, of course, equivalent to providing the last two terms of the sequence (104). Continuing in this fashion, one can build the refinement

$$\bigvee_{k=0}^{\infty} T^k\alpha = \alpha \vee T\alpha \vee T^2\alpha \cdots \vee T^k\alpha \vee \cdots, \tag{107}$$

each element of which corresponds to a particular coarse-grained history (104) up to t=0. The collection (107) is no longer finite, but still a countable partition of Γ.

Now take $\mathcal{A}_0$ to be the σ-algebra generated from the partition $\bigvee_{k=0}^{\infty} T^k\alpha$. Clearly, the events in this algebra are just those whose occurrence is completely decided whenever the coarse-grained history is known. In other words, for all $A \in \mathcal{A}_0$, $\mu(A|C)$ is zero or one, if C is a member of (107). It is easy to see that $T^{-m}\mathcal{A}_0$ is just the σ-algebra generated from $T^{-m}\bigvee_{k=0}^{\infty} T^k\alpha = \bigvee_{k=-m}^{\infty} T^k\alpha$, i.e. from the partition characterizing the coarse-grained histories up to $t = m$. Since the latter partition contains the history up to $t = n$ for all $n < m$, we have:

$$T^{-m}\mathcal{A}_0 \subseteq T^{-n}\mathcal{A}_0 \text{ for all } n < m. \tag{108}$$

This is equivalent to condition 1, but with '$\subset$' replaced by '$\subseteq$'.

Further, to explain condition 2, note that the smallest σ-algebra containing $\cup_{n=1}^{N} T^{-n}\mathcal{A}_0$ is generated by the union of the partitions $\bigvee_{k=-n}^{\infty} T^k\alpha$ for all $n \leq N$, which in view of (108) is just $T^{-N}\mathcal{A}_0$. Thus, condition 2 just says that if we extend the record of the coarse-grained history to later times $t = N > 0$, and let $N \longrightarrow \infty$, the partition eventually becomes sufficiently fine to generate all measurable sets in $\mathcal{A}$. This is a strong property of the dynamics. It means that the entire coarse-grained record, extending from $-\infty$ to ∞, provides all information needed to separate all the measurable sets in $\mathcal{A}$, (except, possibly, if they differ by a measure zero set.)

Similarly, in order to explain condition 3, note that (108) implies that $\cap_{n=1}^{N} T^n \mathcal{A}_0 = T^N \mathcal{A}_0$, which is generated from $\bigvee_{k=0}^{\infty} T^k \alpha$, i.e., the coarse-grained histories up to time $-N$. Thus, condition 3 expresses the demand that, as we let $N \longrightarrow \infty$, the class of events that are settled by the coarse-grained histories up to time $t = -N$ shrinks to the 'trivial' algebra of those sets that have probability one or zero. In other words, for every event $A \in \mathcal{A}$, with $0 < \mu(A) < 1$, the occurrence of A is undecided at some early stage of the coarse-grained history.

Yet the truly remarkable feature of K systems lies in the strict inclusion demanded in condition 1: at any time n, the collection of events decided by the coarse-grained histories up to n, is strictly smaller than the collection of events decided at time $n + 1$. Since the latter is generated from the former by adding the partition $T^{-(n+1)}\alpha$ to the partition $\bigvee T^{-k}\alpha$, this means that at each time n the question which cell of the partition is occupied at time $n + 1$ is not answerable from the knowledge of the previous coarse-grained history. This is quite a remarkable property for a sequence generated by a deterministic law of motion, although, of course, it is familiar for random sequences such as tosses with a die or spins of a roulette wheel.

In this attempt at elucidation, we have presupposed a particular finite partition α. One may ask whether there always is, for each Kolmogorov system, such a partition. The answer is yes, provided the system obeys some mild condition (that $\langle \Gamma, \mathcal{A}, \mu \rangle$ is a *Lebesgue space*[59] Another question is whether the claims made about coarse-grained histories are specific for this particular partition. The answer is no. One may show that, given that they hold for some partition α, they also hold for *any* choice of a finite partition of Γ. (Very roughly speaking: because the partition $\bigvee_n T^n \alpha$ generates the σ-algebra of all events, the coarse-grained histories constructed from another finite partition can be reconstructed in terms of the coarse-grained histories in terms of α.

6.2.3 Bernoulli systems

The strongest property distinguished in the ergodic hierarchy is that of *Bernoulli systems*. To introduce the definition of this type of dynamical systems, it is useful to consider first what is usually known as a 'Bernoulli' scheme. Consider an elementary chance set-up with outcomes $\{A_1, \ldots A_m\}$ and probabilities p_j. A Bernoulli scheme is defined as the probability space obtained from doubly infinite sequences of independent identically distributed repetitions of trials on this elementary set-up. Formally, a Bernoulli scheme for a set (or "alphabet") $\alpha = \{1, \ldots, m\}$ with probabilities $\{p_j\}$ is the probability space $\langle \Gamma, \mathcal{A}, \mu \rangle$, where Γ is the set of all doubly infinite sequences

$$\eta = (\ldots, i_{-2}, i_{-1}, i_o, i_1, i_2 \ldots,) \quad i_k \in \{1, \ldots, m\}; k \in \mathbb{Z} \tag{109}$$

[59] Roughly, this condition means that $\langle \Gamma, \mathcal{A}, \mu \rangle$ is isomorphic (in a measure-theoretic sense) to the interval $[0, 1]$, equipped with the Lebesgue measure. (see (Cornfeld, Fomin & Sinai 1982, p. 449) for the precise definition).

and $\mathcal{A}$ is defined as the smallest σ-algebra on Γ containing the sets:

$$A_k^j := \{\eta \in \Gamma \ : \ i_k = j\}. \tag{110}$$

$\mathcal{A}$ is also known as the cylinder algebra. Further, we require of a Bernoulli scheme that:

$$\mu(A_k^j) = p_j \text{ for all } k \in \mathbb{Z}. \tag{111}$$

One can turn this probability space into a dynamical system by introducing the discrete group of transformations T^m, $m \in \mathbb{Z}$, where T denotes the shift, i.e. the transformation on Γ that shifts each element of a sequence η one place to the left:

$$\text{For all } k \in \mathbb{Z}: \ \ T(i_k) = i_{k-1}. \tag{112}$$

Thus we define:

> BERNOULLI SYSTEM: A dynamical system $\langle \Gamma, \mathcal{A}, \mu, T \rangle$ with a discrete time evolution T is a Bernoulli-system iff there is a finite partition $\alpha = \{A_1, \ldots, A_m\}$ of Γ such that the doubly infinite coarse-grained histories are (isomorphic to) a Bernoulli scheme for α with distribution
>
> $$p_i = \mu(A_i) \quad i \in \{1, \ldots m\}. \tag{113}$$

Thus, for a Bernoulli system, the coarse-grained histories on α behave as randomly as independent drawings from an urn. These histories show no correlation at all, and the best prediction one can make about the location of the state at time $n + 1$, even if we know the entire coarse-grained history from minus infinity to time n, is no better than if we did not know anything at all. One can show that every Bernoulli-system is also a K-system, but that the converse need not hold.

6.2.4 Discussion

Ergodic theory has developed into a full-fledged mathematical discipline with numerous interesting results and many open problems (for the current state of the field, see Cornfeld, Fomin & Sinai 1982, Petersen 1983, Mañé 1987). Yet the relevance of the enterprise for the foundations of statistical mechanics is often doubted. Thus Earman & Rédei (1996) argue that the enterprise is not relevant for explaining 'why phase averaging works' in equilibrium statistical mechanics; Albert (2000, p. 70) even calls the effort poured into rigorous proofs of ergodicity "nothing more nor less —from the standpoint of foundations of statistical mechanics— than a waste of time". (For further discussions, see: (Farquhar 1964, Sklar 1973, Friedman 1976, Malament & Zabell 1980, Leeds 1989, van Lith

2001a, Frigg 2004, Berkovitz et al. 2006))

This judgment is usually based on the problems already indicated above; i.e. the difficulties of ascertaining that even the lowest property on the ergodic hierarchy actually obtains for interesting physical models in statistical mechanics, the empirical inaccessibility of infinite time averages, and the measure zero problem. Also, one often appeals to the Kolmogorov-Arnold-Moser (KAM) results[60] in order to temper the expectations that ergodicity could be a generic property of Hamiltonian systems. These difficulties are serious, but they do not, in my opinion, justify a definitive dismissal of ergodic theory.

Instead, it has been pointed out by (Khinchin 1949, Malament & Zabell 1980, Pitowsky 2001) that further progress may be made by developing the theory in conditions in which (i) the equality of ensemble averages and time averages need not hold for *all* integrable functions, but for only a physically motivated subclass, (ii) imposing conditions that fix the rate of convergence in the infinite time limits in (99) and (103) and (iii) relaxing the conditions on what counts as an equilibrium state. Indeed important progress concerning (i) has been achieved in the 'theory of the thermodynamic limit', described in paragraph 6.3.1. It is clear that further alterations may be mathematically obstreperous; and that any results that might be obtained will not be as simple and general as those of the existing ergodic theory. But there is no reason why progress in these directions should be impossible. See e.g. (Vranas 1998, van Lith 2001b).

The measure zero problem, I would argue, is unsolvable within any "merely" measure-theoretic setting of the kind we have discussed above. The point is, that any measure theoretic discussion of dynamical systems that differ only on measure zero sets are, in measure-theoretical terms, isomorphic and usually identified. Measure theory has no way of distinguishing measure zero sets from the empty set. Any attempt to answer the measure zero problem should call upon other mathematical concepts. One can expect a further light only by endowing the phase space with further physically relevant structure, e.g. a topology or a symplectic form (cf. Butterfield 2006, Belot 2006).

Furthermore, even if ergodic theory has little of relevance to offer to the explanation of 'why phase averaging works' in the case of equilibrium statistical mechanics, this does not mean it is a waste of time. Recall that the equality of phase and time averages was only one of several points on which the Ehrenfests argued that claims by Boltzmann could be substantiated by an appeal to the ergodic hypothesis. Another point was his (1877) claim that a system initially in a non-equilibrium

[60] Quite roughly, the KAM theorems show that some Hamiltonian systems for which trajectories are confined to an invariant set in phase space of small positive measure —and therefore *not* metrically transitive—, will continue to have that property when a sufficiently small perturbation is added to their Hamiltonian (for a more informative introduction, see Tabor 1989). This conclusion spoilt the (once common) hope that non-metrically transitive systems were rare and idealized exceptions among Hamiltonian systems, and that they could always be turned into a metrically transitive system by acknowledging a tiny perturbation from their environment. As we have seen (p. 39), Boltzmann (1868) had already expressed this hope for the ergodic hypothesis.

macrostate should tend to evolve towards the equilibrium macrostate.

It is ironic that some critics of ergodic theory dismiss the attempt to show in what sense and under which conditions the microstate does display a tendency to wander around the entire energy hypersurface as irrelevant, while relying on a rather verbal and pious hope that this will "typically" happen without any dynamical assumption to fall back on. Clearly, the ergodic hierarchy might still prove relevant here.

Still, it is undeniable that many concrete examples can be provided of systems that are not ergodic in any sense of the word and for which equilibrium statistical mechanics should still work. In a solid, say an ice cube, the molecules are tightly locked to their lattice site, and the phase point can access only a minute region of the energy hypersurface. Similarly, for a vapour/liquid mixture in a $\cap$-shaped vessel in a gravity field, molecules may spend an enormously long proportion of time confined to the liquid at the bottom of one leg of the vessel, even though the region corresponding to being located in the other leg is dynamically accessible. And still one would like to apply statistical mechanics to explain their thermal properties.

Summing up, even admitting that ergodic theory cannot provide the whole story in all desired cases does not mean it is irrelevant. I would argue that, on a qualitative and conceptual level, one of the most important achievements of ergodic theory is that it has made clear that strict determinism on the microscopic level is not incompatible with random behaviour on a macroscopic level, even in the strong sense of a Bernoulli system. This implies that the use of models with a stochastic evolution like urn drawings, that Boltzmann used in 1877, or the dog flea model of the Ehrenfests, (cf. §7.2), are not necessarily at odds with an underlying deterministic dynamics.

6.3 Khinchin's approach and the thermodynamic limit

In the 'hard core' version of ergodic theory, described in the previous two paragraphs, one focuses on abstract dynamical systems, i.e. the only assumptions used are about a measure space equipped with a dynamical evolution. It is not necessary that this dynamics arises from a Hamiltonian. Further, it is irrelevant in this approach whether the system has a large number of degrees of freedom. Indeed, the 'baker transformation', an example beloved by ergodic theorists because it provides a dynamical system that possesses *all* the properties distinguished in the ergodic hierarchy, uses the unit square as phase space, and thus has only two degrees of freedom. On the other hand, Hamiltonian systems with large numbers of degrees of freedom, may fail to pass even the lowest step of the ergodic hierarchy, i.e. metric transitivity.

This aspect of ergodic theory is often criticized, because the thermal behaviour of macroscopic systems that the foundations of statistical mechanics ought to explain, arguably appears only when their number of degrees of freedom is huge. As Khinchin puts it:

> All the results obtained by Birkhoff and his followers [...] pertain to the most general type of dynamic systems [...]. The authors of these studies have not been interested in the problem of the foundations of statistical mechanics which is our primary interest in this book. Their aim was to obtain the results in the most general form; in particular all these results pertain equally to the systems with only a few degrees of freedom as well as to the systems with a very large number of degrees of freedom.
>
> From our point of view we must deviate from this tendency. We would unnecessarily restrict ourselves by neglecting the special properties of the systems considered in statistical mechanics (first of all their fundamental property of having a very large number of degrees of freedom) [...]. Furthermore, we do not have any basis for demanding the possibility of substituting phase averages for the time averages of all functions; in fact the functions for which such substitution is desirable have many specific properties which make such a substitution apparent in these cases (Khinchin, 1949, p. 62).

Thus, partly in order to supplement, partly in competition to ergodic theory, Khinchin explored an approach to the ergodic problem that takes the large number of degrees of freedom as an essential ingredient, but only works for a specific class of functions, the so-called *sum functions*.

In particular, consider a Hamiltonian dynamical system $\langle \Gamma, \mathcal{A}, T, \mu \rangle$ of N point particles. That is, we assume: $x = (\vec{q}_1, \vec{p_1}; \ldots; \vec{q}_N, \vec{p}_N) \in \Gamma \subset \mathbb{R}^{6N}$. A function f on Γ is a sum function if

$$f(x) = \sum_{i=1}^{N} \phi_i(x_i) \tag{114}$$

where $x_i = (\vec{p}_i, \vec{q}_i)$ is the molecular state of particle i.[61] Under the further assumption that the Hamiltonian itself is a sum function, Khinchin proved:

> KHINCHIN'S ERGODIC THEOREM: For all sum functions f there are positive constants κ_1, κ_2 such that, for all N:
>
> $$\mu\left(\left\{x \in \Gamma : \left|\frac{\overline{f(x)} - \langle f \rangle_\mu}{\langle f \rangle_\mu}\right| \geq \kappa_1 N^{-1/4}\right\}\right) \leq \kappa_2 N^{-1/4} \tag{115}$$

In words: as N becomes larger and larger, the measure of the set where $\bar{f}$ and $\langle f \rangle$ deviate more than a small amount goes to zero.

This theorem, then, provides an alternative strategy to address the ergodic problem: it says that time average and microcanonical phase average of sum functions will be roughly equal, at least in a very large subset of the energy hypersurface, provided that the number of particles is large enough.

[61] Note that Khinchin does not demand that sum functions are symmetric under permutation of the particles.

Of course, this 'rough equality' is much weaker than the strict equality 'almost everywhere' stated in the von Neumann-Birkhoff ergodic theorem. Moreover, it holds only for the sum functions (114). However, the assumption of metric transitivity is not needed here; nor is any of the more stringent properties of the ergodic hierarchy.

The advantages of this approach to the ergodic problem are clear: first, one avoids the problem that ergodic properties are hard to come by for physically interesting systems. Second, an important role is allotted to the large number of degrees of freedom, which, as noted above, seems a necessary, or at least welcome ingredient in any explanation of thermal behaviour,[62] and thirdly a physically motivated choice for special functions has been made.

However, there are also problems and drawbacks. First, with regard to the "infinite-times" problem (cf. p. 91), Khinchin's approach fares no better or worse than the original ergodic approach. Second, since the rough equality does not hold "almost everywhere" but outside of a subset whose measure becomes small when N is large, the measure-zero problem of ergodic theory (p. 92) is now replaced by a so-called "measure-epsilon problem": if we wish to conclude that in practice the time average and the phase average are (roughly) equal, we should argue that the set for which this does not hold, i.e. the set in the left-hand side of (115) is negligible. This problem is worse than the o measure-zero problem. For example, we cannot argue that ensembles whose density functions have support in such sets are excluded by an appeal to absolute continuity or translation continuity (cf. the discussion on p. 85). Further, if we wish to apply the result to systems that are indeed not metrically transitive, there may be integrals of the equations of motion that lock the trajectory of the system into a tiny subset of Γ for all times, in which case such a set cannot be neglected for practical purposes (cf. Farquhar 1964).

Khinchin argued that the majority of physically important phase functions that one encounters in statistical mechanics are sum functions (cf. Khinchin 1949, p. 63,97). However, this view is clearly too narrow from a physical point of view. It means that all quantities that depend on correlations or interactions between the particles are excluded.

Finally there is the 'methodological paradox' (Khinchin 1949, p. 41–43). This refers to the fact that Khinchin had to assume that the Hamiltonian itself is also a sum function. Let me emphasize that this assumption is *not* made just for the purpose of letting the Hamiltonian be one of the functions to which the theorem applies; the assumption is crucial to the very derivation of the theorem. As Khinchin clearly notes, this is paradoxical because for an equilibrium state to arise at all, it is essential that the particles can interact (e.g. collide), while this possibility is denied when the Hamiltonian is a sum function.

[62] The point can be debated, of course. Some authors argue that small systems can show thermal behaviour too, which statistical mechanics then ought explain. However, the very definition of thermal quantities (like temperature etc.) for such small systems is more controversial (Hill 1987, Feshbach 1987, Rugh 2001, Gross & Votyakov 2000).

In Khinchin's view, the assumption should therefore not be taken literally. Instead, one should assume that there really are interaction terms in the Hamiltonian, but that they manifest themselves only at short distances between the particles, so that they can be neglected, except on a tiny part of phase space. Still, it remains a curious feature of his work that his theorem is intended to apply in situations that are inconsistent with the very assumptions needed to derive it (cf. Morrison 2000, p. 46-47). As we shall see in the next paragraph, later work has removed this paradox, as well as many other shortcomings of Khinchin's approach.

6.3.1 The theory of the thermodynamic limit

The approach initiated by Khinchin has been taken further by van der Linde and Mazur (1963), and merged with independent work of van Hove, Yang and Lee, Fisher, Griffiths, Minlos, Ruelle, Lanford and others, to develop, in the late 60s and early 70s, into what is sometimes called the 'rigorous results' approach or the 'theory of the thermodynamic limit'. The most useful references are (Ruelle 1969, Lanford 1973, Martin-Löf 1979). The following is primarily based on Lanford (1973), which is the most accessible and also the most relevant for our purposes, since it explicitly addresses the ergodic problem, and on (van Lith 2001b).

As in Khinchin's work, this approach aims to provide an explanatory programme for the thermal behaviour of macroscopic bodies in equilibrium by relying mostly on the following central points,

- One adopts the microcanonical measure on phase space.
- the observable quantities are phase functions F of a special kind (see below).
- The number of particles N is extremely large.

It is shown that, under some conditions, in the 'thermodynamic limit', to be specified below, the microcanonical probability distribution for F/N becomes concentrated within an narrow region around some fixed value. This result is similar to Khinchin's ergodic theorem. However, as we shall see, the present result is more powerful, while the assumptions needed are much weaker.

To start of, we assume a Hamiltonian, of the form

$$H(x) = \sum_{i}^{N} \frac{\vec{p}_i^{\,2}}{2m} + U(\vec{q}_1, \ldots, \vec{q}_N). \tag{116}$$

defined on the phase space Γ for N particles. For technical reasons, it is more convenient and simpler to work in the configuration space, and ignore the momenta. Consider a sequence of functions $F(\vec{q}_1, \ldots \vec{q}_n)$, $n = 1, 2, \ldots$ with an indefinite number of arguments, or, what amounts to the same thing, a single function F defined on

$$\cup_{n=1}^{\infty} (\mathbb{R}^3)^n. \tag{117}$$

Such a function is called an 'observable' if it possesses the following properties:

(a). Continuity: For each n, $F(\vec{q}_1, \ldots \vec{q}_n)$ is a continuous function on $\mathbb{R}^{3n}$

(b). Symmetry: For each n, $F(\vec{q}_1, \ldots \vec{q}_n)$ is invariant under permutation of its arguments.

(c). Translation invariance: For each n, and each $\vec{a} \in \mathbb{R}^3$, $F(\vec{q}_1 + \vec{a}, \ldots \vec{q}_n + \vec{a}) = F(\vec{q}_1, \ldots \vec{q}_n)$

(d). Normalization: $F(\vec{q}_1) = 0$

(e). Finite range: There exists a real number $R \in \mathbb{R}$ such that, for each n, the following holds: Suppose we divide the n particles into two clusters labeled by $i = 1, \ldots m$, and $i' = 1, \ldots m'$, where $m + m' = n$. If $|\vec{q}_i - \vec{q}_{i'}| > R$ for all i, i', then $F(\vec{q}_1, \ldots \vec{q}_m; \vec{q}_1, \ldots, \vec{q}_{m'}) = F(\vec{q}_1, \ldots \vec{q}_m) + F(\vec{q}_1, \ldots, \vec{q}_{m'})$.

For the most part, these conditions are natural and self-explanatory. Note that the symmetry condition (b) is very powerful. It may be compared to Boltzmann's (1877b) combinatorial approach in which it was argued that macrostates occupy an overwhelmingly large part of phases space due to their invariance under permutations of the particles (see §4.4). Note further that condition (e) implies that F reduces to a sum function if all particles are sufficiently far from each other. It also means that the observables characterized by Lanford may be expected to correspond to extensive quantities only. (Recall that a thermodynamical quantity is called extensive if it scales proportionally to the size of the system, and intensive if it remains independent of the system size.) In the present approach, intensive quantities (like temperature and pressure) are thus not represented as observables, but rather identified with appropriate derivatives of other quantities, after we have passed to the thermodynamical limit.

Further, it is assumed that the potential energy function U in (116) also satisfies the above conditions. In addition, the potential energy is assumed to be *stable*,[63] i.e.:

(f). Stability: There is a number $B \in \mathbb{R}$, such that, for all n and all $\vec{q_1}, \ldots \vec{q}_n$:

$$U(\vec{q_1}, \ldots \vec{q}_n) \geq -nB. \tag{118}$$

This condition —which would be violated e.g. for Newtonian gravitational interaction— avoids that as n becomes large, the potential energy per particle goes to minus infinity, i.e., it avoids a collapse of the system.

For some results it is useful to impose an even stronger condition:

[63] Strictly speaking, condition (f) is not needed for the existence of the thermodynamic limit for the configurational microcanonical measure. It is needed, however, when these results are extended to phase space (or when using the canonical measure). Note also that the term "stability' here refers to an extensive lower bound of the Hamiltonian. This should be distinguished from thermodynamic concept of stability, which is expressed by the concavity of the entropy function (cf. p. 21).

(f′.) Superstability: The potential energy U is called superstable if, for every continuous function Φ of compact support in $\mathbb{R}^3$:

$$U(\vec{q_1}, \ldots \vec{q}_N) + \lambda \sum_{i \neq j} \Phi(\vec{q_i} - \vec{q_j}) \tag{119}$$

is stable for a sufficiently small choice of $\lambda > 0$. In other words, a stable potential is superstable if it remains stable when perturbed slightly by a continuous finite-range two-body interaction potential.

As in Khinchin's approach, the assumption (f) or (f′) is not just needed because one would like to count the potential energy among the class of observables; rather it is crucial to the proof of the existence of the thermodynamic limit. Of course, the assumption that the interaction potential is continuous and of finite range is still too restrictive to model realistic inter-molecular forces. As Lanford notes, one can weaken condition (e) to a condition of 'weakly tempered' potentials,[64], dropping off quickly with distance (cf. Fisher 1964, p. 386, Ruelle 1969, p. 32) , although this complicates the technical details of the proofs. Again, it is clear, however, that some such condition on temperedness of the long range interactions is needed, if only to avoid another catastrophe, namely that the potential energy per particle goes to $+\infty$ as n increases, so that system might tend to explode. (As could happen, e.g. for a system of charges interacting by purely repulsive Coulomb forces.)

Now, with the assumptions in place, the idea is as follows. Choose a given potential U and an observable F obeying the above conditions. Pick two numbers u and ρ, that will respectively represent the (potential) energy per particle and the particle density (in the limit as N gets large), a bounded open region $\Lambda \subset \mathbb{R}^3$, and a large integer N, such that $\frac{N}{V(\Lambda)} \approx \rho$. (Here, $V(\Lambda)$ denotes the volume of Λ.) Further, choose a small number $\delta u > 0$, and construct the (thickened) energy hypersurface in configuration space, i.e. the shell:

$$\Omega_{\Lambda,N,u,\delta u} = \left\{ (\vec{q_1}, \ldots \vec{q}_N) \in \Lambda^N : \frac{U(\vec{q_1}, \ldots \vec{q}_N)}{N} \in (u - \delta u, u + \delta u) \right\}. \tag{120}$$

Let μ denote the Lebesgue measure on Λ^N; its (normalized) restriction to the above set may then be called the 'thickened configurational microcanonical measure'. Note that

$$\omega^{\mathrm{cf}}(E) := \int_{\Lambda^N} d\vec{q_1} \cdots \vec{q}_N \, \delta(U(\vec{q_1} \ldots \vec{q}_N) - E) \tag{121}$$

[64] If, for simplicity, the potential U is a sum of pair interactions $U = \sum_{i \neq j} \phi(\vec{q_i} - \vec{q_j})$, it is weakly tempered iff there are real constants $R, D, \epsilon > 0$, such that $\phi(\vec{r}) \leq D\|\vec{r}\|^{3+\epsilon}$ when $\|\vec{r}\| \geq R$.

may be considered as the configurational analogue of the structure function (41). Thus

$$\mu(\Omega_{\Lambda,N,u,\delta u}) = \int_{\Lambda^N} d\vec{q}_1 \cdots \vec{q}_N \mathbb{1}_{(u-\delta u, u+\delta u)}(U/N) = \int_{N(u-\delta u}^{N(u+\delta u)} dE\, \omega^{\mathrm{cf}}(E), \tag{122}$$

so that $\frac{1}{2N\delta u}\mu(\Omega_{\Lambda,N,u,\delta u})$ provides a thickened or smoothened version of this configurational structure function. The reason for working with this thickened hypershell instead of the thin hypersurface is of course to circumvent singularities that may appear in the latter. In any case, we may anticipate that, when δu is small, this expression will represent the configurational part of the microcanonical entropy (84. A further factor $1/N!$ may be added to give this entropy a chance of becoming extensive.[65] (See also paragraph §5.2.

We are interested in the probability distribution of F/N with respect to this thickened microcanonical measure on configuration space. For this purpose, pick an arbitrary open interval J, and define

$$\mathcal{V}(\Lambda, N, u, \delta u, F, J) := \frac{1}{N!}\mu\left(\left\{(\vec{q}_1, \ldots, \vec{q}_N) \in \Omega_{u,\delta u} : \ \frac{F(\vec{q}_1, \ldots, \vec{q}_N)}{N} \in J\right\}\right). \tag{123}$$

So,

$$\frac{1}{\mu(\Omega_{\Lambda,N,u,\delta u})}\mathcal{V}(\Lambda, N, u, \delta u, F, J) = \frac{\mathcal{V}(\Lambda, N, u, \delta u, F, J)}{\mathcal{V}(\Lambda, N, u, \delta u, F, \mathbb{R})} \tag{124}$$

gives the probability that F/N lies in the interval J with respect to the above microcanonical measure.

We wish to study the behaviour of this probability in the thermodynamic limit, i.e. as N becomes large, and $V(\Lambda)$ grows proportional to N, such that $N/V(\Lambda) = \rho$. This behaviour will depend on the precise details of the limiting procedure, in particular on the shape of Λ. Lanford chooses to take the limit in the sense of van Hove: A sequence of bounded open regions Λ in $\mathbb{R}^3$ is said to become infinitely large in the sense of Van Hove if, for all $r > 0$, the volume of the set of all points within a distance r from the boundary of Λ, divided by the volume of Λ, goes to zero as N goes to infinity. In other words, the volume of points close to the surface becomes negligible compared to the volume of the interior. This avoids that surface effects could play a role in the limiting behaviour — and eliminates the worry that interactions with the walls of the container should have been taken into account.

Now, the first major result is:

(EXISTENCE OF THE THERMODYNAMIC LIMIT.) As $N \longrightarrow \infty$, and Λ becomes in-

[65] For example, if the system is an ideal gas, i.e. if $U(\vec{q}_1, \ldots, \vec{q}_N) \equiv 0$, one will have $\omega^{\mathrm{cf}}(E) = V^N = \left(\frac{N}{\rho}\right)^N$, so that $\ln \frac{1}{N!}\omega^{\mathrm{cf}}(E)$ scales proportionally to N, but $\ln \omega^{\mathrm{cf}}(E)$ does not.

finitely large in the sense of Van Hove, in such a way that $N/V(\Lambda) = \rho$, then either of the following cases holds:

(α). $\mathcal{V}(\Lambda, N, u, \delta u, F, J)$ goes to zero faster than exponentially in N, or:

(β). $\mathcal{V}(\Lambda, N, u, \delta u, F, J) \approx e^{Ns(\rho,u,\delta u,F,J)}$ where $s(\rho, F, J)$ does not depend on Λ or N, except through the ration $\frac{N}{V(\Lambda)} = \rho$.

In other words, this result asserts the existence of

$$s(\rho, u, \delta u, F, J) := \lim_{N \longrightarrow \infty} \frac{1}{N} \ln \mathcal{V}(\Lambda, N, u, \delta u, F, J) \tag{125}$$

where s is either finite or $-\infty$. (The possibility that $s = -\infty$ for *all* values of the arguments of s is further ignored.) This already gives some clue for how the probability (123) behaves as a function of J. If J_1 and J_2 are two open intervals, N is large, and we suppress the other variables for notational convenience, we expect:

$$\frac{\mu(\frac{F}{N} \in J_1)}{\mu(\frac{F}{N} \in J_2)} = \frac{\mathcal{V}(J_1)}{\mathcal{V}(J_2)} \approx e^{N(s(J_1)-s(J_2))}. \tag{126}$$

If $s(J_2) > s(J_1)$, this ratio goes to zero exponentially in N. Thus, for large systems, the probability $\mu(\frac{F}{N} \in J)$ will only be appreciable for those open intervals J for which $s(J)$ is large.

A stronger statement can be obtained as follows. Associated with the set function $s(J)$ one may define a point function s:

$$s(x) := \inf_{\substack{J \ni x \\ J \text{ open}}} s(J) \tag{127}$$

It can then be shown that, conversely, for all open J:

$$s(J) = \sup_{x \in J} s(x) \tag{128}$$

Moreover, —and this is the second major result— one can show:

$$s(x) \text{ is concave.} \tag{129}$$

Further, $s(x)$ is finite on an open convex subset of its domain (Lanford 1973, p. 26).

Now, it is evident that a concave function $s(x)$ may have three general shapes: It either achieves its maximum value: (i) never; (ii) exactly once, say in some point x_0; or (iii) on some interval. In case (i), F/N 'escapes to infinity' in the thermodynamic limit; this case can be excluded by imposing the superstability condition (f′). Case (ii) is, for our purpose, the most interesting one. In this case, we may consider intervals $J_2 = (x_0 - \epsilon, x_0 + \epsilon)$, for arbitrarily small $\epsilon > 0$ and J_1 any open interval

that does not contain x_0; infer from (127,128) that $s(J_2) > s(J_1)$, and conclude from (126) that the relative probability for F/N to take a value in J_2 rather than J_1 goes to zero exponentially with the size of the system.

Thus we get the desired result: As N becomes larger and larger, the probability distribution of F/N approaches a delta function. Or in other words, the function F/N becomes roughly constant on an overwhelmingly large part of the configurational energy-hypershell:

$$\lim_{N\longrightarrow\infty} \mu\left(\left\{(\vec{q}_1,\ldots,\vec{q}_N)\in\Omega_{\Lambda,N,u,\delta u}:\ |\frac{F(\vec{q}_1,\ldots,\vec{q}_N)}{N}-x_0|>\epsilon\right\}\right)=0 \qquad (130)$$

In case (iii), finally, one can only conclude that the probability distribution becomes concentrated on some interval, but that its behaviour inside this interval remains undetermined. One can show, if this interval is bounded, that this case is connected to phase transitions (but see Lanford 1973, p. 12, 58 for caveats). [66]

6.3.2 Remarks.

1. Phase transitions. First, it is obviously an immense merit of the theory of the thermodynamic limit that, in contrast to ergodic theory, it is, in principle, capable of explaining and predicting the occurrence of phase transitions from a model of the microscopic interaction, in further work often in conjunction with renormalization techniques. Indeed, this capability is its major claim to fame, quite apart from what it has to say about the ergodic problem. What is more, it is often argued that phase transitions are strictly impossible in any finite system, and thus absolutely *require* the thermodynamic limit (Styer 2004, Kadanoff 2000).

This argument raises the problem that our experience, including that of phase transitions in real physical bodies, always deals with finite systems. A theory that presents an account of phase transitions only in the thermodynamic limit, must then surely be regarded as an idealization. This conclusion will not come as a shock many physicists, since idealizations are ubiquitous in theoretical physics. Yet a curious point is that this particular idealization seems to be 'uncontrollable". See (Sklar 2002) and (Liu 1999, Callender 2001, Batterman 2005) for further discussion. I also note that an alternative approach has been proposed recently. In this view phase transitions are associated with topology changes in the microcanonical hypersurface $\{x : H(x) = E\}$ with varying

[66] To see the connection (loosely), note that if one removes the condition $F/N \in J$ from the definition (123) —or equivalently, chooses $J = \mathbb{R}$—, then s in (125) can be interpreted as the (thickened, configurational) microcanonical entropy per particle. Considered now as a function of the open interval $(u-\delta u, u+\delta u)$, s has the same properties as established for $s(J)$, since U itself belongs to the class of observables. Thus, here too, there exists a point function $s(u)$ analogous to (127), and this function is concave (Actually, if we restore one more variable in the notation, and write $s(\rho, u)$, the function is concave in both variables). In case (iii), therefore, this function is constant in u over some interval, say $[u_0'.u_0'']$. This means that there is then a range of thermodynamical states with the same temperature $T = (\frac{\partial s}{\partial u})_\rho^{-1}$, for a range of values of u and ρ, which is just what happens in the condensation phase transition in a van der Waals gas.

E. The crucial distinction from the theory of the thermodynamic limit is, of course, is that such topology changes may occur in finite, —indeed even in small— systems.(cf. Gross 1997, Gross & Votyakov 2000, Casetti et al 2003) However this may be, I shall focus below on the virtues of the thermodynamic limit for the ergodic problem.

2. The ergodic problem revisited. When compared to ergodic theory or Khinchin's approach, the theory of the thermodynamic limit has much to speak in its favour. As in Khinchin's work, the problem of establishing metric transitivity for physically interesting systems does not arise, because the approach does not need to assume it. Further, as in Khinchin's work, the approach works only for special functions. But the class of functions singled out by the assumptions (a–f.) or (a–f′.) above is not restricted to (symmetric) sum functions, and allows for short-range interactions between the particles. Thus, unlike Khinchin, there is no methodological paradox (cf. p.101).

Yet one might still question whether these assumptions are not too restrictive for physically interesting systems. On the one hand, it is clear that some conditions on temperedness and stability are needed to rule out catastrophic behaviour in the thermodynamic limit, such as implosion or explosion of the system. One the other hand, these assumptions are still too strong to model realistic thermal systems. The Coulomb interaction, which according to Lieb & Lebowitz (1973, p. 138) is "the true potential relevant for real matter", is neither tempered nor stable. A tour the force, performed by Lenard, Dyson, Lebowitz and Lieb, has been to extend the main results of the theory of the thermodynamical limit to systems interacting purely by Coulomb forces (if the net charge of the system is zero or small), both classically and quantum mechanically (for fermions) (see Lieb 1976, and literature cited therein). This result, then, should cover most microscopic models of ordinary matter, as long as relativistic effects and magnetic forces can be ignored. But note that this extension is obtained by use of the canonical, rather than the microcanonical measure, and in view of the examples of non-equivalence of these ensembles (cf. p. 5.5) one might worry whether this result applies to ordinary matter in metastable states (like supersaturated vapours, and superheated or supercooled liquids).

Another remarkable point is that, unlike Khinchin's result (115), the result (130) does not refer to time averages at all. Instead, the *instantaneous* value of F/N is found to be almost constant for a large subset of the configurational energy hypersurface. Hence, there is also no problem with the infinite time limit (cf. p. 91. Indeed, dynamics or time evolutions play no role whatsoever in the present results, and the contrast to the programme of ergodic theory is accordingly much more pronounced than in Khinchin's approach.

3. Problems left. What is left, in comparison to those two approaches to the ergodic problem, are two problems. First, there is still the question of how to motivate the choice for the configurational

microcanonical measure (i.e. the normalized Lebesgue measure restricted to the energy hypershell). Lanford is explicit that the theory of the thermodynamic limit offers no help in this question:

> It is a much more profound problem to understand why events which are very improbable with respect to Lebesgue measure do not occur in nature. I, unfortunately, have nothing to say about this latter problem. (Lanford 1973, p. 2).

For this purpose, one would thus have to fall back on other attempts at motivation (cf. p. 84).

Secondly, there is the measure-epsilon problem (cf. p. 101). The desired equality $F/N \approx x$ holds, according to (130), if N is large, outside of a set of small measure. Can we conclude that this set is negligible, or that its states do not occur in nature? In fact, the result (130) instantaneous values is so strong that one ought to be careful of not claiming too much. For example, it would be wrong to claim that for macroscopical systems (i.e. with $N \approx 10^{27}$), the set in the left-hand side of (130) does not occur in nature. Instead, it remains a brute fact of experience that macroscopic systems also occur in non-equilibrium states. In such states, observable quantities take instantaneous values that vary appreciably over time, and thus differ from their microcanonical average. Therefore, their microstate must then be located inside the set of tiny measure that one would like to neglect. Of course, one might argue differently if N is larger still, say $N = 10^{100}$ but this only illustrates the 'uncontrollability' of the idealization involved in this limit, i.e. one still lacks control over how large N must be to be sure that the thermodynamic limit is a reasonable substitute for a finite system.

Further points. Other points, having no counterpart in the approaches discussed previously, are the following. The approach hinges on a very delicately construed sequence of limits. We first have to take the thickened energy shell, then take N, Λ to infinity in the sense of van Hove, finally take δu to zero. But one may ask whether this is clearly and obviously the right thing to do, since there are alternative and non-equivalent limits (the sense of Fisher), the order of the limits clearly do not commute (the thickness of the energy hypershell is proportional to $N\delta u$), and other procedures like the 'continuum limit' (Compagner 1989) have also been proposed.

Finally, in order to make full contact to classical statistical mechanics, on still has to lift restriction to configuration space, and work on phase space. Lanford (1973, p. 2) leaves this as a "straightforward exercise" to the reader. Let's see if we can fill in the details.

Suppose we start from a thickened microcanonical measure on phase space, with the same thickness $2N\delta u$, around a total energy value of $E_0 = Ne_0$. Its probability density is then given by

$$\rho_{Ne_0,N\delta u}(\vec{p}_1,\ldots\vec{p}_N;\vec{q}_1\ldots\vec{q}_N) \quad = \quad \frac{1}{2N\delta u}\int_{E_0-N\delta u}^{E_0+N\delta u}\frac{1}{\omega(E)}\delta(H(x)-E)dE \qquad (131)$$

For the Hamiltonian (116), the integral over the momenta can be performed (as was shown by Boltz-

mann (1868) (cf. Eqn (43). This yields a marginal density

$$\overline{\rho} N e_0, N\delta u(\vec{q}_1, \ldots, \vec{q}_N) = \frac{1}{2N\delta u} \frac{2m\pi^{3N/2}}{\Gamma(\frac{3N}{2})} \int_{E_0 - N\delta u}^{E_0 + N\delta u} \frac{1}{\omega(E)} \left(2m(E - U(q))\right)^{(3N-2)/2} dE \tag{132}$$

This is not quite the normalized Lebesgue measure on configuration space employed by Lanford, but since the factor $(2m(E-U(q))^{(3N-2)/2}$ is a continuous function of U,—at least if $E_0 - N\delta u - U > 0$— it is absolutely continuous with respect to the Lebesgue measure on the shell, and will converge to it in the limit $\delta u \longrightarrow 0$.

But in a full phase space setting, the physical quantities can also depend on the momenta, i.e., they will be functions $F(\vec{p}_1, \ldots \vec{p}_N; \vec{q}_1 \ldots \vec{q}_N)$ and, even if one assumes the same conditions (a–f) as before for their dependence on the second group of arguments, their probability distribution cannot always be determined from the configurational microcanonical measure. For example, let F_1 and F_2 be two observables on configuration space, for which F_1/N and F_2/N converge to different values in the thermodynamical limit, say x_1 and x_2, and let G be any symmetric function of the momenta that takes two different values each with probability $1/2$. For example, take

$$G(\vec{p}_1, \ldots \vec{p}_N) = \begin{cases} 1 & \text{if } \sum_i \vec{p}_i \cdot \vec{n} \geq 0, \\ 0 & \text{elsewhere.} \end{cases} , \tag{133}$$

for some fixed unit vector $\vec{n}$. Now consider the following function on phase space:

$$A(\vec{p}_1, \ldots \vec{p}_N; \vec{q}_1 \ldots \vec{q}_N)) = G(\vec{p}_1, \ldots \vec{p}_N)\, F_1(\vec{q}_1 \ldots \vec{q}_N) + G'(\vec{p}_1, \ldots \vec{p}_N)\, F_2(\vec{q}_1 \ldots \vec{q}_N), \tag{134}$$

where $G' = 1 - G$. If we first integrate over the momenta, we obtain $\tilde{A} = \frac{1}{2}(F_1 + F_2)$, which converges in the thermodynamical limit to $\frac{1}{2}(x_1 + x_2)$. However, it would be wrong to conclude that A is nearly equal to $\frac{1}{2}(x_1 + x_2)$ $(x_1 + x_2)/2$ in an overwhelmingly large part of phase space. Instead, it is nearly equal to x_1 on (roughly) half the available phase space and nearly equal to x_2 on the remaining half.

The extension of (130) to phase space functions will thus demand extra assumptions on the form of such functions; for example, that their dependence on the momenta comes only as some function of the kinetic energy, i.e.

$$A\vec{p}_1, \ldots \vec{p}_N; \vec{q}_1, \ldots \vec{q}_N) = \psi(\sum \frac{\vec{p}_i^2}{2m}) + F(\vec{q}_1, \ldots, \vec{q}_N) \tag{135}$$

for some continuous function ψ.

6.4 Lanford's approach to the Boltzmann equation

We now turn to consider some modern approaches to non-equilibrium statistical mechanics. Of these, the approach developed by Lanford and others (cf. Lanford 1975, Lanford 1976, Lanford 1981, Spohn 1991, Cercignani, Illner &Pulvirenti 1994) deserves special attention because it stays conceptually closer to Boltzmann's 1872 work on the Boltzmann equation and the H-theorem than any other modern approach to statistical physics. Also, the problem Lanford raised and tried to answer is one of no less importance than the famous reversibility and recurrence objections. Furthermore, the results obtained are the best efforts so far to show that a statistical reading of the Boltzmann equation or the H-theorem might hold for the hard spheres gas.

The question Lanford raised is that of the consistency of the Boltzmann equation and the underlying Hamiltonian dynamics. Indeed, if we consider the microstate of a mechanical system such as a dilute gas, it seems we can provide two competing accounts of its time evolution.

(1) On the one hand, given the mechanical microstate x_0 of a gas, we can form the distribution of state $f(\vec{r},\vec{v})$, such that $f(\vec{r},\vec{v})d^3\vec{v}d^3\vec{r}$ gives the relative number of molecules with a position between $\vec{r}$ and $\vec{r}+d^3\vec{r}$ and velocity between $\vec{v}$ and $\vec{v}+d^3\vec{v}$. Presumably, this distribution should be uniquely determined by the microstate x_0. Let us make this dependence explicit by adopting the notation $f^{[x_0]}$. This function, then, should ideally serve as an initial condition for the Boltzmann equation (48), and solving this equation —assuming, that is, that it, that it has a unique solution— would give us the shape of the distribution function at a later time, $f_t^{[x_0]}(\vec{r},\vec{v})$.

(2) On the other hand, we can evolve the microstate x_0 for a time t with the help of the Hamiltonian equations. That will give us $x_t = T_t x_0$. This later state x_t will then also determine a distribution of state $f^{[x_t]}(\vec{r},\vec{v})$.

It is a sensible question whether these two ways of obtaining a later distribution of state from an initial microstate are the same, i.e. whether the two time evolutions are consistent. In other words, the problem is whether the diagram below commutes:

$$\begin{array}{ccc} x_0 & \overset{\text{Hamilton}}{\longrightarrow} & x_t \\ \downarrow & & \downarrow \\ f^{[x_0]} & \overset{\text{Boltzmann}}{\longrightarrow} & f_t^{[x_0]} \overset{?}{=} f^{[x_t]} \end{array} \tag{136}$$

The first issue that has to be resolved here is the precise relation between a microstate and the distribution of state f. It is obvious that, in so far as this function represents the physical property of a gas system, it should be determined by the momentary microstate x. It is also clear, that in so far as it is assumed to be continuous and differentiable in time in order to obey the Boltzmann equation, this cannot be literally and exactly true.

So let us assume, as Boltzmann did, that the gas consists of N hard spheres, each of diameter

d and mass m, contained in some fixed bounded spatial region Λ with volume $|\Lambda| = V$. Given a microstate x of the system one can form the 'exact' distribution of state:

$$F^{[x]}(\vec{r},\vec{v}) := \frac{1}{N}\sum_{i}^{N}\delta^3(\vec{r}-\vec{q}_i)\delta^3(\vec{v}-\frac{\vec{p}_i}{m}). \tag{137}$$

This distribution is, of course, not a proper function, and being non-continuous and non-differentiable, clearly not a suitable object to plug into the Boltzmann equation. However, one may reasonably suppose that one ought to be able to express Boltzmann's ideas in a limit in which the number of particles, N, goes to infinity. However, this limit clearly must be executed with care.

On the one hand, one ought to keep the gas dilute, so that collisions involving three or more particles will be rare enough so that they can safely be ignored in comparison to two-particle collisions. On the other hand, the gas must not be so dilute that collisions are altogether too rare to contribute to a change of f. The appropriate limit to consider, as Lanford argues, is the so-called Boltzmann-Grad limit in which $N \longrightarrow \infty$, and:[67]

$$\frac{Nd^2}{V} = \text{constant} > 0. \tag{138}$$

Denote this limit as "$N \xrightarrow{\text{BG}} \infty$", where it is implicitly understood that $d \propto N^{-1/2}$. The hope is then that in this Boltzmann-Grad limit, the exact distribution $F^{[x^N]}$ will tend to a continuous function that can be taken as an appropriate initial condition for the Boltzmann equation. For this purpose, one has to introduce a relevant notion of convergence for distributions on the μ-space $\Lambda \times \mathbb{R}^3$. A reasonable choice is to say that an arbitrary sequence of distributions f_n (either proper density functions or in the distributional sense) converges to a distribution f, $f_n \longrightarrow f$, iff the following conditions hold:

For each rectangular parallelepiped $\Delta \subset \Lambda \times \mathbb{R}^3$:

$$\lim_{n\longrightarrow\infty}\int_{\Delta} f_N d^3\vec{r}d^3\vec{v} = \int_{\Delta} f d^3\vec{r}d^3\vec{v}, \tag{139}$$

and

$$\lim_{n\longrightarrow\infty}\int \vec{v}^2 f_n d^3\vec{r}d^3\vec{v} = \int \vec{v}^2 f d^3\vec{r}d^3\vec{v}, \tag{140}$$

where the second condition is meant to guarantee the convergence of the mean kinetic energy.

It is also convenient to introduce some distance function between (proper or improper) distributions that quantifies the sense in which one distribution is close to another in the above sense. That

[67] The condition can be explained by the hand-waving argument that Nd^2/V is proportional to the 'mean free path', i.e. a typical scale for the distance traveled by a particle between collisions, or also by noting that the collision integral in the Boltzmann equation is proportional to Nd^2/V, so that by keeping this combination constant, we keep the Boltzmann equation unchanged.

is to say, one might define some distance $d(f, g)$ between density functions on $\Lambda \times R^3$ such that

$$d(f_n, f) \longrightarrow 0 \Longrightarrow f_n \longrightarrow f. \tag{141}$$

There are many distance functions that could do this job, but I won't go into the question of how to pick out a particular one.

The hope is then, to repeat, that $F^{[x^N]} \longrightarrow f$ in the above sense when $N \stackrel{BG}{\longrightarrow} \infty$, where f is sufficiently smooth to serve as an initial condition in the Boltzmann equation, and that with this definition, the Boltzmannian and Hamiltonian evolution become consistent in the sense that the diagram (136) commutes. But clearly this will still be a delicate matter. Indeed, increasing N means a transition from one mechanical system to another with more particles. But there is no obvious algorithm to construct the state x^{N+1} from x^N, and thus no way to enforce convergence on the level of individual states.

Still, one might entertain an optimistic guess, which, if true, would solve the consistency problem between the Boltzmann and the Hamiltonian evolution in an approximate fashion if N is very large.

> OPTIMISTIC GUESS: If $F^{[x_0^N]}$ is near to f then $F^{[x_t^N]}$ is near to f_t for all $t > 0$, and where f_t is the solution of the Boltzmann equation with initial condition f.

As Lanford (1976) points out, the optimistic guess cannot be right. This is an immediate consequence of the reversibility objection: Indeed, suppose it were true for all $x \in \Gamma$, and $t > 0$. (Here, we momentarily drop the superscript N from x^N to relieve the notation.) Consider the phase point Rx obtained from x by reversing all momenta: $R(\vec{q}_1, \vec{p}_1; \ldots; \vec{q}_N, \vec{p}_N) = (\vec{q}_1, -\vec{p}_1; \ldots, ; \vec{q}_N, -\vec{p}_N)$. If $F^{[x]}(\vec{r}, \vec{v})$ is near to some distribution $f(\vec{r}, \vec{v})$, then $F^{[Rx]}(\vec{r}, \vec{v})$ is near to $f(\vec{r}, -\vec{v})$. But as x evolves to x_t, Rx_t evolves to $T_t R x_t = R T_{-t} x_t = Rx$. Hence $F^{[T_t R x_t]}(\vec{r}, \vec{v}) = F^{[Rx]}(\vec{r}, \vec{v})$ is near to $f(\vec{r}, -\vec{v})$. But the validity of the conjecture for Rx_t would require that $F^{[T_t R x_t]}(\vec{r}, \vec{v})$ is near to $f_t(\vec{r}, -\vec{v})$ and these two distributions of state are definitely not near to each other, except in some trivial cases.

But even though the optimistic guess is false in general, one might hope that it is 'very likely' to be true, with some overwhelming probability, at least for some finite stretch of time. In order to make such a strategy more explicit, Lanford takes recourse to a probability measure on Γ, or more precisely a sequence of probability measures on the sequence of Γ_N's.

Apart from thus introducing a statistical element into what otherwise would have remained a purely kinetic theory account of the problem, there is a definite advantage to this procedure. As mentioned above, there is no obvious algorithm to construct a sequence of microstates in the Boltzmann-Grad limit. But for measures this is different. The microcanonical measure, for example is not just a measure for the energy hypersurface of one N-particles-system; it defines an algorithmic sequence

of such measures for each N.

In the light of this discussion, we can now state Lanford's theorem as follows (Lanford 1975, 1976):

> LANFORD'S THEOREM: Let $t \mapsto f_t$ be some solution of the Boltzmann equation, say for $t \in [0, a) \subset \mathbb{R}$. For each N, let Δ_N denote the set in the phase space Γ_N of N particles, on which $F^{[x^N]}$ is near to f_0 (the initial condition in the solution of the Boltzmann equation) in the sense that for some chosen distance function d and for tolerance $\epsilon > 0$:
>
> $$\Delta_N = \{x^N \in \Gamma_N \ : \ d(F^{[x^N]}, f_0) < \epsilon\}. \tag{142}$$
>
> Further, for each N, conditionalize the microcanonical measure μ_N on Δ_N:
>
> $$\mu_{\Delta,N}(\cdot) := \mu_N(\cdot|\Delta_N). \tag{143}$$
>
> In other words, $\mu_{\Delta,N}$ is a sequence of measures on the various Γ_N that assign measure 1 to the set of microstates $x^N \in \Gamma_N$ that are close to f_0 in the sense that $d(F^{[x^N]}, f_0) < \epsilon$. Then: $\exists \tau$, $0 < \tau < a$ such that for all t with $0 < t < \tau$:
>
> $$\mu_{\Delta,N}(\{x^N \in \Gamma_N \ : \ d(F^{[x_t^N]}, f_t) < \epsilon\}) > 1 - \delta \tag{144}$$
>
> where $\delta \longrightarrow 0$ as both $\epsilon \longrightarrow 0$ and $N \overset{BG}{\longrightarrow} \infty$.

In other words: as judged from the microcanonical measure on Γ_N restricted to those states x^N that have their exact distribution of state close to a given initial function f_0, a very large proportion $(1-\delta)$ evolve by the Hamiltonian dynamics in such a way that their later exact distribution of state $F^{[x_t^N]}$ remains close to the function f_t, as evolved from f_0 by the Boltzmann equation.

6.4.1 Remarks

Lanford's theorem shows that a statistical and approximate version of the Boltzmann equation can be derived from Hamiltonian mechanics and the choice of an initial condition in the Boltzmann-Grad limit. This is a remarkable achievement, that in a sense vindicates Boltzmann's intuitions. According to Lanford (1976, p. 14), the theorem says that the approximate validity of the Boltzmann equation, and hence the H-theorem, can be obtained from mechanics alone and a consideration of the initial conditions.

Still the result established has several remarkable features, all of which are already acknowledged by Lanford. First, there are some drawbacks that prevent the result from having practical impact for

the project of justifying the validity of the Boltzmann equation in real-life physical applications. The density of the gas behaves like N/d^3, and in the Boltzmann-Grad limit this goes to zero. The result thus holds for extremely rarified gases. Moreover, the length of time for which the result holds, i.e. τ, depends on the constant in (138), which also provides a rough order of magnitude for the mean free path of the gas . It turns out that, by the same order of magnitude considerations, τ is roughly two fifths of the mean duration between collisions. This is a disappointingly short period: in air at room temperature and density, τ is in the order of microseconds. Thus, the theorem does not help to justify the usual applications of the Boltzmann equation to macroscopic phenomena which demand a much longer time-scale.

Yet note that the time scale is not trivially short. It would be a misunderstanding to say that the theorem establishes only the validity of the Boltzmann equation for times so short that the particles have had no chance of colliding: In two fifths of the mean duration between collisions, about 40 % of the particles have performed a collision.

Another issue is that in comparison with Boltzmann's own derivation no explicit mention seems to have been of the *Stoßzahlansatz*. In part this is merely apparent. In a more elaborate presentation (cf. Lanford 1975, 1976), the theorem is not presented in terms of the microcanonical measure, but an arbitrary sequence of measures ν_N on (the sequence of phase spaces) Γ_N. These measures are subject to various assumptions. One is that each ν_N should be absolutely continuous with respect to the microcanonical measure μ_N, i.e. ν_N should have a proper density function

$$d\nu_N(x) = n_N(x_1, \ldots x_N)dx_1 \cdots x_N \tag{145}$$

where $x_i = (\vec{q_i}, \vec{p_i})$ denotes the canonical coordinates of particle i. Further, one defines, for each N and $m < N$, the reduced density functions by

$$n_N^{(m)}(x_1, \ldots x_m) := \frac{N!}{(N-m)!}\frac{1}{N^m}\int n_N(x_1, \ldots x_N)dx_{m+1} \cdots dx_N \tag{146}$$

i.e. as (slightly renormalized) marginal probability distributions for the first m particles. The crucial assumption is now that

$$\lim_{N \overset{\text{BG}}{\longrightarrow} \infty} n_N^{(m)}(x_1, \ldots x_m) = n^{(1)}(x_1) \cdots n^{(1)}(x_m) \tag{147}$$

uniformly on compact subsets of $(\Lambda \times \mathbb{R}^3)^m$. This assumption (which can be shown to hold for the microcanonical measures) is easily recognized as a measure-theoretic analogy to the *Stoßzahlansatz*. It demands, in the Boltzmann-Grad limit, statistical independence of the molecular quantities for any pair or m-tuple of particles at time $t = 0$. As Lanford also makes clear, it is assumption (146) that

would fail to hold if we run the construction of the reversibility objection; (i.e. if we follow the states x in Δ_N for some time t, $0t < \tau$, then reverse the momenta, and try to apply the theorem to the set $\Delta'_N = \{Rx_t \ : \ x \in \Delta_N\}$).

But another aspect is more positive. Namely: Lanford's theorem does not need to assume explicitly that the *Stoßzahlansatz* holds *repeatedly*. Indeed a remarkable achievement is that once the factorization condition (146) holds for time $t = 0$ it will also hold for $0 < t < \tau$, albeit in a weaker form (as convergence in measure, rather than uniform convergence). This is sometimes referred to as "propagation of chaos" (Cercignani, Illner &Pulvirenti 1994).

But the main conceptual problem concerning Lanford's theorem is where the apparent irreversibility or time-reversal non-invariance comes from. On this issue, various opinions have been expressed. Lanford (1975, p. 110) argues that irreversibility is the result of passing to the Boltzmann-Grad limit. Instead, Lanford (1976) argues that it is due to condition (146) plus the initial conditions (i.e.: $x_N \in \Delta_N$).

However, I would take a different position. The theorem equally holds for $-\tau < t < 0$, with the proviso that f_t is now a solution of the anti-Boltzmann equation. This means that the theorem is, in fact, invariant under time-reversal.

6.5 The BBGKY approach

The so-called BBGKY-hierarchy (named after Bogolyubov, Born, Green, Kirkwood and Yvon) is a unique amalgam of the description of Gibbs and the approach of Boltzmann. The goal of the approach is to describe the evolution of ensembles by means of reduced probability densities, and to see whether a Boltzmann-like equation can be obtained under suitable conditions —and thereby an approach to statistical equilibrium.

First, consider an arbitrary time-dependent probability density ρ_t. The evolution of ρ is determined via the Liouville-equation by the Hamiltonian:

$$\frac{\partial \rho_t}{\partial t} = \{H, \rho\}. \tag{148}$$

Central in the present approach is the observation that for relevant systems in statistical mechanics, this Hamiltonian will be symmetric under permutation of the particles. Indeed, the Hamiltonian for a system of N indistinguishable particles usually takes the form

$$H(\vec{q}_1, \vec{p_1}; \ldots; \vec{q}_N, \vec{p}_N) = \sum_{=1}^{N} \frac{\vec{p_i}^2}{2m} + \sum_{i}^{N} V(\vec{q}_i) + \sum_{i<j}^{N} \phi(\|q_i - \vec{q_j}\|) \tag{149}$$

where V is the potential representing the walls of the bounded spatial region Λ, say:

$$V(\vec{q}) = \begin{cases} 0 & \text{if } \vec{q} \in \Lambda \\ \infty & \text{elsewhere} \end{cases} \tag{150}$$

and ϕ the interaction potential between particle i and j. This is not only symmetric under permutation of the particle labels, but even has the more special property that it is a sum of functions that never depend on the coordinates of more than *two* particles. (cf. the discussion in §6.3.)

Let us again use the notation $x = (\vec{q}_1, \vec{p}_1; \ldots; \vec{q}_N, \vec{p}_N) = (x_1, \ldots, x_N)$; with $x_i = (\vec{q}_i, \vec{p}_i)$, and consider the sequence of reduced probability density functions, defined as the marginals of ρ:

$$\begin{aligned} \rho^{(1)}(x_1) &:= \int \rho_t(x)\, dx_2 \cdots x_N \\ &\vdots \\ \rho^{(m)}(x_1, \ldots, x_m) &= \int \rho_t(x)\, dx_{m+1} \cdots dx_N \end{aligned} \tag{151}$$

Here, $\rho^{(m)}$ gives the probability density that particles $1, \ldots, m$ are located at specified positions $\vec{q}_1, \ldots \vec{q}_m$ and moving with the momenta $\vec{p}_1, \ldots \vec{p}_m$, whereas all remaining particles occupy arbitrary positions and momenta.

Symmetry of the Hamiltonian need not imply symmetry of ρ. But one might argue that we may restrict ourselves to symmetric probability densities if *all* observable quantities are symmetric. In that case, it makes no observable difference when two or more particles are interchanged in the microstate and one may replace ρ by its average under all permutations without changing the expectation values of any observable quantity. However this may be, we now assume that ρ is, in fact, symmetric under permutations of the particle labels. In other words, from now on $\rho^{(m)}$ gives the probability density that any arbitrarily chosen set of m particles have the specified values for position and momentum.

The guiding idea is now that for relevant macroscopic quantities, we do not need the detailed form of the time evolution of ρ_t. Rather, it suffices to focus on no more than just a few marginals from the hierarchy (151). For example, suppose a physical quantity represented as a phase function A is a symmetric sum function on Γ:

$$A(x) = \sum_{i=1}^{N} A(x_i) \tag{152}$$

Then

$$\langle A \rangle = N \int A(x_1) \rho^{(1)}(x_1)\, dx_1 \tag{153}$$

which is a considerable simplification. But this is not to say that we can compute the evolution of

$\langle A \rangle$ in time so easily.

Consider in particular $\rho_t^{(1)}$ in (151). This is the *one-particle distribution function*: the probability that an arbitrary particle is in the one-particle state $(\vec{p}, \vec{q})$. This distribution function is in some sense analogous to Boltzmann's f. But note: ρ_1 is a marginal probability distribution; it characterizes an ensemble, whereas f is (in this context) a stochastic variable, representing a property of a single gas:

$$f(\vec{r}, \vec{v})) = \frac{1}{N} \sum_i \delta(\vec{q}_i - \vec{r}) \delta(\vec{v} - \frac{\vec{p}_i}{m}). \tag{154}$$

How does $\rho_t^{(1)}$ evolve? From the Liouville-equation we get

$$\frac{\partial \rho^{(1)}(x_1)}{\partial t} \quad = \quad \int \{H, \rho\} d^3\vec{p}_2 \cdots \vec{p}_N d\vec{q}_2 \cdots \vec{q}_N. \tag{155}$$

It is convenient here to regard the Poisson bracket as a differential operator on ρ, usually called the Liouville operator $\mathcal{L}$:

$$\mathcal{L}\rho := \sum_{i=1}^{N} \left(\frac{\partial H}{\partial \vec{q}_i} \cdot \frac{\partial}{\partial \vec{p}_i} - \frac{\partial H}{\partial \vec{p}_i} \cdot \frac{\partial}{\partial \vec{q}_i} \right) \rho. \tag{156}$$

For the Hamiltonian (149) this can be expanded as:

$$\mathcal{L} = \sum_i^N \mathcal{L}_i^{(1)} + \sum_{i<j}^N \mathcal{L}_{ij}^{(2)} \tag{157}$$

where

$$\mathcal{L}_i^{(1)} := \vec{p}_i \cdot \frac{\partial}{\partial \vec{q}_i} \tag{158}$$

and

$$\mathcal{L}_{ij}^{(2)} := \frac{\partial \phi_{ij}}{\partial \vec{q}_i} \cdot \left(\frac{\partial}{\partial \vec{p}_i} - \frac{\partial}{\partial \vec{p}_j} \right) \tag{159}$$

The evolution of $\rho^{(1)}$ is then given by:

$$\frac{\partial \rho_t^{(1)}(x_1)}{\partial t} \quad = \quad \mathcal{L}_1^{(1)} \rho_t^{(1)}(x_1) + \int dx_2 \mathcal{L}_{12}^{(2)} \rho^{(2)}(x_1, x_2) \tag{160}$$

More generally, for higher-order reduced distribution functions $\rho^{(m)}$ $(m \geq 2)$, the evolution is governed by the equations:

$$\begin{aligned} \frac{\partial \rho_t^{(m)}(x_1, \ldots, x_m)}{\partial t} \quad = \quad & \sum_{i=1}^{m} \mathcal{L}_i^{(1)} \rho_t^{(m)}(x_1, \ldots x_m) + \sum_{i<j=1}^{m} \mathcal{L}_{ij}^{(2)} \rho_t^{(m)}(x_1, \ldots, x_m) \\ & + \sum_i^m \int dx_{m+1} \mathcal{L}_{i,m+1}^{(2)} \rho_t^{(m+1)}(x_1, \ldots x_{m+1}) \end{aligned} \tag{161}$$

The equations (160,161) form the *BBGKY hierarchy*. It is strictly equivalent to the Hamiltonian formalism for symmetric ρ and H, provided that H contains no terms that depend on three or more particles. As one might expect, solving these equations is just as hard as for the original Hamiltonian equations. In particular, the equations are not closed: in order to know how $\rho_t^{(1)}$ evolves, we need to know $\rho_t^{(2)}$. In order to know how $\rho_t^{(2)}$ evolves, we need to know $\rho_t^{(3)}$ etc.

The usual method to overcome this problem is to cut off the hierarchy, i.e. to assume that for some finite m, $\rho^{(m)}$ is a functional of $\rho^{(\ell)}$ with $\ell < m$. In particular, if we just consider the easiest case ($m = 2$) and the easiest form of the functional, we can take $\rho^{(2)}$ to factorize in the distant past ($t \longrightarrow -\infty$), giving:

$$\rho_t^{(2)}(x_1, x_2) = \rho_t^{(1)}(x_1)\rho_t^{(1)}(x_2); \quad \text{if } t \longrightarrow -\infty \tag{162}$$

i.e., requiring that the molecular states of any pair of particles are uncorrelated *before* their interaction. This is analogous to the *Stoßzahlansatz* (29), but now, of course, formulated in terms of the reduced distribution functions of an ensemble.

It can be shown that for the homogeneous case, i.e. when $\rho^{(2)}$ is uniform over the positions $\vec{q}_1$ and $\vec{q}_2$, i.e. when $\rho^{(2)}(x_1, x_2) = \rho^{(2)}(\vec{p}_1, \vec{p}_2)$ and when ϕ is a interaction potential of finite range, the evolution equation for $\rho^{(1)}$ becomes formally identical to the Boltzmann equation (48). That is to say, in (160) we may substitute $\mathcal{L}_i^{(1)} = 0$ and:

$$\begin{aligned} \frac{\partial \rho_t^{(1)}(\vec{p}_1)}{\partial t} &= \int \mathcal{L}_{12}^{(2)} \rho(\vec{p}_1, \vec{p}_2) d^3\vec{p}_2 \\ &= \frac{N}{m} \int b db d\phi \int d\vec{p}_2 \|\vec{p}_2 - \vec{p}_1\| \left(\rho_t^{(1)}(\vec{p}_1{}')\rho_t^{(1)}(\vec{p}_2{}') - \rho_t^{(1)}(\vec{p}_1)\rho_t^{(1)}(\vec{p}_2) \right) \end{aligned} \tag{163}$$

(See Uhlenbeck and Ford (1963, p. 131) for more details.)

6.5.1 Remarks

The BBGKY approach is thoroughly Gibbsian in its outlook, i.e. it takes a probability density over phase space as its basic conceptual tool. An additional ingredient, not used extensively by Gibbs, is its reliance on permutation symmetry. It gives an enormous extension of Gibbs' own work by providing a systematic hierarchy of evolution equations for reduced (or marginalized) density functions, which can then be subjected to the techniques of perturbation theory. An ensemble-based analogy of the Boltzmann equation comes out of this approach as a first-order approximation for dilute gases with collision times much smaller than the mean free time. The Boltzmann equation for inhomogeneous gases cannot be obtained so easily– as one might expect also on physical grounds that one will need extra assumptions to motivate its validity.

It is instructive to compare this approach to Lanford's. His analogy of the Boltzmann equation is

obtained for a different kind of function, namely the one-particle distribution function $F^{[x]}$, i.e. the exact relative number of particles with molecular state $(\vec{r}, \vec{v})$, instead of $\rho^{(1)}$. Of course, there is a simple connection between the two. Noting that $F^{[x]}$ is a sum function (cf. equation (137), we see that

$$\langle F^{[x]} \rangle = \int \rho^{(1)}(\vec{p}_1, \vec{q}_1) f(\delta(\vec{r} - \vec{q}_1)\delta(\vec{v} - \frac{\vec{p}_1}{m}) dp_1 dq_1 = \rho^{(1)}(\vec{r}, \vec{v}). \tag{164}$$

In other words, the one-particle distribution function $\rho^{(1)}$ is the expected value of the exact distribution of state. It thus appears that where Lanford describes the probability of the evolution of the exact distribution of state, the BBGKY result (163) describes the evolution of the average of the exact distribution of state. Lanford's results are therefore much more informative.

One might be tempted here to argue that one can justify or motivate that that actual particle distribution might be taken equal to its ensemble average by arguments similar to those employed in ergodic theory. In particular, we have seen from Khinchin's work (cf. §6.3) that for large enough systems, the probability that a sum function such as $F^{[x]}$ deviates significantly from its expectation value is negligible. However, an important complication is that this reading of Khinchin's results holds for equilibrium, i.e. they apply with respect to the microcanonical distribution ρ_{mc}, not to an arbitrary time-dependent density ρ_t envisaged here.

The time asymmetry of the resulting equation does not derive from the hierarchy of equations, but from the ensemble-based analogy of the *Stoßzahlansatz* (162). That is to say, in this approach time asymmetry is introduced via an initial condition on the ensemble, i.e. the absence of initial correlations. It can be shown, just like for the original Boltzmann equation, that when the alternative boundary condition is imposed that makes the momenta independent after collisions, (i.e. if (162) is imposed for $t \longrightarrow \infty$ instead) the anti-Boltzmann equation is obtained (see Uhlenbeck and Ford 1963, p. 127).

7 Stochastic dynamics

7.1 Introduction

Over recent decades, some approaches to non-equilibrium statistical mechanics, that differ decidedly in their foundational and philosophical outlook, have nevertheless converged in developing a common unified mathematical framework. I will call this framework 'stochastic dynamics', since the main characteristic feature of the approach is that it characterizes the evolution of the state of a mechanical system as evolving under stochastic maps, rather than under a deterministic and time-reversal invariant Hamiltonian dynamics.[68]

[68] Also, the name has been used in precisely this sense already by Sudarshan and coworkers, cf. (Sudarshan et al. 1961, Mehra & Sudarshan 1972).

The motivations for adopting this stochastic type of dynamics come from different backgrounds, and one can find authors using at least three different views.

1. "Coarse graining"(cf. van Kampen 1962, Penrose 1970): In this view one assumes that on the microscopic level the system can be characterized as a (Hamiltonian) dynamical system with deterministic time-reversal invariant dynamics. However, on the macroscopic level, one is only interested in the evolution of macroscopic states, i.e. in a partition (or coarse graining) of the microscopic phase space into discrete cells. The usual idea is that the form and size of these cells are chosen in accordance with the limits of our observational capabilities. A more detailed exposition of this view is given in §7.5.

On the macroscopic level, the evolution now need no longer be portrayed as deterministic. When only the macrostate of a system at an instant is given, it is in general not fixed what its later macrostate will be, even if the underlying microscopic evolution is deterministic. Instead, one can provide *transition probabilities*, that specify how probable the transition from any given initial macrostate to later macrostates is. Although it is impossible, without further assumptions, to say anything general about the evolution of the macroscopically characterized states, it is possible to describe the evolution of an ensemble or a probability distribution over these states, in terms of a *stochastic process*.

2. "Interventionism", "tracing" or "open systems" (cf. Blatt 1959, Davies 1976, Lindblad 1976, Lindblad 1983, Ridderbos 2002): On this view, one assumes that the system to be described is not isolated but in interaction with the environment. It is assumed that the total system, consisting of the system of interest and the environment can be described as a (Hamiltonian) dynamical system with a time-reversal invariant and deterministic dynamics. If we represent the state of the system by $x \in \Gamma^{(\mathrm{s})}$ and that of the environment by $y \in \Gamma^{(\mathrm{e})}$, their joint evolution is given by a one-parameter group of evolution transformations, generated from the Hamiltonian equations of motion for the combined system: $U_t : (x, y) \mapsto U_t(x, y) \in \Gamma^{(s)} \times \Gamma^{(e)}$. The evolution of the state x in the course of time is obtained by projecting, for each t, to the coordinates of $U_t(x, y)$ in $\Gamma^{(\mathrm{s})}$; call the result of this projection x_t. Clearly, this reduced time evolution of the system will generally fail to be deterministic, e.g. the trajectory described by x_t in $\Gamma^{(\mathrm{s})}$ may intersect itself.

Again, we may hope that this indeterministic evolution can nevertheless, for an ensemble of the system and its environment, be characterized as a stochastic process, at least if some further reasonable assumptions are made.

3. A third viewpoint is to deny (Mackey, 1992, 2001), or to remain agnostic about (Streater 1995), the existence of an underlying deterministic or time-reversal invariant dynamics, and simply regard the evolution of a system as described by a stochastic process as a new fundamental form of dynamics in its own right.

While authors in this approach thus differ in their motivation and in the interpretation they have of its subject field, there is, as we shall see, a remarkable unity in the mathematical formalism adopted for this form of non-equilibrium statistical mechanics. The hope, obviously, is to arrange this description of the evolution of mechanical systems in terms of a stochastic dynamics in such a way that the evolution will typically display 'irreversible behaviour': i.e. an 'approach to equilibrium', that a Boltzmann-like evolution equation holds, that there is a stochastic analogy of the H-theorem, etc. In short, one would like to recover the autonomy and irreversibility that thermal systems in non-equilibrium states typically display.

We will see that much of this can be achieved with relatively little effort once a crucial technical assumption is in place: that the stochastic process is in fact a homogeneous Markov process, or, equivalently, obeys a so-called master equation. Much harder are the questions of whether the central assumptions of this approach might still be compatible with an underlying deterministic time-reversal invariant dynamics, and in which sense the results of the approach embody time-asymmetry. In fact we shall see that conflicting intuitions on this last issue arise, depending on whether one takes a probabilistic or a dynamics point of view towards this formalism.

From a foundational point of view, stochastic dynamics promises a new approach to the explanation of irreversible behaviour that differs in interesting ways from the more orthodox Hamiltonian or dynamical systems approach. In that approach, any account of irreversible phenomena can only proceed by referring to special initial conditions or dynamical hypotheses. Moreover, it is well-known that an ensemble of such systems will conserve (fine-grained) Gibbs entropy so that the account cannot rely on this form of entropy for a derivation of the increase of entropy.

In stochastic dynamics, however, one may hope to find an account of irreversible behaviour that is not tied to special initial conditions, but one that is, so to say, built into the very stochastic-dynamical evolution. Further, since Liouville's theorem is not applicable, there is the prospect that one can obtain a genuine increase of Gibbs entropy from this type of dynamics.

As just mentioned, the central technical assumption in stochastic dynamics is that the processes described have the Markov property.[69] Indeed, general aspects of irreversible behaviour pour out almost effortlessly from the Markov property, or from the closely connected "master equation". Consequently, much of the attention in motivating stochastic dynamics has turned to the assumptions needed to obtain this Markov property, or slightly more strongly, to obtain a non-invertible Markov process (Mackey 1992). The best-known specimen of such an assumption is van Kampen's (1962) "repeated randomness assumption". And similarly, critics of this type of approach (Sklar 1993, Redhead 1995, Callender 1999) have also focused their objections on the question just

[69] Some authors argue that the approach can and should be extended to include non-Markovian stochastic processes as well. Nevertheless I will focus here on Markov processes.

how reasonable and general such assumptions are (cf. paragraph 7.5).

I believe both sides of the debate have badly missed the target. Many authors have uncritically assumed that the assumption of a (non-invertible) Markov process does indeed lead to non-time-reversal-invariant results. As a matter of fact, however, the Markov property (for invertible or non-invertible Markov processes) is time-reversal invariant. So, any argument to obtain that property need not presuppose time-asymmetry. In fact, I will argue that this discussion of irreversible behaviour as derived from the Markov property suffers from an illusion. It is due to the habit of studying the prediction of future states from a given initial state, rather than studying retrodictions towards an earlier state. As we shall see, for a proper description of irreversibility in stochastic dynamics one needs to focus on another issue, namely the difference between backward and forwards transition probabilities.

In the next paragraphs, I will first (§7.2) recall the standard definition of a homogeneous Markov process from the theory of stochastic processes. Paragraph 7.3 casts these concepts in the language of dynamics, introduces the master equation, and discusses its analogy to the Boltzmann equation. In §7.4, we review some of the results that *prima facie* display irreversible behaviour for homogeneous Markov processes. In paragraph 7.5 we turn to the physical motivations that have been given for the Markov property, and their problems, while §7.6 focuses on the question how seemingly irreversible results could have been obtained from a time-symmetric assumptions. Finally, §7.7 argues that a more promising discussion of these issues should start from a different definition of reversibility of stochastic processes.

7.2 The definition of Markov processes

To start off, consider an example. One of the oldest discussions of a stochastic process in the physics literature is the so-called 'dog flea model' of P. and T. Ehrenfest (1907).

Consider N fleas, labeled from 1 to N, situated on either of two dogs. The number of fleas on dog 1 and 2 are denoted as n_1 and $n_2 = N - n_1$. Further, we suppose there is an urn with N lots carrying the numbers $1, \ldots N$ respectively. The urn is shaken, a lot is drawn (and replaced), and the flea with the corresponding label is ordered to jump to the other dog. This procedure is repeated every second.

It is not hard to see that this model embodies an 'approach equilibrium' in some sense: Suppose that initially all or almost all fleas are on dog 1. Then it is very probable that the first few drawings will move fleas from dog 1 to 2. But as soon as the number of fleas on dog 2 increases, the probability that some fleas will jump back to dog 1 increases too. The typical behaviour of, say, $|n_1 - n_2|$ as a function of time will be similar to Boltzmann's H-curve, with a tendency of $|n_1 - n_2|$ to decrease if it was initially large, and to remain close to the 'equilibrium' value $n_1 \approx n_2$ for most of the time. But note that in contrast to Boltzmann's H-curve in gas theory, the 'evolution' is here entirely stochastic,

i.e. generated by a lottery, and that no underlying deterministic equations of motion are provided.

In general, a stochastic process is, mathematically speaking, nothing but a probability measure P on a measure space X, whose elements will here be denoted as ξ, on which there are infinitely many random variables Y_t, with $t \in \mathbb{R}$ (or sometimes $t \in \mathbb{Z}$). Physically speaking, we interpret t as time, and Y as the macroscopic variable(s) characterizing the macrostate —say the number of fleas on a dog, or the number of molecules with their molecular state in some cell of μ-space, etc. Further, ξ represents the total history of the system which determines the values of $Y_t(\xi)$. The collection Y_t may thus be considered as a single random variable Y evolving in the course of time.

At first sight, the name 'process' for a probability measure may seem somewhat unnatural. From a physical point of view it is the *realization*, in which the random variables Y_t attain the values $Y_t(\xi) = y_t$ that should be called a process. In the mathematical literature, however, it has become usual to denote the measure that determines the probability of all such realizations as a 'stochastic process'.

For convenience we assume here that the variables Y_t may attain only finitely many discrete values, say $y_t \in \mathcal{Y} = \{1, \ldots, m\}$. However, the theory can largely be set up in complete analogy for continuous variables.

The probability measure P provides, for $n = 1, 2, \ldots$, and instants $t_1, \ldots, t_n$ definite probabilities for the event that Y_t at these instants attains certain values $y_1, \ldots, y_n$:

$$\begin{array}{c} P_{(1)}(y_1, t_1) \\ P_{(2)}(y_2, t_2; y_1, t_1) \\ \vdots \\ P_{(n)}(y_n, t_n; \ldots; y_1, t_1) \\ \vdots \end{array} \tag{165}$$

Here, $P_{(n)}(y_n, t_n; \ldots; y_1, t_1)$ stands for the joint probability that at times $t_1, \ldots, t_n$ the quantities Y_t attain the values $y_1, \ldots y_n$, with $y_i \in \mathcal{Y}$. It is an abbreviation for

$$P_{(n)}(y_n, t_n; \ldots; y_1, t_1) := P(\{\xi \in X \ : \ Y_{t_n}(\xi) = y_n \ \& \ \cdots \ \& \ Y_{t_1}(\xi) = y_1\}) \tag{166}$$

Obviously the probabilities (165) are normalized and non-negative, and each $P_{(n)}$ is a marginal of all higher-order probability distributions:

$$P_{(n)}(y_n, t_n; \ldots; y_1, t_1) = \sum_{y_{n+m}} \cdots \sum_{y_{n+1}} P_{(n+m)}(y_{n+m}, t_{n+m}; \ldots; y_1, t_1). \tag{167}$$

In fact, the probability measure P is uniquely determined by the hierarchy (165).[70]

Similarly, we may define conditional probabilities in the familiar manner, e.g.:

$$P_{(1|n-1)}(y_n, t_n | y_{n-1}, t_{n-1}; \ldots; y_1, t_1) := \frac{P_{(n)}(y_n, t_n; \ldots; y_1, t_1)}{P_{(n-1)}(y_{n-1}, t_{n-1}; \ldots; y_1, t_1)} \tag{168}$$

provides the probability that Y_{t_n} attains the value y_n, under the condition that $Y_{t_{n-1}}, \ldots, Y_{t_1}$ have the values $y_{n-1}, \ldots, y_1$.

In principle, the times appearing in the joint and conditional probability distributions (165,168) may be chosen in an arbitrary order. However, we adopt from now on the convention that they are ordered as $t_1 < \cdots < t_n$.

A special and important type of stochastic process is obtained by adding the assumption that such conditional probabilities depend only the condition at the last instant. That is to say: for all n and all choices of $y_1, \ldots y_n$ and $t_1 < \ldots < t_n$:

$$P_{(1|n)}(y_n, t_n | y_{n-1}, t_{n-1}; \ldots; y_n, t_n) = P_{(1|1)}(y_n, t_n | y_{n-1}, t_{n-1}) \tag{169}$$

This is the *Markov property* and such stochastic processes are called *Markov processes*.

The interpretation often given to this assumption, is that Markov processes have 'no memory'. To explain this slogan more precisely, consider the following situation. Suppose we are given a piece of the history of the quantity Y: at the instants $t_1, \ldots, t_{n-1}$ its values have been $y_1, \ldots, y_{n-1}$. On this information, we want to make a prediction of the value y_n of the variable Y at a later instant t_n. The Markov-property (169) says that this prediction would not have been better or worse if, instead of knowing this entire piece of prehistory, only the value y_{n-1} of Y at the last instant t_{n-1} had been given. Additional information about the past values is thus irrelevant for a prediction of the future value.

For a Markov process, the hierarchy of joint probability distributions (165) is subjected to stringent demands. In fact they are all completely determined by: (a) the specification of $P_{(1)}(y, 0)$ at one arbitrary chosen initial instant $t = 0$, and (b) the conditional probabilities $P_{(1|1)}(y_2, t_2 | y_1, t_1)$ for all $t_2 > t_1$. Indeed,

$$P_{(1)}(y, t) \quad = \quad \sum_{y_0} P_{(1|1)}(y, t | y_0, 0) P_{(1)}(y_0, 0); \tag{170}$$

[70] At least, when we assume that the σ-algebra of measurable sets in X is the cylinder algebra generated by sets of the form in the right-hand side of (166).

and for the joint probability distributions $P_{(n)}$ we find:

$$\begin{aligned} P_{(n)}(y_n, t_n; \ldots; y_1, t_1) &= P_{(1|1)}(y_n, t_n|y_{n-1}, t_{n-1})\, P_{(1|1)}(y_{n-1}, t_{n-1}|y_{n-2}, t_{n-2}) \times \\ &\quad \times \cdots \times P_{(1|1)}(y_2, t_2|y_1, t_1) P_{(1)}(y_1, t_1). \end{aligned} \tag{171}$$

It follows from the Markov property that the conditional probabilities $P_{(1|1)}$ have the following property, known as the *Chapman-Kolmogorov* equation:

$$P_{(1|1)}(y_3, t_3|y_1, t_1) = \sum_{y_2} P_{(1|1)}(y_3, t_3|y_2, t_2) P_{(1|1)}(y_2, t_2|y_1, t_1) \quad \text{for } t_1 < t_2 < t_3. \tag{172}$$

So, for a Markov process, the hierarchy (165) is completely characterized by specifying $P_{(1)}$ at an initial instant and a system of conditional probabilities $P_{(1|1)}$ satisfying the Chapman-Kolmogorov equation. The study of Markov processes therefore focuses on these two ingredients.[71]

A following special assumption is *homogeneity*. A Markov process is called homogeneous if the conditional probabilities $P_{(1|1)}(y_2, t_2|y_1, t_1)$ do not depend on the two times t_1, t_2 separately but only on their mutual difference $t = t_2 - t_1$; i.e. if they are invariant under time translations. In this case we may write

$$P_{(1|1)}(y_2, t_2|y_1, t_1) = T_t(y_2, y_1) \tag{173}$$

such conditional probabilities are also called *transition* probabilities.

Is the definition of a Markov process time-symmetric? The choice in (169) of conditionalizing the probability distribution for Y_{t_n} on *earlier* values of Y_t is of course special. In principle, there is nothing in the formulas (165) or (168) that forces such an ordering. One might, just as well, ask for the probability of a value of Y_t in the past, under the condition that part of the *later* behaviour is given (or, indeed, conditionalize on the behaviour at both earlier and later instants.)

At first sight, the Markov property makes no demands about these latter cases. Therefore, one might easily get the impression that the definition is time-asymmetric. However, this is not the case. One can show that (169) is equivalent to:

$$P_{(1|n-1)}(y_1, t_1|y_2, t_2; \ldots; y_n, t_n) = P_{(1|1)}(y_1 t_1|y_2, t_2) \tag{174}$$

where the convention $t_1 < t_2 < \cdots < t_n$ is still in force. Thus, a Markov process does not only have 'no memory' but also 'no foresight'. Some authors (e.g. Kelly 1979) adopt an (equivalent)

[71] Note, however, that although every Markov process is fully characterized by (i) an initial distribution $P_{(1)}(y, 0)$ and (ii) a set of transition probabilities $P_{(1|1)}$ obeying the Chapman-Kolmogorov equation and the equations (171), it is *not* the case that every stochastic process obeying (i) and (ii) is a Markov process. (See (van Kampen 1981, p. 83) for a counterexample). Still, it is true that one can define a unique Markov process from these two ingredients by stipulating (171).

definition of a Markov process that is explicitly time-symmetric: Suppose that the value y_i at an instant t_i somewhere in the middle of the sequence $t_1 < \cdots < t_n$ is given. The condition for a stochastic process to be Markov is then

$$P_{(n|1)}(y_n, t_n; \ldots; y_1, t_1|y_i, t_i) = \\ P_{(n-i|1)}(y_n, t_n; \ldots; y_{i+1}, t_{i+1}|y_i, t_i)\, P_{(i-1|1)}(y_{i-1}, t_{i-1}; y_1, t_1|y_i, t_i) \tag{175}$$

for all $n = 1, 2, \ldots$ and all $1 \leq i \leq n$. In another slogan: The future and past are independent if one conditionalizes on the present.

7.3 Stochastic dynamics

A homogeneous Markov process is for $t > 0$ completely determined by the specification of an initial probability distribution $P_{(1)}(y, 0)$ and the transition probabilities $T_t(y_2|y_1)$ defined by (173). The difference in notation (between P and T) also serves to ease a certain conceptual step. Namely, the idea is to regard T_t as a stochastic evolution operator. Thus, we can regard $T_t(y_2|y_1)$ as the elements of a matrix, representing a (linear) operator T that determines how an initial distribution $P_{(1)}(y, 0)$ will evolve into a distribution at later instants $t > 0$. (In the sequel I will adapt the notation and write $P_{(1)}(y, t)$ as $P_t(y)$.)

$$P_t(y) = (T_t P)(y) := \sum_{y'} T_t(y|y') P_0(y'). \tag{176}$$

The Chapman-Kolmogorov equation (172) may then be written compactly as

$$T_{t+t'} = T_t \circ T_{t'} \quad \text{for } t, t' \geq 0 \tag{177}$$

where $\circ$ stands for matrix multiplication, and we now also extend the notation to include the unit operator:

$$\mathbb{1}(y, y') = T_0(y, y') := \delta_{y,y'} \tag{178}$$

where δ denotes the Kronecker delta.

The formulation (177) can (almost) be interpreted as the group composition property of the evolution operators T. It may be instructive to note how much this is due to the Markov property. Indeed, for arbitrary conditional probabilities, say, if A_i, B_j and C_k denote three families of complete and mutually exclusive events (i.e. $\cup_i A_i = \cup_j B_j = \cup_k C_k = \mathcal{Y}$; $A_i \cap A_{i'} = B_j \cap B_{j'} = C_k \cap C_{k'} = \emptyset$ for $i \neq i', j \neq j'$ and $k \neq k'$), the rule of total probability gives :

$$P(A_i|C_k) = \sum_j P(A_i|B_j, C_k) P(B_j|C_k). \tag{179}$$

In general, this rule can *not* be regarded as ordinary matrix multiplication or a group composition! But the Markov property makes $P(A_i|B_j, C_k)$ in (179) reduce to $P(A_i|B_j)$, and then the summation in (179) coincides with familiar rule for matrix multiplication.

I wrote above: 'almost', because there is still a difference in comparison with the normal group property: in the Chapman-Kolmogorov-equation (177) all times must be positive. Thus, in general, for $t > 0$, T_t may not even be defined and so it does *not* hold that

$$T_{-t} \circ T_t = \mathbb{1}. \tag{180}$$

A family of operators $\{T_t, t \geq 0\}$ which is closed under a operation $\circ$ that obeys (177), and for which $T_0 = \mathbb{1}$ is called a *semigroup*. It differs from a group in the sense that its elements T_t need not be *invertible*, i.e., need not have an inverse. The lack of an inverse of T_t may be due to various reasons: either T_t does not possess an inverse, i.e. it is not a one-to-one mapping, or T_t does possess an inverse matrix T_t^{inv}, which however is itself non-stochastic (e.g. it may have negative matrix-elements). We will come back to the role of the inverse matrices in Sections 7.4 and 7.7.

The theory of Markov processes has a strong and natural connection with linear algebra. Sometimes, the theory is presented entirely from this perspective, and one starts with the introduction of a semigroup of *stochastic matrices*, that is to say, m by m matrices T with $T_{ij} \geq 0$ and $\sum_i T_{ij} = 1$. Or, more abstractly, one posits a class of states P, elements of a Banach space with a norm $\|P\|_1 = 1$, and a semigroup of stochastic maps T_t, ($t \geq 0$), subject to the conditions that T_t is linear, positive, and preserves norm: $\|T_t P\|_1 = \|P\|_1$, (cf. Streater 1995).

The evolution of a probability distribution P (now regarded as a vector or a state) is then particularly simple when t is discrete ($t \in \mathbb{N}$):

$$P_t = T^t P_0, \quad \text{where } T^t = \underbrace{T \circ \cdots \circ T}_{t \text{ times}}. \tag{181}$$

Homogeneous Markov processes in discrete time are also known as *Markov chains*.

Clearly, if we consider the family $\{T_t\}$ as a semigroup of stochastic evolution operators, or a stochastic form of dynamics, it becomes attractive to look upon $P_0(y)$ as a contingent initial state, chosen independently of the evolution operators T_t. Still, from the perspective of the probabilistic formalism with which we started, this might be an unexpected thought: both $P_{(1)}$ and $P_{(1|1)}$ are aspects of a single, given, probability measure P. The idea of regarding them as independent ingredients that may be specified separately doesn't then seem very natural. But, of course, there is no formal objection against the idea, since every combination of a system of transition probabilities T_t obeying the Chapman-Kolmogorov equation, and an arbitrary initial probability distribution $P_0(y) = P_{(1)}(y, 0)$ defines a unique homogeneous Markov process (cf. footnote 71). In fact, one

sometimes even goes one step further and identifies a homogeneous Markov process completely with the specification of the transition probabilities, without regard of the initial state $P_0(y)$; just like the dynamics of a deterministic system is usually presented without assuming any special initial state.

For Markov chains, the goal of specifying the evolution of $P_t(y)$ is now already completely solved in equation (181). In the case of continuous time, it is more usual to specify evolution by means of a differential equation. Such an equation may be obtained in a straightforward manner by considering a Taylor expansion of the transition probability for small times (van Kampen 1981, p.101–103)— under an appropriate continuity assumption.

The result (with a slightly changed notation) is:

$$\frac{\partial P_t(y)}{\partial t} = \sum_{y'} \left(W(y|y') P_t(y') - W(y'|y) P_t(y) \right) \tag{182}$$

Here, the expression $W(y|y')$ is the transition probability from y' to y *per unit of time*. This differential equation, first obtained by Pauli in 1928, is called the *master equation*. (This name has become popular because an equation of this type covers a great variety of processes.)

The interpretation of the equation is suggestive: the change of the probability $P_t(y)$ is determined by making up a balance between gains and losses: the probability of value y increases in a time dt because of the transitions from y' to y, for all possible values of y'. This increase per unit of time is $\sum_{y'} W(y|y') P_t(y')$. But in same period dt there is also a decrease of $P_t(y)$ as a consequence of transitions from the value y to all other possible values y'. This provides the second term.

In this "balancing" aspect, the master equation resembles the Boltzmann equation (48), despite the totally different derivation, and the fact that $P_t(y)$ has quite another meaning than Boltzmann's $f_t(v)$. (The former is a probability distribution, the latter a distribution of particles.) Both are first-order differential equations in t. A crucial mathematical distinction from the Boltzmann equation is that the master equation is linear in P, and therefore much easier to solve.

Indeed, any solution of the master equation can formally be written as:

$$P_t = e^{tL} P_0, \tag{183}$$

where L represents the operator

$$L(y|y') := W(y|y') - \sum_{y''} W(y''|y') \delta_{y,y'}. \tag{184}$$

The general solution (183) is similar to the discrete time case (181), thus showing the equivalence of the master equation to the assumption of a homogeneous Markov process in continuous time.

A final remark(not needed for later paragraphs). The analogy with the Boltzmann equation can

even be increased by considering a Markov process for particle pairs, i.e. by imagining a process where pairs of particles with initial states (i, j) make a transition to states (k, l) with certain transition probabilities (cf. Alberti & Uhlmann 1982, p. 30) Let $W(i, j|k, l)$ denote the associated transition probability per unit of time. Then the master equation takes the form:

$$\frac{\partial P_t(i,j)}{\partial t} = \sum_{k,l} \left(W(i,j|k,l)P_t(k,l) - W(k,l|i,j)P_t(i,j)\right). \tag{185}$$

Assume now that the transitions $(i, j) \longrightarrow (k, l)$ and $(k, l) \longrightarrow (i, j)$ are equally probable, so that the transition probability per unit of time is symmetric: $W(i, j|k, l) = W(k, l|i, j)$, and, as an analogue to the *Stoßzahlansatz*, that $P(i, j)$ in the right-hand side may be replaced by the product of its marginals:

$$P(i,j) \longrightarrow \sum_j P(i,j) \cdot \sum_i P(i,j) = P'(i)P''(j) \tag{186}$$

Summing the above equation (185) over j, we finally obtain

$$\frac{\partial P'_t(i)}{\partial t} = \sum_j \frac{\partial P_t(i,j)}{\partial t} = \sum_{j,k,l} T(i,j|k,l)\big(P'_t(k)P''_t(l) - P'_t(i)P''_t(j)\big), \tag{187}$$

i.e., an even more striking analogue of the Boltzmann equation (48). But note that although (185) describes a Markov process, the last equation (187) does not: it is no longer linear in P, as a consequence of the substitution (186).

7.4 Approach to equilibrium and increase of entropy?

What can we say in general about the evolution of $P_t(y)$ for a homogeneous Markov process? An immediate result is this: the *relative entropy* is monotonically non-decreasing. That is to say, if we define

$$H(P,Q) := -\sum_{y\in\mathcal{Y}} P(y) \ln \frac{P(y)}{Q(y)} \tag{188}$$

as the relative entropy of a probability distribution P relative to Q, then one can show (see e.g. Moran 1961; Mackey 1991, p. 30):

$$H(P_t, Q_t) \geq H(P, Q) \tag{189}$$

where $P_t = T_t P$, $Q_t = T_t Q$, and T_t are elements of the semigroup (181) or (183).

One can also show that a non-zero relative entropy increase for at least some pair probability distributions P and Q, the stochastic matrix T_t must be non-invertible.

The relative entropy $H(P|Q)$ can, in some sense, be thought of as a measure of how much P and

Q "resemble" each other.[72] Indeed, it takes its maximum value (i.e. 0) if and only if $P = Q$; it may become $-\infty$ if P and Q have disjoint support, (i.e. when $P(y)Q(y) = 0$ for all $y \in \mathcal{Y}$.) Thus, the result (189) says that if the stochastic process is non-invertible, pairs of distributions P_t and Q_t will generally become more and more alike as time goes by.

Hence it seems we have obtained a general weak aspect of "irreversible behaviour" in this framework. Of course, the above result does not yet imply that the 'absolute' entropy $H(P) := -\sum_y P(y) \ln P(y)$ of a probability distribution is non-decreasing. But now assume that the process has a *stationary state*. In other words, there is a probability distribution $P^*(y)$ such that

$$T_t P^* = P^*. \tag{190}$$

The intention is, obviously, to regard such a distribution as a candidate for the description of an equilibrium state. If there is such a stationary distribution P^*, we may apply the previous result and write:

$$H(P, P^*) \leq H(T_t P, T_t P^*) = H(P_t, P^*). \tag{191}$$

In other words, as time goes by, the distribution $T_t P$ will then more and more resemble the stationary distribution than does P. If the stationary distribution is also uniform, i.e.:

$$P^*(y) = \frac{1}{m}, \tag{192}$$

then not only the relative but also the absolute entropy $H(P) := -\sum_y P(y) \ln P(y)$ increases, because

$$H(P, P^*) = H(P) - \ln m. \tag{193}$$

In order to get a satisfactory description of an 'approach to equilibrium' the following questions remain:

(i) is there such a stationary distribution?

(ii) If so, is it unique?

(iii) does the monotonic behaviour of $H(P_t)$ imply that $\lim_{t \longrightarrow \infty} P_t = P^*$?

Harder questions, which we postpone to the next subsection 7.5, are:

(iv) how to motivate the assumptions needed in this approach or how to make judge their (in)compatibility with an underlying time deterministic dynamics; and

(v) how this behaviour is compatible with the time symmetry of Markov processes.

[72] Of course, this is an asymmetric sense of "resemblance" because $H(P, Q) \neq H(Q, P)$.

Ad (i). A stationary state as defined by (190), can be seen as an eigenvector of T_t with eigenvalue 1, or, in the light of (183), an eigenvector of L for the eigenvalue 0. Note that T or L are not necessarily Hermitian (or, rather, since we are dealing with real matrices, symmetric), so that the existence of eigenvectors is not guaranteed by the spectral theorem. Further, even if an eigenvector with the corresponding eigenvalue exists, it is not automatically suitable as a probability distribution because its components might not be positive.

Still, it turns out that, due to a theorem of Perron (1907) and Frobenius (1909), every stochastic matrix indeed has a eigenvector, with exclusively non-negative components, and eigenvalue 1 (see e.g. Gantmacher 1959, Van Harn & Holewijn 1991). But if the set $\mathcal{Y}$ is infinite or continuous this is not always true.

A well-known example of the latter case is the so-called Wiener process that is often used for the description of Brownian motion. It is characterized by the transition probability density:

$$T_t(y|y') = \frac{1}{\sqrt{2\pi t}} \exp \frac{(y-y')^2}{2t}, \quad y, y' \in \mathbb{R}. \tag{194}$$

The evolution of an arbitrary initial probability density ρ_0 can be written as a convolution:

$$\rho_t(y) = \int T_t(y|y')\rho_0(y')dy'; \tag{195}$$

which becomes gradually lower, smoother and wider in the course of time, but does not approach any stationary probability density. Because this holds for every choice of ρ_0, there is no stationary distribution in this case.

However, it is not reasonable to see this as a serious defect. Indeed, in thermodynamics too one finds that a plume of gas emitted into free space will similarly diffuse, becoming ever more dilute without ever approaching an equilibrium state. Thermodynamic equilibrium is only approached for systems enclosed in a vessel of finite volume.

However, for continuous variables with a range that has finite measure, the existence of a stationary distribution is guaranteed under the condition that the probability density ρ_y is at all times bounded, i.e. $\exists M \in \mathbb{R}$ such that $\forall t\ \rho_t \leq M$; (see Mackey 1992, p. 36).

Ad (ii). The question whether stationary solutions will be unique is somewhat harder to tackle. This problem exhibits an analogy to that of metric transitivity in the ergodic problem (cf. paragraph 6.1).

In general, it is very well possible that the range $\mathcal{Y}$ of Y can be partitioned in two disjoint regions, say A and B, with $\mathcal{Y} = A \cup B$, such that there are no transitions from A to B or vice versa (or that such transitions occur with probability zero). That is to say, the stochastic evolution T_t might have

the property

$$T_t(Y \in A|Y \in B) = T_t(Y \in B|Y \in A) = 0 \tag{196}$$

In other words, its matrix may, (perhaps after a conventional relabeling of the outcomes) be written in the form:

$$\begin{pmatrix} T_A & 0 \\ 0 & T_B \end{pmatrix}. \tag{197}$$

The matrix is then called (completely) *reducible*. In this case, stationary distributions will generally not be unique: If P_A^* is a stationary distribution with support in the region A, and P_B^* is a stationary distribution with support in B, then every convex combination

$$\alpha P_A^*(y) + (1-\alpha)P_B^*(y) \quad \text{with } 0 \leq \alpha \leq 1. \tag{198}$$

will be stationary too. In order to obtain a unique stationary solution we will thus have to assume an analogue of metric transitivity. That is to say: we should demand that every partition of $\mathcal{Y}$ into disjoint sets A and B for which (196) holds is 'trivial' in the sense that $P(A) = 0$ or $P(B) = 0$.

So, one may ask, is the stationary distribution P^* unique if and only if the transition probabilities T_τ are not reducible? In the ergodic problem, as we saw in 6.1, the answer is positive (at least if P^* is assumed to be absolutely continuous with respect to the microcanonical measure). But not in the present case!

This has to do with the phenomenon of so-called 'transient states', which has no analogy in Hamiltonian dynamics. Let us look at an example to introduce this concept. Consider a stochastic matrix of the form:

$$\begin{pmatrix} T_A & T' \\ 0 & T_B \end{pmatrix} \tag{199}$$

where T' is a matrix with non-negative entries only. Then:

$$\begin{pmatrix} T_A & T' \\ 0 & T_B \end{pmatrix} \begin{pmatrix} P_A \\ 0 \end{pmatrix} = \begin{pmatrix} T_A P_A \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} T_A & T' \\ 0 & T_B \end{pmatrix} \begin{pmatrix} 0 \\ P_B \end{pmatrix} = \begin{pmatrix} T' P_B \\ T_B P_B \end{pmatrix} \tag{200}$$

so that here transitions of the type $a \longrightarrow b$ have probability zero, but transitions of the type $b \longrightarrow a$ occur with positive probability. (Here, a, b stand for arbitrary elements of the subsets A and B.) It is clear that in such a case the region B will eventually be 'sucked empty'. That is to say: the total probability of being in region B (i.e. $\|T^t P_B\|$) will go exponentially to zero. The distributions with support in B are called 'transient' and the set A is called 'absorbing' or a 'trap'.

It is clear that these transient states will not play any role in the determination of the stationary distribution, and that for this purpose they might be simply ignored. Thus, in this example, the only

stationary states are those with a support in A. And there will be more than one of them if T_A is reducible.

A matrix T that may be brought (by permutation of the rows and columns) in the form (199), with T_A reducible is called *incompletely reducible* (van Kampen 1981, p. 108). Further, a stochastic matrix is called *irreducible* if it is neither completely or incompletely reducible. An alternative (equivalent) criterion is that all states 'communicate' with each other, i.e. that for every pair of $i, j \in \mathcal{Y}$ there is some time t such that $P_t(j|i) > 0$.

The Perron-Frobenius theorem guarantees that as long as T irreducible, there is a unique stationary distribution. Furthermore, one can then prove an analogue of the ergodic theorem:(Petersen 1983, p. 52)

> ERGODIC THEOREM FOR MARKOV PROCESSES: If the transition probability T_t is irreducible, the time average of P_t converges to the unique stationary solution:
>
> $$\lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau T_t P(y) dt = P^*(y). \tag{201}$$

Ad (iii). If there is a unique stationary distribution P^*, will $T_t P$ converge to P^*, for every choice of P? Again, the answer is not necessarily affirmative. (Even if (201) is valid!) For example, there are homogeneous and irreducible Markov chains for which P_t can be divided into two pieces: $P_t = Q_t + R_t$ with the following properties (Mackey 1992, p. 71):

1. Q_t is a term with $\|Q_t\| \longrightarrow 0$. This is a transient term.
2. The remainder R_t is periodic, i.e. after some finite time τ the evolution repeats itself: $R_{t+\tau} = R_\tau$.

These processes are called *asymptotically periodic*. They may very well occur in conjunction with a unique stationary distribution P^*, and show strict monotonic increase of entropy, but still not converge to P^*. In this case, the monotonic increase of relative entropy $H(P_t, P^*)$ is entirely due to the transient term. For the periodic piece R_t, the transition probabilities are permutation matrices, which, after τ repetitions, return to the unit operator.

Besides, if we arrange that P^* is uniform, we can say even more in this example: The various forms R_t that are attained during the cycle of permutations with period τ all have the same value for the relative entropy $H(R_t, P^*)$, but this entropy is strictly lower than $H(P^*, P^*) = 0$. In fact, P^* is the average of the R_t's, i.e.: $P^* = \frac{1}{\tau} \sum_{t=1}^{t=\tau} R_t$, in correspondence with (201).

Further technical assumptions can be introduced to block examples of this kind, and thus enforce a strict convergence towards the unique stationary distribution, e.g. by imposing a condition of 'exactness' (Mackey 1992). However, it would take us too far afield to discuss this in detail.

In conclusion, it seems that a weak aspect of "irreversible behaviour", i.e. the monotonic non-decrease of relative entropy is a general feature for all homogeneous Markov processes, (and indeed for all stochastic processes), and non-trivially so when the transition probabilities are non-invertible. Stronger versions of that behaviour, in the sense of affirmative answers to the questions (i), (ii) and (iii), can be obtained too, but at the price of additional technical assumptions.

7.5 Motivations for the Markov property and objections against them

(ad iv). We now turn to the following problem: what is the motivation behind the assumption of the Markov property? The answer, of course, is going to depend on the interpretation of the formalism that one has in mind, and may be different in the 'coarse-graining' and the 'open systems' or interventionist approaches (cf. Section 7.1). I shall discuss the coarse-graining approach in the next paragraph below, and then consider the similar problem for the interventionist point of view .

7.5.1 Coarse-graining and the repeated randomness assumption

In the present point of view, one assumes that the system considered is really an isolated Hamiltonian system, but the Markov property is supposedly obtained from a partitioning of its phase space. But exactly how is that achieved?

One of the clearest and most outspoken presentations of this view is (van Kampen 1962). As in paragraph 5.4, we assume the existence of some privileged partition of the Hamiltonian phase space Γ —or of the energy hypersurface Γ_E— into disjoint cells: $\Gamma = \omega_1 \cup \cdots \cup \omega_m$. Consider an arbitrary ensemble with probability density ρ on this phase space. Its evolution can be represented by an operator

$$U_t^* \rho(x) := \rho(U_{-t}x), \tag{202}$$

where, —in order to avoid conflation of notation— we now use U_t to denote the Hamiltonian evolution operators, previously denoted as T_t, e.g. in (68) and throughout section 6. Let transition probabilities between the cells of this partition be defined as

$$T_t(j|i) := P(x_t \in \omega_j | x \in \omega_i) = P(U_t x \in \omega_j | x \in \omega_i) = \frac{\int_{(U_{-t}\omega_j)\cap\omega_i} \rho(x)dx}{\int_{\omega_i} \rho(x)dx}, \tag{203}$$

Obviously such transition probabilities will be homogeneous, due to the time-translation invariance of the Hamiltonian evolution U_t. Further, let $\hat{p}_0(i) := P(x \in \omega_i) = \int_{\omega_i} \rho(x)dx$, $i \in \mathcal{Y} = \{1, \ldots, m\}$, be an arbitrary initial coarse-grained probability distribution at time t=0.

Using the coarse-graining map defined by (92), one may also express the coarse-grained distri-

bution at time t as

$$\mathcal{CGU}_t^* \rho(x) = \sum_{ji} T_t(j|i)\hat{p}_0(i) \frac{1}{\mu(\omega_j)} \mathbb{1}_{\omega_j}(x) \tag{204}$$

where μ is the canonical measure on Γ, or the microcanonical measure on Γ_E. This expression indicates that, as long as we are only interested in the coarse grained history, it suffices to know the transition probabilities (203) and the initial coarse grained distributions.

But in order to taste the fruits advertised in the previous paragraphs, one needs to show that the transition probabilities define a Markov process, i.e., that they obey the Chapman-Kolmogorov equation (172),

$$T_{t'+t}(k|i) = T_{t'}(k|j)T_t(j|i); \text{ for all } t, t' > 0. \tag{205}$$

Applying (204) for times t, t' and $t + t'$, it follows easily that the Chapman-Kolmogorov equation is equivalent to

$$\mathcal{CGU}_{t'+t}^* = \mathcal{CGU}_{t'}^* \mathcal{CGU}_t^*, \text{ for all } t, t' > 0. \tag{206}$$

In other words, the coarse-grained probability distribution at time $t + t'$ can be obtained by first applying the Hamiltonian dynamical evolution during a time t, then performing a coarse-graining operation, next applying the dynamical evolution during time t', and then coarse-graining again. In comparison to the relation $U_{t'+t}^* = U_{t'}^* U_t^*$, we see that the Chapman-Kolmogorov condition can be obtained by demanding that it is allowed to apply a coarse-graining, i.e. to reshuffle the phase points within each cell at any intermediate stage of the evolution. Of course, this coarse-graining halfway during the evolution erases all information about the past evolution apart from the label of the cell where the state is located at that time; and this ties in nicely with the view of the Markov property as having no memory (cf. the discussion on p. 125).

What is more, the *repeated* application of the coarse-graining does lead to a monotonic non-decrease of the Gibbs entropy: If, for simplicity, we divide a time interval into m segments of duration τ, we have

$$\rho_{m\tau} = \underbrace{\mathcal{CGU}_\tau^* \mathcal{CGU}_\tau^* \cdots \mathcal{CGU}_\tau^*}_{m \text{ times}} \rho \tag{207}$$

and from (96):

$$\sigma[\rho_{m\tau}] \geq \sigma[\rho_{(m-1)\tau}] \geq \ldots \geq \sigma[\rho_\tau] \geq \sigma[\rho_0]. \tag{208}$$

But since the choice of τ is arbitrary, we may conclude that $\sigma[\rho_t]$ is monotonically non-decreasing.

Thus, van Kampen argues, the ingredient to be added to the dynamical evolution is that, at any stage of the evolution, one should apply a coarse-graining of the distribution. It is important to note that it is not sufficient to do that just once at a single instant. At every stage of the evolution we need to coarse-grain the distribution again and again. Van Kampen (1962, p. 193) calls this the *repeated*

randomness assumption.

What is the justification for this assumption? Van Kampen points out that it is "not unreasonable" (ibid., p. 182), because of the brute fact of its success in phenomenological physics. Thermodynamics and other phenomenological descriptions of macroscopic systems (the diffusion equation, transport equations, hydrodynamics, the Fokker-Planck equation, etc.) all characterize macroscopic systems with a very small number of variables. This means that their state descriptions are very coarse in comparison with the microscopic phase space. But their evolution equations are autonomous and deterministic: the change of the macroscopic variables is given in terms of the instantaneous values of those very same variables. The success of these equations shows, apparently, that the precise microscopic state does not add any relevant information beyond this coarse description. At the same time, van Kampen admits that the coarse-graining procedure is clearly not always successful. It is not difficult to construct a partition of a phase space into cells for which the Markov property fails completely.

Apparently, the choice of the cells must be "just right" (van Kampen 1962, p. 183). But there is as yet no clear prescription how this is to be done. Van Kampen (1981, p. 80) argues that it is "the art of the physicist" to find the right choice, an art in which he or she succeeds in practice by a mixture of general principles and ingenuity, but where no general guidelines can be provided. The justification of the repeated randomness assumption is that it leads to the Markov property and from there onwards to the master equation, providing a successful autonomous, deterministic description of the evolution of the coarse-grained distribution.

It is worth noting that van Kampen thus qualifies the 'usual' point of view (cf. p. 58 above, and paragraph 5.4) on the choice of the cells; namely, that the cells are chosen in correspondence to our finite observation capabilities. Observability of the macroscopic variables is not sufficient for the success of the repeated randomness assumption. It is conceivable (and occurs in practice) that a particular partition in terms of observable quantities does not lead to a Markov process. In that case, the choice of observable variables is simply inadequate and has to be extended with other (unobservable) quantities until we (hopefully) obtain an exhaustive set, i.e. a set of variables for which the evolution can be described autonomously. An example is the spin-echo experiment: the (observable) total magnetization of the system does not provide a suitable coarse-grained description. For further discussion of this theme, see: (Blatt 1959, Ridderbos & Redhead 1998, Lavis 2004, Balian 2005).

Apart from the unsolved problem for which partition the repeated randomness assumption is to be applied, other objections have been raised against the repeated randomness assumption. Van Kampen actually gives us not much more than the advice to accept the repeated randomness assumption bravely, not to be distracted by its dubious status, and firmly keep our eyes on its success. For

authors as Sklar (1993), who refers to the assumption as a "rerandomization posit", this puts the problem on its head. They request a justification of the assumption that would *explain* the success of the approach. (Indeed, even van Kampen (1981, p. 80) describes this success as a "miraculous fact"!). Such a request, of course, will not be satisfied by a justification that relies on its success. (But that does not mean, in my opinion, that it is an invalid form of justification.)

Another point that seems repugnant to many authors, is that the repeated coarse-graining operations appear to be added 'by hand', in deviation from the true dynamical evolution provided by U_t. The increase of entropy and the approach to equilibrium would thus apparently be a consequence of the fact that *we* shake up the probability distribution repeatedly in order to wash away all information about the past, while refusing a dynamical explanation for this procedure. Redhead (1995, p. 31) describes this procedure as "one of the most deceitful artifices I have ever come across in theoretical physics" (see also Blatt (1959) Sklar (1993) and Callender (1999) for similar objections).

One might ask whether the contrast between the repeated randomness assumption and the dynamical evolution need so bleak as Van Kampen and his critics argue. After all, as we have seen in paragraph 6.2.3, there are dynamical systems so high in the ergodic hierarchy that they possess the Bernoulli property for some partition of phase space (cf. paragraph 6.2.3). Since the Markov property is weaker than the Bernoulli property, one may infer there are also dynamical systems whose coarse grained evolutions define a homogeneous Markov process.[73] Thus one might be tempted to argue that the Markov property, or the repeated randomness assumption proposed to motivate it, need not require a miraculous intervention from an external 'hand' that throws information away; a sufficiently complex deterministic dynamics on the microscopic phase space of the system might do the job all by itself. However, the properties distinguished in the ergodic hierarchy all rely on a given measure-preserving evolution. Thus, while some dynamical systems may have the Markov property, they only give rise to *stationary* Markov processes. Its measure-preserving dynamics still implies that the Gibbs entropy remains constant. Thus, the result (208) can only be obtained in the case when all inequality signs reduce to equalities. To obtain a non-trivial form of coarse-graining, we should indeed suspend the measure-preserving dynamics.

In conclusion,!!! although the choice of a privileged partition remains an unsolved problem, there need not be a conflict between the repeated randomness assumption and the deterministic character of the dynamics at the microscopic level. However, whether the assumption (206) might actually hold for Hamiltonian systems interesting for statistical mechanics is, as far as I know, still open.

[73] Strictly speaking this is true only for discrete dynamical systems. For continuous time, e.g. for Hamiltonian dynamics, the Markov property can only be obtained by adding a time smoothing procedure to the repeated randomness assumption (Emch 1965),(Emch & Liu 2001, pp. 484–486).

7.5.2 Interventionism or 'open systems'

Another approach to stochastic dynamics is by reference to open systems. The idea is here that the system in continual interaction with the environment, and that this is responsible for the approach to equilibrium.

Indeed, it cannot be denied that in concrete systems isolation is an unrealistic idealization. The actual effect of interaction with the environment on the microscopic evolution can be enormous. A proverbial example, going back to Borel (1914), estimates the gravitational effect caused by displacing one gram of matter on Sirius by one centimeter on the microscopic evolution of an earthly cylinder of gas. Under normal conditions, the effect is so large, that, roughly and for a typical molecule in the gas, it may be decisive for whether or not this molecule will hit another given molecule after about 50 intermediary collisions. That is to say: microscopic dynamical evolutions corresponding to the displaced and the undisplaced matter on Sirius start to diverge considerably after a time of about 10^{-6} sec. In other words, the mechanical evolution of such a system is so extremely sensitive for disturbances of the initial state that even the most minute changes in the state of the environment can be responsible for large changes in the microscopic trajectory. But we cannot control the state of environment. Is it possible to regard irreversible behaviour as the result of such uncontrollable disturbances from outside?[74]

Let (x, y) be the state of a total system, where, as before, $x \in \Gamma^{(\mathrm{s})}$ represents the state of the object system and $y \in \Gamma^{(\mathrm{e})}$ that of the environment. We assume that the total system is governed by a Hamiltonian of the form

$$H_{\mathrm{tot}}(x, y) = H_{(\mathrm{s})} + H_{(\mathrm{e})} + \lambda H_{\mathrm{int}}(x, y), \tag{209}$$

so that the probability density of the ensemble of total systems evolves as

$$\rho_t(x, y) = U_t^* \rho_0(x, y) = \rho\left(U_{-t}(x, y)\right) \tag{210}$$

i.e., a time-symmetric, deterministic and measure-preserving evolution.

At each time, we may define marginal distributions for both system and environment:

$$\rho_t^{(\mathrm{s})}(x) = \int dy\, \rho_t(x, y), \tag{211}$$

$$\rho_t^{(\mathrm{e})}(x) = \int dx\, \rho_t(x, y). \tag{212}$$

[74] Note that the term 'open system' is employed here for a system in (weak) interaction with its environment. This should be distinguished from the notion of 'open system' in other branches of physics where it denotes a system that can exchange particles with its environment.

We are, of course, mostly interested in the object system, i.e. in (211). Assume further that at time $t = 0$ the total density factorizes:

$$\rho_0(x, y) = \rho_0^{(\mathrm{s})}(x)\rho_0^{(\mathrm{e})}(y). \tag{213}$$

What can we say about the evolution of $\rho_t^{(\mathrm{s})}(x)$? Does it form a Markov process, and does it show increase of entropy?

An immediate result (see e.g. Penrose & Percival 1962) is this:

$$\sigma[\rho_t^{(\mathrm{s})}] + \sigma[\rho_t^{(\mathrm{e})}] \geq \sigma[\rho_0^{(\mathrm{s})}] + \sigma[\rho_0^{(\mathrm{e})}], \tag{214}$$

where σ denotes the Gibbs fine-grained entropy (90). This result follows from the fact that $\sigma[\rho_t]$ is conserved and that the entropy of a joint probability distribution is always smaller than or equal to the sum of the entropies of their marginals; with equality if the joint distribution factorizes. This gives a form of entropy change for the total system, but it is not sufficient to conclude that the object system itself will evolve towards equilibrium, or even that its entropy will be monotonically increasing. (Notice that (214) holds for $t \leq 0$ too.)

Actually, this is obviously not to be expected. There are interactions with an environment that may lead the system away from equilibrium. We shall have to make additional assumptions about the situation. A more or less usual set of assumptions is:

(a). The environment is very large (or even infinite); i.e.: the dimension of $\Gamma^{(\mathrm{e})}$ is much larger than that of $\Gamma^{(\mathrm{s})}$, and $H_{(\mathrm{s})} \ll H_{(\mathrm{e})}$.

(b). The coupling between the system and the environment is weak, i.e. λ is very small.

(c). The environment is initially in thermal equilibrium, e.g., $\rho^{(\mathrm{e})}(y)$ is canonical:

$$\rho_0^{(\mathrm{e})} = \frac{1}{Z(\beta)} e^{-\beta H^{(\mathrm{e})}} \tag{215}$$

(d). One considers time scales only that are long with respect to the relaxation times of the environment, but short with respect to the Poincaré recurrence time of the total system.

Even then, it is a major task to obtain a master equation for the evolution of the marginal state (211) of the system, or to show that its evolution is generated by a semigroup, which would guarantee that this forms a Markov process (under the proviso of footnote 71). Many specific models have been studied in detail (cf. Spohn 1980). General theorems were obtained (although mostly in a quantum mechanical setting) by (Davies 1974, Davies 1976a, Lindblad 1976, Gorini et al. 1976). But there is a similarly to the earlier approach: it seems that, here too, an analogue of 'repeated randomness' must be introduced. (Mehra & Sudarshan 1972, van Kampen 1994, Maes & Netočný 2003).

At the risk of oversimplifying and misrepresenting the results obtained in this analysis, I believe they can be summarized as showing that, in the so-called 'weak coupling' limit, or some similar limiting procedure, the time development of (211) can be modeled as

$$\rho_t^{(\mathrm{s})}(x) = T_t \rho^{(\mathrm{s})}(x) \quad t \geq 0, \tag{216}$$

where the operators T_t form a semigroup, while the environment remains in its steady equilibrium state:

$$\rho_t^{(\mathrm{e})}(y) = \rho_0^{(\mathrm{e})}(y) \quad t \geq 0. \tag{217}$$

The establishment of these results would also allow one to infer, from (214), the monotonic non-decrease of entropy of the system.

To assess these findings, it is convenient to define, for a fixed choice of $\rho_0^{(e)}$ the following linear map on probability distributions of the total system:

$$\mathcal{TR} : \rho(x,y) \mapsto \mathcal{TR}\rho(x,y) = \int \rho(x,y)dy \cdot \rho_0(y) \tag{218}$$

This map removes the correlation between the system and the environment, and projects the marginal distribution of the environment back to its original equilibrium form.

Now, it is not difficult to see that the Chapman-Kolmogorov equation (which is equivalent to the semigroup property) can be expressed as

$$\mathcal{TR}U^*_{t+t'} = \mathcal{TR}U^*_{t'}\mathcal{TR}U^*_t \quad \text{for all } t, t' \geq 0 \tag{219}$$

which is analogous to (206).

There is thus a strong formal analogy between the coarse-graining and the open-systems approaches. Indeed, the variables of the environment play a role comparable to the internal coordinates of a cell in the coarse graining approach. The exact microscopic information about the past is here translated into the form of correlations with the environment. This information is now removed by assuming that at later times, effectively, the state may be replaced by a product of the form (213), neglecting the back-action of the system on the environment. The mappings $\mathcal{CG}$ and $\mathcal{TR}$ are both linear and idempotent mappings, that can be regarded as special cases of the projection operator techniques of Nakajima and Zwanzig, which allows for a more systematical and abstract elaboration, sometimes called *subdynamics*.

Some proponents of the open systems approach, (e.g. Morrison 1966, Redhead 1995), argue that in contrast to the coarse-graining approach, the present procedure is 'objective'. Presumably, this means that there is supposed to be a fact of the matter about whether the correlations are indeed

'exported to the environment'. However, the analogy between both approaches makes one suspect that any problem for the coarse-graining approach is translated into an analogous problem of the open systems approach. Indeed, the problem of finding a privileged partition that we discussed in the previous paragraph is mirrored here by the question where one should place the division between the 'system' and 'environment'. There is no doubt that it practical applications this choice is also arbitrary.

7.6 Can the Markov property explain irreversible behaviour?

Ad (v). Finally, I turn to what may well be the most controversial and surprising issue: is the Markov property, or the repeated randomness assumption offered to motivate it, responsible for the derivation of time-reversal non-invariant results?

We have seen that every non-invertible homogeneous Markov process displays "irreversible behaviour" in the sense that different initial probability distributions will tend to become more alike in the course of time. Under certain technical conditions, one can obtain stronger results, e.g. an approach to a unique equilibrium state, monotonic non-decrease of absolute entropy, etc. All these results seem to be clearly time-asymmetric. And yet we have also seen that the Markov property is explicitly time symmetric. How can these be reconciled?

To start off, it may be noted that it has often been affirmed that the Markov property is the key towards obtaining time-asymmetric results. For example, Penrose writes:

> " . . . the behaviour of systems that are far from equilibrium is not symmetric under time reversal: for example: heat always flows from a hotter to a colder body, never from a colder to a hotter. If this behaviour could be derived from the symmetric laws of dynamics alone there would, indeed, be a paradox; we must therefore acknowledge the fact that some additional postulate, non-dynamical in character and asymmetric under time reversal must be adjoined to the symmetric laws of dynamics before the theory can become rich enough to explain non-equilibrium behaviour. In the present theory, this additional postulate is the Markov postulate" (Penrose 1970, p. 41).

In the previous paragraph, we have already questioned the claim expressed here that the Markov property is "non-dynamical". But now we are interested in the question whether postulating the Markov property would be asymmetric under time-reversal. Many similar statements, e.g. that the repeated randomness assumption is "the additional element by which statistical mechanics has to be supplemented in order to obtain irreversible equations" (van Kampen 1962, p. 182), or that the non-invertibility of a Markov process provides the origin of thermodynamic behaviour (Mackey 1992) can be found in the works of advocates of this approach.

But how can this be, given that the Markov property is explicitly time-symmetric? In order to probe this problem, consider another question. How does a given probability distribution $P(y,0)$ evolve for negative times? So, starting again from (170), let us now take $t \leq 0$. We still have:

$$P(y,t) = \sum_{y'} P(y,t,|y',0)P(y',0). \tag{220}$$

These conditional probabilities $P(y,t,|y',0)$ satisfy the 'time-reversed' Markov property (174), that says that extra specification of later values is irrelevant for the retrodiction of earlier values. As a consequence, we get for $t \leq t' \leq t'', 0$:

$$P(y,t|y'',t'') \quad = \quad \sum_{y'} P(y,t|y',t')P(y',t'|y'',t'') \tag{221}$$

i.e., a time-reversed analogue of the Chapman-Kolmogorov equation.

We may thus also consider these conditional probabilities for negative times as backward evolution operators. If we could assume their invariance under time translation, i.e. that they depend only on the difference $\tau = t - t'$, we could write

$$S_\tau(y|y') := P(y,t|y,t') \quad \text{with } \tau = t - t' \leq 0, \tag{222}$$

and obtain a second semigroup of operators S_τ, obeying

$$S_{\tau+\tau'} = S_\tau \circ S_{\tau'} \quad \tau, \tau' \leq 0 \tag{223}$$

that generate stochastic evolutions towards the past.

Further, these backward conditional probabilities are connected to the forward conditional probabilities by means of Bayes' theorem:

$$P_{(1|1)}(y,t|y',t') = \frac{P_{(1|1)}(y',t'|y,t)P(y,t)}{P(y',t')}; \tag{224}$$

and if the process, as before, is homogeneous this becomes

$$P_{(1|1)}(y,t|y',t') = \frac{T_{-\tau}(y'|y)P_t(y)}{P_{t'}(y')} \;\; ; \;\; \tau = t - t' < 0. \tag{225}$$

The matrix $P_{(1|1)}(y,t|y',t')$ always gives for $t < t'$ the correct 'inversion' of T_t. That is to say:

$$\sum_{y'} P(y,t|y',t')(T_{t'-t}P_t)(y') = P_t(y) \tag{226}$$

Note firstly that (225) is *not* the matrix-inverse of T_t! Indeed, the right-hand side of (225) depends on P_t and $P_{t'}$ as well as T. Even if the matrix-inverse $T^{(\mathrm{inv})}$ does not exist, or is not a bona fide stochastic matrix, the evolution towards the past is governed by the Bayesian inversion, i.e. by the transition probabilities (225).

Note also that if the forward transition probabilities are homogeneous, this is not necessarily so for the backward transition probabilities. For example, if in (225) one translates both t and t' by δ, one finds

$$P(y, t+\delta|y', t'+\delta) \quad = \quad \frac{T_{-\tau}(y'|y)P(y, t+\delta)}{P(y', t'+\delta)}.$$

Here, the right-hand side generally still depends on δ. In the special case that the initial distribution is itself stationary, the backward transition probabilities are homogeneous whenever the forward ones are. If $P(y, t)$ is not stationary, we might still reach the same conclusion, as long as the non-stationarity is restricted to those elements y or y' of $\mathcal{Y}$ for which $T_t(y|y') = 0$ for all t. Otherwise, the two notions become logically independent.

This gives rise to an unexpected new problem. Usually, an assumption of homogeneity (or time translation invariance) is seen as philosophically innocuous, as compared to time reversal invariance. But here we see that assuming time translation invariance for a system of *forward* transition probabilities is not equivalent to assuming the same invariance for the *backward* transition probabilities. If one believes that one of the two is obvious, how will one go about explaining the failure of the other? And how would one explain the preference for which one of the two is obvious, without falling into the "double standards" accusation of the kind raised by (Price 1996)?

But what about entropy increase? We have seen before that for every non-invertible Markov process the relative entropy of the distribution P with respect to the equilibrium distribution increases, and that the distribution evolves towards equilibrium. (Homogeneity of the process is not needed for this conclusion.) But the backward evolution operators form a Markov process too, for which exactly the same holds. This seems paradoxical. If $T_t P_0 = P_t$, we also have $P_t = S_{-t} P_0$. The entropy of P_t can hardly be both higher and lower than that of P_0! An example may clarify the resolution of this apparent problem: namely, the stationary solutions of S are not the same as the stationary solutions of T!

Example Consider a Markov chain with $\mathcal{Y} = \{1, 2\}$ and let

$$T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{227}$$

Choose an initial distribution $P_0 = \begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix}$. After one step we already get

$$TP = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \tag{228}$$

which is also the (unique) stationary distribution P^*. The backward transition probabilities are given by Bayes' theorem, and one finds easily:

$$S = \begin{pmatrix} \alpha & \alpha \\ 1-\alpha & 1-\alpha \end{pmatrix}. \tag{229}$$

The stationary distribution for this transition probability is

$$\tilde{P}^* = \begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix}. \tag{230}$$

That is to say: for the forward evolution operator the transition

$$\begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix} \xrightarrow{T} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \tag{231}$$

is one for which a non-stationary initial distribution evolves towards a stationary one. The relative entropy increases: $H(P_0, P^*) \leq H(TP, P^*)$. But for the backward evolution, similarly:

$$\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \xrightarrow{S} \begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix} \tag{232}$$

represents an evolution from a non-stationary initial distribution to the stationary distribution $\tilde{P}^*$ and, here too, relative entropy increases: $H(P_1, \tilde{P}^*) \leq H(P_0, \tilde{P}^*)$.

The illusion that non-invertible Markov processes possess a built-in time-asymmetry is (at least partly) due to the habit of regarding T_τ as a fixed evolution operator on an independently chosen distribution P_0. Such a view is of course very familiar in other problems in physics, where deterministic evolution operators generally *do* form a group and may be used, at our heart's desire, for positive and negative times.

Indeed, the fact that these operators in general have no inverse might seem to reflect the idea that Markov processes have no memory and 'loose information' along the way and that is the cause of

the irreversible behaviour, embodied in the time-asymmetric master equation, increase of relative or absolute entropy or approach to equilibrium. But actually, every Markov process has apart from a system of forward, also a system of backward transition probabilities, that again forms a semigroup (when they are homogeneous). If we had considered *them* as given we would get all conclusions we obtained before, but now for negative times.

I conclude that irreversible behaviour is not built into the Markov property, or in the non-invertibility of the transition probabilities, (or in the repeated randomness assumption[75], or in the Master equation or in the semigroup property). Rather the appearance of irreversible behaviour is due to the choice to rely on the forward transition probabilities, and not the backward. A similar conclusion has been reached before (Edens 2001) in the context of proposals of Prigogine and his coworkers. My main point here is that the same verdict also holds for more 'mainstream' approaches as coarse-graining or open systems.

7.7 Reversibility of stochastic processes

In order not to end this chapter on a destructive note, let me emphasize that I do not claim that the derivation of irreversible behaviour in stochastic dynamics is impossible. Instead, the claim is that motivations for desirable properties of the forward transition probabilities are not enough; one ought also show that these properties are lacking for the backward transitions.

In order to set up the problem of irreversibility in this approach to non-equilibrium statistical mechanics for a more systematic discussion, one first ought to provide a reasonable definition for what it means for a stochastic process to be (ir)reversible; a definition that would capture the intuitions behind its original background in Hamiltonian statistical mechanics.

One general definition that seems to be common (cf. Kelly 1979 p. 5) is to call a stochastic process reversible iff, for all n and $t_1, \ldots, t_n$ and τ:

$$P_{(n)}(y_1, t_1; \ldots; y_n, t_n) = P_{(n)}(y_1, \tau - t_n; \ldots; y_n, \tau - t_n). \tag{233}$$

See Grimmett & Stirzaker 1982, p. 219) for a similar definition restricted to Markov processes) The immediate consequence of this definition is that a stochastic process can only be reversible if the single-time probability $P_{(1)}(y, t)$ is stationary, i.e. in statistical equilibrium. Indeed, this definition seems to make the whole problem of reconciling irreversible behaviour with reversibility disappear.

[75] In recent work, van Kampen acknowledges that the repeated randomness assumption by itself does not lead to irreversibility: "This repeated randomness assumption [...] breaks the time symmetry by explicitly postulating the randomization *at the beginning* of the time interval Δt. There is no logical justification for this assumption other than that it is the only thing one can do and that it works. If one assumes randomness at the end of each Δt coefficients for diffusion, viscosity, etc. appear with the wrong sign; if one assumes randomness at the midpoint no irreversibility appears" (van Kampen 2002, p.475, original emphasis).

As Kelly (1979, p. 19) notes in a discussion of the Ehrenfest model: "there is no conflict between reversibility and the phenomenon of increasing entropy — reversibility is a property of the model in equilibrium and increasing entropy is a property of the approach to equilibrium"

But clearly, this view trivializes the problem, and therefore it is not the appropriate definition for non-equilibrium statistical mechanics. Recall that the Ehrenfest dog flea model (§7.2) was originally proposed in an attempt of showing how a tendency of approaching equilibrium from a initial non-equilibrium distribution (e.g. a probability distribution that gives probability 1 to the state that all fleas are located on the same dog) could be reconciled with a stochastic yet time-symmetric dynamics.

If one wants to separate considerations about initial conditions from dynamical considerations at all, one would like to provide a notion of (ir)reversibility that is associated with the stochastic dynamics alone, independent of the initial distribution is stationary.

It seems that an alternative definition which would fulfill this intuition is to say that a stochastic process is reversible if, for all y and y' and $t' > t$,

$$P_{(1|1)}(y, t|y', t') = P_{(1|1)}(y, t'|y', t). \tag{234}$$

In this case we cannot conclude that the process must be stationary, and indeed, the Ehrenfest model would be an example of a reversible stochastic process. I believe this definition captures the intuition that if at some time state y' obtains, the conditional probability of the state one time-step earlier being y is equal to that of the state one time-step later being y.

According to this proposal, the aim of finding the "origin" of irreversible behaviour or "time's arrow", etc. in stochastic dynamics must then lie in finding and motivating conditions under which the forward transition probabilities are different from the backwards transition probabilities, in the sense of a violation of (234). Otherwise, irreversible behaviour would essentially be a consequence of the assumptions about initial conditions, a result that would not be different in principle from conclusions obtainable from Hamiltonian dynamics.

References

Albert D.Z. (2000). *Time and Chance* Cambridge, Mass.: Harvard University Press.

Alberti. P.M., & Uhlmann, A. (1982). *Stochasticity and partial order. Doubly stochastic maps and unitary mixing*. Dordrecht: Reidel.

Balescu, R. (1997). *Statistical dynamics*. London: Imperial College Press.

Balian, R. (2005). Information in statistical physics. *Studies In History and Philosophy of Modern Physics, 36*, 323–353.

Batterman, R.W. (1990). Irreversibility and statistical mechanics: A new approach? *Philosophy of Science, 57*, 395–419.

Batterman, R.W. (1991). Randomness and probability in dynamical theories: On the proposals of the Prigogine school. *Philosophy of Science, 58*, 241–263.

Batterman, R. W. (2002). *The devil in the details: asymptotic reasoning in explanation, reduction, and emergence* Oxford: Oxford University Press.

Batterman, R.W. (2005). Critical phenomena and breaking drops: Infinite idealizations in physics. *Studies in History and Philosophy of Modern Physics, 36*, 225–244.

Becker, R. (1967). *Therory of heat* New York: Springer.

Belot, G. (2006). This volume, chapter 2.

Bennett, C.H. (2003). Notes on Landauer's principle, reversible computation, and Maxwell's demon. *Studies In History and Philosophy of Modern Physics, 34*, 501–510.

Berkovitz, J., Frigg, R., & Kronz. F. (2005). The ergodic hierarchy and Hamiltonian chaos. *Studies in History and Philosophy of Modern Physics*, to appear.

Bertrand, J. (1889). *Calcul des Probabilités*. Paris: Gauthier-Villars. Reissued New York: Chelsea, no date.

Bierhalter, G. (1992). Von L. Boltzmann bis J.J. Thomson: die Versuche einer mechanischen Grundlegung der Thermodynamik (1866-1890). *Archive for History of Exact Sciences, 44*, 25–75.

Birkhoff, G.D. (1931). Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences of the United States of America, 17*, 656–660.

Bishop, R.C. (2004). Nonequilibrium statistical mechanics Brussels-Austin style *Studies In History and Philosophy of Modern Physics, 35*, 1–30.

Blatt, J.M. (1959). An alternative approach to the ergodic problem. *Progress in Theoretical Physics, 22*, 745–756.

Boltzmann, L. (1866). Über die Mechanische Bedeutung des Zweiten Hauptsatzes der Wärmetheorie. *Wiener Berichte, 53*, 195–220; in (Boltzmann 1909) Vol. I, paper 2.

Boltzmann, L. (1868a). Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten. *Wiener Berichte, 58*, 517–560; in (Boltzmann 1909) Vol. I, paper 5.

Boltzmann, L. (1868b). Lösung eines mechanischen Problems. *Wiener Berichte, 58*, 1035–1044. In (Boltzmann 1909) Vol. I, paper 6.

Boltzmann, L. (1871a). Über das Wärmegleichgewicht zwischen mehratomigen Gasmolekülen. *Wiener Berichte, 63*, 397–418. In (Boltzmann 1909) Vol. I, paper 18.

Boltzmann, L. (1871b). Einige allgemeine Sätze über Wärmegleichgewicht. *Wiener Berichte, 63*, 679–711. In (Boltzmann 1909) Vol. I, paper 19.

Boltzmann, L. (1871c). Analytischer Beweis des zweiten Haubtsatzes der mechanischen Wärmetheorie aus den Sätzen über das Gleichgewicht der lebendigen Kraft. *Wiener Berichte, 63*, 712–732. In (Boltzmann 1909) Vol. I, paper 20.

Boltzmann, L. (1872). Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Wiener Berichte, 66*, 275–370. In (Boltzmann 1909) Vol. I, paper 23.

Boltzmann, L. (1877a). Bermerkungen über einige Probleme der mechanische Wärmetheorie *Wiener Berichte, 75*, 62–100. In (Boltzmann 1909) Vol. II, paper 39.

Boltzmann, L. (1877b). Über die Beziehung zwisschen dem zweiten Haubtsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung resp. dem Sätzen über das Wärmegleichgewicht *Wiener Berichte, 76*, 373–435. In (Boltzmann 1909) Vol. II, paper 42.

Boltzmann, L. (1878). Weitere Bemerkungen über einige Probleme der mechanischen Wärmetheorie. *Wiener Berichte, 78*, 7–46. In (Boltzmann 1909) Vol. II paper 44.

Boltzmann, L. (1881a). Über einige das Wärmegleichgewicht betreffende Sätze. *Wiener Berichte, 84* 136–145. In (Boltzmann 1909) Vol. II paper 62.

Boltzmann, L. (1881b). Referat über die Abhandlung von J.C. Maxwell: "Über Boltzmann's Theorem betreffend die mittlere verteilung der lebendige Kraft in einem System materieller Punkte". *Wiedemann's Annalen Beiblätter, 5*, 403–417. In (Boltzmann 1909) Vol. II paper 63.

Boltzmann, L. (1884). Über die Eigenschaften Monocyklischer und andere damit verwandter Systeme. *Crelle's Journal für die reine und angewandte Mathematik, 98*, 68–94 1884 and 1885. In (Boltzmann 1909) Vol III, paper 73.

Boltzmann, L. (1887a). Ueber die mechanischen Analogien des zweiten Hauptsatzes der Thermodynamik. *Journal für die reine und angewandte Mathematik, 100*, 201–212. Also in (Boltzmann 1909), Vol. III, paper 78.

Boltzmann, L. (1887b). Neuer Beweis zweier Sätze über das Wärmegleichgewicht unter mehratomigen Gasmolekülen. *Wiener Berichte* **95**, 153–164, in (Boltzmann 1909) Vol. III, paper 83.

Boltzmann, L. (1887c). Über einige Fragen der Kinetische Gastheorie. *Wiener Berichte* **96**, 891–918. In (Boltzmann 1909) Vol. III, paper 86.

Boltzmann, L. (1892). III. Teil der Studien über Gleichgewicht der lebendigen Kraft *Münch. Ber. 22*, 329–358. In (Boltzmann 1909) Vol. III, paper 97.

Boltzmann, L.& C.H. Bryan (1894a). Über die mechanische Analogie des Wärmegleichgewichtes zweier sich berührende Körper. *Wiener Berichte, 103*, 1125–1134. In (Boltzmann 1909) Vol. III, paper 107.

Boltzmann, L. (1894). On the application of the determinantal relation to the kinetic theory of gases. Appendix C to an article by G.H. Bryan on thermodynamics. *Reports of the British Association for the Advancement of Science*, pp. 102–106. In (Boltzmann 1909) Vol. III, paper 108.

Boltzmann, L. (1895). On certain questions in the theory of gases. *Nature* **51**, 413–415. In (Boltzmann 1909), Vol. III, paper 112.

Boltzmann, L. (1895b). On the minimum theorem in the theory of gases. *Nature* **52**, 221. Also in (Boltzmann 1909) Vol. III, paper 114.

Boltzmann, L. (1896). *Vorlesungen über Gastheorie* Vol I. Leipzig, J.A. Barth, 1896.Translated, together with (Boltzmann 1898) by S.G. Brush, *Lecture on Gas Theory* Berkeley: University of California Press, 1964.

Boltzmann, L. (1896b). Entgegnung an die wärmetheoretischen Betrachtungen des Hrn. E., Zermelo. *Wiedemann's Annalen, 57*, 772–784. In (Boltzmann 1909) Vol. III, paper 119.

Boltzmann, L. (1897a). Zu Hrn Zermelos Abhandlung "Über die mechanische Erklärung irreversibler Vorgänge". *Wiedemann's Annalen, 60* 392–398. in (Boltzmann 1909), Vol. III paper 120.

Boltzmann, L. (1897b). Über einige meiner weniger bekannte Abhandlungen über Gastheorie und deren Verhältnis zu derselben *Verhandlungen des 69. Versammlung Deutscher Naturforscher und Ärzte*, Braunschweig 1897, pp. 19–26. *Jahresberichte der Deutsche Mathematiker-Vereinigung*, **6**, (1899), 130–138. Also in (Boltzmann 1909) Vol. III, paper 123.

Boltzmann, L. (1898) *Vorlesungen über Gastheorie* Vol II. Leipzig, J.A. Barth, 1898. Translated, together with (Boltzmann 1896) by S.G. Brush, *Lecture on Gas Theory* Berkeley: University of California Press, 1964.

Boltzmann, L. (1898). Über die sogenannte H-Kurve. *Mathematische Annalen* **50**, 325–332. In (Boltzmann 1909) vol. III, paper 127.

Boltzmann, L. & Nabl, J. Kinetisch theorie der Materie. *Encyclopädie der Mathematischen Wisenschaften*, Vol V-1, pp. 493–557.

Boltzmann, L. (1905). *Populäre Schriften*. Leipzig: J.A. Barth. Re-issued Braunschweig: F. Vieweg, 1979.

Boltzmann, L. (1909). *Wissenschaftliche Abhandlungen* Vol. I, II, and III. F. Hasenöhrl (ed.) Leipzig. Reissued New York: Chelsea, 1969.

Borel, E. (1914). *Le Hasard*. Paris: Alcan.

Boyling, J.B. (1972). An axiomatic approach to classical thermodynamics. *Proceedings of the Royal Society of London, 329*, 35–71.

Bricmont, J. (1995). Science of chaos or chaos in science? *Physicalia, 17*, 159–208; also in P.R. Gross, N.Levitt and M.W.Lewis (Eds.),*The flight from science and reason*, New York: New York Academy of Sciences, pp. 131–176.

Brown, H., & Uffink, J. (2001). The origins of time-asymmetry in thermodynamics: the minus first law. *Studies in History and Philosophy of Modern Physics, 32*, 525–538.

Brush, S.G. (1966). *Kinetic theory* Vol. 1 and 2. Oxford: Pergamon.

Brush, S.G. (1976). *The Kind of motion we call heat*. Amsterdam: North Holland.

Brush, S.G. (1977). Statistical mechanics and the philosophy of science: some historical notes. In F. Supper & P.D. Asquith (Eds.) *PSA 1976 Proceedings of the Biennial meeting of the Philsopophy of Science Association 1976, Vol 2*, East Lansu=ing, Mich.: Philosophy of Science Association, pp. 551–584.

Brush, S.G. (1999). Gadflies and geniuses in the history of gas theory. *Synthese 119*, 11–43. Also in (Brush 2003), pp. 421–450.

Brush, S.G. (2003), *The kinetic theory of gases*. London: Imperial College Press.

Bryan, G.H. (1894). Letter to the editor. *Nature, 51*, 175.

Bryan, G. H. (1895). The Assumption in Boltzmann's Minimum Theorem, *Nature, 52*, 29–30.

Burbury, S.H. (1894a). Boltzmann's minimum theorem. *Nature, 51*, 78–79.

Burbury, S.H. (1894b). The kinetic theory of gases. *Nature, 51* 175–176.

Butterfield, J.N. (2006). This volume, chapter 1.

Callender, C. (1999). Reducing thermodynamics to statistical mechanics: the case of entropy. *Journal of Philosophy, 96*, 348–373.

Callender, C. (2001). Taking Thermodynamics Too Seriously. *Studies In History and Philosophy of Modern Physics, 32*, 539–553.

Callender, C. (2004). Measures, explanations and the past: Should 'special' initial conditions be explained? *British Journal for Philosophy of Science, 55*, 195–217.

Campisi, M. (2005). On the mechanical foundations of thermodynamics: The generalized Helmholtz theorem *Studies In History and Philosophy of Modern Physics, 36*, 275–290.

Casetti, L., Pettini, M. & Cohen, E.G.D (2003). Phase Transitions and Topology Changes in Configuration Space. *Journal of Statistical Physics, 111* , 1091–1123 .

Carathéodory, C. (1909). Untersuchungen über die Grundlagen der Thermodynamik, *Mathematische Annalen, 67*, 355–386. Translation by J. Kestin,"Investigation into the foundations of thermodynamics" in (Kestin 1976), pp. 229–256. This translation is not quite accurate.

Carnot S. (1824). *Réflexions sur la puissance motrice du feu*. Paris: Bachelier. Re-edited and translated in E. Mendoza (Ed.) Reflections on the motive power of fire, New York: Dover, 1960.

Cercignani, C. (1998). *Ludwig Boltzmann, the man who trusted atoms* Oxford: Oxford University Press.

Cercignani, C. Illner, R. & Pulvirenti, M. (1994). *The mathematical theory of dilute gases*. New York: Springer-Verlag.

Chang, H. (2003). Preservative realism and its discontents: revisiting caloric. *Philosophy of Science, 70*, 902–912.

Chang, H. (2004). *Inventing temperature. Measurement and scientific progress*. Oxford: Oxford University Press.

Clausius, R. (1857). Über die Art von Bewegung die wir Wärme nennen, *Poggendorff's Annalen, 100*, 253-280. English translation in (Brush 1966), Vol.1. pp. 111–134.

Clausius, R. (1862). 'Ueber die Anwendung des Satzes von der Aequivalenz der Verwandlungen auf die innere Arbeit', *Viertelsjahrschrift der Züricher naturforschenden Gesellschaft, 7*, 48. Also in *Annalen der Physik und Chemie, 116*: 73–112 (1862), English translation in *Philosophical Magazine, 24*, 81–97, 201–213, also in (Kestin 1976), pp. 133–161.

Clausius, R. (1865). Über verschiedene für die Anwendung bequeme Formen der Haubtgleichungen der mechanische Wärmetheorie, *Poggendorff's Annalen, 100*, (253). English translation in (Kestin 1976).

Cohen, E.G.D. (1996). Boltzmann and statistical mechanics, cond-mat/9608054v2.

Compagner, A. (1989). Thermodynamics as the continuum limit of statistical mechanics *American Journal of Physics, 57*, 106–117.

Cornfeld, I.P. Fomin, S.V. & Sinai, Ya. G. (1982). *Ergodic theory* New York: Springer-Verlag.

Culverwell, E.P. (1894). Dr. Watson's proof of Boltzmann's theorem on permanence of distributions *Nature, 50* 617.

Culverwell, E.P. (1895). Boltzmann's minimum theorem *Nature, 51* 246.

Curd, M. (1982). Popper on the direction of time. In R. Sexl and J. Blackmore (eds.), *Ludwig Boltzmann, internationale Tagung: anlässlich des 75. Jahrestages seines Todes, 5.-8. September 1981: ausgewählte Abhandlungen*, pp. 263-303. Graz: Akademische Druck- und Verlagsanstalt.

Davies, E.B. (1974). Markovian master equations. *Communications in Mathematical Physics, 39* 91-110.

Davies, E.B. (1976a). Markovian master equations II *Mathematische Annalen, 219* 147–158.

Davies, E.B (1976b). *Quantum theory of open systems*. New York: Academic Press

Denbigh, K.G. & Denbigh J. (1985). *Entropy in relation to incomplete knowledge*. Cambridge: Cambridge University Press.

Dias, P.M.C. (1994). Will someone say exactly what the H-theorem proves? A study of Burbury's Condition A and Maxwell's Proposition II. *Archive for History of Exact Sciences 46*, 341–366.

Dugas, R. (1959). *La théorie physique au sens de Boltzmann et ses prolongements modernes*. (Neuchâtel: Griffon).

Earman, J. (2006). The past-hypothesis: not even false. *Studies in History and Philosophy of Modern Physics*, to appear.

Earman, J. and Rédei, M. (1996). Why ergodic theory does not explain the success of equilibrium statistical mechanics. *British Journal for the Philosophy of Science, 45* 63–78.

Earman, J. & Norton, J.D. (1998). Exorcist XIV: The wrath of Maxwell's demon. Part I. From Maxwell to Szilard. *Studies in History and Philosophy of Modern Physics, 29*, 435–471.

Earman, J. & Norton, J.D. (1999). Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond *Studies in History and Philosophy of Modern Physics, 30*, 1–40.

Edens, B. (2001). Semigroups and symmetry: an investigation of Prigogine's theories. http://philsci-archive.pitt.edu/archive/00000436/.

Ehrenfest, P. & Ehrenfest-Afanassjewa, T. (1912). Begriffliche Grundlagen der Statistischen Auffassung in der Mechanik. *Enzyclopädie der Mathematischen Wissenschaften* Vol. 4. F. Klein and C. Müller (eds.). Leibzig: Teubner, pp. 3–90. English translation *The conceptual foundations of the statistical approach in mechanics*. Ithaca N.Y.: Cornell University Press, 1959.

Einstein, A. (1902). Kinetische Theorie des Wärmegleichgewicht und des zweiten Hauptstzes der Thermodynamik. *Annalen der Physik, 9*, 417–433.

Einstein, A. (1910). Theorie der Opaleszenz von homogenen Flüssigkeiten und Flüssigkeitsgemischen in der Nähe des kritischen Zustandes *Annalen der Physik, 33*, 1275–1298.

Ellis, R.L. (1850). Remarks on an alleged proof of the method of least squares, contained in a late number of the Edinburgh Review, in W.Walton (Ed.), Mathematical and other Writings of R.L. Ellis. Cambridge: Cambridge University Press, 1863, pp. 53–61.

Emch, G.G. (1965). On the Markov character of master-equations. *Helvetica Physica Acta, 38*, 164–171.

Emch, G.G. (2005). Probabilistic issues in statistical mechanics. *Studies In History and Philosophy of Modern Physics 36*, 303–322.

Emch, G.G. (2006). This volume, chapter 10.

Emch, G.G. & Liu, C. (2001). *The Logic of thermostatistical physics*. Berlin: Springer.

Farquhar, I.E. (1964). *Ergodic theory in statistical mechanics*, London, Interscience.

Feshbach, H. (1987). Small systems: when does thermodynamics apply? *Physics Today, 40*, 9–11.

Fine, T.L. (1973). *Theories of probability: An examination of foundations*. New York: Academic Press.

Fisher, M.E. (1964). The free energy of a macroscopic system. *Archive for Rational Mechanics and Analysis 17* 377–410.

Fowler, R.H. & Guggenheim, E. (1939). *Statistical Thermodynamics*. Cambridge: Cambridge University Press

Fox, R. (1971). *The caloric theory of gases: from Lavoisier to Regnault*. Oxford: Clarendon Press.

Friedman, K.S. (1976). A Partial vindication of ergodic theory. *Philosophy of Science, 43*, 151-162.

Frigg, R. (2004). In what sense is the Kolmogorov-Sinai entropy a measure for chaotic behaviour?—bridging the gap between dynamical systems theory and communication theory. *British Journal for the Philosophy of Science, 55*, 411–434.

Galavotti, M.C. (2004). *A philosophical introduction to probability*. Chicago: University of Chicago Press.

Gallavotti, G. (1994). Ergodicity, ensembles, irreversibility in Boltzmann and beyond http: arXiv:chao-dyn/9403004. *Journal of Statistical Physics, 78*, 1571–1589.

Gallavotti, G. (1999). *Statistical mechanics: A short treatise*. Berlin: Springer.

Gantmacher, F.R. (1959). *Matrizenrechnung* Vol 2. Berlin: Deutscher Verlag der Wissenschaften.

Garber, E. Brush, S.G., & Everitt C.W.F. (Eds.) (1986). *Maxwell on molecules and gases* Cambridge Mass.: MIT Press.

Garber, E. Brush, S.G., & Everitt, C.W.F, (Eds.) (1995). *Maxwell on heat and statistical mechanics* Bethlehem: Lehigh University Press.

Gearhart C.A. (1990). Einstein before 1905: The early papers on statistical mechanics *American Journal of Physics, 58*, 468–480.

Gibbs, J.W. (1875). On the equilibrium of heterogenous substances. *Transactions of the Connecticut Academy, 3*, 103–246 and 343–524 (1878). Also in (Gibbs 1906, pp. 55–353).

Gibbs, J.W. (1902). *Elementary principles in statistical mechanics*, New York, Scribner etc.

Gibbs, J.W. (1906). *The Scientific Papers of J. Willard Gibbs*, Vol. 1, Thermodynamics, Longmans, London. Reissued by Ox Bow Press, Woodbridge, Connecticut, 1993.

Giles, R. (1964). *Mathematical foundations of thermodynamics* Oxford: Pergamon.

Gold, T. (1956), Cosmic processes and the nature of time. In R. Colodny (Ed.), *Mind and Cosmos*. Pittsburgh: University of Pittsburgh Press, pp. 311–329.

Goldstein, S. (2001). Boltzmann's approach to statistical mechanics. In J. Bricmont, D. Dürr, M.C. Galavotti, G. Ghirardi, F. Petruccione, and N. Zanghi (Eds.) *Chance in physics: foundations and perspectives*, Lecture Notes in Physics 574. Berlin: Springer-Verlag, pp. 39–54. Also as e-print cond-mat/0105242.

Gorini, V., Kossakowski, A. & Sudarshan, E.C.G. Completely positive dynamical semigroups of N-level systems. *Journal of Mathematical Physics 17* 8721–825.

Grimmett, G. R. & Stirzaker, D.R. (1982). *Probability and random processes*. Oxford: Clarendon Press.

de Groot, S. (1951). *Thermodynamics of irreversible processes*. Amsterdam: North-Holland.

de Groot, S.R., & Mazur, P. (1961). *Non-equilibrium thermodynamics*. Amsterdam: North-Holland.

Gross, D.H.E. (1997). Microcanonical thermodynamics and statistical fragmentation of dissipative systems. The topological structure of the N-body phase space. *Physics Reports, 279*, 119–201.

Gross, D.H.E. & Votyakov, E.V. (2000). Phase transitions in "small" systems. *European Physical Journal B 15*, 115–126.

Grünbaum, A. (1973). Is the coarse-grained entropy of classical statistical mechanics an anthropomorphism? In *Philosophical problems of space and time*, 2nd ed., Dordrecht: Reidel, pp. 646–665.

Guthrie, F. (1875). Molecular motion. *Nature, 10* 123. Also in (Garber, Brush & Everitt 1995, p.143–145).

Greven, A. Keller, G. & Warnecke, G. (Eds.) (2003). *Entropy*. Princeton: Princeton University Press.

ter Haar, D. (1955). Foundations of statistical mechanics. *Reviews of Modern Physics, 27*, 289–338.

Hacking, I. (1975). *The emergence of probability*. Cambridge: Cambridge University Press.

Hacking I. (1990). *The taming of Chance*. Cambridge: Cambridhge University Press.

van Harn, K. & Holewijn, P.J. (1991). *Markov-ketens in diskrete tijd*. Utrecht: Epsilon.

Herschel, J.F.W. (1850). Quetelet on Probabilities *Edinburgh Review*. Also in *Essays from the Edinburgh and Quarterly Reviews with addresses and other pieces*, London: Longman, Brown, Green, Longmans and Roberts, 1857, pp. 365–465.

Hertz, P. (1910). Über die mechanischen Grundlagen der Thermodynamik. *Annalen der Physik, 33*, 225–274; 537–552.

Hill, T.L. (1963). *Thermodynamics of small systems*. New York: Benjamin.

Höflechner, W. (1994). *Ludwig Boltzmann Leben und Briefe*. Graz: Akademische Druck- und Verlagsanstalt.

Huang, K. (1987). *Statistical mechanics*. New York: Wiley.

Huggett, N. (1999). Atomic metaphysics. *Journal of Philosophy, 96*, 5–24.

Hutter, K, & Wang, Y. (2003). Phenomenological thermodynamics and entropy principles. In (Greven et al. 2003, pp. 55–78).

Illner, R. & Neunzert, H. (1987). the concept of irreversibility in kinetic theory of gases. *Transport Theory and Statistical Physics, 16*, 89–112.

Janssen, M. (2002). Dog fleas and tree trunks: the Ehrenfests marking the territory of Boltzmann's H-theorem. Unpublished.

Jauch, J. (1972). On a new foundation of equilibrium thermodynamics. *Foundations of Physics, 2*, 327–332.

Jauch, J. (1975). Analytical thermodynamics. Part 1. Thermostatics–general theory. *Foundations of Physics, 5*, 111–132.

Jaynes, E.T, (1965). Gibbs vs. Boltzmann entropies. *American Journal of Physics, 33*, 391–398. Also in (Jaynes 1983, pp. 77–86).

Jaynes, E.T. (1967). Foundations of probability theory and statistical mechanics. In: M. Bunge (Ed.) *Delaware seminar in the foundations of physics*. Berlin: Springer-Verlag, pp. 77–101. Also in (Jaynes 1983, pp. 89–113).

Jaynes, E.T. (1983). *Probability, statistics and statistical physics*. R. Rosenkrantz (Ed.) Dordrecht: Reidel.

Jaynes, E.T. (1992). The Gibbs paradox. In C.R. Smith, G.J. Erickson, & P.O. Neudorfer, (Eds.) *Maximum entropy and Bayesian methods*. Dordrecht: Kluwer, pp. 1–22.

Jepps, O.G., Ayton,G. & Evans D.J. (2000). Microscopic expressions for the thermodynamic temperature. *Physical Review E, 62*, 4757–4763.

Kadanoff, L.P. (2000). *Statistical physics : statics, dynamics and renormalization* Singapore: World Scientific.

van Kampen, N.G. (1962). Fundamental problems in the statistical mechanics of irreversible processes. In E.G.D. cohen (Ed.), *Fundamental problems in statistical mechanics* Amsterdam: North-Holland, pp.173–202.

van Kampen, N.G. (1981). *Stochastic processes in chemistry and physics*. Amsterdam: North-Holland.

van Kampen, N.G. (1984). The Gibbs paradox. In W.A. Parry (ed.), *Essays in theoretical physics in honour of Dirk ter Haar*, Oxford: Pergamon Press, pp. 303–312.

van Kampen, N.G. (1994) Models for dissitpation in quantumn mechanics. In J.J.Brey, J. Marro, J.M. Rubi, M. San Miguel (Eds.) *25 years of non-equilibrium statistical mechanics* Berlin: Springer.

van Kampen, N.G. (2002) The road from molecules to Onsager. *Journal of Statistical Physics, 109*, 471–481.

van Kampen, N.G. (2004). A new approach to noise in quantum mechanics *Journal of Statistical Physics, 115*, 1057–1072.

Karakostas, V. (1996). On the Brussels school's arrow of time in quantum theory. *Philosophy of Science, 63*, 374–400.

Kelly F.P. (1979). *Reversibility and stochastic networks* Chichester: Wiley. Also at http://www.statslab.cam.ac.uk/ afrb2/kelly_book.html.

Kestin, J. (1976). *The second law of thermodynamics*. Stroudsburg, Pennsylvania: Dowden, Hutchinson and Ross.

Keynes, J.M. (1921). *A treatise on probability*. London: Macmillan.

Khinchin, A. (1949). *Mathematical foundations of statistical mechanics*. (New York: Dover).

Kirchhoff, G. (1894). *Vorlesungen über mathematische Physik, Vol IV: Theorie der Wärme* M. Planck (Ed.). Leipzig: Teubner.

Kirsten, C. & H.-G. Körber, (1975). *Physiker über Physiker*, Berlin: Akademie-Verlag.

Klein, M.J. (1970). Maxwell, his demon and the second law of thermodynamics, *American Scientist 58*, 82–95, 1970; also in (Leff & Rex 1987, pp. 59–72).

Klein, M.J. (1972). Mechanical explanation at the end of the nineteenth century. *Centaurus, 17*, 58–82.

Klein, M.J. (1973). The Development of Boltzmann's Statistical Ideas. In E.G.D. Cohen & W. Thirring (eds.), *The Boltzmann equation*, Wien: Springer, pp. 53–106.

Klein, M.J. (1974). Boltzmann, monocycles and mechanical explanation. In R.J. Seeger & R.S. Cohen (Eds.), *Philsophical foundations of science; Boston Studies in the Philosophy of Science XI* Dordrecht: Reidel, pp. 155–175.

Klein, M.J. (1978). The thermostatics of J. Willard Gibbs: atyransformation of thermoidynamics. In E.G. Forbes (Ed.) *Human implications of scientific advance*. Edinburgh: Edinburgh University Press, pp. 314–330.

Kroes, P.A. (1985). *Time: its structure and role in physical theories*. Dordrecht: Reidel.

Kurth, R. (1960). Axiomatics of classical statistical mechanics. Oxford: Pergamon Press.

Ladyman, J., Presnell, S., Short, T. & Groisman, B. (2006). The connection between logical and thermodynamical irreversibility. http://philsci-archive.pitt.edu/archive/00002374/.

Landau, L.D. & Lifshitz, E.M. (1987). *Fluid Mechanics* 3rd Edition, Oxford: Butterworth-Heinemann.

Lanford, O.E. (1973). Entropy and equilibrium states in classical statistical mechancs. In A. Lenard (Ed.) *Statistical mechanics and mathematical problems*. Berlin: Springer-Verlag, pp. 1–113.

Lanford, O.E. (1975). Time evolution of large classical systems. In J. Moser (Ed.) *Dynamical Systems, Theory and Applications*, Lecture Notes in Theoretical Physics Vol. 38, Berlin: Springer, pp. 1–111.

Lanford, O.E. (1976). On a derivation of the Boltzmann equation. *Astérisque, 40* 117–137. Also in J.L Lebowitz & E.W. Montroll (Eds.) *Nonequilibrium phenomena I: the Boltzmann Equation*. Amsterdam: North-Holland, 1983.

Lanford, O.E. (1981). The hard sphere gas in the Boltzmann-Grad limit. *Physica, 106A*, 70–76.

Lavis, D.A. (2004). The spin-echo system reconsidered. *Foundations of Physics, 34*, 669–688.

Lavis, D.A. (2005). Boltzmann and Gibbs: An attempted reconciliation. *Studies In History and Philosophy of Modern Physics, 36*, 245–273.

Lebowitz, J. (1994). Time's arrow and Boltzmann's entropy. In J.J. Halliwell, J. Pérez-Mercader & W.H. Zurek (Eds.). *Physical origins of time asymmetry*. Cambridge: Cambridge University Press, pp. 131–146.

Lebowitz, J.L. (1999). Statistical mechanics: A selective review of two central issues. *Reviews of Modern Physics, 71*, S346–S357. math-ph/0010018.

Lebowitz, J., & Penrose, O.(1973). Modern ergodic theory. *Physics Today, 26* 23–29.

Leeds, S. (1989). Discussion: D. Malament and S. Zabell on Gibbs phase averaging. *Philosophy of Science, 56*, 325–340.

Leff, H.S. & Rex, A.F. (2003) *Maxwell's Demon 2*. Bristol: Institute of Physics.

Leinaas J.M. & Myrheim, J. (1977). On the theory of identical particles. *Il Nuovo Cimento, 37B*, 1–23.

Leff, H.S. & Rex, A.F. (2003). *Maxwell's Demon 2*. Bristol: Institute of Physics.

Lieb, E.H. (1976). The stability of matter. *Reviews of Modern Physics, 48*, 553–569.

Lieb, E.H. & Lebowitz, J.L. Lectures on the thermodynamic limit for Coulomb systems. In A. Lenard (Ed.) *Statistical mechanics and mathematical problems*. Berlin: Springer-Verlag, pp. 136–162.

Lieb, E. & Yngvason, J.: 1999, The phyics and mathematics of the second law of thermodynamics, *Physics Reports, 310*, 1–96; erratum 314, (1999), p. 669. Also as e-print: http://xxx.lanl.gov/abs/cond-mat/9708200.

Lindblad, G. (1976). On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics, 48*, 119–130.

Lindblad, G. (1983). *Non-equilibrium entropy and irreversibility* Dordrecht: Reidel.

van Lith, J. (2001a). Ergodic theory, interpretations of probability and the foundations of statistical mechanics *Studies in History and Philosophy of Modern Physics, 32*, 581–594.

van Lith, J. (2001b). *Stirr in Stillness*. Ph.D. Thesis, Utrecht University.

Liu, C. (1999). Explaining the emergence of cooperative phenomena. *Philosophy of Science 66 (Proceedings)*, S92–S106.

Lorentz, H.A. (1887). Über das Gleichgewicht der lebendige Kraft unter Gasmolekülen. *Wiener Berichte, 95*, 115–152, 187. Also in: *Collected Papers* The Hague: Martinus Nijhoff 1938, pp. 74–111.

Lorentz, H.A. (1916). *Les theories statistiques en thermodynamique*. Leipzig: Teubner.

Loschmidt, J. (1876). Über die Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft. *Wiener Berichte, 73*, 128–142, 366–372 (1876), 75, 287–298, 76, 209–225 (1877).

Maes, C. & Netočný, K. (2003). Time-reversal and entropy. *Journal of Statistical Physics, 110*, 269–310.

Mackey, M.C. (1992). *Time's arrow: the origins of thermodynamic behavior*. New York: Springer.

Mackey, M.C. (2001). Microscopic dynamics and the second law of thermodynamics. In: *Time's Arrows, quantum Measurements and Superluminal Behavior*. C. Mugnai, A. Ranfagni & L.S. Schulman (Eds.) (Roma: Consiglio Nazionale delle Ricerche).

Malament, D.B. & Zabell, S.L. (1980). Why Gibbs Phase Averages Work—The Role of Ergodic Theory. *Philosophy of Science, 56*, 339–349.

Mandelbrot, B. (1956). An outline of a purely phenomenological theory of statistical thermodynamics: 1. canonical ensembles. *IRE Transactions on Information Theory, IT-2*, 190–203 (1956).

Mandelbrot, B. (1962). The role of sufficiency and of estimation in thermodynamics. *Annals of Mathematical Statistics, 33*, 1021–1038.

Mandelbrot, B. (1964). On the derivation of statistical thermodynamics from purely phenomenological principles. *Journal of Mathematical Physics, 5*, 164–171.

Mañé, R. (1987). *Ergodic theory and differentiable dynamics*. Berlin: Springer.

Maroney, O.J.E. (2005). The (absence of a) relationship between thermodynamic and logical reversibility. *Studies In History and Philosophy of Modern Physics, 36*, 355–374.

Martin-Löf, A. (1979). *Statistical mechanics and the foundations of thermodynamics*. Berlin: Springer-Verlag.

Maxwell, J.C. (1860). Illustrations of the dynamical theory of gases. *Philosophical Magazine, 19*, 19–32; 20, 21–37. Also in (Garber, Brush & Everitt 1986, pp. 285–318).

Maxwell, J.C. (1867). On the dynamical theory of gases. *Philosophical Transactions of the Royal Society of London, 157* 49-88 (1867). Also in (Brush 1966) and (Garber, Brush & Everitt 1986, pp. 419–472)

Maxwell, J.C. (1872). *Theory of heat*.(2nd ed.) London: Longmans, Green and Co.

Maxwell, J.C. (1873a). Molecules. *Nature, 8*, 437–441 (1873). Also in (Garber, Brush & Everitt 1986, pp. 138–154)

Maxwell, J.C. (1873b). On the final state of a system of molecules in motion subject to forces of any kind. *Nature, 8*, 537-538 (1867). Also in (Garber, Brush & Everitt 1995, pp. 138–143)

Maxwell, J.C. (1875). On the dynamical evidence of the molecular constitution of bodies. *Nature, 11* 357-359, 374–377, (1875). Also in (Garber, Brush & Everitt 1986, pp. 216–237).

Maxwell, J.C. (1877). A treatise on the kinetic theory of gases by Henry William Watson, *Nature, 18*, 242–246. Also in (Garber, Brush & Everitt 1995, p. 156–167).

Maxwell, J.C. (1878a). Constitution of bodies. *Encyclopaedia Brittanica*, ninth edition, Vol. 6, pp. 310–313. Also in (Garber, Brush & Everitt 1986, pp. 246–254).

Maxwell, J.C. (1878b). Diffusion. *Encylopedia Brittanica* ninth edition, Vol. 7, pp. 214–221; also in (Garber, Brush & Everitt 1986, pp. 525–546).

Maxwell, J.C. (1878c). On stresses in rarified gases arising from inequalities of temperature *Proceedings of the Royal Society of London, 27* 304–308; *Philosophical Transactions of the Royal Society of London, 170*, (1880) 231–256. Also in (Garber, Brush & Everitt 1995, pp. 357–386).

Maxwell, J.C., (1879). On Boltzmann's theorem on the average distribution of energy in a system of material points. *Transactions of the Cambridge Philosophical Society, 12*, 547–570, 1879. Also in (Garber, Brush & Everitt 1995, pp. 357–386).

Mehra, J. (1998). Josiah Willard Gibbs and the Foundations of Statistical Mechanics *Foundations of Physics, 28*, 1785–1815.

Mehra, J. & Sudarshan, E.C.G. (1972). Some reflections on the nature of entropy, irreversibility and the second law of thermodynamics, *Nuovo Cimento B, 11*, 215–256.

Meixner, J. (1969). Processes in simple thermodynamic materials *Archive for Rational Mechanics and Analysis 33*, 33–53.

Meixner, J. (1970). On the foundations of thermodynamics of processes. In B. Gal-Or, E.B. Stuart & A. Brainard (Eds.), *A Critical Review of Thermodynamics*. Baltimore: Mono Book Corporation, pp. 37–47. Also in (Kestin 1976), pp. 313–323.

Moran, P.A.P. (1961). Entropy, Markov processes and Boltzmann's H-theorem. *Proceedings of the Cambridge Philosophical Society, 57*, 833–842.

Morrison, M. (2000). *Unifying scientific theories: physical concepts and mathemathical structures*. Cambridge: Cambridge University Press.

Morrison, P. (1966). Time's arrow and external perturbations. In A. de Shalit, H. Feshbach, & L. van Hove (Eds.), *Preludes in Theoretical Physics in honor of V. F. Weisskopf*. Amsterdam: North Holland, pp. 347–.

Müller, I. (2003). Entropy in non-equilibrium. In (Greven et al. 2003, pp. 79–107).

Navarro, L. (1998). Gibbs, Einstein and the foundations of statistical mechanics. *Archive for History of Exact Science, 53*, 147–180.

Nemytskii, V.V., & Stepanov V.V. (1960). *Qualitative theory of differential equations*. Princeton: Princeton University Press.

Norton, J.D. (2005). Eaters of the lotus: Landauer's principle and the return of Maxwell's demon. *Studies In History and Philosophy of Modern Physics, 36*, 375–411.

von Neumann, J., (1932). Proof of the quasi-ergodic hypothesis. *Proceedings of the National Academy of sciences of the United States of America, 18*, 70–82 and 263–266.

Obcemea, Ch., & Brändas, E. (1983). Analysis of Prigogine's theory of subdynamics. *Annals of Physics, 147*, 383–430.

Olsen, E.T. (1993). Classical mechanics and entropy. *Foundations of Physics Letters, 6*, 327–337.

Penrose, O. (1970). *Foundations of statistical mechanics: a deductive treatment*. Oxford: Pergamon Press.

Penrose, O. (1979). Foundations of statistical mechanics. *Reports on Progress in Physics, 42*, 1937–2006.

Penrose, O., & Percival I. (1962). The direction of time. *Proceedings of the Physical Society, 79*, 605–616.

Pešić, P.D. 1991 The principle of identicality and the foundations of quantum theory. I. The Gibbs paradox. *American Journal of Physics, 59*, 971–974.

Petersen, K. (1983). *Ergodic theory*. Cambridge: Cambridge University Press.

Pitowsky, I. (2001. Local fluctuations and local observers in equilibrium statistical mechanics. *Studies In History and Philosophy of Modern Physics, 32*, 595–607.

Plancherel, M. (1913). Foundations of statistical mechanics. *Annalen der Physik, 42*, 1061–1063.

von Plato, J. (1991). Boltzmann's ergodic hypothesis. *Archive for History of Exact Sciences, 42*, 71–89.

von Plato, J. (1994). *Creating modern probability*. Cambridge: Cambridge University Press.

Poincaré, H. (1889). Sur les tentatives d'explication mécanique des principes de la thermodynamique. *Comptes Rendus de l'Académie des Sciences (Paris), 108*, 550–553. English translation in (Olsen 1993).

Poincaré, H. (1893). Le mécanisme et l'expérience. *Revue de Métaphysique et de Morale, 1*, 534–537. English translation in (Brush 1966).

Poincaré, H. (1896). *Calcul des Probabilités*. Paris: Carré.

Popper, K. (1982). *Quantum theory and the schism in physics*. London: Hutschinson.

Price, H. (1996). *Time's arrow and Archimedes' point*. New York: Oxford University Press.

Prigogine, I. (1955). *Introduction to the thermodynamics of irreversible processes*. New York: Interscience.

Ray, J.R. (1984). Correct Boltzmann counting. *European Journal of Physics, 5* 219–224.

Redhead, M. (1995). *From physics to metaphysics*. Cambridge: Cambridge University Press.

Reichenbach, H. (1956). *The direction of Time*. Berkeley: University of California Press.

Ridderbos, T.M. (2002). The coarse-graining approach to statistical mechanics: how blissful is our ignorance? *Studies in History and Philosophy of Modern Physics, 33*, 65–77.

Ridderbos, T.M & Redhead, M.L.G. (1998). The spin-echo experiment and the second law of thermodynamics. *Foundations of Physics, 28*, 1237–1270.

Rosenthal, A. (1913). Beweis der Unmöglichkeit ergodischer mechanischer Systeme. *Annalen der Physik, 42*, 796–806. English translation in (Brush 2003, p.505–523).

Rovelli, C. (2006). This volume, chapter 12.

Ruelle, D. (1969). *Statistical mechanics: rigorous results*. New York: Benjamin.

Rugh, H.H. (1997). Dynamical approach to temperature *Physical Review Letters 78*, 772–774.

Rugh, H.H. (2001). Microthermodynamic formalism *Physical Review E 64*, 055101.

Saunders, S. (2006). On the explanation for quantum statistics. *Studies in History and Philosophy of Modern Physics, 37*, to appear.

Schrödinger, E., 1950. Irreversibility. *Proceedings of the Royal Irish Academy 53*, 189–195.

Simányi, N. and Szász, D. (1999). Hard balls systems are completely hyperbolic. *Annals of Mathematics, 149*, 35–96.

Sinai, Ya. G. (1963). On the foundation of the ergodic hypothesis for a dynamical system of statistical mechanics *Soviet Mathematics Doklady, 4*, 1818–1822.

Sinai, Ya.G. and Chernov, N.I. (1987). Ergodic properties of certain systems of $2-D$ discs and $3-D$ balls. *Russian Mathematical Surveys, 42* 181–207. Also in Ya.G. Sinai (Ed.) *Dynamical systems; collection of papers* Singapore: Wold Scientific (1991) pp. 345–371.

Shenker, O. (2000). Interventionism in statistical mechanics: some philosophical remarks. http://philsci-archive.pitt.edu/archive/00000151/.

Sklar, L. (1973). Statistical explanation and ergodic theory. *Philosophy of Science, 40*, 194–212.

Sklar, L. (1993). *Physics and Chance. Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press.

Sklar, L. (2002). Sklar, L. (2002). *Theory and thruth*. Oxford: Oxford University Press.

Spohn, H. (1980). Kinetic equations from Hamiltonian dynamics: Markovian limits. *Reviews of Modern Physics 52*, 569–615.

Spohn, H. (1991). *Large Scale Dynamics of interacting Particles*. Berlin: Springer.

Streater, R.F. (1995). *Statistical dynamics; a stochastic approach to non-equilibrium thermodynamics*. London: Imperial College Press.

Styer, D.F. (2004). What good is the thermodynamic limit?" *American Journal of Physics 72*, 25–29; Erratum p. 1110.

Sudarshan, E.C.G., Mathews, P.M., & Rau, J. (1961). Stochastic dynamics of quantum-mechanical systems, *Physical Review, 121*, 920–924.

Szász, D. (1996). Boltzmann's ergodic hypothesis, a conjecture for centuries? *Studia Scientiarum Mathematicarum Hungaria, 31*, 299–322.

Szilard, L. (1925). Über die Ausdehnung der phänomenologischen Thermodynamik auf die Schwankungserscheinungen. *Zeitschrift für Physik, 32*, 753–788.

Tabor, M. (1989). *Chaos and integrability in nonlinear dynamics: an introduction*. New York: Wiley.

Thomsen, J.S. & Hartka, Th.J. (1962). Strange Carnot cycles; thermodynamics of a system with a density extremum *American Journal of Physics 30*, 26–33.

Thompson, C. (1972). *Mathematical Statistical Mechanics*. Princeton: Princeton University Press.

Tisza, L. (1966). *Generalized Thermodynamics*. Cambridge, Mass.: M.I.T. Press.

Tisza, L. & Quay, P.M. (1963). The statistical thermodynamics of equilibrium. *Annals of Physics, 25*, 48–90. Also in (Tisza 1966, pp. 245–287).

Tolman, R.C. (1938). *The principles of statistical mechanics*. London: Oxford University Press.

Touchette, H. (2003). *Equivalence and nonequivalence of the microcanonical and canonical ensembles: a large deviations study*. Ph.D Thesis, McGill University, Montréal.

Touchette, H., Ellis, R.S. & Turkington, B. (2004). An introduction to the thermodynamic and macrostate levels of nonequivalent ensembles. *Physica A 340*, 138-146.

Truesdell, C. (1961). Ergodic theory in classical statistical mechanics. In P. Caldirola (ed.), *Ergodic theories*, New York: Academic Press.

Truesdell, C. (1969). *Rational thermodynamics*. New York: McGraw-Hill.

Truesdell, C. (1980). *The tragicomical history of thermodynamics 1822–1854*. New York: Springer.

Uffink, J. (1995). Can the maximum entropy principle be regarded as a consistency requirement? *Studies in History and Philosophy of Modern Physics, 26*, 223–261.

Uffink, J. (1996). The constraint rule of the maximum entropy principle. *Studies in History and Philosophy of Modern Physics, 27*, 47–79.

Uffink, J. (1996b). Nought but molecules in motion. *Studies In History and Philosophy of Modern Physics, 27*, 373–387.

Uffink, J. (2001). Bluff your way in the second law of thermodynamics. *Studies in History and Philosophy of Modern Physics, 32*, 305–394.

Uffink, J. (2003). Irreversibility and the second law of thermodynamics. In (Greven et al. 2003, pp. 121–146).

Uffink, J. (2004). Boltzmann's Work in Statistical Physics. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2004 Edition), http://plato.stanford.edu/archives/win2004/entries/statphys-Boltzmann.

Uffink, J. (2005). Rereading Ludwig Boltzmann. In P. Hájek, L. Valdés-Villanueva & D. Westerståhl (Eds.) *Logic, Methodology and Philosophy of Science. Proceedings of the twelfth international congress*. London: King's College Publications, pp. 537–555.

Uffink, J. (2006). Insuperable difficulties: Einstein's statistical road to molecular physics. *Studies in History and Philosophy of Modern Physics*, to appear.

Uffink, J. & van Lith, J. (1999). Thermodynamic uncertainty relations. *Foundations of Physics, 29*, 655–692.

Uhlenbeck, G.E. and Ford, G.W. (1963). *Lectures in statistical mechanics*. Providence, Rhode Island: American Mathematical Society.

Vranas, P.B.M. (1998). Epsilon-ergodicity and the success of equilibrium statistical mechanics. *Philosophy of Science, 65*, 688–708.

Wald, R.M. (2001). The thermodynamics of black holes. *Living Reviews in Relativity, 4*, 6. http://www.livingreviews.org/lrr-2001-6.

Winsberg, E. (2004). Laws and statistical mechanics. *Philosophy of Science, 71*, 707–718.

Yourgrau, W., van der Merwe, A., & Raw, G. (1966). *Treatise on irreversible thermophysics*, New York: Macmillan.

Zermelo, E. (1896a). Ueber einen Satz der Dynamik und die mechanische Wärmetheorie. *Annalen der Physik, 57*, 485–494. English translation in (Brush 2003, pp. 382–391).

Zermelo, E. (1896b). Ueber mechanische Erklärungen irreversibler Vorgänge. *Annalen der Physik, 59*, 793–801. English translation in (Brush 2003, pp. 403–411).