Chapter 4
Ehrenfest and Ehrenfest-Afanassjewa on Why Boltzmannian and Gibbsian Calculations Agree

Charlotte Werndl and Roman Frigg

4.1 Introduction

The relation between the Boltzmannian and the Gibbsian formulations of statistical mechanics (SM) is one of the major conceptual issues in the foundations of the discipline. In their celebrated review of SM, Paul Ehrenfest and Tatiana Ehrenfest-Afanassjewa discuss this issue and offer an argument for the conclusion that Boltzmannian equilibrium values agree with Gibbsian phase averages.[1] In this paper, we analyse their argument, which is still important today, and point out that its scope is limited to dilute gases.

[1]The original paper was published in German under the title 'Begriffiche Grundlagen der Statistischen Auffassung in der Mechanik' in 1911. Throughout this paper, we quote the English translation that came out in 1959 under the title 'The conceptual foundations of the statistical approach in mechanics'.

C. Werndl (✉)
Department of Philosophy (KGW), University of Salzburg, Franziskanergasse 1, 5020 Salzburg, Austria
e-mail: charlotte.werndl@sbg.ac.at

C. Werndl · R. Frigg
Department of Philosophy, Logic and Scientific Method, London School of Economics, Houghton Street, London WC2A 2AE, UK
e-mail: r.p.frigg@lse.ac.uk

J. Uffink et al. (eds.), *The Legacy of Tatjana Afanassjewa*, Women in the History of Philosophy and Sciences 7, https://doi.org/10.1007/978-3-030-47971-8_4

4.2 Boltzmannian and Gibbsian Statistical Mechanics

In statistical mechanics (SM) there are two main theoretical frameworks, namely Boltzmannian and Gibbsian SM.[2] Consider a system S consisting of the following: X is the set of all *possible states* (the state space), μ_X is the *probability measure* on X (that is assumed to be invariant under the dynamics) and $T_t(x)$ is the *dynamics* specifying the state of the system after t time steps given that it started in x.[3]

At the beginning of *Boltzmannian SM* stands the introduction of macro-states M_j, $j = 1, \ldots, m$, which are characterised by the values of a set of *macro-variables* $\{f_1, \ldots, f_k\}$ (where both m and k are in $\mathbb{N}$). A macro-variable $f_i : X \to \mathbb{R}$ is a function that associates a value with each $x \in X$. Capital letters F_i denote the values of the f_i. A *macro-state* M_i is defined by a particular set of values $\{F_1, \ldots, F_k\}$. Macro-states are assumed to supervene on micro-states, and hence there corresponds a micro-region $X_{M_j} \subseteq X$ to each M_j, which consists of all $x \in X$ for which the macroscopic variables assume the values characteristic for M_j. The X_{M_i} together form a partition of X, meaning that they do not overlap and jointly cover X. One of the macro-states is then singled out as the equilibrium state, and the *equilibrium values* of the f_i are the values F_i that the macro-variables assume in the equilibrium macro-state. The standard line on how to single out the equilibrium state is that size is the determining factor: the equilibrium state is the state for which $\mu_X(X_{M_i})$ assumes the highest value. As we will see in Sect. 4.5, this definition stands in need of qualification, but since it is widely used, we work with it for now and see how far it takes us.

The most important method to determine the largest macro-state is Boltzmann (1877) combinatorial argument, which Ehrenfest and Ehrenfest-Afanassjewa discuss in detail (1959, 26–30). The argument runs as follows. The state of one particle is given by a point in the six-dimensional state space X_1, and thus the state of the system (of the N particles) is given by N points in X_1. Because the system is confined to a finite container and the energy is constant, only a certain finite part of X_1 is accessible. This accessible part of X_1 is then divided into cells of equal size $\delta\omega$ whose dividing lines run parallel to the position and momentum axes. The result is a finite partition $\Omega := \{\omega_1, \ldots, \omega_l\}, l \in \mathbb{N}$. The cell in which a particle's state lies is referred to as the particle's coarse-grained micro-state. The specification of the coarse-grained micro-state for all particles is called an arrangement. Finally, a specification of the number of particles in each cell is referred to as a distribution $D = (N_1, N_2, \ldots, N_l)$ (N_i is the number of particles in cell ω_i). Each distribution is compatible with several arrangements, and the number of arrangements corresponding to a given distribution D is $G(D) = N! \,/\, N_1!N_2! \ldots, N_l!$.

---

[2]We briefly review both frameworks in this section. More extensive presentations can be found in Frigg (2008) and Uffink's (2007). See Frigg and Werndl (2019) for a discussion of the Gibbs formalism in particular.

[3]In this paper, we mostly follow Ehrenfest and Ehrenfest-Afanassjewa and consider deterministic systems. In our (2017) we discuss stochastic systems and show that the main results carry over to the stochastic context. We consider an explicitly stochastic system below in Sect. 4.6.

Ehrenfest and Ehrenfest-Afanassjewa now associate macro-states with distributions (1959, 49–50): each distribution defines a macro-state. This assumption is motivated by the fact that the macro-properties of a system are a function of the micro-properties, and hence a given macro-variable will assume different values for different distributions (we come back to this assumption below in Sect. 4.4). Clearly, every micro-state x of X corresponds to exactly one distribution $D(x)$. The macro-region X_D is then simply defined as the set of all x that are associated with the macro-state D.

The equilibrium macro-region is the region X_D with the largest measure. To determine this largest macro-region, Boltzmann (1877) provided a classical argument, which Ehrenfest and Ehrenfest-Afanassjewa discuss in detail (1959, 27–31). Boltzmann assumed that the energy e_i of particle i is only dependent on the cell in which it is located (and *not* on the location of the other particles), implying that the total energy of the system is $E = \sum_{i=1}^{l} N_i e_i$. With the further assumption that the number of cells in Ω is small compared to the number of particles, Boltzmann showed that $\mu_X(X_D)$ has a maximum when

$$N_i = \gamma e^{\lambda e_i}, \tag{4.1}$$

where γ and λ are parameters which depend on N and E. Equation (4.1) is now known as the *discrete Maxwell–Boltzmann distribution*. The equilibrium macro-state, therefore, corresponds to the Maxwell–Boltzmann distribution.

However, as Ehrenfest and Ehrenfest-Afanassjewa rightly emphasise (1959, 30), there is a last step missing. The X_D as defined above are $6N$-dimensional, and Eq. (4.1) gives us is the distribution for the cell of largest size relative to the Lebesgue measure μ_X (or more precisely, relative to the $6N$-dimensional subset X_{ES} of X defined by the condition that $E = \sum_{i=1}^{l} N_i e_i$). However, by assumption, the system has constant energy, and so we know that the system's motion takes place on the $6N$-1-dimensional energy hypersurface X_E. Hence, the relevant macro-regions are ones that lie in X_E rather than in X. A quick fix is the following: define the relevant $6N$-1-dimensional macro-regions as the intersection of the $6N$-dimensional X_D with X_E, and use the restriction μ_E, the restriction of μ_X to X_E, to measure their size.

Ehrenfest and Ehrenfest-Afanassjewa are careful to point out that this is not enough to give us what is needed, namely the macro-region of largest size relative to the measure μ_{X_E} on the $6N-1$-dimensional set X_E. Standard presentations of the combinatorial argument simply assume that the possible distributions and the proportion of the different distributions would not change if macro-states were instead defined on X_E, which yields the desired result that the equilibrium region is the largest region on X_E. Ehrenfest and Ehrenfest-Afanassjewa (1959, 30) are more careful. While they also adopt this assumption, they stress that it is in need of further justification.

So the conclusion Ehrenfest and Ehrenfest-Afanassjewa arrive at is that *in the Boltzmannnian framework the observed value in equilibrium for the observable f is the value of f in the macro-region corresponding to the Maxwell–Boltzmann distribution.*

Gibbsian SM studies *ensembles*, infinite collections of independent systems that are all governed by the same equations but start in different initial states. Formally, an ensemble is a probability density $\rho(x, t)$, $x \in X$, describing the probability of finding the state of a system chosen at random from the ensemble in a certain region of X at time t.

Given an ensemble ρ, the Gibbs entropy is

$$S_G[\rho] = -k_B \int_X \rho(x, t) \log[\rho(x, t)] dx, \tag{4.2}$$

where k_B is the Boltzmann constant. An ensemble $\rho(x, t)$ is called *stationary* if and only if it does not depend on time, i.e. $\rho(x, t) = \rho(x)$ for all t. In Gibbsian SM equilibrium is a property of an ensemble. More specifically, the ensemble is in equilibrium if and only if it is stationary, and sometimes it is also required that it has maximum Gibbs entropy given the constraints imposed on the system. The most common constraints give rise to the microcanonical, canonical and grand-canonical distributions (1959, 46–47).

As in Boltzmannian SM, physical observables correspond to a set of real-valued functions f_i, and the *phase average* of such a function in equilibrium is defined as

$$\langle f_i \rangle = \int_X f_i(x)\rho(x) dx. \tag{4.3}$$

According to the canonical understanding of Gibbsian SM, *what is observed in experiments on systems in equilibrium are such phase averages* (1959, 47 and 49). There is, however, a question about the scope of this claim: according to Gibbsian SM, does one *always* observe phase averages or are phase averages only observed in certain situations? The answer to this question is a matter of dispute which depends on how exactly Gibbsian SM is interpreted (for a discussion see Frigg and Werndl 2019). It is not entirely clear what reading of Gibbsian SM Ehrenfest and Ehrenfest-Afanassjewa endorse (though it seems to us that they rather endorse the claim and that, according to Gibbsian SM, always phase averages are observed). Fortunately, this issue does not matter in what follows.

Now, we are in a curious situation. Two different frameworks make predictions for the same experimental values. The Boltzmannian account says that the observed equilibrium value for the observable f_i is the value that it assumes in the macro-region corresponding to the Maxwell–Boltzmann distribution, while the Gibbsian account says that that the equilibrium value is $\langle f_i \rangle$. Do these values coincide? If so, why? If not, which of the values, if any, is correct?

4.3 Ehrenfest and Ehrenfest-Afanassjewa on Gibbs Versus Boltzmann

Ehrenfest and Ehrenfest-Afanassjewa opt for the first solution and set out to show that Boltzmannian equilibrium values and Gibbsian phase averages coincide. Their argument is an important one, and similar points have been made more recently by Davey (2009), Myrvold (2016). They begin by discussing the Gibbsian treatment of the gas with the observable f.[4] According to the Gibbsian framework, what is observed in equilibrium is the phase average. Because energy is conserved, it would be natural to consider the phase average relative to the micro-canonical ensemble (because this is the stationary distribution of maximum Gibbsian entropy under the constraint of constant energy). However, Ehrenfest and Ehrenfest-Afanassjewa do not do this and instead consider the phase average with respect to the canonical ensemble. The canonical ensemble is the stationary distribution of maximum entropy when the energy is allowed to vary:

$$\rho_c(q, p) = e^{\frac{\Psi - E(q,p)}{\Theta}}, \tag{4.4}$$

where $E(q, p)$ is the total energy, Θ is an constant, and Ψ is determined by the constraint that $\int_X \rho_c(q, p) = 1$.

The reason why they consider the phase average with respect to the canonical ensemble is unclear. A possible motivation might be that they want to show that it does not matter which distribution is chosen: Gibbsian SM leads to the same result as Boltzmannian SM regardless of whether one works with the microcanonical or the canonical ensemble.

As a first step they appeal to the well-known result, often referred to as the equivalence between the microcanonical and canonical distributions that holds when the number of particles of a gas is extremely large:

> In an ensemble which is canonically distributed with the modulus $\Theta = \Theta_0$, an overwhelming majority of individuals will have nearly the same total energy $E = E_0$ (Ehrenfest and Ehrenfest-Afanassjewa 1959, 48).

(Here Θ_0 is the fixed value of Θ in Eq. (4.4) of the canonical distribution above and E_0 is the energy value that nearly all individuals will have for the fixed value Θ_0).

Based on this result Ehrenfest and Ehrenfest-Afanassjewa (1959, 48–49) argue that it is plausible that $\int_X f(x) d\rho_c$, the phase average with respect to the canonical distribution ρ_c on X, is approximately equal to $\int_{X_E} f(x) d\rho_m$, the phase average with respect to the microcanonical distribution ρ_m on X_E (when f is restricted to X_E).

The next step is the vital move in the argument. Recall that the combinatorial argument shows that the equilibrium macro-region is the largest macro-region. So the macro-value corresponding to the Maxwell–Boltzmann distribution is the macro-value that is taken by more microstates than any other macro-value on X_E.[5] It is

[4] For ease of notation, we suppress the subscript 'i' from now.

[5] Strictly speaking, this is true only under an additional assumption that we discuss in the next section.

crucial to be clear on the sense of ‘large’ that is being used here. What the combinatorial argument shows is that the equilibrium macro-region is larger than *any other macro-region*. It does not show that the equilibrium macro-region is large in an absolute sense, i.e. that it occupies the largest part of X_E. The latter does not follow from the former. A macro-region can be larger than any other macro-region without being large relative to X_E. Ehrenfest and Ehrenfest-Afanassjewa bridge the gap between a relative and the absolute sense of ‘large’ by referring to results due to Jeans (1904, Sects. 46–56), who argues that *nearly all* states in X_E are in the macro-region corresponding to the Maxwell–Boltzmann distribution. Hence, f assumes the equilibrium value on almost all states in X_E. From this, they infer that this value is approximately equal to the Gibbsian phase average derived in the previous paragraph.

So their conclusion is that in a system in which the combinatorial argument applies, the Boltzmannian equilibrium value and the Gibbsian phase average with respect to the macro-variable f are approximately the same.

4.4 Assessment of Ehrenfest and Ehrenfest-Afanassjewa’s Argument

The considerations we make to assess the Ehrenfest and Ehrenfest-Affanassjewa argument fall into two groups. Considerations in the first group concern the combinatorial argument and its limitations; considerations in the second group concern the identity argument in the last section. We will focus mainly on the second group but will begin by making a few observations about the first.

As has been pointed out previously,[6] a core assumption of the combinatorial argument, namely that $E = \sum_{i=1}^{l} N_i e_i$, is very restrictive. In essence, this assumption implies that the argument only applies (even in an approximate form) to *dilute gases*. So it is unsurprising that Ehrenfest and Ehrenfest-Afanassjewa (1911, 36–60) talk about gas systems when presenting the combinatorial argument. However, it remains unclear from the text whether they are clear on the fact that it *only* applies to dilute gases.

Second, the conclusion that the macro-value of f in the Maxwell–Boltzmann distribution is the macro-value that is taken by more micro-states than any other macro-value on X_E follows only under the strong assumption that f assumes a different value for *every* macro-region. However, Lavis (2005, 2008) pointed out that this need not always be the case.[7] Macro-regions can show *degeneracy* in the sense that f can assume the same value in several regions. It is possible that a number of such (non-equilibrium) macro-regions *taken together* are larger than the equilibrium region, and so f assumes the equilibrium value in a region of the state space that is smaller than the union of the degenerate macro-regions. Lavis (2005,

[6] See, for instance, Uffink (2007) and Werndl and Frigg (2015b).

[7] Lavis (2005, 2008) discussed the case of the Boltzmann entropy, but the point obviously generalises to phase functions.

2008) shows that this happens in the case of the baker's gas, thereby driving home the point that degeneracies causing difficulties is more than just a theoretical possibility.

Let us set these concerns aside and assume, for the sake of argument, that we are dealing with a dilute gas and a 'well-behaved' function f (we will discuss what happens if these assumptions fail in Sect. 4.5). Does Ehrenfest and Ehrenfest-Afanassjewa's equivalence argument hold under these assumptions? It is obvious that their argument contains a gap. They conclude from the fact that f assumes the equilibrium value on nearly all states in X_E that the average of f over X_E is approximately equivalent to that value. This, however, is true only if the non-equilibrium values are not disproportionately far away from the equilibrium value. If the non-equilibrium values differ significantly from the equilibrium values, their contribution to the average can be significant and the average need no longer be equal to the equilibrium value of the function, not even approximately.

To rule out such a scenario one needs to assume that f satisfies some kind of 'small fluctuation condition'. The most common condition of this kind is now known as the *Khinchin Condition*. The condition plays a crucial role in the work of Khinchin (1960 [1949]) and variants of it have been appealed to in the foundational literature on SM, for instance by Malament and Zabell (1980), Myrvold (2016). This condition requires that the observable f equals the phase average nearly everywhere on phase space. Formally:

> There is a $\bar{X} \subseteq X$ with $\mu_X(\bar{X}) = 1 - \delta$ for a small $\delta \geq 0$ such that $|f(x) - \langle f(x) \rangle| \leq \varepsilon$ for all $x \in \bar{X}$ and a very small $\varepsilon \geq 0$.

Under Ehrenfest and Ehrenfest-Afanassjewa's assumptions the Boltzmannian equilibrium macro-region satisfies the condition on $\bar{X}$. Let F_{equ} be the value of f in that macro-region. It then follows that $|\langle f(x) \rangle - F_{equ}| \leq \varepsilon$, and therefore the Boltzmannian value and the Gibbsian average agree, at least approximately.

Ehrenfest and Ehrenfest-Afanassjewa, however, do not appeal to this formulation of the condition, but to a variant of the Khinchin condition that we call the *Ehrenfest-Afanassjewa Condition*. The condition is that the observable f is approximately equal to the Boltzmannian equilibrium value nearly everywhere on phase space and that the observable does not take extreme values on the rest of the phase space. Formally, the Ehrenfest-Afanassjewa Condition can be formulated as follows[8]:

> Consider a system of the kind introduced in Sect. 4.2 endowed with an observable f. Further assume that the system has a Boltzmannian equilibrium with equilibrium macro-value F_{equ}.

[8] A variant of the Ehrenfest-Afanassjewa Condition requires that the observable f is constant nearly everywhere on phase space and does not take extreme values on the rest of the phase space:

> There is an constant $C \in \mathbb{R}$ and a $\bar{X} \subseteq X$ with $\mu_X(\bar{X}) = 1 - \delta$ (for a small $\delta \geq 0$) such that (i) $|f(x) - C| \leq \varepsilon$ for all $x \in \bar{X}$ for a very small $\varepsilon \geq 0$ and (ii) $|\int_{X \setminus \bar{X}} f(x) d\mu_X - C\delta)| \leq \gamma$ (for a very small $\gamma \geq 0$).

Because the Boltzmannian equilibrium macro-value F_{equ} takes up more than δ of phase space, it follows that F_{equ} is very close to C. Therefore, $|f(x) - F_{equ}| \leq \epsilon_1$ for a small $\varepsilon_1 \geq 0$ for all $x \in \bar{X}$ and $|\int_{X \setminus \bar{X}} f(x) d\mu_X - F_{equ}\delta)| \leq \gamma_1$ (for a very small $\gamma_1 \geq 0$). This is in fact the original Ehrenfest-Afanassjewa Condition and so the variant is in fact equivalent to the original Ehrenfest-Afanassjewa Condition.

Then there is an $\bar{X} \subseteq X$ with $\mu_X(\bar{X}) = 1 - \delta$ (for a small $\delta \geq 0$) such that (i) $|f(x) - F_{equ}| \leq \varepsilon$ for all $x \in \bar{X}$ (for a very small $\varepsilon \geq 0$) and (ii) $|\int_{X \setminus \bar{X}} f(x) d\mu_X - F_{equ}\delta)| \leq \gamma$ (for a very small $\gamma \geq 0$).

A simple calculation shows that for systems that satisfy the Ehrenfest-Afanassjewa Condition with respect to f, the phase average is approximately equal to the Boltzmannian equilibrium macro-value F_{equ}:

$$|\langle f(x) \rangle - F_{equ}| \leq$$
$$|\int_{\bar{X}} f(x) d\mu_X - F_{equ}(1-\delta)| + |\int_{X \setminus \bar{X}} f(x) d\mu_X - F_{equ}\delta| \leq$$
$$\varepsilon(1-\delta) + \gamma \text{ (because of (i) and (ii) of the Khinchin condition).}$$

It is interesting to discuss both the Khinchin and the Ehrenfest-Afanassjewa conditions because, depending on the context, one or the other may turn out to be more useful. There is, however, a slight mismatch between the Ehrenfest-Afanassjewa Condition and the calculations of Ehrenfest and Ehrenfest-Afanassjewa: they perform Gibbsian phase space averaging with the canonical and not the micro-canonical distribution. However, because of the equivalence of the micro-canonical and macro-canonical ensemble as discussed above this difference does not matter; and if for some reason it did, one could simply perform the Gibbsian calculations with the microcanonical ensemble.

It is important to note that neither of the two conditions is in any way trivially true. Khinchin could prove his condition only for the special case of sum functions in non-interacting systems (sum functions are functions in many-particle systems that can be written as a sum over one-particle functions). The generalisation of this result to the case interacting system is a veritable challenge and no general solution has been found to date.[9]

Ehrenfest and Ehrenfest-Afanassjewa argue in their survey that the Ehrenfest-Afanassjewa condition is satisfied. Their argument is valid but only subject to a change in one of the assumptions and an additional assumption in their argument. Namely, first, as outlined above, they assume (by referring to Jeans 1904, Sects. 46–56) that nearly all states in X_E are in the macro-region corresponding to the Maxwell–Boltzmann distribution. We have seen above that this need not always be the case. Furthermore, a closer look at Jeans' text reveals that he does not actually offer a proof of the claim. What Jeans shows is that the nearly all of phase space X is taken up by macro-regions with a distribution D *very close* to the Maxwell–Boltzmann distribution. Hence the assumption that the macro-region corresponding to the exact Maxwell–Boltzmann distribution is large in absolute terms has to be given up. Fortunately, a weaker assumption provides what we need. All that is required for the argument to go through is that the observable f is such that macro-regions with distribution D very close to the Maxwell–Boltzmann distribution have approximately the

[9]See Uffink's (2007, 1020–1028) for a discussion.

same macro-value as the macro-regions with the Maxwell–Boltzmann distribution.[10] Note that this amounts to conditions imposed on the Boltzmannian macro-structure f.

With this new assumption in place, Jeans' (1904, Sects. 46–56) calculations indeed imply that condition (i) of the Ehrenfest-Afanassjewa Condition is satisfied. Second, Jeans (1904, Sects. 46–56) shows that the states whose macro-values are not very close to the Maxwell–Boltzmann distribution take up a tiny fraction of phase space, i.e. $X \setminus \bar{X}$ is extremely small. But what is still needed is the further condition that f does not take extremely large or extremely low values on $X \setminus \bar{X}$ (and this again is a condition imposed on f). With this new assumption in place, (ii) of the Ehrenfest-Afanassjewa condition is satisfied. Hence, we conclude that *with the modifications just outlined the Ehrenfest-Afanassjewa condition is satisfied and the Boltzmannian equilibrium value and the Gibbsian phase average lead to approximately the same result*.

To sum up, Ehrenfest and Ehrenfest-Afanassjewa identify an important case where the Boltzmannian equilibrium values and the Gibbsian phase averages agree. However, their argument relies on strong assumptions, and while these assumptions are satisfied for certain observables in the case of dilute gases, the assumptions need not hold in general. In fact, in the remainder of this paper, we discuss cases that do not fit Ehrenfest and Ehrenfest-Afanassjewa's mould. First, there are cases where the Boltzmannian equilibrium value is different from the Gibbsian phase average. This shows that it is an important task for foundational debates to find out under what conditions the Boltzmannian equilibrium value and the Gibbsian phase average agree or disagree. Examples of disagreement will be discussed in Sect. 6. Second, there are cases where the Boltzmannian equilibrium value and the Gibbsian phase averages agree but where the Ehrenfest-Afanassjewa condition is not satisfied. The Ehrenfest-Afanassjewa condition and the Khinchin condition provide one condition where there is an agreement (cf. also Werndl and Frigg 2017a; 2017b, 2020).

For instance, consider the Kac ring, consisting of an even number N of sites distributed equidistantly around a circle. On each site, there is a spin, which can be in states up (u) or down (d). A *micro-state* x^{kr} of the Kac ring is a specific combination of up and down spin for all sites and the full state space $Z = K^{kr}$ consist of all combinations of up and down spins (i.e. of 2^N elements). There are s, $1 \leq s \leq N-1$, spin flippers distributed at some of the midpoints between the spins. The dynamics rotates the spins one spin site in the clockwise direction every second (or whichever unit of time one chooses), and when the spins pass through a spin flipper, they change their direction. The measure that is usually considered is the uniform measure $\mu_{X^{kr}}$ on X^{kr} (Lavis 2008). The *macro-states* usually considered are the *total number of up spins*, conveniently labelled as M_i^K, where i denotes the total number of up spins, $0 \leq i \leq N$. The Kac-ring with the standard macro-state structure is a paradigm example where Boltzmannian equilibrium values and Gibbsian phase averages agree.

[10]Given a certain macro-variable f and an allowable difference between the Gibbsian phase average and the Boltzmannian equilibrium macro-value, one could precisely quantify what notion of 'approximately the same macro-value as the Maxwell–Boltzmann distribution' would be needed in order for the Khinchin theorem to go through by making use of the calculations in Jeans (1904, Sects. 46–56).

However, it is not an instance of the Ehrenfest-Afanassjewa-condition because, as shown in Lavis (2005, 2008), the equilibrium macro-region corresponding to an equal number of up and down spins only takes up less than half of state space (the rest is taken up by macro-states that are macroscopically distinguishable from the Boltzmannan equilibrium macro-state). Other examples where the Boltzmannian equilibrium value and the Gibbsian phase average agree but where the Ehrenfest-Afanassjewa condition the Khinchin condition does not apply include the baker's gas with the standard macro-state structure and the ideal gas with the standard macro-state structure (cf. Werndl and Frigg 2017a; 2017b, 2020). The reason why the Boltzmannian equilibrium value and the Gbbsian phase average agree in these cases will be discussed later in Sect. 4.7.

4.5 Beyond Dilute Gases

As we have seen above, the combinatorial argument is restricted to dilute gases. Most systems of interest in SM are not of this kind and so this is a serious restriction. In two recent papers, we have discussed this problem at length and proposed an alternative Boltzmannian definition of equilibrium (2015a, 2015b). On this definition, it is not size but 'residence time' that defines equilibrium: the macro-state in which the system spends most of its time is the equilibrium macro-state. More specifically, define LF_R to be the fraction of time a system spends in region $R \subseteq X$ in the long run:

$$LF_R(x) = \lim_{t \to \infty} \frac{1}{t} \int_0^t 1_A(T_\tau(x)) \mathrm{d}\tau, \tag{4.5}$$

where $1_A(x)$ is the characteristic function of R: $1_A(x) = 1$ for $x \in R$ and 0 otherwise.

'Most' is interpreted as requiring that the system spends more time in equilibrium than in any other macro-state, leading to the notion of an γ-ε-equilibrium[11]:

> Let $\gamma > 0$ and let ε be a very small positive real number, $\varepsilon < \gamma$. If there is a macro-state $M_{F_1^*,\ldots,F_l^*}$ satisfying the following condition, then it is the γ-ε-equilibrium state of S: There exists a set $Y \subseteq X$ such that $\mu_X(Y) \geq 1 - \varepsilon$, and all initial states $x \in Y$ satisfy $LF_{X_{M_{F_1^*,\ldots,F_l^*}}}(x) \geq LF_{X_{M_{F_1,\ldots,F_l}}}(x) + \gamma$ for all macro-states $M \neq M_{F_1^*,\ldots,F_l^*}$

Clearly, *the value observed in equilibrium is simply the value associated with the equilibrium macro-state*. Further, it should be mentioned that one can prove that equilibrium states defined in this way correspond to the largest macro-region in the sense that their measure is $\gamma - \epsilon$ larger than any other macro-region (Werndl and

[11] Alternatively, 'most' can also be understood as referring to the fact that the system spends at least $\alpha > 1/2$ of its time in equilibrium, leading to the different notion of an α-ε-equilibrium. Nothing in what follows hinges on which notion of equilibrium is adopted (cf. Werndl and Frigg 2015b and forthcoming references).

Frigg 2015b). This provides a notion of equilibrium that is fully general in that it does not depend on the system's dynamics and is hence applicable also to strongly interacting systems like solids and fluids.

4.6 An Example Where Boltzmannian Equilibrium Values and Gibbsian Phase Averages Differ

In this section, we see that Ehrenfest and Ehrenfest-Afanassjewa's result fails to generalise: in strongly interacting systems like solids and fluids the Boltzmannian equilibrium value and the Gibbsian phase average can differ. The six-vertex model with energy as the relevant macro-variable will serve as an example of a case where the Boltzmannian and Gibbsian equilibrium values differ. Consider a two-dimensional quadratic lattice with N sites on a torus (the choice of a torus ensures that every grid point has exactly four nearest neighbours, thus allowing to neglect border effects). Each site is connected to its four nearest neighbours by edges. Each edge carries an arrow that either points towards or away from the site. The so-called 'ice-rule' restricts the allowable arrangements of the arrows: the arrows have to be distributed in a way such that at each site in the lattice there are exactly two inward and two outward pointing arrows. It is easy to see that there are exactly six configurations of the arrows that satisfy the ice-rule, and they are shown in Fig. 4.1. The name 'six-vertex model' is motivated by the existence of these six configurations.

The reason for the name 'ice-rule' is that in frozen water each oxygen atom is connected to four other oxygen atoms. So the sites can be thought of as representing oxygen atoms and the edges as representing their bonds. For each bond, there is a hydrogen atom that does not sit in the middle between the two oxygen atoms but instead occupies a position closer to one of the oxygen atoms. Thus, the arrows can be interpreted as indicating to which oxygen atom the hydrogen atom is closer. The ice-rule then corresponds to the requirement that each oxygen atom has two close and two remote hydrogen atoms. Not only water ice but also several crystals, in particular potassium dihydrogen phosphate, satisfy the ice-rule (cf. Baxter 1982; Lavis and Bell 1999; Slater 1941).

The micro-states of the six-vertex model $\xi = (\xi_1, \ldots, \xi_N)$ are given by assigning one of the six types of configurations of the arrows permitted by the ice rule to each site in the model. Each of the six configurations has a certain energy ϵ_j, $1 \le j \le 6$.

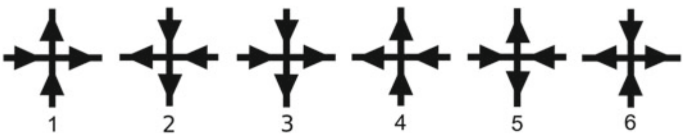

Fig. 4.1 The configurations of the six-vertex model

Denote by $\epsilon(\xi_j)$ the energy of the jth configuration. Then the energy of the state ξ is given by:

$$E(\xi) = \sum_{i=1}^{N} \epsilon(\xi_i). \tag{4.6}$$

We now assume that the energy of the different configurations is $\epsilon_1 = \epsilon_2 = 0$ and $\epsilon_3 = \epsilon_4 = \epsilon_5 = \epsilon_6 = 1$. The probability of the micro-states is given by the canonical distribution $p(\xi) = e^{-E(\xi)/kT}/Z$, with $Z = \sum_\xi e^{-E(\xi)/kT}$. Note that this is merely the probability measure over the micro-states, and is per se neither Boltzmannian nor Gibbsian. For the six-vertex model, one usually works with a stochastic dynamics. More specifically, the underlying dynamics is assumed to be an irreducible Markov chain (Baxter 1982; Lavis and Bell 1999; Werndl and Frigg 2020). The probability $p(\xi)$ is then invariant under the Markov dynamics and is thus a stationary probability measure.

We now study the six-vertex model with the *internal energy* E as defined in Eq. (4.6) as the relevant macro-variable for low temperatures. The lowest energy value is $E = 0$, which defines a macro-state M_0 with macro-region $X_{M_0} = \{\xi^*, \xi^+\}$ (here ξ^* is the state where all vertices are in the first configuration, and ξ^+ is the state where all vertices are in the second configuration). Note that the lower the temperature, the larger the probability of the lower energy states; and the higher the temperature, the more uniform the probability measure. Hence for sufficiently low temperatures, the probability mass is concentrated on low-energy states. For this reason, X_{M_0} is the largest macro-region. Because the dynamics is an irreducible Markov chain, the model spends most of its time in M_0. It follows that M_0 is the Boltzmannian equilibrium state and $E = 0$ is the Boltzmannian equilibrium value (cf. Werndl and Frigg 2020).

Let us now turn to the Gibbsian treatment. Here, $p(\xi)$ is the stationary measure of maximum entropy, and E is observable. E will assume its lowest value $E = 0$ only for two specific micro-states, namely ξ^* and ξ^+. For all other states (and they all have positive probability), the value of E will be higher. From this, we conclude that the Gibbsian phase average $\langle E \rangle$ is greater than zero and hence higher than the Boltzmannian equilibrium value. Thus, the Boltzmannian equilibrium value and the Gibbsian phase average differ.

Now, of course, the question is whether this difference can be significant. To see that this can be so, choose a T such that $\{\xi^*, \xi^+\}$ is still the largest macro-region but that the probability of this macro-region is equal or less than 0.5.[12] Clearly, the Boltzmannian equilibrium value is still $E = 0$. Yet the second lowest macro-value is $E = \sqrt{N}$, which is the energy corresponding to micro-states where all columns of the lattice except one are taken up by states which are in the first or the second configuration, and the states in the exceptional row are all states in the third or

[12] As we have seen, for sufficiently low temperatures $\{\xi^*, \xi^+\}$ is the largest macro-region. The higher the temperature, the more uniform is the probability measure. Hence, for sufficiently high temperatures, the largest macro-region will differ from $\{\xi^*, \xi^+\}$. Because the canonical distribution is continuous in T, there exists a T such that $\{\xi^*, \xi^+\}$ is the largest macro-region but its probability is ≤ 0.5.

fourth configuration.[13] It follows that $\langle E \rangle$ is higher than $\sqrt{N}/2$. Consequently, the Gibbsian phase average and the Boltzmannian equilibrium value will differ by more than $\sqrt{N}/2$, which is not a difference that is negligible (especially when N is large). Note also that the Boltzmannian macro-value that is closest to the value obtained from Gibbsian phase space averaging is larger or equal to $\sqrt{N}$. But this Boltzmannian macro-value is *different* from the Boltzmannian equilibrium macro-value, which is zero. This again underlines that Gibbsian phase space averaging results in a different outcome than the Boltzmannian calculations.[14]

4.7 When Boltzmann and Gibbs Agree

Boltzmannian equilibrium values and Gibbsian phase averages can come apart. This raises the question under what conditions the two coincide. We have already seen in Sect. 4.4 that one situation where there is agreement is when the Ehrenfest-Afanassjewa condition is satisfied. However, as already noted then, there are important cases including the baker's gas with the standard macro-state structure, the KAC-ring with the standard macro-state structure and the ideal gas with the standard macro-state structure, that do not, in general, satisfy the Ehrenfest-Afanassjewa condition or the Khinchin condition.

In our (2017b, 2020) we present *another* set of conditions under which the Boltzmannian equilibrium value and the Gibbsian phase average coincide. Intuitively speaking, the conditions are: (i) the measure on phase space is the product measure of the one-constituent space; (ii) the macro-variable considered is the sum of a one-constituent observable; and (iii) this one-constituent observable takes finitely many values with the same probability. With these conditions in place, the average equivalence theorem then shows that, if a Boltzmannian equilibrium exists, the Boltzmannian equilibrium value and the Gibbsian phase average coincide:

> **Average Equivalence Theorem (AET)**. Suppose that a system with phase space X, dynamics T_t and measure μ_X is composed of $N \geq 1$ constituents. That is, the state $x \in X$ is given by the N coordinates $x = (x_1, \ldots, x_N)$; $X = X_1 \times X_2 \ldots \times X_N$, where $X_i = X_{oc}$ for all i, $1 \leq i \leq N$ (X_{oc} is the one-constituent space). Let μ_X be the product measure $\mu_{X_1} \times \mu_{X_2} \ldots \times \mu_{X_N}$, where $\mu_{X_i} = \mu_{X_{oc}}$ is the measure on X_{oc}. Suppose that an observable κ is defined on the one-particle space X_{oc} and takes the values $\kappa_1, \ldots, \kappa_k$ with equal probability $1/k$, $k \leq N$.[15] Suppose that the macro-variable K is the sum of the one-component observable, i.e. $K(x) = \sum_{i=1}^{N} \kappa(x_i)$. Then the value corresponding to the largest macro-region as well as the value obtained by phase space averaging is $\frac{N}{k}(\kappa_1 + \kappa_2 + \ldots \kappa_N)$.

[13]Such micro-state corresponds to the smallest possible departure from the macro-state with zero energy because the number of downward pointing arrows is the same for all rows. From this, then follows that there has to be a perturbation in each row and that $\sqrt{N}$ has to be the second lowest value of the internal energy (Lavis and Bell 1999).

[14]Further examples where the Gibbsian phase average and the Boltzmannian equilibrium value come apart can be found in our (2017b and 2020).

[15]It is assumed that N ia a multiple of k, i.e. $N = k * s$ for some $s \in \mathbb{N}$.

This theorem applies to the KAC-ring and the other examples (baker's system, ideal gas) mentioned above as cases where the Boltzmannian equilibrium value agrees with the Gibbsian phase average but where the Ehrenfest-Afanassjewa condition does not apply. Hence, it explains in these cases why the Boltzmannian equilibrium value and the Gibbsian phase average coincide. As it should be, the theorem does not apply to the six vertex model with the energy macro-variable because conditions (i) and (iii) are not satisfied (the measure is not the product measure of the one-constituent space, and the macro-variable considered is not the sum of a one-constituent observable, taking values with equal probability).

Note that the conditions of the Average Equivalence Theorem are not necessary for Boltzmannian equilibrium values and Gibbsian phase averages to coincide. In particular, that the macro-variable is a sum of the variables on the one-component space, that the macro-variable on the one-component space corresponds to a partition into cells of equal probability, or that the measure on state space is the product measure of the measure on the one-component space are strong conditions that are often not satisfied. This is illustrated by our example of the dilute gas with the macro-variables we discussed above. As we have seen, this example is an instance of the Ehrenfest-Afanassjewa condition. However, it is *not* an instance of the AET. More specifically, it is *not* the case that all sums of possible values of the one-component variable are possible values of the macro-variable f (because of the requirement that the total energy is constant, only certain sums of values of the one-component variable are possible macro-values). Hence the condition that the macro-variable K is the sum of the one-component variable where all sums of possible values of the one-component variable are possible values of the macro-variable is violated.

To conclude, the Ehrenfest-Afanassjewa condition instead and the Khinchin condition and the conditions of the AET provide sufficient but not necessary conditions. So they just identify two cases where the Boltzmannian equilibrium values and Gibbsian phase averages agree. We suspect that there will be other conditions where the Boltzmannian equilibrium values and Gibbsian phase averages agree.

4.8 Conclusion

We have considered Ehrenfest and Ehrenfest-Afanassjewa's argument for the conclusion that Boltzmannian equilibrium values and Gibbsian phase averages agree. We pointed out that their argument is true only under special circumstances. This is not a shortcoming of their proof but an inherent limitation of the claim: it is not generally the case that Boltzmannian equilibrium values and Gibbsian phase averages agree. We discussed the example of the six-vertex model and showed that in that model the two values come apart. We then offered a general theorem providing conditions for the equivalence of Boltzmannian equilibrium values and Gibbsian phase averages. The conditions of the theorem are sufficient but not necessary. This raises the important question under what other conditions Boltzmannian equilibrium values and Gibbsian phase averages agree.

References

Baxter, R. J. (1982). *Exactly solved models in statistical mechanics*. London: Academic Press.

Boltzmann, L. (1877). Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung resp. den Sätzen über das Wärmegleichgewicht. *Wiener Berichte*, *76*, 373–435.

Davey, K. (2009). What is Gibbs's canonical distribution? *Philosophy of Science*, *76*, 970–983.

Ehrenfest, P., & Ehrenfest-Afanassjewa, T. (1959). *The conceptual foundations of the statistical approach in mechanics*. Ithaca, N.Y.: Cornell University Press.

Frigg, R. (2008). A field guide to recent work on the foundations of statistical mechanics. In D. Rickles (Ed.), *The ashgate companion to contemporary philosophy of physics* (pp. 99–196). London: Ashgate.

Frigg, R., & Werndl, C. (2019). Can somebody please say what Gibbsian statistical mechanics says? *The British Journal for the Philosophy of Science*, online first, https://doi.org/10.1093/bjps/axy057.

Jeans, J. H. (1904). *The dynamical theory of gases*. Cambridge: Cambridge University Press.

Khinchin, A. I. (1960) [1949]. *Mathematical foundations of statistical mechanics*. Mineola/NY: Dover Publications.

Lavis, D. (2005). Boltzmann and Gibbs: An attempted reconciliation. *Studies in History and Philosophy of Modern Physics*, *36*, 245–73.

Lavis, D. (2008). Boltzmann, Gibbs and the concept of equilibrium. *Philosophy of Science*, *75*, 682–96.

Lavis, D., & Bell, G. M. (1999). *Statistical mechanics of lattice systems, Volume 1: Closed form and exact solutions*. Berlin and Heidelberg: Springer.

Malament, D., & Zabell, S. L. (1980). Why Gibbs phase averages work. *Philosophy of Science 47*, 339–49.

Myrvold, W. C. (2016). Probabilities in statistical mechanics. In C. Hitchcock & A. Hájek (Eds.), *The Oxford handbook of probability and philosophy* (pp. 573–600). Oxford: Oxford University Press.

Slater, J. C. (1941). Theory of the transition in KH^2PO^4. *Journal of Chemical Physics*, *9*, 16–33.

Uffink, J. (2007). Compendium of the foundations of classical statistical physics. In J. Butterfield & J. Earman (Eds.), *Philosophy of physics* (pp. 923–1047). Amsterdam: North Holland.

Werndl, C., & Frigg, R. (2015a). Rethinking Boltzmannian equilibrium. *Philosophy of Science*, *82*, 1224–35.

Werndl, C., & Frigg, R. (2015b). Reconceptionalising equilibrium in Boltzmannian statistical mechanics. *Studies in History and Philosophy of Modern Physics*, *49*, 19–31.

Werndl, C., & Frigg, R. (2017a). Boltzmannian equilibrium in stochastic systems. In M. Michela & R. Jan-Willem (Eds.), *Proceedings of the EPSA15 conference* (pp. 243–254). Berlin and New York: Springer.

Werndl, C., & Frigg, R. (2017b). Mind the gap: Boltzmannian versus Gibbsian equilibrium. *Philosophy of Science*, **84**, 1289–1302.

Werndl, C., & Frigg, R. Forthcoming. When does a Boltzmannian equilibrium exist?. In D. Bedingham, O. Maroney, C. Timpson (Eds.), *Quantum foundations of statistical mechanics*. Oxford: Oxford University Press.

Werndl, C., & Frigg, R. (2020). When do Gibbsian phase averages and Boltzmannian equilibrium values agree? Studies in History and Philosophy of Modern Physics, DOI:https://www.sciencedirect.com/science/article/abs/pii/S1355219820300903.