When do Gibbsian Phase Averages and Boltzmannian Equilibrium Values Agree?

Abstract

This paper aims to shed light on the relation between Boltzmannian statistical mechanics and Gibbsian statistical mechanics by studying the Mechanical Averaging Principle, which says that, under certain conditions, Boltzmannian equilibrium values and Gibbsian phase averages are approximately equal. What are these conditions? We identify three conditions each of which is individually sufficient (but not necessary) for Boltzmannian equilibrium values to be approximately equal to Gibbsian phase averages: the Khinchin condition, and two conditions that result from two new theorems, the Average Equivalence Theorem and the Cancelling Out Theorem. These conditions are not trivially satisfied, and there are core models of statistical mechanics, the six-vertex model and the Ising model, in which they fail.

Contents

1 Introduction

Characterising equilibrium and determining properties of systems in equilibrium are central issues in statistical mechanics (SM). These tasks are, however, complicated by the fact that there are two competing theoretical approaches in SM, one associated with Boltzmann and the other with Gibbs. This would not be a problem if the two formalisms were equivalent, or at least inter-translatable (for instance in a similar way as the Schrödinger and the Heisenberg pictures in quantum mechanics). But they are not. Boltzmannian SM (BSM) and Gibbsian SM (GSM) offer different characterisations of equilibrium and distinct procedures for determining equilibrium properties.

The relation between BSM and GSM is a somewhat understudied topic. The aim of this paper is to shed light on the relation between BSM and GSM by exploring the so-called Mechanical Averaging Principle (MAP). The principle says that, under certain conditions, Boltzmannian equilibrium values and Gibbsian phase averages are approximately equal. What are these conditions? We identify three conditions, each individually sufficient (but not necessary) for Boltzmannian equilibrium values to be approximately equal to Gibbsian phase averages: the Khinchin condition (which comes in two different versions), and two conditions that result from two new theorems, the Average Equivalence Theorem and the Cancelling Out Theorem. These theorems cover some paradigmatic models such as the dilute gas and provide a rationale for why Boltzmannian equilibrium values and Gibbsian phase averages agree in these cases. An important insight is that agreement depends both on the model and the macro-variables, and it can happen that in the same model there is agreement for one macro-variable and disagreement for another macro-variable.

These conditions are, however, not trivially satisfied. We provide several examples in which Boltzmannian equilibrium values and Gibbsian phase averages come apart. We also show that Boltzmannian and Gibbsian equilibria do not exist under the same conditions: there are systems with a Gibbsian equilibrium that fail to have a Boltzmannian equilibrium. Some of our examples, in particular the Ising model and the six-vertex model, are core models of SM. Discrepancies between Boltzmannian equilibrium values and Gibbsian phase averages therefore cannot be dismissed as 'mathematical contrivances' that are irrelevant to the practice of the discipline.

The paper is organised as follows. In Section 2 we define deterministic and stochastic models, and we introduce the Boltzmannian and the Gibbsian characterisations of equilibrium. In Section 3 we articulate the Mechanical Averaging Principle and explain how this principle relates to the averaging principle usually appealed to in GSM. In Section 4 we identify and discuss in some detail three conditions under which Boltzmannian equilibrium values and Gibbsian phase averages coincide. In Section 5 we turn to examples where these conditions fail and, as a result, Boltzmannian equilibrium values and Gibbsian phase averages come apart. We also show that equilibria exist under different conditions in the two frameworks. Section 6 concludes.

2 Defining Equilibrium

In this section we first introduce deterministic and stochastic models, and we then discuss the Boltzmannian and the Gibbsian characterisations of equilibrium.

2.1 Deterministic and Stochastic Models

From a mathematical point of view, SM studies models that are equipped with a measure that is invariant over time. The dynamics of such a model can be either deterministic or stochastic, and so we speak of *deterministic models* and *stochastic models* respectively; we speak of *models* (without qualifications) if it does not matter whether the dynamics is deterministic or stochastic.[1] The formal treatment of these models is different and for this reason we introduce them one at a time. We do not take a stance on the question whether deterministic or stochastic models are more important or fundamental and stick with scientific practice where both types are used side by side.

A deterministic model has a *state space* X, which contains all possible *micro-states* the model can be in. In a typical mechanical N-particle model the state space has $6N$ dimensions, three dimensions for the position of each particle and three dimensions for the corresponding momenta, but the notion of a state space is in no way restricted to such cases and any set of relevant micro-variables can form a state space. This space is equipped with a σ-algebra Σ_X of subsets of X and a measure μ_X on (X, Σ_X). The dynamics of the model is given by an *evolution function* $T_t : X \to X$, where $t \in \mathbb{R}$ if time is continuous and $t \in \mathbb{Z}$ if time is discrete. T_t is assumed to be measurable in (t, x) and to satisfy the requirement $T_{t_1+t_2}(x) = T_{t_2}(T_{t_1}(x))$ for all $x \in X$ and all $t_1, t_2 \in \mathbb{R}$ or $\mathbb{Z}$. If the model is Hamiltonian, T_t is given by the solution of Hamilton's equation of motion. However, SM is not restricted to Hamiltonian models and, as we shall see, there are cases where T_t is specified directly (without first formulating an equation of motion). The measure μ_X is required to be invariant under the dynamics, meaning that $\mu_X(T_t^{-1}(A)) = \mu_X(A)$ for all $A \in \Sigma_X$ and all t.[2] The *solution* (or *trajectory)* through a point x in X is the function $s_x : \mathbb{R} \to X$, $s_x(t) = T_t(x)$ (and *mutatis mutandis* for discrete time). Gathering the various elements together, we can now define a *deterministic model* as the quadruple $(X, \Sigma_X, \mu_X, T_t)$ (cf. Berger 2001). In the foundations of statistical mechanics the dynamical condition of ergodicity is important. Intuitively speaking, a deterministic model is ergodic if a system eventually visits all regions of state space (of nonzero measure). Formally, $(X, \Sigma_X, \mu_X, T_t)$ is ergodic iff (if and only if) for any $A \in \Sigma_X$ with $T_t^{-1}(A) = A$ for all t, it follows that $\mu_X(A) = 0$ or $\mu_X(A) = 1$.

To define a stochastic model, we first have to introduce the notion of a random variable. Consider a set $\bar{X}$, which consists of all possible outcomes of a probabilistic experiment. $\bar{X}$ is equipped with a σ-algebra $\Sigma_{\bar{X}}$ of subsets of $\bar{X}$. The tuple $(\bar{X}, \Sigma_{\bar{X}})$ is known as the *outcome space*. Intuitively, a *random variable* R gives the outcome of a probabilistic experiment, where the distribution $p\{R \in A\}$ specifies the probability that the outcome will be in A. Formally, a random variable is a measurable function $R : \Omega \to \bar{X}$, where $(\bar{X}, \Sigma_X)$ is the outcome space. $(\Omega, \Sigma_\Omega, \nu)$ is a probability space, i.e. Ω is a set that encodes the outcomes of the probabilistic experiment, Σ_Ω is a σ-algebra on Ω and ν is a probability measure that defines the probability distribution $p\{R \in A\} := \nu(R^{-1}(A))$ for all $A \in \Sigma_{\bar{X}}$.

A stochastic model consists of a string of the kind of probabilistic experiments that are described by a random variable. Formally, a *stochastic model* $\{R_t\}$ (in mathematical parlance a 'stochastic process'), $t \in \mathbb{R}$ for continuous time or $t \in \mathbb{Z}$ for discrete time, is a family of random variables, which

[1] We talk about 'models' rather than 'systems' because we reserve the term 'system' for the parts or aspects of the physical world that are represented by mathematical structures, and we refer to the structures themselves as 'models'. We note, however, that terminology varies. The mathematics and physics literature speaks of 'dynamical systems' rather than 'models'. A similar issue arises in connection with the term 'process', where the mathematical literature refers to models with a stochastic dynamics as 'stochastic processes'. We avoid 'processes' talk altogether because outside mathematics a process is usually taken to be a part of the world represented by a model rather than a mathematical object. For a discussion of how models represent systems see Frigg and Nguyen (2017).

[2] At this point there is no requirement that the measure be normalised.

are defined on the *same* probability space $(\Omega, \Sigma_\Omega, \nu)$ and take values in the *same* measurable space $(\bar{X}, \Sigma_{\bar{X}})$ such that $R(t, \omega) := R_t(\omega)$ is jointly measurable in (t, ω). The crucial difference between a stochastic model and a 'mere' random variable is that a random variable describes the outcome of one experiment and a stochastic model describes the outcome of a *sequence* of experiments. This is reflected in Ω. While in the case of a 'mere' random variable Ω only encodes the outcome of one experiment, in the case of a stochastic model Ω is defined so that it encodes the outcomes of the sequence of experiments and thus encodes *all possible histories of the entire process*. In the example of an infinite sequence of coin tosses, for instance, Ω could consist of all the bi-infinite sequences $\omega = (..., \omega_{-1}, \omega_0, \omega_1, ...)$ where ω_i encodes the outcome of the probabilistic experiment at time $t = i$. An example of such a sequence is $(..., 0, 1, 1, 0...)$ (where '1' encodes the outcome Heads and '0' encodes the outcome Tails). R_i then picks the element of a sequence at $t = i$ and maps onto an element of $\bar{X} = \{\text{Heads, Tails}\}$ (namely, '1' to Heads and '0' to Tails). Hence R_i gives the outcome of the coin toss at time $t = i$. A *realisation* is a possible path of the model. That is, it is a function $r_\omega : \mathbb{R} \to \bar{X}$, $r_\omega(t) = R(t, \omega)$, for $\omega \in \Omega$. The difference between ω and r_ω is simply that while r_ω gives a possible path of the model in terms of sequences of elements of $\bar{X}$, ω *encodes* such a possible history (cf. Doob 1953, 4-46).

The probability of outcome A at time t is given by $P\{R_t \in A\} := R_t^{-1}(A)$. But how is this probability determined by $(\Omega, \Sigma_\Omega, \nu)$? Ω contains bi-infinite sequences and ν, as a measure on Ω, assigns probabilities to such *sequences*, yet $R_t^{-1}(A)$ is the probability of a particular outcome A at a particular time t. The point to realise is that probabilities on sequences determine probabilities of outcomes at time t in a straightforward manner. Take again the example of the coin toss. To determine the probability of Tails at time $t = 5$ we group the sequences in Ω into two sets. The first set, B, contains all sequences that do *not* have 0 at $t = 5$; the second set, G, contains sequences that *do* have a 0 at position $t = 0$. The probability of Tails at $t = 5$ then is $\nu(G)$.

If the stochastic model does not depend explicitly on time (if, for instance, the outcome does not depend on when you toss a coin), then we have a stationary stochastic model, and in what follows all stochastic models we will be working with are assumed to be stationary. In formal terms, a stochastic model $\{R_t\}$ is *stationary* iff the distributions of the multi-dimensional random variable $(R_{t_1+h}, \ldots, R_{t_n+h})$ is the same as the one of $(R_{t_1}, \ldots, R_{t_n})$ for all $t_1, \ldots, t_n \in \mathbb{R}$ or $\mathbb{Z}$, $n \in \mathbb{N}$, and all $h \in \mathbb{Z}$ or $\mathbb{R}$.

In what follows *Markov models* play a crucial role (in the mathematical literature they are referred to as 'Markov chains'). Intuitively, a Markov model describes probabilistic laws where the probability distribution of the next state of the model only depends on the previous state of the model and no other past states. Formally, $\{R_t;\, t \in \mathbb{Z}\}$ is a Markov model iff the following three conditions are satisfied: (i) The model's outcome space consists of a finite number of states $\bar{X} := \{s_1, \ldots, s_N\}$, $N \in \mathbb{N}$, and $\Sigma_{\bar{X}} = \mathbb{P}(\bar{M})$; (ii) $P\{R_{t+1} = s_j \,|\, R_t, R_{t-1} \ldots, R_k\} = P\{R_{t+1} = s_j \,|\, R_t\}$ for any t, any $k \in \mathbb{Z}, k \leq t$, and any $s_j \in \bar{M}$; and (iii) $P\{R_{t+1} = s_j \,|\, R_t = s_i\}$ for any $s_i, s_j \in \bar{M}$ is independent of t, $t \in \mathbb{Z}$. Clearly, such a model is stationary. Define $P^k(s_i, s_j) := P\{R_{n+k} = s_i \,|\, R_n = s_j\}$ for $k \in \mathbb{Z}$. The Markov model is *irreducible* iff it cannot be split into two models because each state can be reached from all other states, i.e. for any $s_i, s_j \in \bar{M}$ there is a $k \in \mathbb{N}$ such that $P^k(s_i, s_j) > 0$. Since in an irreducible Markov model any state is accessible from any other state, irreducibility can be seen as the stochastic equivalent to ergodicity (Berger 2001).

It is illustrative to compare deterministic and stochastic models. There are some obvious correspondences: the state space X of a deterministic model corresponds to the outcome space $\bar{X}$ of a stochastic model; the sigma algebra Σ_X corresponds to $\Sigma_{\bar{X}}$; the measure μ_X corresponds to the

distribution $p\{R \in A\}$; and a solution s_x corresponds to a realisation r_ω. But what corresponds to the evolution function T_t? The not entirely obvious answer is: the measure ν. Ω contains all possible histories and ν assigns probabilities to them. By making some of these histories more likely than others, ν in effect provides dynamical information. Take again the example of coin tosses and consider two finite strings of outcomes $\omega'_1 = (0,1,0,1,0,1,0,1)$ and $\omega'_2 = (0,0,1,1,0,0,1,1)$. These are not themselves elements of Ω (which is a space of bi-infinite sequences), but they are the middle segments of great many sequences in Ω. Let $M(\omega'_1)$ be the set of all bi-infinite sequences with the middle segment ω'_1, and likewise for $M(\omega'_2)$. If the coin is fair and the dynamics is such that the outcome at time t is independent from the outcome at any earlier (or future) time, then $\nu(M(\omega'_1)) = \nu(M(\omega'_2))$. If however, the process is such that successive results anticorrelate (i.e. if the outcome at t is unlikely to be H if the outcome at $t-1$ was H, and likewise for T), then $\nu(M(\omega'_1)) > \nu(M(\omega'_2))$. In this way the measure ν enshrines the dynamics of the model.

2.2 Boltzmannian Statistical Mechanics

Presentations of the conception of equilibrium in Boltzmannian statistical mechanics (BSM) often begin with what is now known as the combinatorial argument, and then present the result of these combinatorial considerations as a definition of equilibrium. However, it is now recognised that combinatorial considerations do not provide a general definition of equilibrium (see Uffink, 2007, and Werndl and Frigg, 2015a, for discussions). We therefore work with the time-average conception of equilibrium, which has recently been proposed by Werndl and Frigg (2015a, 2015b, 2017b, forthcoming-a). This conception is free of the restriction faced by the combinatorial argument and provides a fully general definition of equilibrium.

Consider a physical system S like a gas, a magnet or a crystal. At the micro-level such a system is described either by a deterministic or a stochastic model of the kind introduced in Subsection 2.1. At the macro-level S is characterised by a set of *macro-variables* $\{v_1, \ldots, v_l\}$ ($l \in \mathbb{N}$). The choice of macro-variables depends on a number of factors, including the purpose and aim of an investigation and the availability of certain measurement procedures that are eventually used to perform observations on the system. These macro-variables are measurable functions $v_i : X \to \mathbb{V}_i$ in the deterministic case and $v_i : \bar{X} \to \mathbb{V}_i$ in the stochastic case, associating a value with each point in state space or outcome space. We use capital letters V_i to denote the values of v_i. These values can now be used to define macro-states. The standard case is when a macro-state is defined by a particular set of values $\{V_1, \ldots, V_l\}$: the model is in macro-state $M_{V_1,\ldots,V_l}$ iff $v_1 = V_1, \ldots, v_l = V_l$. This definition formalises the intuitive idea behind the notion of a macro-state: all models that are macroscopically indistinguishable are in the same macro-state. Central to BSM is that macro-states *supervene* on micro-states, implying that a system's micro-state uniquely determines its macro-state. This determination relation usually is many-to-one, and therefore every macro-state M is associated with a macro-region consisting of all micro-states for which the system is in M.

An important yet often neglected issue is on what space macro-regions are defined. The obvious option would be X (or $\bar{X}$), but often this is not what happens. In general, macro-regions are defined on a subset $Z \subset X$ (or $\bar{Z} \subset \bar{X}$). Intuitively speaking, Z (or $\bar{Z}$) is a subset whose states have the same equilibrium macro-state (share the same equilibrium properties). In the case of the dilute gas with N particles, for instance, X is the $6N$-dimensional space of all position and momenta while Z is the $6N-1$ dimensional hypersurface of constant energy E. This is because the equilibrium state is dependent on the energy E. We refer to X (or $\bar{X}$) as the *full state space* and to Z (or $\bar{Z}$) as the *effective state space* of the system. The *macro-region* Z_M (or $\bar{Z}_M$) corresponding to

macro-state M can then be defined as the set of all $x \in Z$ (or $\bar{Z}$) for which M supervenes on x. The macro-regions on Z (or $\bar{Z}$) form a partition of Z (or $\bar{Z}$), meaning that they do not overlap and jointly cover Z (or $\bar{Z}$). The correct choice of Z (or $\bar{Z}$) depends on the system, and has to be determined on a case-by-case basis (cf. Werndl and Frigg 2015b). Since a system can never leave the partition of macro-regions, Z must be mapped onto itself under the model's time evolution. The sigma algebra can then be restricted to Z and so that one ends up considering a measure on Z which is both invariant under the dynamics and normalised (i.e. $\mu_Z(Z) = 1$).[3] In this way one obtains what the dynamical systems literature refers to as a measure-preserving dynamical system $(Z, \Sigma_Z, \mu_Z, T_t)$ with a normalised measure μ_Z. We call $(Z, \Sigma_Z, \mu_Z, T_t)$ the *effective deterministic model* (as opposed to the *full deterministic model* $(X, \Sigma_X, \mu_Z, T_t)$). Similarly, for stochastic models $\bar{Z}$ is invariant under all $R(t)$, and this gives rise to a stochastic model $\{S_t\}$ consisting of a family of random variables from $(\hat{\Omega}, \Sigma_{\hat{\Omega}}, \hat{\nu})$ to $(\bar{Z}, \Sigma_{\bar{Z}})$ where $\hat{\Omega}$ is the subset of Ω that encodes histories with outcomes in $\bar{Z}$, $\Sigma_{\hat{\Omega}}$ is the sigma algebra Σ_Ω restricted to $\bar{\Omega}$, and $\hat{\nu}$ is the measure ν restricted to $\hat{\Omega}$, and the probability $\bar{P}\{S(t) \in A\} := S_t^{-1}(A)$. We call $\{S_t\}$ the *effective stochastic model*, which contrasts with the the *full stochastic model* $\{R_t\}$.

Sometimes the standard macro-states (defined through exact values of the macro-variables) are too fine. In such situations one can turn to macro-states that are defined by the macro-variables taking values in a certain range. One can then say, for instance, that the model is in macro-state $M_{[A_1,B_1],\ldots,[A_l,B_l]}$ iff $V_1 \in [A_1, B_1], \ldots, V_l \in [A_l, B_l]$ for suitably chosen intervals. This can be a useful move, for instance, if the v_i are continuous variables and one wants to say that the model is in a particular macro-state if the values of the v_i lie within a certain range (which could be defined, for instance, by the measurement precision of the available laboratory equipment).

To introduce the notion of an equilibrium macro-state, let us have a brief look at the notion of equilibrium in thermodynamics. Recall that a system is part of the physical world. A system is in *thermal equilibrium* 'when none of its thermodynamic properties are changing with time' (Reiss 1996, 3; cf. Fermi 2000, 4).[4] This might suggest defining the equilibrium macro-state of a model as the macro-state whose macro-region is such that every initial condition eventually moves into the region and then stays there indefinitely. Unfortunately, this is unattainable in SM: due to Poincaré recurrence and time-reversal invariance trajectories will not remain indefinitely in any macro-region, and there may always be a few initial conditions that lie on trajectories that avoid equilibrium altogether. To get around these problems while still saving the basic intuition of thermal equilibrium, it is natural to postulate that the equilibrium state is the state in which the system spends most of the time in the long run.

To make this notion precise, we need the concept of the long-run fraction of time $LF_A(x)$ (for the deterministic case) and $LF_A(\omega)$ (for the stochastic case) that a model spends in a subset A (of either

[3] The dynamics is given by the evolution equations restricted to Z, and we follow the dynamical systems literature in denoting it again by T_t.

[4] Being in thermal equilibrium is an intrinsic property of the system, which offers a notion of 'internal equilibrium' (Guggenheim 1967, 7). It contrasts with 'mutual equilibrium' (*ibid.*, 8), which is the relational property of *being in equilibrium with each other* that two systems eventually reach after being put into thermal contact with each other. Mutual equilibrium is often also referred to as 'thermal equilibrium'. It is the notion that figures in the zeroth law of thermodynamics, which effectively says that thermal equilibrium is an equivalence relation. In the current context, we use the phrase 'thermal equilibrium' to refer to the internal notion of equilibrium.

Z or $\bar{Z}$):[5]

$$\text{Deterministic model} \quad : \quad LF_A(x) = \lim_{t\to\infty} \frac{1}{t} \int_0^t 1_A(T_\tau(x))d\tau, \tag{1}$$

$$\text{Stochastic model} \quad : \quad LF_A(\omega) = \lim_{t\to\infty} \frac{1}{t} \int_0^t 1_A(S_\tau(\omega))d\tau, \tag{2}$$

where $1_A(x)$ is the characteristic function of A: $1_A(x) = 1$ for $x \in A$ and 0 otherwise. Note that long-run fractions depend on the initial condition in the deterministic case and on the history in the stochastic case.

The notion of 'most of the time' can be read in two different ways, leading to two different notions of equilibrium. The first introduces a lower bound of 1/2 for the fraction of time, and stipulates that whenever a model spends more than half of the time in a particular macro-state, then that state is the equilibrium state of the model. Formally, let α be a real number in the interval $(\frac{1}{2}, 1]$, and let ε be a very small positive real number. If there is a macro-state $M_{V_1^*,\ldots,V_l^*}$ satisfying the following condition, then it is an *α-ε-equilibrium state*:

> Deterministic model: There exists a set $Y \subseteq Z$ such that $\mu_Z(Y) \geq 1-\varepsilon$, and all initial states $x \in Y$ satisfy $LF_{Z_{M_{V_1^*,\ldots,V_l^*}}}(x) \geq \alpha$. x_t (the state of the deterministic model at time t) is then said to be in equilibrium iff $x_t \in Z_{M_{V_1^*,\ldots,V_l^*}}$.
>
> Stochastic model: There exists a set $\Omega^* \subseteq \hat{\Omega}$ such that $\hat{\nu}(\Omega^*) \geq 1-\varepsilon$, and all $\omega \in \Omega^*$ satisfy $LF_{\bar{Z}_{M_{V_1^*,\ldots,V_k^*}}}(\omega) \geq \alpha$. $S(t)$ (the state of the stochastic model at time t) is then said to be in equilibrium iff $S(t) \in \bar{Z}_{M_{V_1^*,\ldots,V_l^*}}$.

The second reading takes 'most of the time' to refer to the fact that the model spends more time in the equilibrium state than in any other state (allowing that this can be less than 50% of its time). Formally, let γ be a real number in $(0, 1]$ and let ε be a small positive real number. If there is a macro-state $M_{V_1^*,\ldots,V_l^*}$ satisfying the following condition, then it is a *γ-ε-equilibrium state*:

> Deterministic model: There exists a set $Y \subseteq Z$ such that $\mu_Z(Y) \geq 1-\varepsilon$ and for all initial conditions $x \in Y$: $LF_{Z_{M_{V_1^*,\ldots,V_l^*}}}(x) \geq LF_{Z_M}(x) + \gamma$ for all macro-states $M \neq M_{V_1^*,\ldots,V_l^*}$. Again, x_t is said to be in equilibrium iff $x_t \in Z_{M_{V_1^*,\ldots,V_l^*}}$.
>
> Stochastic model: There exists a set $\Omega^* \subseteq \hat{\Omega}$ such that $\hat{\nu}(\Omega^*) \geq 1-\varepsilon$, and all $\omega \in \Omega^*$ satisfy $LF_{\bar{Z}_{M_{V_1^*,\ldots,V_k^*}}}(\omega) \geq LF_{\bar{Z}_M}(\omega) + \gamma$ for all $M \neq M_{V_1^*,\ldots,V_k^*}$. Again, $S(t)$ is said to be in equilibrium iff $S(t) \in \bar{Z}_{M_{V_1^*,\ldots,V_l^*}}$.

These two notions are not equivalent. In fact, an α-ε-equilibrium is strictly stronger than γ-ε-equilibrium in the sense that the existence of the former implies the existence of the latter but not *vice versa*. Recall that macro-states, and hence also equilibrium states, are defined in terms of the values taken by macro-variables. For brevity we talk of 'equilibrium values', which should be understood as meaning 'the value of the relevant physical quantities when the system is in the equilibrium state'. Note that the equilibrium state and equilibrium value depends on the measure defined on the effective state space (in particular, different measures might lead to different equilibrium states). One can speak of an *observed* equilibrium state at the macro-level if solutions end

[5] We state the definitions for continuous time. The corresponding definitions for discrete time are obtained simply by replacing the integrals by sums.

up spending most of their time in the same macro-state when the system is repeatedly prepared in various initial states. This happens if repeated state preparation leaves the system in a micro-state that is in Y, which one expects to happen.[6]

These definitions are about the *time* a model spends in the equilibrium state, and as such they remain silent about the *size* of the equilibrium macro-regions. There is nothing in the above definitions that would, in principle, preclude equilibrium macro-regions from being small. We call a macro-region β-*dominant* if its measure is greater or equal to β for a particular $\beta \in (\frac{1}{2}, 1]$. We say that a macro region is δ-*prevalent* if its measure is larger than the measure of any other macro-region by a margin of at least $\delta > 0$. One can then prove the following theorems (Werndl and Frigg 2015b, 2017b).

> *Dominance Theorem*: If $M_{\alpha\text{-}\varepsilon\text{-}eq}$ is an α-ε-equilibrium, then the following holds for $\beta = \alpha(1-\varepsilon)$: $\mu_Z(Z_{M_{\alpha\text{-}\varepsilon\text{-}eq}}) \geq \beta$ (for deterministic models) and $p\{\bar{Z}_{M_{\alpha\text{-}\varepsilon\text{-}eq}}\} \geq \beta$ (for stochastic models).[7]

> *Prevalence Theorem*: If $M_{\gamma\text{-}\varepsilon\text{-}eq}$ is a γ-ε-equilibrium, then the following holds for $\delta = \gamma - \varepsilon$: $\mu_Z(Z_{M_{\gamma\text{-}\varepsilon\text{-}eq}}) \geq \mu_Z(Z_M) + \delta$ (for deterministic models) and $p\{\bar{Z}_{M_{\varepsilon\text{-}eq}}\} \geq p\{\bar{Z}_M\} + \delta$ (for stochastic models) for all macro-states M such that $M \neq M_{\gamma\text{-}\varepsilon\text{-}eq}$.[8]

These theorems establish that equilibrium macro-regions are large in one of the two senses specified.

The Prevalence theorem shows that the equilibrium macro-region of γ-ε-equilibrium can take less than 50% of the effective state space (the equilibrium macro-region just needs to be larger than any other macro-region). Some may regard this as *reductio* of the notion of a γ-ε-equilibrium, because, as matter of principle, an equilibrium state must take at least 50% (or even much more) of the state space. We want to remain agnostic about this issue here, but note the following. First, γ-ε-equilibria play a role in the practice of physics: for instance, the equilibrium of the widely discussed KAC-ring model (Thompson 1972, 23; Lavis 2005) and the ideal gas with the macro-state structure given by the combinatorial argument (Werndl and Frigg 2015b, 27-28; Swendsen 2012, 11) are γ-ε-equilibria (and no α-ε-equilibria). Hence dismissing γ-ε-equilibria on conceptual grounds requires a certain degree of revisionism. Second, the claims in this paper are conditional in that they explicate what happens *if* one accepts γ-ε-equilibria. Those who do not recognise γ-ε-equilibria as *bona fide* equilibria will say that systems that have no α-ε-equilibrium in fact have no Boltzmannian equilibrium at all. We discuss what happens in such cases in Subsection 5.4.

As we will see in Section 5 (when we discuss the polarisation of the six-vertex model and the magnetisation of the Ising model), there are models in which the relevant macro-variables fluctuate between two values, assuming one or the other value most of the time. The values of the magnetisation m of a spin system, for instance, can jump back and forth between two extremal values C and $-C$. How should one think about such a situation? A radical suggestion would be to say that such a model has no equilibrium and leave it at that. Another possibility would be to further broaden the definition of a macro-state and say that macro-states can not only be defined by intervals but in fact by any set η of values. One could then take $\eta = \{C, -C\}$ and define M_η, the macro-state in which the magnetisation is either C or $-C$. Both proposals have their merits, but in the physics

[6] One way to motivate this assumption is to study the properties of the functions describing the preparation of initial states. E.g. a common assumption is that such preparation functions are absolutely continuous w.r.t. Z, which, if Y is of measure 1, immediately gives the desired result that when preparing the system one usually ends up in Y (for further details, see Werndl 2013).

[7] We assume that ε is small enough so that $\alpha(1-\varepsilon) > \frac{1}{2}$.

[8] We assume that $\varepsilon < \gamma$.

literature this situation is usually dealt with in another way, namely by saying that the model has *two equilibria*, one associated with $m = C$ and one with $m = -C$.[9] We adopt this point of view and integrate it into our framework.[10]

However, having two equilibria is incompatible with the above definitions of equilibrium, and to accommodate models with two equilibria these definitions need to be generalised. A natural way to go is to require that nearly all initial states spend most of their time in either one or the other of the two macro-states. If this is the case we speak of a *dual equilibrium* (and we call an equilibrium of the kind we discussed so far a *single equilibrium* if there is any ambiguity about which type of equilibrium it is referred to). As in the case of the single equilibrium, there are two versions. With α and ε as above, we say that a model has a *dual α-ε-equilibrium* iff there are two macro-states $M_{V_1^*,...,V_l^*}$ and $M_{V_1^+,...,V_l^+}$ such that:

> Deterministic model dual equilibrium condition: There is no γ-ε-equilibrium, and there exists a set $Y \subseteq Z$ such that $\mu_Z(Y) \geq 1 - \varepsilon$ and all initial states $x \in Y$ satisfy $LF_{Z_{M_{V_1^*,...,V_l^*}} \cup Z_{M_{V_1^+,...,V_l^+}}}(x) \geq \alpha$.[11] Again, $x(t)$ is then said to be in one of the equilibrium states iff $x(t) \in Z_{M_{V_1^*,...,V_l^*}} \cup Z_{M_{V_1^+,...,V_l^+}}$.
>
> Stochastic model dual equilibrium condition: There is no γ-ε-equilibrium, and there exists a set $\Omega^* \subseteq \hat{\Omega}$ such that $\hat{\nu}(\Omega^*) \geq 1-\varepsilon$, and all $\omega \in \Omega^*$ satisfy $LF_{\bar{Z}_{M_{V_1^*,...,V_l^*}} \cup \bar{Z}_{M_{V_1^+,...,V_l^+}}}(x) \geq \alpha$.[12] Again, $S(t)$ is then said to be in one of the equilibrium states iff $S(t) \in \bar{Z}_{M_{V_1^*,...,V_l^*}} \cup \bar{Z}_{M_{V_1^+,...,V_l^+}}$.

The requirement that there is no γ-ε-equilibrium is added to rule out that models can have single and dual equilibria simultaneously.[13] So either a system has just a single equilibrium, or it has just a dual equilibrium. The γ-ε-equilibrium generalises in the same way. With γ and ε as above, we say that a model has a *dual γ-ε-equilibrium* iff there are two macro-states $M_{V_1^*,...,V_l^*}$ and $M_{V_1^+,...,V_l^+}$ such that:

> Deterministic model: There exists no γ-ε-equilibrium, and there exists a set $Y \subseteq Z$ such that $\mu_Z(Y) \geq 1-\varepsilon$ and for all initial conditions $x \in Y$ we have $LF_{Z_{M_{V_1^*,...,V_l^*}} \cup Z_{M_{V_1^+,...,V_l^+}}}(x) \geq LF_{Z_M \cup Z_{M'}}(x) + \gamma$ for all macro-states $M, M' \neq M_{V_1^*,...,V_l^*}$ and $M, M' \neq M_{V_1^+,...,V_l^+}$. Again, $x(t)$ is said to be in one of the equilibrium states iff $x(t) \in Z_{M_{V_1^*,...,V_l^*}} \cup Z_{M_{V_1^+,...,V_l^+}}$.

[9] See, for instance, Baxter (1982, Ch. 1), Cassandro et al. (1973, 153), Gonsalves (2007), Lavis and Bell (1999, 307), Levis (2012, Section IV.3), Sekular (Unpublished, 2), van Enter and van Hemmen (1984, 258), and Venaille and Bouchet (2009, 1).

[10] In this paper we will only deal with cases where there are two equilibria. In principle the framework can be extended to conceptualise what it means that a system has three or more equilibria.

[11] It is furthermore required that all macro-states that satisfy the dual equilibrium condition can be labelled as $M1$ and $M2$ such that $M1 = M_{V_1^*,...,V_l^*}$ and $M2 = M_{V_1^+,...,V_l^+}$. This is to make sure that the dual equilibria are unique; we thank an anonymous reviewer for making us aware of the necessity of this condition.

[12] Again, to make sure that the dual equilibria are unique, it is furthermore required that for all macro-states that satisfy the dual equilibrium condition can be labelled as $M1$ and $M2$ that satisfy $M1 = M_{V_1^*,...,V_l^*}$ and $M2 = M_{V_1^+,...,V_l^+}$.

[13] If a model has no γ-ε-equilibrium, it does not have a α-ε-equilibrium either, so the clause in effect rules out both kinds of single equilibria. This requirement also implies that the system spends an equal amount of time in $Z_{M_{V_1^*,...,V_l^*}}$ and $Z_{M_{V_1^+,...,V_l^+}}$. If it did not spend an equal amount of time in $Z_{M_{V_1^*,...,V_l^*}}$ and $Z_{M_{V_1^+,...,V_l^+}}$, then a γ-ε-equilibrium would exist.

Stochastic model: There is no γ-ε-equilibrium, and there exists a set $\Omega^* \subseteq \hat{\Omega}$ such that $\hat{\nu}(\Omega^*) \geq 1-\varepsilon$, and all $\omega \in \Omega^*$ satisfy $LF_{\bar{Z}_{M_{V_1^*,\ldots,V_l^*}} \cup \bar{Z}_{M_{V_1^+,\ldots,V_l^+}}}(x) \geq LF_{\bar{Z}_M \cup \bar{Z}_{M'}}(x) + \gamma$ for all macro-states $M, M' \neq M_{V_1^*,\ldots,V_l^*}$ and $M, M' \neq M_{V_1^+,\ldots,V_l^+}$. Again, $S(t)$ is then said to be in one of the equilibrium states iff $S(t) \in \bar{Z}_{M_{V_1^*,\ldots,V_l^*}} \cup \bar{Z}_{M_{V_1^+,\ldots,V_l^+}}$.

Analogues of the Dominance and Prevalence theorems also hold for dual equilibria. More specifically, the two macro-regions corresponding to a dual α-ε-equilibrium take up at least $\alpha(1-\varepsilon)$ of state space, and the two macro-regions corresponding to an dual γ-ε-equilibrium together are at least $\gamma - \varepsilon$ larger than the union of the macro-regions of any two macro-states other than $M_{V_1^*,\ldots,V_l^*}$ and $M_{V_1^+,\ldots,V_l^+}$.

2.3 Gibbsian Statistical Mechanics

In Gibbsian statistical mechanics (GSM) the object of study is an *ensemble*, an infinite collection of independent models that are all governed by the same laws of motion but are in different states.[14] The ensemble is described by a probability density $\rho(x,t)$, $x \in Z$, over the effective state space if the model is deterministic or a probability density $\rho(\bar{x},t)$, $\bar{x} \in \bar{Z}$ over the outcome space if the model is stochastic (recall that, intuitively speaking, Z (or $\bar{Z}$) is the subset of X (or $\bar{X}$) whose states share the same equilibrium properties.) For brevity we only state definitions and results for the deterministic case; their stochastic cousins can be obtained simply by replacing x by $\bar{x}$ and Z by $\bar{Z}$.

The probability density reflects the probability of finding the state of a model chosen at random from the ensemble in a region $R \subseteq Z$ at time t:

$$p_t(R) = \int_R \rho(x,t)dx. \tag{3}$$

The probability must be conserved over time, meaning that for every $R(t) \subseteq Z$ that is moving forward under the time evolution of the model the probability must be constant:

$$\frac{d}{dt}\int_{R(t)} \rho(x,t)dx = 0 \tag{4}$$

It is a necessary and sufficient condition for this to hold that ρ satisfies the Liouville equation (Tolman 1938). For this reason equation 4 holds in all Hamiltonian systems.

In his discussion of ensembles Gibbs introduces what he calls the *condition of statistical equilibrium* (1902, 8). An ensemble is in statistical equilibrium iff ρ is stationary. A distribution is stationary iff it does not change over time, meaning that it is invariant under the dynamics: $\rho(x,t) = \rho(x)$ for all t.

There will usually be a large number of stationary distributions for a certain dynamics,[15] and so the question arises which of these distributions should be chosen to characterise a given situation. Gibbs

[14] We follow Gibbs' (1902) original presentation of GSM. Alternative presentations endeavour to avoid reference to ensembles and regard GSM as probabilistic algorithm. For a discussion of different interpretations of GSM see Frigg and Werndl (forthcoming-a).

[15] If a system is Hamiltonian, then every distribution of the form $\rho(x) = \rho(H(x))$ is stationary, where $H(x)$ is the system's Hamiltonian. An interesting result in this context is also the theorem that if the system (X, Σ_X, ρ, T_t) is ergodic and a stationary distribution ξ is absolutely continuous w.r.t. ρ, then it follows that $\rho = \xi$.

discusses this issue at length and proposes the so-called microcanonical distribution if the system is completely isolated from its environment (and thus the number of particles and the energy are both constant). The microcanonical distribution is defined on the hypersurface of constant energy Z given by $H(x) = E$ relative to a fixed energy value of E (where H is the Hamiltonian of the model):

$$\rho_m(x) = \frac{1}{\omega(E)}\delta(H(x) - E), \tag{5}$$

where $\omega(E)$ is the area of the surface of constant energy E (cf. Uffink 2007, equations (41) and (42)). If the system is in contact with a heat bath of a certain temperature and the number of particles is constant but the energy varies, Gibbs proposes as the equilibrium distribution the so-called *canonical distribution*[16] (here Z equals the full phase space X):

$$\rho_c(x) = \frac{e^{-H(x)/kT}}{\zeta_T}, \tag{6}$$

where, again, H is the Hamiltonian, T is the temperature, k is the Boltzmann constant, and ζ_T is the so-called *partition function*

$$\zeta_T = \int_X e^{-H(x)/kT} dx. \tag{7}$$

The microcanonical and canonical distributions will be needed for the examples that will be discussed later in the paper. The justification of these distributions as the correct distributions for certain situation can proceed along different lines, and a number of suggestions have been made (see Myrvold, 2016, and Frigg and Werndl, forthcoming-b, for reviews). In connection with the canonical distribution for thermal systems, Szilard's (1925) argument deserves note. The core idea of his argument is that the canonical distribution can be derived from only two requirements, namely that (i) if two systems in isolation are at the same temperature, then there can be no mean flow of energy between them ('no spontaneous flow of energy') and that (ii) when two systems at the same temperature are combined, the joint system is at the same temperature ('composition') (cf. Maroney 2008). We do not discuss the justification of the distributions in more detail here because nothing in what follows depends on how the choice of a particular distribution is justified. The crucial point to bear in mind is that statistical equilibrium pertains to an ensemble and hence provides a notion of *ensemble-equilibrium* which is the standard view of Gibbsian equilibrium (see, for instance, Hill 1987, 8, Myrvold 2016, 588-589, and Tolman 1938, 63).

As we have seen previously, a system is in thermal equilibrium if none of its thermodynamic properties change over time, and we have seen that BSM captures this notion by defining the equilibrium macro-state of the model as the macro-state in which the model spends most of its time. How does GSM's notion of statistical equilibrium relate to thermal equilibrium? A common way to relate the two is to appeal to phase averages. Assume that the relevant physical variable is associated with a real-valued function $f : Z \to \mathbb{R}$.[17]. Examples of such functions are the total magnetisation or the total polarisation of the system (which we will discuss in Section 5). The *phase average* of f is

$$\langle f \rangle = \int_Z f(x)\rho(x,t)dx. \tag{8}$$

[16] One might worry that the canonical distribution is dependent on the temperature T. We thank a referee to point out a simple resolution to this. Take the canonical distribution to be specified not by temperature but by expected energy, with T understood purely formally as the Lagrange multiplier obtained in extremising Gibbs entropy subject to a fixed expected-energy. An analysis of thermodynamics in the Gibbsian framework would have to tell us why this Lagrange multiplier is to be identified with thermodynamic temperature. Yet this is something that does not need to be answered before one use the canonical ensemble.

[17] The variables we consider in this paper are not explicitly time-dependent. One can consider time-dependent variables simply by replacing the above definition by $f : X \times t \to \mathbb{R}$.

If the model is in statistical equilibrium, then $\langle f \rangle$ does not depend on time because ρ is stationary. The practice of calculating phase averages is called *Gibbsian phase averaging*. It pays noting that the results of Gibbsian phase averaging crucially depend on the measure chosen because different measures can give different values for $\langle f \rangle$.

Assume that a system in thermal equilibrium is represented by a Gibbsian model in statistical equilibrium. The *Averaging Principle* (AP) is then the proposition that, under certain conditions, when measuring the quantity that is associated with f on the system, then the measured equilibrium value of f is equal to the phase average $\langle f \rangle$ of the Gibbsian model. In other words: if conditions C are true, then the outcome of a measurement of a property associated with function f on a system in equilibrium is $\langle f \rangle$.

We will discuss this principle and the associated conditions in more detail in the next section. Before getting into this discussion it is worth mentioning that there is a question about what counts as 'measured' or 'observed' value. There are two options. In BSM one considers *instantaneous measurements*, which see a measurement as happening at a particular instant of time. Penrose describes a measurement of that kind as 'an instantaneous act, like taking a snapshot' (2005, 17-18). This option is also available in GSM. If this notion is adopted, then AP is interpreted as making a statement about a *single* and instantaneous measurements. AP then should be read as saying that the outcome of a single measurement of property f on a system in system-equilibrium is $\langle f \rangle$. The second option is to see observation and measurement as process that is carried out over an extended period of time, and the outcome of a measurement as some sort of aggregate over instantaneous values. In this vein Chandler considers the option that an observed value of quantity 'is actually the average over very many independent observations' (1987, 58). Much can be said about these options, but the question of the nature of observation need not occupy us here because, as we will see in the next section, the problem we discuss in this paper does *not* depend on which notion of measurement is adopted.

Finally, we note that Gibbsian statistical mechanics can be interpreted in various ways, and in this paper we need not commit ourselves to any particular interpretation because the points we make are independent of particular interpretations (for a review of various interpretations, see Frigg and Werndl 2019). However, because of its prominence, we would like to briefly discuss the fluctuation interpretation of GSM. The core idea here is to use the probabilities of GSM as given in equation (3) to calculate the probability that a fluctuation of a certain magnitude away from the phase average occurs. In more detail: the fluctuation for a micro-state $x \in Z$ is the difference between between the value $f(x)$ (the true value if the model is in state x) and the Gibbsian phase average $\langle f \rangle$::

$$\Delta(t) = f(x(t)) - \langle f \rangle. \tag{9}$$

and likewise for a micro-state $\bar{x} \in \bar{Z}$ if the system is stochastic.

Given an interval $\delta := [\delta_1, \delta_2]$, where δ_1 and δ_2 are real numbers such that $0 \leq \delta_1 \leq \delta_2$, equation (3) can then be used to arrive at the probability for a fluctuation of a magnitude between δ_1 and δ_2 to occur:

$$p(\delta) = \int_D \rho(x) dx, \tag{10}$$

where $D = \{x \in X \mid \delta_1 \leq |\Delta(t)| \leq \delta_2\}$. This equation gives the probability that the system exhibits fluctuations of a certain magnitude at a certain time. We will discuss such fluctuations and their probabilities later in Section 5.

It is important to interpret the scope of this equation correctly. Sometimes the probabilities in equation (3) are interpreted as holding universally. That is, ρ is seen as providing the correct probabilities for the state of a system to be in region R at time t for *all* R in Z and for *any* time t. Under such an interpretation the fluctuation probabilities in equation (10) are then seen as universal in the sense that for any magnitude and for any time t, $p(\delta)$ gives the probability for a fluctuation of a certain magnitude to occur at t.

Yet universality of this kind is a very strong demand and fails in general. A careful study of GSM reveals that at least one of two conditions have to be met in order for this universality to hold (for more on those two conditions, see Frigg and Werndl 2019). First, suppose that the probability of fluctuations describe observations on a single system when one traces the system's state over time. The *masking condition* requires either that the model has access to all parts of state space, or, if that is not the case, that f must be such that the proportion of states for which f assumes a particular value is the same in each invariant subset of Z. If the masking condition is satisfied, then the probability of fluctuations describing observations on a single system are correctly described by equation (10).Second, suppose that the ensemble is like an urn of balls in the sense that the distribution ρ specifies the probability of finding that the state x of the system lies in a certain part of the state space in much the same way in which the fraction of red balls in the urn specifies the probability of drawing a red ball. The condition of *f-independence* then (roughly) states that the dynamics of the model must be such that the probability of finding a specific value of f in two consecutive yet sufficiently temporally distant measurements have to be (approximately) independent of each other. If this is the case, then the probability of fluctuations, when probabilities are interpreted like draws from an urn, are correctly described by equation (10). The Gibbsian ρ can be used to calculate correct fluctuation probabilities only if at least one of these conditions is satisfied. These conditions limit the scope of the fluctuation interpretation of GSM because both conditions are strong requirements on the dynamics and the macro-variables and their satisfaction cannot be taken for granted.

3 The Mechanical Averaging Principle

This paper is an investigation of the practice of averaging, but rather than investigating AP, we investigate a related principle that we call the *Mechanical Averaging Principle* (MAP). In this section we derive MAP and explain how MAP relates to AP.

Consider a physical quantity that is associated with a real-valued function f. Let F_T be the value that the quantity assumes when the system is in thermal equilibrium (see Section 2.2). In our context this an *empirical value*: F_T is the outcome of a measurement of f on a system in thermal equilibrium. Then let F_B be the Boltzmannian equilibrium value for the same quantity in the same system.[18] The fundamental claim of BSM then is that $F_B = F_T$. Let us call this the *Boltzmannian Equilibrium Principle* (BEP). We can then formulate what we call the *Equilibrium Argument*:

BEP: $F_B = F_T$

AP: $\langle f \rangle = F_T$ under condition C

[18] In terms of the formalism introduced in Section 2.2, this means that f is one of the macro-variables v_i and F_B is the value that v_i assumes when the system is in state $M_{V_1^*,\ldots,V_l^*}$, i.e. V_i^*.

Conclusion: $F_B = \langle f \rangle$ under condition C

The conclusion of the Equilibrium Argument is what we call the *Strict Mechanical Averaging Principle*. The principle is 'strict' because it postulates a strict identity between the two values. The qualification 'mechanical' highlights that the principle equates two mechanical quantities, namely the Boltzmannian equilibrium value and the Gibbsian phase average of f. This contrasts with AP, which concerns the connection between a mechanical and an observational quantity.

The requirement of a strict identity between the two may be too stringent a requirement in practice and so it is advisable to investigate a more permissive principle. Such a principle can be reached by introducing an error term. Doing so leads us to the Mechanical Averaging Principle (MAP):

$$F_B = \langle f \rangle \pm \chi_f \text{ under conditions } C, \tag{11}$$

where χ_f is very small relative to $\langle f \rangle$ and F_B. We call the equation in the principle, $F_B = \langle f \rangle \pm \chi_f$, the Mechanical Averaging Equation (MAE).

It is crucial to stress that χ_f depends on f. More specifically, given a macro-variable f it can then be decided what a very small observational difference relative to values of f amounts to. We do not believe that there are absolute criteria for what counts as very small. In addition to the dependence on f, how large χ_f is allowed to be will depend on the problem at hand and on the context of the discussion. So there is some flexibility concerning what 'very small' means. No harm is done to our discussion by the absence of absolute standards because our arguments in what follows neither depend on a particular choice of χ_f, nor on a particular understanding of 'very small'. [19] Note also that the conditions we formulate in Section 4 are either based on exact results that hold for $\chi_f = 0$ (AET and COT), or on results that resolve this issue by offering explicit bounds for χ_f (Khinchin condition). In the examples in Section 5, the differences between F_B and $\langle f \rangle$ are so large that for any halfway reasonable understanding of 'very small' some of the examples would have to count as counterexamples to the MAP. In Section 2.2 we noted that in some cases macro-states are defined by macro-variables taking a value in a certain range or interval $[a, b]$. In such cases the question whether the Gibbsian phase average is in agreement with the Boltzmannian equilibrium value amounts to the question of whether the Gibbsian phase average is contained in the interval $[a - \chi_f, b + \chi_f]$.[20]

The question that this paper is concerned with now comes clearly into focus: what are the conditions C for which MAE holds true? And let us be absolutely clear on what has to be the case for MAE to

[19] A referee pointed out that one might want to say that the difference between F_B and $\langle f \rangle$ can be regared as small just in case $\frac{|F_B - \langle f \rangle|}{|\langle f \rangle|}$ is negligible. This is a possible view, and it is compatible with our account (because χ_f can be estimated from the requirement that $\frac{|F_B - \langle f \rangle|}{|\langle f \rangle|}$ has to be negligible). There remains a question, however, why one would want to accept this as the only criterion. In fact, there are good reasons not to do so because whenever F_B is zero (which is the case for some of the examples discussed in this paper), then $\frac{|F_B - \langle f \rangle|}{|\langle f \rangle|}$ is 1, and hence not informative.

[20] In stating MAP, and indeed the Equilibrium Argument, we took f to play the role both of a Boltzmannian macro-variable v_i and of a Gibbsian functions f_i. This is legitimate because in many cases the Boltzmannian and the Gibbsian functions are indeed identical (as we will see in the examples in Section 5). If they are not identical, they are very similar in the following way. To end up with a finite set of macro-states, BSM needs a coarse graining. One way to get a coarse graining is, as we have noted in Section 2.2, to define macro-states through the requirement that the values of f lie in a certain interval. An alternative option is to coarse-grain the variable itself, meaning that it is changed to assume constant values on macro-regions of finite measure. If an originally continuous variable is coarse grained in this way, it is still similar to the original variable, and it can still be used for a comparison between BSM and GSM. One can of course also use the coarse-grained macro-variables in GSM, which facilitates a direct comparison.

hold. Implicit in the plea to find conditions C that make MAE true is that the conditions have to be such that $F_B = \langle f \rangle \pm \chi_f$ holds under C in the *same* model, where 'model' is used in the technical sense defined in Section 2.1. We are not interested in situations in which F_B in one model is equal to $\langle f \rangle$ in *another* model. What we have to find are conditions that guarantee that $F_B = \langle f \rangle \pm \chi_f$ whenever F_B and $\langle f \rangle$ are calculated in one and the same model. A fortiori this means that the we have to compare values that are obtained for models that have the same measures (i.e. $\rho = \mu_Z$ in the deterministic case and $\rho = \bar{P}\{S \in A\}$ in the stochastic case), because altering the measure amounts to altering the model.[21]

It is important to bear in mind that enquiring into the truth of MAP is not tantamount to enquiring into the truth of AP. The logical relation between the two principles is indirect. If MAP is false for a certain C, this implies that the Strict Mechanical Averaging Principle is also false for that C. This, in turn, implies that AP is false for C or BEP is false (or both). In Section 5 we will encounter examples for which the conclusion of the Equilibrium Argument is false, which raises the interesting question whether one sees the reason for this in the failure of AP or in the failure of BEP (or, indeed, in the failure of both). We briefly discuss these alternatives in Section 6. Conversely one cannot infer from the truth of MAP for a certain C to the truth of AP for the same C. In fact, such an inference would be a fallacy of the affirmation of the consequent. It could, in principle, be the case that BEP and AP are both false, but in a way that 'cancels out' and makes MAP come out true even though both premises are false. In fact, MAP only provides 'negative' information about AP: if MAP is false for conditions C *and* BEP is true, then it follows that AP is also false for conditions C.

The project for this paper is to investigate MAP by discussing various candidates for condition C. That is, we are looking for Cs that make it that case that $F_B = \langle f \rangle \pm \chi_f$ whenever F_B and $\langle f \rangle$ are calculated on the same model. Our strategy for this is to start with Cs that have been proposed for AP and ask whether they work for MAP. This is a good *heuristic* because intuitively one expects MAP to hold for conditions that make AP true. But, for the reasons discussed in the previous paragraph, this is not more than a heuristic: neither does the truth of AP guarantee the truth of MAP, not can we infer back from truth of MAP to the truth of AP.[22]

The structure of the two theories makes it clear that C will, to a large extent, be concerned with the kinds of variables that one should consider. This choice is a subtle matter because there are important differences between the Boltzmannian and the Gibbsian conceptions of equilibrium. The existence of a Boltzmannian equilibrium (the largest macro-region relative to the values of certain macro-variables) depends on the choice of macro-variables: there may be an equilibrium for one variable but not for another variable. There is no such dependence in the Gibbsian conception of equilibrium (a stationary distribution), even though the phase averages of course also depend on the variables chosen. So agreement and disagreement between these two notions will depend on what kind of macro-variables one focuses on, and much of the effort to articulate conditions will be concerned with identifying the 'right' kind of variables.

We emphasise that an enquiry into MAP involves the comparison of *phase averages* with Boltz-

[21] On the face of it, $\bar{P}\{S \in A\}$ and $\rho(\bar{x})$ seem to be different mathematical objects and so one may wonder how they can be identical. But the difference is merely in the notation. Since $\bar{P}\{S \in A\}$ is the probability distribution over $\bar{Z}$ one could just as well write $p(\bar{x})$, which makes the connection to $\rho(\bar{x})$ obvious.

[22] Incidentally, this is also why it does not matter for our investigation which notion of measurement one adopts in GSM (as we noted at the end of the previous section). We can consider candidate conditions C for both kinds of measurement and ask whether MAP holds true under C without committing to C making AP true. If a particular C would make AP true under, say, the time average notion of measurement but not under an instantaneous value interpretation of measurement, we can remain agnostic about which notion of measurement is the correct one.

mannian equilibrium; it does not involve a comparison of 'Gibbsian equilibrium values' with Boltzmannian equilibrium values. This is important because there is no agreement over what 'Gibbsian equilibrium values' would be. In BSM the system can (and occasionally will) leave the equilibrium macro-region, but still spend most of the time in the equilibrium region. Hence BSM characterises equilibrium through a specific set of values (or intervals of values) and conceptualises fluctuations as taking the system *away* from equilibrium. The situation is subtly but importantly different in GSM. In GSM equilibrium is characterised by a stationary distribution. Under the fluctuation interpretation of GSM introduced in Section 2.3, this distribution provides the probabilities for fluctuations away from $\langle f \rangle$, but such fluctuations appear *within* equilibrium. So under that interpretation, fluctuations are *inherent* to equilibrium, and do not count as departures from equilibrium. The quantity f can then assume a whole array of values *in Gibbsian equilibrium*, which makes it meaningless to speak of a 'Gibbsian equilibrium value' (and hence there is simply nothing one could meaningfully compare to the Boltzmannian equilibrium value). We avoid this difficulty here by not taking a stand either on the interpretation of GSM or on the issue of what a Gibbsian equilibrium value might be. Instead we focus on the phase average $\langle f \rangle$, which is a mathematically well-defined object no matter what one's interpretation of GSM.

Finally, we note that identifying such conditions is only a first step toward a better understanding of the relation between BSM and GSM, and that these conditions by no means offer a full account of the relation between BSM and GSM. Finding such an account is of course a much broader problem and would also include the discussion of aspects other than equilibrium values, most notably how BSM and GSM understand the approach to equilibrium and how the two understandings relate. But such a discussion is beyond the scope of this paper.

4 Demarcating the Validity of the Mechanical Averaging Equation

We now turn to the core question of this paper: for what conditions C is MAP true? In Subsection 4.1 we briefly comment on two common but unsuccessful attempts. In Subsection 4.2 we discuss conditions that impose restrictions on the fluctuation of a variable. In Subsection 4.3 we discuss the *Average Equivalence Theorem*, and in Subsection 4.4 we discuss and prove a new theorem, which we call the *Cancelling Out Theorem*. All theorems offer sufficient (but not necessary) conditions for the agreement of Boltzmannian equilibrium values and Gibbsian phase averages. In Subsection 4.5 we discuss the relations between these conditions and point out that they are independent of one another.

4.1 Unsuccessful Attempts

SM is habitually introduced as theory of large systems. Commenting on the systems that fall within the scope of SM Baxter, for instance, says that '[s]uch systems are made up of a huge number of individual components (usually molecules)' (1982, 1). This might be understood as suggesting that AP holds for C that says that the system is large.[23] Following the heuristic outlined in Section 3, let us now ask whether MAP is true with $C = \{\text{the system is large}\}$. Unfortunately this is obviously wrong because consisting of large number of molecules is neither necessary nor sufficient for it to be the case that F_B and $\langle f \rangle$ are approximately equal. It is not sufficient because the equation can fail in large systems, no matter how large the systems are. We will see examples of this in the next

[23] We do *not* attribute this claim to Baxter.

section. Conversely, the condition is not necessary because $F_B = \langle f \rangle \pm \chi_f$ can hold in small systems like a single harmonic oscillator for particular variables f (it holds exactly, for instance, for $f(x) = c$ for all x and c constant).

Another possible condition is $C = \{$the system is ergodic$\}$. Hill begins his discussion of AP by associating observed values with *finite* time averages over a period τ during which the 'experimental measurement' lasts (1987, 3). He notes that a 'direct computation' of this value cannot be carried out and that 'Gibbs' alternative suggestion' was 'that an ensemble average be used in place of a time average' (*ibid.* 5). This instates a 'correspondence' between time averages and phase averages. But he immediately adds that '[n]o completely general and rigorous proof of the validity of this correspondence is available' (*ibid.* 8). What would it take to get such a proof? Hill states that establishing the correspondence for '$\tau \to$ large' is known as the 'ergodic problem', and there have been 'many attempts' to solve the problem, 'none completely successful' (*ibid.* 16). Hence, the limitation on AP are grounded in the limitations of establishing ergodicity. Similarly, Kittel states AP and anchors it in equating phase averages and time averages. He notes that to justify such a move the system has to be ergodic but immediately adds a cautionary note: 'It is certainly plausible that the two averages might be equivalent, but it has not been proved in general that they are exactly equivalent' (1958, p. 8).[24] Ruelle (1969, 2-3) also discusses the justification of AP in terms of time-averages and ergodicity and notes that 'it should be stressed that a more satisfactory argument should involve the fact that we deal with large systems. In particular, a large system may have from the physical viewpoint a completely normal thermodynamic behaviour without being, strictly speaking, ergodic' (Ruelle 1969, 3). By not being strictly ergodic he seems to have the idea in mind that a system is nearly ergodic; formally, this is usually captured by the condition of epsilon-ergodicity (cf. Vranas 1998). So the view expounded here seems to be that $F_T = \langle f \rangle \pm \chi_f$ holds if the system is ergodic, or the slightly weaker condition that the system is epsilon-ergodic, i.e. $C = \{$the system is epsilon-ergodic$\}$.

This viewpoint is unsuccessful for several reasons. Sklar (1973, 211; 1993, 176-9) and Malament and Zabell (1980, 342-3) argue that from the fact that measurements take some time it does not follow that what is actually measured are time averages, and the association of measurement results with time averages is unjustified. One might try to circumvent this objection by adopting Chandler's point of view, mentioned in the last paragraph of Subsection 2.3, that the observed value is a average over many independent observations. But this would still have to be the average of observations made over a finite interval of time, and this does not sit well with ergodicity with requires a limit for time towards infinity.

Even if one could, somehow, set all these issues aside and legitimately associate measured values with *infinite* time averages, ergodicity would not provide a condition which makes MAP true because ergodicity is neither necessary nor sufficient for $F_B = \langle f \rangle \pm \chi_f$. First, there are system in which this equation holds despite them not being ergodic or epsilon-ergodic. An example of such a system is the Kac ring, which we will encounter later in the next subsection. Second, there are ergodic and epsilon-ergodic systems in which $F_B = \langle f \rangle \pm \chi_f$ fails, which happens trivially when f assumes values on non-equilibrium states that are vastly different from F_B.

What we have seen is that being large and being ergodic or epsilon-ergodic *by themselves* do not provide the C needed. This does, however, not rule out that they play a role in a larger package of conditions that do fit the bill.

[24] Similar arguments can found in Chandler (1987, 55-59) and Isihara (1971, 23-30).

4.2 The Khinchin Condition

In the previous subsection we noted that $F_B = \langle f \rangle$ holds trivially if the function f is constant. This is of course an uninteresting case because functions that are of interest in physics will vary. But it provides an interesting perspective on the problem because we can now ask: how far can we move away from the trivial case of f being constant and still retain the result that $F_B = \langle f \rangle \pm \chi_f$. Intuitively this is the case when fluctuations of f away from $\langle f \rangle$ are small. This suggestion has been articulated by a number of authors. Hill submits that the validity of the identification of observable values with ensemble averages is legitimate only when the fluctuations of f away from $\langle f \rangle$ are small (1987, 9-10). Schrödinger declares phase averaging works in cases where the 'distribution becomes infinitely sharp. Mean values, most probable values, any values that occur with non-vanishing probability — all become the same thing' (1989, p. 35). And Landau and Lifshitz note that 'if, by means of the function $\rho(p, q)$ we construct the probability distribution of the function for the various values of the quantity $f(p, q)$, this function will have an extremely sharp maximum for $f = \bar{f}$, and will be appreciably different from zero only in the immediate vicinity of this point' (1980, p. 5).

The challenge then is to articulate what it means for f to have small fluctuations. An important way to do this is by dint of the Khinchin condition. There are two versions of the Khinchin condition, one formulated in terms of Gibbsian phase averages and the other in terms of Boltzmannian equilibrium values. The idea underlying the first version is that the values of the variable f are approximately equal to the Gibbsian phase average $\langle f \rangle$ almost everywhere on the state space. This formulation is often appealed to in the literature, e.g. Wallace 2015, 289; Unpublished; Malament and Zabell 1980, 344-345; Vranas 1998, 693; Lavis 2005, 267-268, also endorses this condition and, in particular, seems to have in mind that this condition should apply to densities like the entropy density, internal energy density and magnetisation density).

To make this idea precise, recall the notion of a fluctuation as introduced in equation (9) and assume, to begin with, that macro-states are defined by certain specific values of the macro-variables (see Section 2.2). Then the *Khinchin-condition (Version A)* is then satisfied iff:

> There is a $\hat{X} \subseteq Z$ with $\mu_X(\hat{X}) = 1 - \delta$ for a very small $\delta \geq 0$ such that $|\Delta f(x)| = 0$ for all $x \in \hat{X}$.

If macro-states are defined via ranges of macro-values (see again Section 2.2), which is how they seem to be defined in the references given in the second paragraph of this section, then the *Khinchin-condition (Version A)* is satisfied iff:

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

The underlying assumption here is that one macro-state is defined by exactly those values within ε of $\langle f \rangle$.

Suppose now that a Boltzmannian equilibrium exists and let F_{equ} be the value of f in the Boltzmannian equilibrium macro-region (for standardly defined macro-states) or let F_{equ} be one of the values in the range of values defining the equilibrium macro-state (when macro-states are defined via ranges of values). There are only very few states of at most measure δ that have macro-values that differ from $\langle f(x) \rangle$ (for standardly defined macro-states) or that differ by more than ε from $\langle f(x) \rangle$ (for macro-states defined via ranges of values). Hence these states cannot form the Boltzmannian equilibrium macro-state. Therefore, for standard macro-states F_{equ} must be equal to $\langle f(x) \rangle$ (for standard macro-states); for macro-states defined via ranges of values F_{equ} must be within ε of $\langle f(x) \rangle$,

i.e.,

$$|\langle f(x)\rangle - F_{equ}| \leq \varepsilon, \tag{12}$$

and the Boltzmannian equilibrium values and the Gibbsian phase average agree approximately.

The second formulation of the Khinchin condition is also often appealed to in the physics and philosophy literature (e.g. Ehrenfest and Ehrenfest-Afanassjewa 1959, 46-52). Here the idea is that the variable f is equal to the Boltzmannian equilibrium value nearly everywhere on state space and that the variable does not take extreme values on the rest of the state space. That is, as remarked above, Version B is formulated in terms of Boltzmannian equilibrium values and not, as Version A, in terms of the phase average. Also, Version B is only formulated for a standard macro-states structure (and not for macro-states defined via ranges of values). Formally, the *Khinchin condition (Version B)* is satisfied iff:

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

The idea behind (ii) is that macro-values of non-equilibrium states should not be extremely high or extremely low such that their contribution to the phase average causes a significant difference between the Boltzmannian equilibrium value and the phase average.

A simple calculation shows that for systems that satisfy Version B of the Khinchin condition with respect to f, the phase average is approximately equal to the Boltzmannian equilibrium macro-value F_{equ}:

$$\begin{aligned} &|\langle f(x)\rangle - F_{equ}| \\ &\leq |\int_{\bar{X}} f(x)d\mu_X - F_{equ}(1-\delta)| + |\int_{X\setminus\bar{X}} f(x)d\mu_X - F_{equ}\delta \\ &| \leq 0 + \gamma = \gamma, \text{ (because of (i) and (ii) of Version B of the the Khinchin condition).} \end{aligned} \tag{13}$$

In what follows, dependent on what is more suitable, we will sometimes focus on Version 1 and sometimes on Version 2.

The name of the condition is owed to the fact that Khinchin instigated a systematic study of functions that satisfy strong symmetry requirements and therefore have small fluctuations for systems with a large number of constituents.[25] As discussed in this section, and also in Section 1, arguments for the conclusion that Gibbsian phase averages are the correct equilibrium values and hence approximately agree with Boltzmannian values (or lie within the Boltzmannian equilibrium interval) typically assume that the Khinchin condition is satisfied.The importance of the Khinchin condition (Version B) can be illustrated with two examples.

4.2.1 The Dilute Gas With Distributions as Macro-States

The first example for a system in which the Khinchin condition furnishes the required justification is the dilute gas as discussed by Ehrenfest and Ehrenfest-Afanassjewa (1959). Consider a dilute gas with N particles. The state of one particle is given by a point in the 6-dimensional state space X_1^{dg} (consisting of the three position and the three momentum coordinates of the particle). Thus the state x^{dg} of the entire gas is given by N points in X_1^{dg}, which is equivalent to specifying one point in X^{dg} – N times the Cartesian product of X_1^{dg}. Since the gas is confined to a finite container and

[25]The Khinchin condition should not be conflated with Khinchin's (1949) theorem, to which we turn shortly.

its energy is constant, only a certain finite part of X_1^{dg} is accessible. This accessible part of X_1^{dg} is then divided into cells of equal size whose dividing lines run parallel to the position and momentum axes. The result is a finite partition $\Omega^{dg} := \{\omega_1^{dg}, ..., \omega_l^{dg}\}$, $l \in \mathbb{N}$, on X_1^{dg}. 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* $D^{dg} = (N_1, N_2, \ldots, N_l)$ (N_i is the number of particles in cell ω_i^{dg}). Each distribution is compatible with several arrangements, and the number of arrangements corresponding to a given distribution D^{dg} is $G(D^{dg}) = N! \,/\, N_1!N_2! \ldots, N_l!$.

Since the energy is constant, the effective state space of the system is the *hypersurface of constant energy* $Z = X_E^{dg} = \{x^{dg} \in X^{dg} \mid E(x^{dg}) = E\}$, where $E(x^{dg})$ is the energy of x^{dg} and E is a certain energy value. The measure on this state space is the microcanonical measure μ_E^{dg}. Under the assumption that the energy e_i of particle i only depends on the cell in which it is located (and *not* on the location of the other particles),[26] one can show that the most common distribution (in terms of the measure μ_E^{dg}) is the (discrete) Maxwell-Boltzmann distribution

$$N_i = \gamma e^{\lambda e_i}, \tag{14}$$

where γ and λ are parameters which depend on N and E. This is the result of the calculations in Boltzmann's (1877) classical combinatorial argument).

Now suppose that, based on the distributions D^{dg}, a macro-variable f^{dg} is defined on Z as follows: it assigns the value F_{equ} to states x^{dg} that are in the Maxwell-Boltzmann distribution or in a distribution that is very close to the Maxwell-Boltzmann distribution. It assigns different values for all other distributions, and it is assumed that these values do not differ dramatically from F_{equ}. Ehrenfest and Ehrenfest-Afanassjewa then assume that a Boltzmannian equilibrium exists.[27] More specifically, they assume that the macro-region in which the macro-variable has value F_{equ} is an α-ε-equilibrium (i.e. that for all initial states x (except possibly for a set of measure ε) $LF_{X^{dg}_{M_{F_{equ}}}}(x) \geq \alpha$ for a certain $\alpha > 1/2$). By the dominance theorem, it follows that this region takes up most of the state space. In fact, as Ehrenfest and Ehrenfest-Afanassjewa (1959, 46-52) show with reference to Jeans' (1916, §46-§56) calculations, in this case nearly all of state space is taken up by the equilibrium macro-region $M_{F^{equ}}$. Because nearly all of state space is taken up by the equilibrium macro-region and the macro-variable f^{dg} does not take extreme values outside equilibrium, the Khinchin condition (Version B) is satisfied. Hence the Gibbsian phase average $\int f^{dg} d\mu_E$ is approximately equal to the Boltzmannian equilibrium value F_{equ}. To conclude, for dilute gases with variables f^{dg} of the kind considered here, the Boltzmannian equilibrium value and the Gibbsian phase average are approximately equal.

4.2.2 The Kac Ring With Coarser Macro-States

As another example consider the Kac ring with a non-standard macro-state structure. The Kac ring consists 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). Hence the one spin state space is $X_1^{kr} = \{u, d\}$. A *micro-state* x^{kr} of the ring is a specific combination of up and down spin for all sites, and the full

[26] Strictly speaking this amounts to assuming that the gas is ideal.

[27] More precisely, they state that the equilibrium macro-region is given by the macro-region of largest measure. This assumes that a Boltzmannian equilibrium exists. In general, it does not follow that the macro-region of largest measure automatically corresponds to a Boltzmannian equilibrium. For instance, if the dynamics is the identity function, the macro-region of largest measure does not correspond to a Boltzmannian equilibrium – cf. Werndl and Frigg (2015a, 2015b).

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 T^{kr} 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 usually considered is the uniform measure $\mu_{X^{kr}}$ on X^{kr}. $(X^{kr}, P(X^{kr}), T_t^{kr}, \mu_{X^{kr}})$, where T_t^{kr} is the t-th iterate of T^{kr} and $P(X^{kr})$ is the power set of X^{kr}, is a deterministic model describing the behaviour of the spins (cf. Bricmont 2001; Lavis 2008). The *macro-states* usually considered are the *total number of up spins* and will be labelled as M_i^K, where i denotes the total number of up spins, $0 \leq i \leq N$. The usual macro-state structure will be discussed in the next section (because it does *not* provide an instance of the Khinchin condition). Here we will discuss instead a different macro-state structure also discussed in Lavis (2008, 686). Namely, let $N = 10.000$ and let one macro-state be defined by the union of all M_i^K for $5000 - 221 \leq i \leq 5000 + 221$ and assign the macro-value 0 to it (this is an example where new macro-variables are defined by previously considered macro-variables as discussed in Section 2.2; in this case the new macro-variables are defined by the previously considered standard macro-variables of the KAC ring). It can be shown that the macro-region corresponding to this macro-state takes up 99.999% of state space. Now assume that there are other positive macro-values that are different from zero but that are not exorbitantly large or exorbitantly small. Then, clearly, the Khinchin condition (Version B) is satisfied for this macro-state structure of the Kac ring and the Boltzmannian equilibrium value is 0 and the value obtained by Gibbsian phase averaging is approximately 0.

4.2.3 Fluctuation Theorems and Khinchin's Theorem

The requirement that $\Delta f(x)$ is small in Version A of the Khinchin condition is in effect a condition on f and so the question is: what functions satisfy the requirement of zero (small) fluctuation? When put in this way, two places to look for conditions suggest themselves: the subfield of SM working on so-called *fluctuation theorems*, and Khinchin's own research programme. We now look at each in turn. Our sober conclusion will be that while both offer interesting results, these results concern different issues and provide no clear-cut criteria for f to have small fluctuations. Yet *what remains is that both Version A and Version B of the Khinchin condition identify some of the most important conditions under which the Boltzmannian equilibrium value and the Gibbsian phase average coincide.*

The core idea of fluctuation theorems is to show that under certain conditions (typically involving the limit $N \to \infty$) fluctuations vanish. But rather than working with $\Delta f(x)$, fluctuation theorems typically concern *relative* fluctuations:

$$\Delta_r f(x) = \frac{\sqrt{\langle (\Delta f(x))^2 \rangle}}{|\langle f \rangle|}, \tag{15}$$

which quantify the average size of fluctuations – as defined in Equation (9) – *relative* to the absolute value of the Gibbsian phase average: $\Delta_r f(x)$ is small if the average fluctuations $\Delta f(x)$ are small compared to $\langle f \rangle$. A typical fluctuation theorem then shows for a certain function f that $\Delta_r f(x)$ vanishes as $N \to \infty$ (Lavis and Bell 1999, Section 2.5). As an example consider the internal energy U (which will be defined in Section 4.4.1). One can show that $\Delta_r U(x) \approx \sqrt{2/(3N)}$, and hence $\Delta_r U(x)$ goes to zero as N goes to infinity (*ibid.*).

If a fluctuation theorem holds, in many cases the Boltzmannian equilibrium value is approximately equal to the Gibbsian phase average. So that a fluctuation theorem holds can give a hint that there might be approximate euqality between Gibbsian phase averages and Boltzmannian equilibrium

values. Yet logically the question is whether a fluctuation theorem provides us with a sufficient condition for approximate equality. That is, assuming that a fluctuation theorem for a certain quantity f holds, can we infer that Gibbsian phase averages and Boltzmannian equilibrium values coincide? Unfortunately not. The problem is that results about *relative* fluctuations $\Delta_r f(x)$ imply nothing about absolute fluctuations $\Delta f(x)$. The numerator in $\Delta_r f(x)$ is the expectation value of the square of $\Delta f(x)$. But from the fact that $\Delta_r f(x) \to 0$ as $N \to \infty$ one cannot infer that $\Delta f(x) \to 0$. In fact $\Delta_r f(x) \to 0$ is compatible with $\Delta f(x)$ diverging for $N \to \infty$. The internal energy of the six-vertex model mentioned above is a case in point. As just noted, it can be shown that the relative fluctuation for the internal energy is proportional to $\sqrt{N}/N$. So the relative fluctuation tends toward zero as $N \to \infty$. However, as we will see in Section 4.4.1, the absolute fluctuation is at least $\sqrt{N}/2$ and in fact diverges for $N \to \infty$. So a zero relative fluctuation is compatible with large absolute fluctuations, which shows that Boltzmannian equilibrium values and Gibbsian phase average can be very different even if there is a fluctuation theorem.

Can we at least draw the converese conclusion and take the failure of a fluctuation theorem to indicate that Boltzmannian equilibrium value and the Gibbsian phase average *are* different? Unfortunately this inference does not hold either. First, there could be cases where the numerator in the relative fluctuation tends to zero (implying that Boltzmannian equilibrium value and the Gibbsian phase average agree approximately) while the relative fluctuation does not tend toward zero (e.g. because the denominator of the relative fluctuation is or goes to zero). Second, note that the numerator does not concern what we are interested in, i.e. the difference between the Boltzmannian equilibrium value and the phase average. Rather, it concerns the *expectation* value of the difference between the value of the macro-variable $f(x)$ and the phase average $\langle f \rangle$. There are cases where the Boltzmannian equilibrium value is approximately equal to the phase average while at the same time the numerator of the relative fluctuation is *not* small because the expectation value of $f(x)$ and the phase average is not small. Examples include cases where the Cancelling Out Theorem (to be discussed in Section 4.4) holds and hence the Boltzmannian equilibrium value equals the phase average but where there are large differences between $f(x)$ and $\langle f \rangle$ on macro-regions different from the equilibrium region. So the sober conclusion is that fluctuation theorems *per se* tell us nothing about the relation between Boltzmannian and Gibbsian equilibrium calculations.

We are well aware of the prominent role of fluctuation theorems in the literature and that they are often taken to tell us when Boltzmannian and Gibbsian calculations agree. Yet, it is simply a misapprehension that fluctuation theorems tell us something about the most common formulations (formulation 1 and formulation 2 above) of the Khinchin condition. They do not because they concern relative fluctuations while the Khinchin condition concerns absolute fluctuations.[28]

The core idea of Khinchin's (1949) programme is to focus attention on a specific class of variables, so called sum functions, and to study these in the context of large models.[29] Sum functions are functions f that can be written as a sum of one-particle functions $f = \sum_{i=1}^{N} f_i$, where $a < f_i < b$ for some $a, b \in \mathbb{R}$. A simple example is the internal energy of a model of noninteracting particles, which is just the sum of the energy of the individual particles. Khinchin furthermore assumed that the model was isolated from the environment and that the Hamiltonian of the model was also a sum function. Under these assumptions Khinchin could prove that for all sum functions f there exist

[28] The Khinchin conditions will, in general, be much easier to satisfy if the relevant quantities are intensive because the relevant calculations are then often similar to those in the fluctuation theorems. (Recall that the macro-variables f that figure in the Khinchin condition can be extensive or intensive – whatever one is interested in. As already argued above, we want to be flexible and noncommittal about the choice of macro-variables).

[29] For detailed discussion of this programme see Badino (2006), Batterman (1998) and van Lith (2001).

positive constants k_1 and k_2 such that

$$\lambda\Big(\Big\{ x \in X_E : \Big|\frac{f^*(x) - \langle f\rangle}{\langle f\rangle}\Big| \geq k_1\, N^{-1/4}\Big\}\Big) \;\leq\; k_2\, N^{-1/4}, \tag{16}$$

where X_E is hypersurface of constant energy, λ is the microcanonical measure and N the number of constituents of the system. $f^*(x)$ denotes the infinite time average of the function f along the trajectory that starts in initial condition x (the average is infinite in that the limit $t \to \infty$ is taken). This result is now known as Khinchin's ergodic theorem. In effect the theorem says that as N becomes larger, the measure of those regions on the hypersurface of constant energy where the infinite time average and the phase average differ by more than a small amount when compared to the phase average tends towards zero.

The numerator of the crucial term in the theorem, $(f^*(x) - \langle f\rangle)/\langle f\rangle$, looks similar to the definition of $\Delta f(x)$ and so one might hope that Khinchin's theorem offers an answer to the question about the circumstances in which fluctuations are small. Unfortunately, this hope is in vein. Most importantly, there is a crucial difference between $\Delta f(x)$ and $f^*(x) - \langle f\rangle$: the former is defined in terms of the value of f at x while the latter is defined in terms of the *infinite time average* of a trajectory starting in x. There is a tradition in the foundation of GSM that associates time averages with measurement outcomes, and therefore sees time averages as the true equilibrium values. However, as many commentators have pointed out, it is far from clear why observations on a model should yield time averages[30] (much less why they should yield *infinite* time averages).[31] So bringing time averages back into the account would be a regress rather than progress. Even if this problem could be circumvented, we would have to face up to the intrinsic limitations of Khinchin's theorem, namely the unrealistic (and therefore constraining) assumption that both the observable and the Hamiltonian are sum-functions (effectively limiting the theorem to non-interacting particles) and the restriction to the microcanonical ensemble.

4.3 The Average Equivalence Theorem

In addition to the Khinchin condition, another important case of agreement between the Boltzmann and Gibbsian equilibrium calculations is given by the Average Equivalence Theorem (Werndl and Frigg 2017).[32] The conditions of this theorem will be referred to as the 'Average Equivalence Conditions'.

> **Average Equivalence Theorem (AET)**. Suppose that a model is composed of $N \geq 1$ components. That is, for the deterministic case the state $x \in Z$ is given by the N coordinates $x = (x_1, \ldots, x_N)$; $Z = Z_1 \times Z_2 \ldots \times Z_N$, where $Z_i = Z_{oc}$ for all i, $1 \leq i \leq N$ (Z_{oc} is the one-component space). Let μ_Z be the product measure $\mu_{Z_1} \times \mu_{Z_2} \ldots \times \mu_{Z_N}$, where $\mu_{Z_i} = \mu_{Z_{oc}}$ is the measure on Z_{oc}.[33] Suppose that the macro-variable K is the sum of the one-component variable, i.e. $K(x) = \sum_{i=1}^{N} \kappa(x_i)$ (it is assumed here that all

[30] See Frigg (2008, 146-147) for a review of this discussion.

[31] Lavis' (2005, 2007) conditions for the agreement of Gibbsian and Boltzmannian values also rests on the assumption that what is observed in equilibrium are time averages of the observable, and hence his approach suffers from the same difficulty.

[32] Werndl and Frigg (2017) refer to this result as the 'equilibrium equivalence theorem'. This name could be misleading because the theorem concerns the largest macro-region and *not* the a Boltzmannian equilibrium *per se*. For this reason we prefer the label 'average equivalence theorem'. If the theorem is used to make claims about Boltzmannian equilibrium, dynamical assumptions have to come in. Specifically, if a Boltzmannian equilibrium state exists, then, by the dominance/prevalence theorem, the Boltzmannian equilibrium value equals the value of the largest macro-region which, by the theorem, is equal to the Gibbsian phase average.

[33] N is assumed to be a multiple of k, i.e. $N = k * s$ for some $s \in \mathbb{N}$.

sums of possible values of the one-component variable are possible values of the macro-variable). Then the value corresponding to the largest macro-region as well as the value obtained by Gibbsian phase averaging is $\frac{N}{k}(\kappa_1 + \kappa_2 + \ldots \kappa_N)$. It is obvious how to formulate the theorem for the stochastic case by making the substitutions discussed in Subsection 2.1.

The proof for the deterministic case can be found in Werndl and Frigg (2017) (and it is clear how it carries over to the stochastic case). Note that the AET equally applies to deterministic and stochastic models and makes no assumptions about the dynamics of the model. While it is true that the existence of a Boltzmannian equilibrium crucially depends on the dynamics, the AET is not a claim about a Boltzmannian equilibrium but about the largest-macro-region. Of course, if the AET is used to compare the Boltzmannian equilibrium value and the Gibbsian phase average (and the AET will usually be used in this way), dynamical assumptions play a role. Namely, then the dynamics has to be such that a Boltzmannian equilibrium exists. Finally, note that even if a Boltzmannian equilibrium exists, the theorem only offers sufficient but not necessary conditions for the Boltzmannian equilibrium value and Gibbsian phase averages to agree (indeed, in Subsection 4.5 we show that the Khinchin condition, the Average Equivalence Condition and the Cancelling Out Condition are independent).

The crucial assumptions of the theorem are (i) that the macro-variable is a sum of a variable on the one-component space (where all sums of possible values of the one-component variable are possible values of the macro-variable); (ii) that the macro-variable on the one-component space corresponds to a partition with cells of equal probability; and (iii) that the measure on state space is the product measure of the measure on the one-component space. These assumptions may seem restrictive and to some extent they are. Nevertheless a number of standard applications of SM fall within the scope of the theorem.[34]

4.3.1 The Baker's Gas with Distributions as Macro-states

One instance of the AET is the baker's gas. The baker's gas is a deterministic model, consisting of N identical particles that evolve independently according to the baker's transformation (Lavis 2005). The model's state space is $X^{bg} = [0,1]^{2N}$ (which is at once the full and the effective state space of the system), and its micro-states are of the form $x^{bg} = (b_1, c_1, \ldots, b_N, c_N)$, where $b_i \in [0,1]$

[34]There seems to be a similarity between the AET and the weak law of large numbers (LLN), which states that given independent and identically distributed random variables (which we consider in the AET theorem) for any $\varepsilon > 0$ (cf. Meester 2003, Section 4.1):

$$\mu\left(\left\{x : \left|\frac{\sum_{i=1}^{N}\kappa(x_i)}{N} - \frac{(\kappa_1 + \kappa_2 + \ldots \kappa_N)}{k}\right| < \varepsilon\right\}\right) \geq 1 - \frac{\sigma^2}{\varepsilon^2 N}. \quad (17)$$

This similarity is superficial and the theorems are different. We list two major differences here. First of all, the LLN only states that the values of $\sum_{i=1}^{N}\kappa(x_i)/N$ and $(\kappa_1 + \kappa_2 + \ldots \kappa_N)/k$ are within ε. It does *not* say whether the values of the *extensive* macro-variables we consider in the AET, $\sum_{i=1}^{N}\kappa(x_i)$, are close to $N(\kappa_1 + \kappa_2 + \ldots \kappa_N)/k$. All one obtains from the LLN is that their values are within $N\varepsilon$, but $N\varepsilon$ can be very large. By contrast, the AET states that the value of the largest macro-region of the macro-variable $\sum_{i=1}^{N}\kappa(x_i)$ *equals* $N(\kappa_1 + \kappa_2 + \ldots \kappa_N)/k$. Second, AET and the LLN are results about *different* macro-variables. AET is a result about the macro-variable $\sum_{i=1}^{N}\kappa(x_i)$ or, if it is divided by N, about $\sum_{i=1}^{N}\kappa(x_i)/N$. By contrast, LLN is a statement about the probability of states that are close or equal to $(\kappa_1 + \ldots + k_k)/k$. Hence it can tell us something about the *different* macro-variables that are defined by assigning the same macro-value to all states that are close or equal to $(\kappa_1 + \ldots + k_k)/k$ (or about macro-states that are defined by a range of values, and one macro-state is given by values that are close or equal to $(\kappa_1 + \ldots + \kappa_k)/k$). Yet, note that these new macro-variables are *not* the macro-variables considered in the AET theorem, viz. $\sum_{i=1}^{N}\kappa(x_i)$. So the LLN does not tell us about what happens for the macro-variables considered in the AET.

is the momentum and $c_i \in [0,1]$ the position coordinate of the i-th particle. Time is discrete and the evolution to the next time step is given by applying to each coordinate the baker's transformation. That is, $x^{bg} = (\ldots b_i, c_i \ldots)$ evolves into $\Lambda(x^{bg}) = (\ldots \theta(b_i, c_i) \ldots)$, where

$$\theta(b_i, c_i) = 2b_i, \frac{c_i}{2} \text{ if } 0 \leq b_i \leq \frac{1}{2} \text{ and } 2b_i - 1, \frac{c_i + 1}{2} \text{ otherwise.} \tag{18}$$

The state space X^{bg} is endowed with a uniform probability measure $\mu_{X^{bg}}$, the $2N$-dimensional Lebesgue measure, which is invariant under the dynamics.

Next we need a macro-variable. So we partition the unit square (the state space for one particle) into cells of equal size $\delta\omega$ whose dividing lines run parallel to the position and momentum axes. This results in a finite partition $\Omega^{bg} := \{\omega_1^{bg}, ..., \omega_k^{bg}\}$, $k \in \mathbb{N}$. The coarse-grained micro-state of a particle is the cell in which a particle's state lies. An *arrangement* is given by a specification of the coarse-grained micro-state of all the particles. A *distribution* is a specification of how many particles' states lie in a given cell. Consider the distribution $D^{bg} = (N_1, N_2, \ldots, N_k)$, where N_i is the number of particles in cell ω_i^{bg}. As shown in Werndl and Frigg's (2017), for these macro-variables the AET theorem applies and the Boltzmannian equilibrium value and the Gibbsian phase average are both $(N/k, N/k, \ldots, N/k)$.

4.3.2 The Kac Ring With Distributions as Macro-States

Another instance of the AET is the Kac ring (Section 4.2.2) with the standard macro-state structure. The *macro-states* that are usually considered are the *total number of up spins* M_i^K. Clearly, the total number of up spins is a sum of the variable of the one-component space $\{u, d\}$ (namely, it is 1 for u and 0 for d). Also, the macro-variable on the one-component space corresponds to a partition with cells of equal probability (each state u and d has probability 1/2). Further, the uniform measure on X^{kr} is the product measure of the measures on the one-spin space $\{u, d\}$. It follows that the Kac ring is an instance of the AET and the Boltzmannian equilibrium value and the Gibbsian phase average are both $N/2$.

4.3.3 The Ideal Gas With the Coarse-Grained Position

As a further example consider the ideal gas with N particles. For simplicity, we assume that the particles are moving on a three-dimensional torus (implying that the momentum of each particle stays constant). The state of one particle is given by a point in the 6-dimensional state space X_1^{ig} (of all possible three momentum and position coordinates). The state x^{ig} of the system is given by N points in X_1^{ig}, and X^{ig} is then the set of all possible states of the system. Now as macro-variable let us consider the coarse-grained position of the particles. More specifically, consider a partition X_1^{ig} into cells of equal size whose dividing lines run parallel to the momentum axes, i.e. the partition tells you about the coarse-grained position of the particle (for an ideal gas on a torus the momentum of any particle stays constant; hence it only makes sense to coarse-grain the position coordinate). This results into a finite partition $\Omega^{ig} := \{\omega_1^{ig}, ..., \omega_l^{ig}\}$, $l \in \mathbb{N}$, on X_1^{ig}. The specification of where each particle's state lies is called an arrangement. Finally, a specification of the number of particles in each cell is referred to as a distribution $D^{ig} = (N_1, N_2, \ldots, N_l)$ (N_i is the number of particles in cell ω_i^{ig}), and the distributions are the possible macro-states. The number of arrangements corresponding to a given distribution D^{ig} is $G(D^{ig}) = N! \,/\, N_1!N_2! \ldots, N_l!$. It is then easy to see that the most common distribution is the uniform distribution $D_u^{ig} = (N/l, \ldots, N/l)$[35]. The calculations in Lavis (2005)[36]

[35] We are assuming that N is divisible by l, i.e. there is a $m \in \mathbb{N}$ such that $N = m * l$

[36] The calculations performed for the baker's gas carry over to the ideal gas.

then show that the largest macro-region corresponds to the uniform distribution and that this largest macro-region takes up less than half of state space. It can be shown that the motion confined to the hypersurface X_p determined by the constant momenta of the particles is ergodic for nearly all values of p (namely, when the coordinates of $p = (p_1, \ldots, p_{3N})$ are linearly independent over Z – then the dynamics corresponds to an irrational rotation on a torus). Hence D_u^{ig} corresponds to an α-0-equilibrium. Clearly, the distributions are a sum of an variable on the one-component space X_1^{ig} (namely the l-dimensional vector, whose j^{th} coordinate is 1 for the cell the particle is in and whose other coordinates are all 0). Also, by construction, the macro-variable on the one-constituent space corresponds to a partition with cells of equal probability. Further, the uniform measure on X^{ig} is the product measure of the uniform measure on the one-constituent space. It follows that the ideal gas with the distributions as macro-variables is an instance of the AET and the Boltzmannian equilibrium value and the value provided by Gibbsian phase averaging is $(N/l, N/l, \ldots, N/l)$.

4.4 The Cancelling Out Theorem

A third important condition under which the Boltzmannian and Gibbsian equilibrium calculations agree is given by the Cancelling Out Theorem (and the conditions of this theorem will be refereed to as the 'Cancelling Out Conditions').

> **Cancelling Out Theorem (COT).** Consider a deterministic or stochastic model with Boltzmannian equilibrium macro-state M_{equ} with equilibrium value V_{equ} and other macro-states $M_1, \ldots, M_q$, $q \in \mathbb{N}$, with corresponding macro-values $V_{M_1}, \ldots, V_{M_q}$. Further, suppose that for any macro-state $M_i \neq M_{equ}$ there is a macro-state M_j such that (i) $\mu_Z(Z_{M_i}) = \mu_Z(Z_{M_j})$ (for deterministic models) or $\bar{P}\{S(t) \in \bar{Z}_{M_i}\} = \bar{P}\{S(t) \in \bar{Z}_{M_j})\}$ (for stochastic models) and (ii) $V_{M_i} + V_{M_j} = 2V_{M_{equ}}$. Then the Boltzmannian equilibrium value as well as the value obtained by phase averaging is V_{equ}.

The proof of this theorem is given in the Appendix. The intuitive reasoning behind the proof is as follows. If the state space is divided up in such a way that next to the largest macro-region (which corresponds to the Boltzmannian equilibrium) there are always two macro-states of equal size whose average equals the Boltzmannian equilibrium value, then the Boltzmannian equilibrium value is equal the value obtained by Gibbsian phase averaging.

Note that, as the other conditions, the COT only offers sufficient but not necessary conditions for the Boltzmannian equilibrium value and Gibbsian phase average to agree (and in Subsection 4.5 we will show that the Khinchin condition, Average Equivalence Conditions and Cancelling Out Conditions are independent). Note also that the proof does not make any assumptions about the dynamics of the model (other than that a Boltzmannian equilibrium exists), and it applies equally to deterministic and stochastic models. The assumptions of the theorem are, of course, to some extent restrictive (the macro-state structure needs to be of a special kind). Nevertheless, a number of standard applications of SM fall within the scope of the theorem. We will now discuss two important instances of the COT.

4.4.1 The Six-Vertex Model With the Polarisation for High Temperatures

First, consider a two-dimensional quadratic lattice with N grid points. We assume that the lattice lies on a two-dimensional torus, which ensures that every grid point has exactly four nearest neighbours and allows us to neglect border effects. The grid points are referred to as 'vertices'. Each vertex is connected to its four nearest neighbours by an edge. Each edge carries an arrow either pointing towards or away from the vertex. A rule is now imposed restricting the allowable

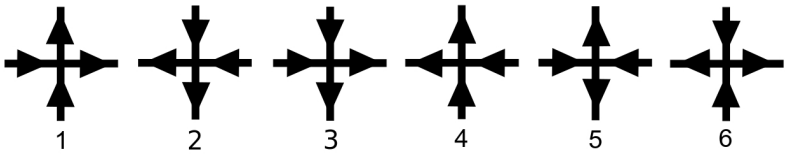

Figure 1: The configurations of the six-vertex model.

arrangements of arrows: the arrows have to be distributed such that at each vertex in the lattice there are exactly two inward and two outward pointing arrows. This rule is known as the 'ice-rule', and models based on this rule are known as 'ice-type models'. At every vertex there are exactly six configurations of the arrows that satisfy the ice-rule. These are shown in Figure 1. The existence of exactly six configurations is what motivates the name 'six-vertex model'.

The physical motivation of the ice-rule is that in frozen water each oxygen atom is connected to four other oxygen atoms. So one can think of vertices as representing oxygen atoms and the edges as representing their bonds, where each bond contains a hydrogen atom. This hydrogen atom does not sit in the middle between the two oxygen atoms; it occupies a position closer to either one or the other oxygen atoms. The arrow can be seen as indicating to which oxygen atom the hydrogen atom is closer. The ice-rule is then the requirement that each oxygen atom has two close and two remote hydrogen atoms (Baxter 1982; Lavis and Bell 1999). The ice rule is satisfied not only by water ice, but also by several crystals and, in particular, potassium dihydrogen phosphate (Slater 1941).

The micro-state of the model $\kappa = (\kappa_1, \ldots, \kappa_N)$ is given by assigning one of the six types of configurations of the arrows permitted by the ice rule to each vertex in the model, where the assignment of configurations to each vertex have to fit together (e.g. one cannot assign vertex 3 to one vertex and vertex 4 to the vertex immediately to its right because they disagree on the direction of the arrow on the line joining them). Each of the six configurations has a certain energy ϵ_i, $1 \leq i \leq 6$. Let $\epsilon(\kappa_j)$ be the energy or the j^{th} vertex (hence all $\epsilon(\kappa_j)$ range over the ϵ_i). The energy of the state κ is then given by:

$$E(\kappa) = \sum_{j=1}^{N} \epsilon(\kappa_j). \tag{19}$$

In what follows we assume that the energy of the different configurations is $\epsilon_1 = \epsilon_2 = 0$ and $\epsilon_3 = \epsilon_4 = \epsilon_5 = \epsilon_6 = 1$ (this is an important case of parameter values that is often discussed – cf. Lavis and Bell 1999, 299). It is common to take the canonical distribution

$$p(\kappa) = e^{-E(\kappa)/kT}/\zeta \tag{20}$$

with

$$\zeta = \sum_{\kappa} e^{-E(\kappa)/kT} \tag{21}$$

to be the outcome distribution. The canonical distribution implies that the lower the temperature, the larger the probability of the lower energy states; and the higher the temperature, the more uniform the probability distribution becomes.

There are many versions of the six-vertex model, but most versions work with a stochastic dynamics that is assumed to be an irreducible Markov model (cf. Section 2). There are several specific algorithms that can be used to generate such a model (see, e.g., Allison and Reshetikin 2005, Barkema

and Newman 1998, Levis 2012; Syljuasen and Zvonarev 2004). A popular family of algorithms are loop algorithms (cf. Levis 2012, 101). The core idea of these algorithms is that, given an initial state κ, an arrow is chosen at random from κ. The vertex where it points to is denoted by v_0. Then randomly an arrow is chosen among the two outgoing arrows attached to v_0. This procedure is repeated until the path encounters a vertex that already belongs to it, creating a closed loop. Then all spins along the loop are reversed to obtain a state κ^*. A Metropolis rule then decides whether this is accepted as the new state. That is, the energy difference between κ^* and κ is calculated. If the energy difference is smaller than zero, the new state is always accepted. If it is larger than zero, it is accepted with probability $e^{\Delta E/kT}$.

A comment on the choice of the canonical distribution as the outcome distribution is in place. The canonical distribution is historically associated with Gibbs, and the choice of this distribution might therefore give the impression that we treat the model in a Gibbsian way right from the start. This impression is mistaken. Its historical origins notwithstanding, the canonical distribution *per se* is simply a probability distribution, playing the role of the outcome distribution of a stochastic model. Given the state space, the canonical distribution and a deterministic or stochastic dynamics, we have a deterministic or stochastic model. And a deterministic or stochastic model can be studied either from the Boltzmannian perspective (by looking at macro-states and the macro-values the system takes over time) or phase averages can be calculated for this system (as done in Gibbsian statistical mechanics). So at this point we simply specify a stochastic model (in the sense of Subsection 2.1), and such a model is conceptually prior to either the Boltzmannian and or the Gibbsian approach, and indeed to any consideration of SM!

Consider now two polarisation macro-variables, the vertical polarisation π_v and the horizontal polarisation π_h. They are conveniently written as a vector $\vec{\pi}$:

$$\vec{\pi}(\kappa) = (\pi_v(\kappa), \pi_h(\kappa)) = (\frac{N-2n}{N}, \frac{N-2m}{N}), \tag{22}$$

where n is the number of arrows pointing downward and m is the number of arrows pointing to the left. Note that $-1 \leq \pi_v \leq 1$ and $-1 \leq \pi_h \leq 1$.

As just noted, the higher the temperature, the more the micro-states become equally likely. From this it follows that for sufficiently high temperatures, the largest macro-region corresponds to the polarisation $(0, 0)$ (because the most frequent micro-states will be the ones with an equal number of sites pointing up and pointing down, and an equal number of sites pointing left and pointing right). Because the dynamics is given by an irreducible Markov model (which corresponds to ergodicity in the deterministic case), $(0, 0)$ is the Boltzmannian equilibrium value V_{equ}. Now note that to any micro-state κ there corresponds a micro-state κ^* that results from a 180° rotation of all the arrows of κ. One easily finds that $(\pi_v(\kappa), \pi_h(\kappa)) + (\pi_v(\kappa^*), \pi_h(\kappa^*)) = 0 = 2V_{equ}$ and that the probabilities of κ and κ^* are the same. By grouping the micro-states together into macro-states, it is obvious that conditions (i) and (ii) of COT are satisfied. Hence Gibbsian phase averaging also yields the value $(0, 0)$ (cf. Lavis and Bell 1999, Chapter 10).

4.4.2 The Ising-Model With the Magetisation for High Temperatures

The Ising model is another paradigm model of SM, and there are many versions of the model. In this section we consider a two-dimensional version with nearest neighbour interactions and a stochastic dynamics.[37] Despite being only two-dimensional, this model provides a realistic description of crys-

[37] Our presentation of the Ising model follows Baxter's (1982).

tals that have strong horizontal and weak vertical interactions such as K_2NF_4 and RB_2MnF_4.

Consider again a two-dimensional quadratic lattice with N grid points. At every grid point there is a spin. The state of the i^{th} spin is described by a variable σ_i that can take two values: $\sigma_i = 1$ if the spin points up and $\sigma_i = -1$ if the spin points down. The model's micro-state σ is then given by a specification of the state of every spin on the lattice:

$$\sigma = \{\sigma_1, \ldots, \sigma_N\}. \tag{23}$$

The model's energy is given by its Hamiltonian:

$$H(\sigma) = E(\sigma) = -J \sum_{nn} \sigma_i \sigma_j + L \sum_i \sigma_i, \tag{24}$$

where the first sum ranges of over all nearest neighbour pairs and the second over $i = 1, ..., N$. The first term describes the contribution of intermolecular forces to the energy and the constant $J \geq 0$ is the energy associated with the nearest-neighbour interaction. The second term is the contribution of the interaction of the spins with an external magnetic field of strength L. The zero field case ($L = 0$) plays a particularly important role in the discussion, not least because it is the only case that has been solved analytically, and in what follows we assume $L = 0$.

The standard choice for the outcome distribution is again the canonical distribution

$$p(\sigma) = \frac{e^{-H(\sigma)/kT}}{\zeta}, \tag{25}$$

which specifies the probability of finding the model in a certain configuration σ. ζ is the partition function

$$\zeta_T = \sum_{\sigma} e^{-H(\sigma)/kT}, \tag{26}$$

where the sum is taken over all possible configurations σ of the model. As in the case of the six-vertex model, the underlying dynamics is assumed to be an irreducible Markov model. The probability $p(\sigma)$ is invariant under this dynamics and is therefore the stationary measure of the model. As for the six-vertex model: $p(\sigma)$ is simply the outcome distribution of a stochastic model. Given the state space, the canonical distribution and a deterministic or stochastic dynamics, we have a deterministic or stochastic model. And a deterministic or stochastic model can be studied either from the Boltzmannian perspective (by looking at macro-states and the macro-values the system takes over time) or phase averages can be calculated for this system (as done in Gibbsian statistical mechanics). For what follows it is important to note that if the field is zero and the temperatures are low, the probabilities of the lower energy states are dominant. For high temperatures the probability distribution is flattened out and all configurations are more or less equally likely (Baxter 1982, 9 and 21; Cipra 1987, 942).

The assumption that the dynamics of the Ising model is a Markovian and the stationary outcome distribution is the canonical distribution is a standard assumption in the literature. The dynamics is often specified to be either the Glauber dynamics or the Metropolis algorithm (Adler 2016; Ferrenberg and Landau 1991). The idea underlying the Metropolis algorithm is that, starting from an initial configuration σ, one of the sites is randomly chosen and its spin is flipped. This gives rise to a new state σ_f. The energy difference between the new state and the old state $\Delta E : \sigma_f - \sigma$ is then calculated. If the energy difference is smaller than zero, the new state is always accepted. If it is larger than zero, it is accepted with probability $e^{\Delta E/kT}$. The idea underlying the Glauber dynamics

is that, starting from an initial configuration σ, one of the sites of the lattice i is randomly chosen. Then the spin of the site is assigned a positive spin with a probability that equals the p-conditional probability of a positive spin at site i, given that that all spins agree with σ, at sites different from i, and in this way the new state is obtained.

Consider now the magnetisation as the relevant macro-variable:

$$m(\sigma) = \sum_i -s\sigma_i, \tag{27}$$

where s is a constant. As just noted, the higher the temperature, the more the micro-states are more or less equally likely. Hence, for sufficiently high temperatures T the largest macro-region is the one with macro-value 0 (because the most frequent micro-states will be the ones with an equal number of up and down spins). Because the dynamics is given by an irreducible Markov model, 0 is the Boltzmannian equilibrium value V_{equ}. Now for any state $\sigma = (\sigma_1, \ldots, \sigma_n)$ we can define its conjugate state $\sigma^* = (-\sigma_1, \ldots, -\sigma_n)$, and it is clear that $m(\sigma) + m(\sigma^*) = 0 = 2V_{equ}$ and that both σ and σ^* have the same probability. By grouping the micro-states together into macro-states, it is obvious that conditions (i) and (ii) of the COT are satisfied, and thus the value of Gibbsian phase average is also 0 (cf. Cipra 1987; Lavis and Bell 1999, Chapter 8).

4.5 Independence of the Three Conditions

We have identified three conditions (the Khinchin condition, the Average Equivalence Conditions and the Cancelling Out Conditions) under which the Mechanical Averaging Equation holds, i.e. the Gibbsian phase average agrees with the Boltzmannian equilibrium value. We will now show that these three conditions are independent.

Let us first show the independence of the Khinchin condition and the Average Equivalence Conditions. The baker's gas with the macro-structure that is usually considered (as discussed in Subsection 4.3) is an instance of the AET but it is not an instance of the Khinchin condition. As Lavis (2008, 685-688) showed, the equilibrium macro-region corresponding to $(N/k, N/k, \ldots, N/k)$ only takes up less than a half of state space. The rest of state space is taken up by macro-regions with for which the macro-variables assume values which are distinguishable from the equilibrium value. Thus it is obvious that this cannot be an instance of the Khinchin condition. Likewise, the Kac ring with the standard macro-state structure usually considered (cf. Section 4.3) is an instance of the AET. Yet it is not an instance of the Khinchin condition because, as Lavis (2005, 2008) has shown, 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 states that are macroscopically distinguishable from the Boltzmannan equilibrium state). Finally, the ideal gas with the coarse-grained distributions as macro-variables is also an instance of the AET (cf. Section 4.3) with the uniform distribution as equilibrium distribution. Yet the Khinchin condition is again not satisfied (the calculations in Lavis (2005, 2008) carry over and show that the equilibrium macro-region does not take up nearly all of state space). These examples illustrate the general point that the conditions of the AET theorem are not equivalent to those of the Khinchin condition because the AET theorem does *not* require that the equilibrium macro-region takes up nearly all of state space.

Conversely, the Khinchin condition does not imply that the AET theorem is satisfied. For the Khinchin condition none of the specific requirements of the AET theorem needs to be satisfied. In particular, the Khinchin condition does not require that the macro-variable is a sum of variable 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. This is illustrated by our example of the dilute gas with the macro-variables we discussed above in the Introduction and again in Section 4.1. As we have seen, this example is an instance of the Khinchin 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.

Second, let us now show the independence of the Khinchin condition and the Cancelling Out Conditions. Recall the example of the magnetisation of the Ising model for sufficiently high temperatures, which is an instance of the COT. This example is, however, not an instance of the Khinchin condition because the equilibrium region (where there are an equal number of up and down spins) does not take up nearly all of state space (here again the calculations in Lavis (2005) carry over to show that the largest macro-region does not take up nearly all of state space). In general, the Cancelling Out Conditions do not imply the Khinchin condition because the COT does *not* require that the equilibrium macro-region takes up nearly all of state space.

Conversely, the Khinchin condition also does not imply the Cancelling Out Conditions. Intuitively speaking, this is clear because the Cancelling Out Conditions require a very specific macro-state structure (that, next to the equilibrium macro-state, pairs of macro-regions have the same size and that their averaged macro-value equals the Boltzmannian equilibrium value) that is not required by the Khinchin condition. An example illustrating this point is the dilute gas with the macro-value f as discussed by Ehrenfest and Ehrenfest-Afanassjewa (1959), which, as we have seen above, is an instance of the Khinchin condition. Yet the Cancelling Out Conditions do not necessarily apply here. Recall that the macro-variable f defined on the hypersurface of constant energy X_E^{dg} is such that it assigns the same value to states x^{dg} that are in the Maxwell-Boltzmann distribution or in a distribution that is very close to the Maxwell-Boltzmann distribution and it assigns different values for all other distributions. Now suppose that the values assigned by f for macro-states that are not very close to the Maxwell-Boltzmann distribution are all larger than the macro-value of the Maxwell-Boltzmann distribution. Then it cannot be that condition (ii) of the COT is satisfied.

Finally, let us show that the Average Equivalence Conditions and the Cancelling Out Conditions are independent. Consider the magnetisation of the Ising model for sufficiently high temperatures or the polarisation of the six-vertex model for sufficiently high temperatures, which, as we have seen, are instances of the COT. However, they are not examples of the AET because the measure on state space is not the product measure of the measure of the one-component space and hence (ii) and (iii) of the AET are not satisfied. In general, the Cancelling Out Conditions do not require that the macro-variable is a sum of the variable on the one-component space, that the macro-variable on the one-component space corresponds to a partition with cells of equal probability, or that the measure on state space is the product measure of the measure on the one-constituent space. Hence the Cancelling Out Conditions do not imply the conditions of the AET.

For the other direction, consider again the baker's gas with the distributions $D^{bg} = (N_1, N_2, \ldots, N_k)$ (where N_i is the number of particles in cell ω_i^{bg}) as macro-variables as discussed in Subsection 4.3. Suppose that the macro-value is given by the sum function $\sum_{i=1}^{k} 1N_1 + 2N_2 + \ldots (k-1)N_{k-1} + (k+1)N_k$. Because the dynamics is ergodic (cf. footnote 2.1), $D_{equ}^{bg} = (N/k, \ldots, N/k)$ is the

Boltzmannian equilibrium macro-state with equilibrium macro-value $V_{D^{bg}_{equ}} = \sum_{i=1}^{k} 1\frac{N}{k} + 2\frac{N}{k} + \ldots (k-1)\frac{N}{k} + (k+1)\frac{N}{k}$. Clearly, the baker's gas with these macro-variables is a instance of the AET and the Boltzmannian and Gibbsian phase average is $\sum_{i=1}^{k} 1\frac{N}{k} + 2\frac{N}{k} + \ldots (k-1)\frac{N}{k} + (k+1)\frac{N}{k}$. But now consider the macro-state $D^{bg}_{h} = (0, 0, \ldots, N)$ with macro-value $V_{D^{bg}_{h}} = N(k+1)$ (the highest possible macro-value). For this macro-state there cannot be another macro-state such that condition (ii) of the COT is satisfied. This is so because even if one pairs it with the distribution with the lowest macro-value $D^{bg}_{l} = (N, 0, \ldots, 0)$, the combined values $V_{D^{bg}_{l}} + V_{D^{bg}_{h}}$ will be larger than $2V_{D^{bg}_{equ}}$. To show this, recall that $1 + 2 + \ldots k - 1 = k(k-1)/2$. Hence:

$$\begin{aligned} 2V_{D^{bg}_{equ}} = 2\big(1\frac{N}{k} + 2\frac{N}{k} + \ldots (k-1)\frac{N}{k} + (k+1)\frac{N}{k}\big) &= \\ 2\frac{N}{k}\frac{k(k-1)}{2} + \frac{2N(k+1)}{k} &= N(k-1) + \frac{2N(k+1)}{k}. \end{aligned} \tag{28}$$

Because

$$3N > 2N\frac{(k+1)}{k}, \tag{29}$$

also

$$Nk + 2N > Nk - N + 2N\frac{(k+1)}{k}. \tag{30}$$

Hence from equations (28) and (30) it follows that:

$$V_{D^{bg}_{l}} + V_{D^{bg}_{h}} = N + N(k+1) = N(k+2) > N(k-1) + \frac{2N(k+1)}{k} = 2V_{D^{bg}_{equ}}. \tag{31}$$

Yet if the value of $V_{D^{bg}_{l}} + V_{D^{bg}_{h}}$ is larger than $2V_{D^{bg}_{equ}}$, then for every macro-value V^{bg}_{M} the value of $V_{D^{bg}_{h}} + V^{bg}_{M}$ will be higher than $2V_{D^{bg}_{equ}}$. This implies that condition (ii) of the COT cannot be satisfied. In general, this example illustrates that the Average Equivalence Conditions do not imply the Cancelling Out Conditions because the COT requires a very specific macro-state structure (where pairs of macro-values, averaged over, "cancel out" to yield the Boltzmannian equilibrium value).

5 When the Mechanical Averaging Equation Fails

The conditions in the last section are not just consolations for philosophers. They can fail. We now consider three examples in which MAE fails. The first example, the Baker's gas (Subsection 5.1), is a toy model that we discuss for its simplicity and intuitive appeal. The other two examples, the six-vertex model (Subsection 5.1) and the Ising model (Subsection 5.2), occupy centre stage in SM, and show that one cannot take MAP for granted.

5.1 The Baker's Gas With the Evenness Macro-Variable

Recall the baker's gas as introduced in Subsubsection 4.3.1. Assume we are interested in how homogeneous the particles are distributed, both in terms of location and momentum, and we introduce a variable e for 'evenness'. The variable measures to what extent the gas molecules depart from an even distribution, with $e = 0$ indicating that the distribution is perfectly even and with e assuming higher values for more pronounced inhomogeneities. To define e, consider again the distributions $D^{bg} = (N_1, N_2, \ldots, N_k)$, where N_i is the number of particles in cell ω^{bg}_{i}. A distribution is perfectly

even if all N_i assume the same values; the wider the spread of the values of the N_i the more uneven the distribution. One can now define a partition on X^{bg} by grouping together all points that have the same distribution. Let X^{bg}_u be the subset of all points of X^{bg} which correspond to the uniform distribution $N_i = N/k$ for all $i = 1, ..., k$.[38] One can now define the variable e as follows: $e(x^{bg}) = 10^5 \times \sqrt{(N_1 - N/k)^2 + \ldots + (N_k - N/k)^2}$, where $(N_1, N_2, \ldots, N_k)$ is the distribution corresponding to x^{bg}. It follows that $e(x^{bg}) = 0$ for $x^{bg} \in X^{bg}_u$. For a distribution where only one particle is displaced we have $e(x^{bg}) = 10^5\sqrt{2}$, and for distribution with more that one displaced particles the value of $e(x^{bg})$ is even higher. Therefore $e(x^{bg}) \geq 10^5\sqrt{2}$ for all states that do not have an even distribution.

In the *Boltzmannian treatment* of the baker's gas the model has one macro-variable, which is e: $v_1 = e$. The elements of the partition of X^{bg} are then the macro-regions of e that we have seen in the previous paragraph. The number $G(D^{bg})$ of arrangements that lead to the same distribution D^{bg} is $G(D^{bg}) = N! \,/\, N_1!N_2! \ldots, N_k!$ and it is therefore clear that X^{bg}_u is larger than any other element of the partition. The baker's gas is ergodic (Lavis 2005), and hence spends more time in X_u than in any other cell. For this reason, the long run fraction of time for which the value of e is 0 is larger than the long run fraction for any other value. Hence the macro-state defined by $e = 0$ is a γ-ε-equilibrium, and $e = 0$ is the Boltzmannian equilibrium value.[39]

In the *Gibbsian treatment* of the baker's gas the model has one variable f^{bg}, namely $f^{bg} = e$. Since the gas is isolated from the its environment its equilibrium distribution is the microcanonical distribution, i.e. the uniform distribution $\rho(x^{bg}) = 1$ for all $x^{bg} \in X$ is the stationary equilibrium distribution. The phase average $\langle e \rangle$ for the macro-variable e will be greater than (1-$\mu_{X^{bg}}(X^{bg}_u)) \times 10^5$. Lavis (2005, table on page 275) showed that $\mu_{X^{bg}}(X^{bg}_u) \leq 0.385$ (for large N), and hence $(1 - 0.385) \times 100.000\sqrt{2} = \sqrt{2} \times 62.500$ is a lower bound for $\langle e \rangle$.

So the Boltzmannian equilibrium value and the Gibbsian phase averages are very different! We note that this difference will not disappear in the thermodynamic limit when N goes to infinity: for any arbitrary large N the phase average will always be at least $\sqrt{2} * 62.500$, which is different from the Boltzmannian value 0.

Four comments are in order. First, the equilibrium macro-region of the baker's gas is small. The baker's gas shares this feature with other toy models that are standardly discussed in the literature, for instance the Kac ring (e.g. Lavis 2008). Some may want to dig in their heels and say that a small region cannot be an equilibrium region. This in effect amounts to rejecting the notion of γ-ε-equilibrium state altogether. As we noted in Subsection 2.2, this is a possible but revisionary move. We here proceed under the assumption that γ-ε-equilibria are genuine equilibria.

Second, the size of macro-regions depend on the macro-variables chosen, which raises the question of some macro-variables are more natural than others and whether 'unnatural' ones can be dismissed as irrelevant. A discussion of what counts as a 'natural' macro-variable would take us too far away from our main concerns and will have to be left for another occasion. However, we readily admit that the macro-variable e is somewhat contrived and so one might hope that the problem does not arise once such variable are ruled out. We will see in the next two sections that this hope is in vein: problems similar to those we have seen in this subsection crop up also in the case of perfectly 'natural' macro-variables like magnetisation. The grain of truth in this remark is that much depends on

[38] We assume that $N = k \times r$ for some $r \in \mathbb{N}$.

[39] Lavis (2005) showed that for large N the equilibrium macro-region takes up less than half of the state space, and therefore $e = 0$ is not an α-ε-equilibrium.

the choice of the macro-variable and, as we have seen in Section 4, for sufficiently restrictive classes of macro-variables the problem can indeed be avoided. But there are serious physical models that do *not* fall into these classes. Thus, the relevant contrast is not between 'mathematical contrivance' and 'natural physics'.

Third, a referee urged us to emphasise that just because MAE fails for evenness macro-variable fo the baker's gas, this does not imply that GSM cannot make sense of this example. Consider the fluctuation interpretation of GSM as discussed in Section 2.3. Since the baker's gas is ergodic, the masking condition applies and hence the baker's gas can be interpreted according to the fluctuations interpretation of GSM. Under that interpretation, fluctuations appear within equilibrium and the uniform distribution of the baker's gas provides the probabilities for fluctuations away from $\langle e \rangle$ to occur.

Finally, let us explain why for the baker's gas with the evenness macro-variable the Khinchin condition, the AET and the COT do not apply. First, note that for the baker's gas significant chunks of state space are taken up by non-equilibrium states whose macro-values differ considerably from the Gibbsian phase average. Hence the Khinchin-condition is not satisfied. Second, the evenness macro-variable for the baker's gas is not the sum of a one-component macro-variable whose outcomes have equal probability. So the first condition of the AET is not satisfied and hence the AET does not apply. Third, for the evenness macro-variable of the baker's gas the macro-values different from the equilibrium value are all higher than the equilibrium macro-variable, implying that (ii) of the COT fails and hence COT does not apply.

5.2 The Six-Vertex Model

5.2.1 Internal Energy Macro-Variable

Consider again the six vertex model as introduced in Subsection 4.4.1. Let us now study the *internal energy* as defined in Equation (19) as the relevant macro-variable. In the Boltzmannian treatment, the full state space (which is at the same time the effective state space) consists of all possible states κ that satisfy the ice rule. The lowest energy value is $E = 0$, which defines a macro-state M_0 with the associated macro-region $\bar{X}_{M_0} = \{\kappa^*, \kappa^+\}$, where κ^* is the state where all vertices are in the first configuration, and κ^+ is the state where all vertices are in the second configuration. Recall that N (the size of the system) is an arbitrarily large but finite number. Then, for sufficiently low temperatures T the probability mass is concentrated on the two lowest energy states, and therefore $\bar{X}_{M_0}$ is the largest macro-region. Since the dynamics is given by an irreducible Markov model and the outcome distribution is stationary, the model spends more time in the largest macro-region, i.e. M_0, than in any other macro-region provided that the temperature is sufficiently low. For this reason M_0 is a Boltzmannian γ-0-equilibrium, and $E = 0$ is the Boltzmannian equilibrium value.

In the Gibbsian treatment $p(\kappa)$ is the stationary equilibrium measure and the variable f is the internal energy. By definition the internal energy assumes its lowest value $E = 0$ only for two specific micro-states, namely κ^* and κ^+, and will assume higher values for all other micro-states. Furthermore, for any $T > 0$ there will be a non-zero probability that the model is in a state of higher energy. Therefore, the phase average $\langle E \rangle$ is greater than 0 and hence higher than the Boltzmannian equilibrium value.

To see that this difference can be significant and hence that this is a case where MAP fails, choose

a T such that $\{\kappa^*, \kappa^+\}$ is the largest macro-region while its probability is less than 0.5.[40] In this case the Boltzmannian equilibrium value is still $E = 0$. The second lowest macro-value is $E = \sqrt{N}$, which is the energy of micro-states where all columns, except one, are taken up by states with the first or the second configuration, and the states in the exceptional row are all states of the third or fourth configuration.[41] For this reason $\langle E \rangle$ is higher than $\sqrt{N}/2$, and therefore the Gibbsian phase average and the Boltzmannian equilibrium value will differ by at least $\sqrt{N}/2$, which is not at all a negligible difference, in particular for large N. Note that this argument holds for any arbitrary large N. It is important that the Boltzmannian macro-value that is closest to the value obtained from Gibbsian phase averaging is higher or equal to $\sqrt{N}$; and the Boltzmannian macro-value of higher or equal to $\sqrt{N}$ is different from the Boltzmannian macro-value in equilibrium which is 0. This underscores that the Gibbsian phase average is different from the Boltzmannian equilibrium value.

Finally, note that the macro-variable of the internal energy considered above is *extensive*, i.e. it depends on the number of constituents of the system. Extensive variables are standardly used in statistical mechanics (e.g. Baxter 1982; Lavis and Bell 1999). It is interesting to point out what happens if the corresponding intensive macro-variable – the energy density (the internal energy divided by N) – is considered instead (*intensive* macro-variables do not depend on the number of constituents). One readily finds that for the intensive variable the difference between the Gibbsian and Boltzmannian equilibrium calculations tends toward zero as $N \to \infty$ because $\sqrt{N}/2N \to 0$. This illustrates the point that whether or not the Gibbsian phase average agrees with the Boltzmannian equilibrium value depends on the macro-variable. We also expect that the lesson of this example generalises and that it is usually easier to achieve agreement for intensive than extensive variables. Yet the problems we are discussing cannot simply be dismissed by exorcising extensive variables and only working with intensive variables instead because, at least to many authors, extensive variables simply are very important. Indeed, one of the best currently available rationalisations of thermodynamics construes a system's state space as consisting only of extensional variables (Lieb and Yngvason 1999). Furthermore, as we will see in the next subsection, in some cases differences between Gibbsian and Boltzmannian appear both for intensive and extensive variables. In general, we want to leave it open what kind of macro-variables one wants to consider as this will depend on the context.

In our case one first fixes N and then chooses a sufficiently low T. One might wonder what happens if one first chooses the temperature and then takes the limit for N toward infinity. In this case the situation changes. More specifically, if the temperature is below the critical temperature, then the Boltzmannian equilibrium value and the Gibbsian phase average agree (the equilibrium values is then $(0, 0)$) (Lavis and Bell 1989, 307). However, we think that what we do in the example above (i.e. first fixing N and then choosing a sufficiently low T) is physically realistic insofar as real systems are finite and then for sufficiently low temperatures the disagreement between Gibbsian phase averages and Boltzmannian equilibrium values arises. The equality between Gibbsian phase averages and Boltzmannian equilibrium values for the internal energy only arises when N is taken to infinity after fixing T, something one can never do with real – finite! – systems. In any case, in all the examples that follow the difference between Gibbsian phase averages and Boltzmannian equilibrium values will persist even if first T is fixed and then N is taken to infinity.

[40] Such a choice is possible because the higher the temperature, the more uniform the probability distribution becomes. Hence for sufficiently high temperature values, the largest macro-region will be different from $\{\kappa^*, \kappa^+\}$ Since the canonical distribution is continuous, there has to be a T such that $\{\kappa^*, \kappa^+\}$ is the largest equilibrium macro-region while its probability is less than 0.5.

[41] These states mark the smallest possible departure from states with zero energy: it can be shown that the number of downward pointing arrows is the same for all rows, and from this follows that there has to be a perturbation in each row and that $\sqrt{N}$ is the second lowest value of the internal energy (Lavis and Bell 1999, Chapter 10).

5.2.2 Polarisation Macro-Variable

Consider now again the polarisation macro-variables $\vec{\pi}(\kappa)$ as defined in equation (22). As above, we suppose that the dynamics is given by an irreducible Markov model. For sufficiently low values of T it is then the case that the value of $\vec{\pi}$ will be flipping back and forth between $(1,1)$ and $(-1,-1)$, with other values occurring only for very brief intermittent periods. In the long run the system spends the same fraction of time in both polarisation states, with flips becoming less frequent as the model gets larger (cf. Levis 2012).

In BSM this situation is conceptualised as the system having a dual equilibrium, with $M_{(1,1)}$ and $M_{(-1,-1)}$ being the two equilibrium states. Recall that N is an arbitrarily large finite number. Then, as above, for sufficiently low temperatures T the probability mass is concentrated on micro-states with minimal energy. It is obvious that the micro-states with extremal polarisation are also the micro-states with minimal energy and so the union of the macro-regions of $M_{(1,1)}$ and $M_{(-1,-1)}$, $\bar{X}_{M_{(1,1)}} \cup \bar{X}_{M_{(-1,-1)}}$, is in fact identical with the macro-region of $\{\kappa^*, \kappa^+\}$ above. Therefore, for these sufficiently low values of T, the model spends more time in that macro-region than in any other macro-region. Hence $\bar{X}_{M_{(1,1)}} \cup \bar{X}_{M_{(-1,-1)}}$ is a dual γ-ε-equilibrium. Baxter (1982, 151) express that this is a dual equilibrium by saying that '[...] either all arrows point up and to the right or down and to the left. Thus at low temperatures the system is ferroelectrically ordered'; and Lavis and Bell (1999, 307), commenting on vertical polarisation, say that 'the equilibrium state is thus one of perfect long-range order with either $p = 1$ or $p = -1$'.

The result of Gibbsian phase averaging is $\langle \vec{\pi} \rangle = (0,0)$. This is because to any micro-state κ there corresponds a micro-state κ^r that results from κ by a 180° rotation of all arrows. One easily finds $(\pi_v(\kappa), \pi_h(\kappa)) = (-\pi_v(\kappa^r), -\pi_h(\kappa^r))$, and the probabilities of κ and κ^r are the same. The Gibbsian phase average $(0,0)$ is different from the Boltzmannian equilibrium values of $(1,1)$ or $(-1,-1)$, and the differences remain as N goes to infinity. The polarisation is an intensive macro-variable. Nothing changes in substance if the corresponding extensive macro-variable $(N-2n, N-2m)$ is considered. For the sufficiently low values of T one finds a disagreement between Gibbsian phase averaging and the Boltzmannian equilibrium values also in the extensive case because here the Gibbsian phase average is $(0,0)$ while there is a dual Boltzmannian equilibrium with equilibrium values (N,N) and $(-N,-N)$. Hence MEA fails, which shows that MAP is wrong. As in the case of internal energy, this argument applies for any arbitrary large N. As presented here, one first fixes N and then choses T. One might wonder what happens if we first choose T and then take limit $N \to \infty$. The result is that if the temperature is below the critical temperature, then the two Boltzmannian equilibrium states are again $(1,1)$ and $(-1,-1)$ (Lavis and Bell 1989, 307), and the Gibbsian phase average is zero. Hence the differences remain.

One might have the idea to remedy this situation by describing this as a case where there are two Gibbsian equilibria. However, while for the Boltzmannian concept a generalisation to two equilibria is straightforward (as we have seen in section 2.2), it is not clear how the idea of a stationary ensemble should be generalised to make sense of a dual equilibrium.[42] We note that the generalisation in the Boltzmannian case does not presuppose that we know in general how to calculate equilibrium

[42] One idea would be that a dual equilibrium corresponds to two stationary distributions. Yet, then there is the problem that there are always several stationary distributions, and GSM lacks a criterion for determining whether there is one equilibrium or two equilibria. Another idea would be to say that there are two equilibrium states if there are exactly two maximum entropy stationary distributions, or if there are exactly two stationary distributions where every state has a positive probability. But neither of these work in the current case because the six-vertex model has just one stationary maximum entropy distribution and just one stationary distribution where every state has a positive probability.

values in the Boltzmannian framework (in particular, for systems with a broken symmetry). The generalisation relies only the existence of equilibrium values, not on our ability to calculate them (for discussion of this point see Frigg and Werndl (2019)). Indeed, whether Boltzmannian equilibrium values can actually be calculated will depend on the mathematical knowledge of the example in question.

The conclusion that the Gibbsian phase average of the polarisation is $(0, 0)$ can be avoided by appealing to limits. Consider a system that arises by adding a small positive (or negative) electric field E to the six-vertex model. Then two limits are taken at a certain temperature T: first the limit for N to infinity and then the limit for E to zero (from the right had side for the positive electric field and the left hand side for the negative electric field). The idea here is to find out how the system behaves when the electric field is turned off. One considers two different limits for $E > 0$ and $E < 0$ because one expects a discontinuity at $E = 0$ where the right-hand side and left-hand side limits are different. That is, for a positive electric field one considers

$$P_+ := \lim_{E \to 0, E>0} \lim_{N \to \infty} \langle \vec{\pi}(\kappa) \rangle, \tag{32}$$

and for negative electric field one has

$$P_- := \lim_{E \to 0, E<0} \lim_{N \to \infty} \langle \vec{\pi}(\kappa) \rangle. \tag{33}$$

It turns out that the two phase average limits P_+ and P_- are no longer $(0, 0)$. In fact, for a T smaller than the critical temperature, they are in approximate agreement with the two Boltzmannian equilibrium values. Hence it is sometimes stipulated that the results of Gibbsian phase averaging at zero external field are P_+ and P_- (e.g. Baxter 1982, 153). Thus, by considering these limits one can avoid the conclusion that the Gibbsian phase average of the polarisation is $(0, 0)$ and that there is disagreement between the Gibbsian phase average and the Boltzmannian equilibrium value.

While this procedure has a certain formal elegance, is common in physics, and leads, at least approximately, to the desired values, the addition of the external field in the first place seems unmotivated from the point of GSM. There is nothing in the Gibbsian framework that would suggest, let along prescribe, that the equilibrium values of systems without an external field have to be calculated as the limit of systems with an external field. This procedure is *ad hoc* and its main justification is that it produces a result that is in line with the Boltzmannian values. The *ad-hocness* of the procedure is further underscored by the fact that the order of the limits is crucial: if one first takes limit $E \to 0$ and only then the limit $N \to 0$, one again obtains $(0, 0)$ as an equilibrium value. But there is nothing in the physics of the situation that would favour any particular order in which the limits have to be taken. Furthermore, the conclusion of a zero phase average can only be avoided when the limit system as $N \to \infty$ is considered; for any arbitrary finite N the phase average will always be zero. This is problematic because in reality systems are finite and hence that the calculations should work for large but finite N. In sum, calculating the polarisation at zero external field by adding an external field and then taking the limit seems ad hoc: it is done just to obtain the correct results but an independent physical motivation is missing

A referee suggested that we discuss the case in which the Boltzmannian measure is the uniform measure. The reason to do so is that many undergraduate texts tell students that the Boltzmannian equilibrium state is the macro-state that can be realised in the largest number of different ways. For models with discrete state spaces (such as the six-vertex or later the Ising model), this can be interpreted as the prescription to simply count the number of micro-states that give rise to a

certain macro-state, which in effect amounts to assuming that the system has a uniform measure. It is certainly interesting to study the six-vertex model (or indeed the Ising model) with the uniform measure and it is also interesting to explore whether the uniform measure is more natural than other measures in the Boltzmannian context. However, a lack of space prevents us from getting into a full discussion of this case. Such a discussion would take us rather far away from the cases we have discussed so far because the uniform measure is not invariant under the usual Markovian dynamics of the six-vertex model discussed in the literature on the six vertex model, and, similarly, the uniform measure is not invariant either for the usual dynamics of the Ising model discussed in the literature on the Ising model (details will follow in the next subsection). Hence the uniform measure violates the basic assumption that the models considered must have an invariant measure. Intuitively, this is so because the dynamics of the six-vertex model (or the Ising model) for low temperatures is such that the lowest energy states are assigned the bulk of the probability distribution. Hence, the uniform measure gets distorted under the dynamics and is not invariant. If one wants to study the six-vertex model (or the Ising model) with the uniform measure as invariant measure, one has to choose a different dynamics. Yet if the model has a different dynamics, then this clearly is a different system than the one studied in the standard literature on the six vertex model (and the Ising model). Since we want to concentrate on the dynamics that is standardly studied in the literature, we pass over the case of the uniform measure as invariant measure in this paper.

Another referee urged us to stress that just because MAE fails for polarisation of the six vertex model, this does not imply that GSM cannot make sense of this example. A Gibbsian account of this example can be given in the fluctuations interpretation of GSM as discussed in Section 2.3. Since the dynamics oft the six vertex model is an irreducible Markov model, the masking condition applies. Hence the six vertex model can be interpreted according to the fluctuations interpretation of GSM. Under that interpretation, fluctuations appear within equilibrium and the canonical distribution provides the probabilities for fluctuations away from $\langle p \rangle$ to occur. This treatment also explains why that phase average does not coincide with the Boltzmannian equilibrium value.

Finally, let us briefly explain why the examples just discussed are not instances of the Khinchin condition, the AET and the COT. First, for the six-vertex model with the macro-variable of the internal energy the non-equilibrium values that differ considerably from the Gibbsian phase average are not negligible (for the sufficiently low temperature values discussed). Hence for these examples the Khinchin-condition is not satisfied. For the six-vertex model with the polarisation macro-variable the Khinchin condition cannot be satisfied because most of the state space (namely the regions corresponding to the two equilibrium states) is taken up by macro-values that are very different from the phase average. Second, the internal energy for the six-vertex model is not the sum of a one-component macro-variable whose outcomes have equal probability. So it does not satisfy the first condition in the AET. Also, while the polarisation macro-variable in the six-vertex model is a sum of one-component macro-variables, and hence satisfies the first condition of the AET, the probability on the full state space is not the product measure of the measure on the one-component space (indeed there is no measure on the one-component space in terms of which the measure of the state space is defined). Hence the second and third conditions of the AET are not satisfied and the AET does not apply. Third, for the internal energy of the six-vertex model the non-equilibrium macro-values are all higher than the equilibrium macro-value, implying again that (ii) of the COT fails. And for the polarisation macro-variable in the six-vertex model there are two Boltzmannian equilibrium states and hence the structure of the macro-states is not as required by the COT, implying that COT does not apply.

5.3 The Ising Model

Recall the Ising model as introduced in Subsection 4.4.2.The behaviour of the internal energy macro-variable has already been discussed in another paper (Werndl and Frigg 2017). There are differences between the Boltzmannian equilibrium value and the Gibbsian phase average, but they are rather small. More specifically, if first a certain N is fixed and then a sufficiently low temperature is chosen, the differences are larger than $4J$, which, if J is sufficiently large, seems significant. Yet, the differences for the internal energy and polarisation of the six-vertex model are larger and hence, in our opinion, while the internal energy of the Ising model is also an interesting example, the internal energy and polarisation of the six-vertex model provide a clearer example for how Boltzmannian equilibrium values and Gibbsian phase averages can come apart. A macro-variable that has not been previously discussed in Werndl and Frigg (2017a) for the Ising model is the magnetisation.

5.3.1 Magnetisation Macro-Variable

Consider the magnetisation as the relevant macro-variable:

$$m(\sigma) = \sum_i -s\sigma_i, \tag{34}$$

where s is a constant. Supposing again an irreducible Markov model operating at a sufficiently low temperature, the value of m will be flipping back and forth between its extremal values sN and $-sN$ and other values are assumed only for very brief intermittent periods. In the long run the model will spend the same fraction of time in both macro-states, with flips getting less frequent the larger the system becomes.

Let us start with BSM. The relevant macro-variable is $v = m$, and the behaviour of the system is of the kind described at the end of Subsection 2.2 where the model has two equilibrium states. Enter and van Hemmen (1984, 258) express that this is a dual equilibrium state by writing that 'below T_c the model has two equilibrium states, a (+) state with positive magnetisation and a (-) state with negative magnetisation'. Similarly, Baxter (1982, 20) talks about two equilibrium states when he says that '[a]s T approaches T_c from below, the two equilibrium states become the same' (see also Cassandro et al. 1973, 153; Gonsalves 2007; Sekular Unpublished, 2).

Recall that N is an arbitrarily large finite number. As we have seen above, for sufficiently low temperature values the probabilities are centred on the two lowest energy states. These are in fact the states in which all spins point in the same direction, i.e. the states with macro-value sN and $-sN$. Thus, for sufficiently low temperatures (and under the assumption that the dynamics is an irreducible Markov model), the model spends most of it time in the union of the two macro-regions M_{sN} and M_{-sN}, $\bar{X}_{M_{sN}} \cup \bar{X}_{M_{-sN}}$. For this reason the system has a dual γ-ε-equilibrium with M_{sN} and M_{-sN} as its equilibrium states (with $\varepsilon = 0$). In this way BSM is able to offer a coherent and empirically adequate description of the situation.

GSM associates equilibrium with the stationary distribution given in equation (25). The relevant variable is $f = m$, and the phase average is $\langle m \rangle$. It now turns out that $\langle m \rangle = 0$ because for any state $\sigma = (\sigma_1, \ldots, \sigma_n)$ we can define its conjugate state $\sigma^* = (-\sigma_1, \ldots, -\sigma_n)$, and it is then obvious that $m(\sigma) = -m(\sigma^*)$ and that both σ and σ^* have the same probability. For this reason the Gibbsian phase average is 0. So the Gibbsian phase average is different from the two Boltzmannian equilibrium values sN and $-sN$. Hence this is a case where MEA does not hold. Again, it is unclear how to remedy this situation by describing this as a case where there are two Gibbsian equilibria

because, as discussed above, it is unclear how the idea of a stationary ensemble should be generalised to makes sense of a dual Gibbsian equilibrium.

Also as above, the differences between the Gibbsian and Boltzmannian calculations will not disappear for large N. For any arbitrary large N, the value obtained from Gibbsian phase averaging will be 0. Yet, for the same N (and for sufficiently high temperatures) the values of the magnetisation in equilibrium will be either sN or $-sN$. Hence our argument applies for any arbitrary large N one starts with and hence, in a sense, in the infinite limit. In our argument first N is fixed and then T is chosen. One might again wonder what happens when first T is chosen and then the limit as N goes to infinity is considered. What one obtains here is as follows. In the Boltzmannian framework for a nonzero temperature below the critical temperature there are again two equilibrium states, but with a value of a magnetisation that is smaller than sN and $-sN$ but clearly different from zero. The Gibbsian phase average remains zero (for zero temperature, the Boltzmannian equilibrium state corresponds to $\{\sigma', \hat{\sigma}\}$ and the Gibbsian phase average is again zero) (cf. Baxter 1982, Chapter 7). Hence the differences between the Boltzmannian equilibrium value and the Gibbsian phase average remain. Note that the magnetisation is an extensive variable and that, for the sufficiently low temperature values discussed, the difference between the Boltzmannian equilibrium value and the Gibbsian phase average will persist if the corresponding intensive variable of the magnetisation per site $m(\sigma)/N$ is considered: for the magnetisation per site at sufficiently low temperatures there are again two Boltzmannian equilibria corresponding to the macro-values s and $-s$, but the Gibbsian phase average is always zero.

As in the case of the polarisation of the six-vertex model, there is a trick to avoid the conclusion that the Gibbsian phase average of the magnetisation is zero: add a small external positive (or negative) magnetic field L to system (cf. equation (19) and the text below for an explanation of the role of L) and then consider the right-hand side and left-hand side limits as L goes to zero (e.g. Baxter 1982, 118). Again, these two limits are no longer zero and are in in approximate agreement with the two Boltzmannian equilibrium values. For this reason it is sometimes stipulated that these are the values obtained in the Gibbsian framework. However, as in the case of the polarisation of the six-vertex model, we think that this procedure is ad-hoc and unmotivated from the point of GSM: it is done just to obtain the correct results but an independent physical motivation is missing.

As above, referees urged us to stress that just because MAE fails for magnetisation of the Ising model, this does not imply that GSM cannot make sense of this example. For instance, consider again the fluctuations account of GSM as discussed in Section 2.3. Because the dynamics oft the Ising model is an irreducible Markov model, the masking condition applies. Hence the Ising model can be interpreted according to the fluctuations interpretation of GSM, and the canonical distribution provides the probabilities for fluctuations away from $\langle m \rangle$ to occur.

The six-vertex model and the Ising model are often discussed in the context of phase transitions. Hence the question might arise whether there is any relation between phase transitions and the failure of MAE. The answer is no: MAE can fail away from phase transition points, at phase transition points, and also when the system shows no phase transition at all. An example where MAE fails but there is no phase transition is the baker's gas with the evenness macro-variable (Subsubsection 5.1). The six-vertex model with the internal energy macro-variable is an example of a system in which the averaging principle fails away from a temperature where a phase transition occurs (Subsubsection 5.2.1).[43] The Ising model with varying external magnetic field provides an example of

[43]For this examples there is a range of temperature values where MAE fails and hence the failure of MAE does not only happen at a phase transition point of the temperature – cf. footnote 5.2.1

system in which MAE fails at a phase transition point. That is, consider the Ising model in an external magnetic field L. Given a *fixed sufficiently low temperature*, one can then show that if one varies the external field (from negative values through zero to positive values), there is a phase transition at $L = 0$ when the external field goes from negative to positive values. As we have seen, when $L = 0$ MAE fails, and so this is an example for the failure of MAE at a phase transition point.[44]

Let us briefly comment on why the Khinchin condition, the AET and the COT do not apply to the examples just discussed. First, for the Ising model with the magnetisation macro-variable the Khinchin condition cannot be satisfied because most of the state space (namely the regions corresponding to the two equilibrium states) is taken up by macro-values that are very different from the phase average. Second, while the magnetisation macro-variable of the Ising model is a sum of one-component macro-variables, and hence satisfies the first condition of the AET, the probability on the full state space is not the product measure of the measure on the one-component space (indeed there is no measure on the one-component space in terms of which the measure of the state space is defined). Hence the second and third conditions of the AET are not satisfied and the AET does not apply. Third, for the magnetisation macro-variable in the Ising model there are two Boltzmannian equilibrium states and hence the structure of the macro-states is not as required by the COT, implying that COT does not apply.

5.4 Existence Under Different Conditions

In the last three subsections we have seen examples of models in which Boltzmannian equilibrium values and Gibbsian phase averages are markedly different. And things get even more interesting: Boltzmannian and Gibbsian equilibria can exist under different conditions. While it is true that the existence of a Boltzmannian equilibrium implies the existence of a Gibbsian equilibrium, the reverse implication fails: there are models that have a Gibbsian equilibrium but fail to have a Boltzmannian equilibrium.[45]

To see how the existence of a Gibbsian equilibrium fails to imply the existence of a Boltzmannian equilibrium consider the Ising model with the magnetisation macro-variable and the six-vertex model with the polarisation macro-variable. As we have seen, these models oscillate back and forth between two macro-states, spending the same fraction of time in each of the two macro-states in the long run. This implies that there is no single Boltzmannian equilibrium.[46] As noted above, this problem cannot be resolved by introducing something like a 'dual Gibbsian equilibrium'. Furthermore, one can construct models with a Gibbsian equilibrium that have neither a single nor dual Boltzmannian equilibrium (this is the case, for instance, if the system spends the same amount of time in each macro-state of a set of macro-states with more than two elements). In sum, the existence of a Gibbsian equilibrium has no bearing on the existence of a Boltzmannian equilibrium (of any type).

Conversely, however, the existence of a Boltzmannian equilibrium (either single or dual) implies the existence of a Gibbsian equilibrium. As we have seen in Section 2, all models in SM have a stationary measure. A fortiori every model with a Boltzmannian equilibrium has a stationary measure. Recall

[44]Note that if one instead considers the magnetisation of the Ising model for a *fixed* positive or negative external magnetic field when the *temperature is varied*, then there is no phase transition.

[45]In Werndl and Frigg's (2017) it is argued that the implication fails in both directions, i.e. that there are also models that have a Boltzmannian but not a Gibbsian equilibrium. We note that this argument employs a different notion of equilibrium than the one used here and so there is no contradiction between these claims.

[46]There is nothing special about the Ising and the six-vertex models. The Baker's gas with a mass-imbalance macro variable ($v = 1$ if there are more particles in the left half of the container than in the right half, and $v = 0$ otherwise) exhibits the same behaviour.

that in GSM the relevant notion of equilibrium is statistical equilibrium, which simply corresponds to a stationary measure. Hence, trivially, the existence of a Boltzmannian equilibrium implies also the existence of a Gibbsian equilibrium.

6 Conclusion

We have identified three conditions for MAP. These conditions are individually sufficient, but not necessary, to guarantee that Boltzmannian equilibrium values and Gibbsian phase averages agree: the (well-known) Khinchin condition, the conditions given by the recently-proven Average Equivalence Theorem and the conditions given by the new Cancelling Out Theorem. Since these conditions are sufficient but not necessary, there could (and probably will) be other conditions that make MAP true. Uncovering such conditions is a challenge for future research.

As we have seen in the previous section, these conditions are by no means always satisfied and there are cases in which Gibbsian phase averages and Boltzmannian equilibrium values come apart. The cases in which this happens include core models of statistical mechanics such as the six-vertex model and the Ising model, and hence such cases cannot be dismissed as irrelevant mathematical contrivances. We have also seen that the Boltzmannian and Gibbsian equilibria need not even exist under the same conditions.

This raises the important question of which of the two approaches (if any) is correct when then they disagree. This is a complex matter that raises many difficult issues, and a conclusive discussion is a task for a future project. At this point we would to like to articulate the tentative proposal that in cases in which Gibbsian phase averages and Boltzmannian equilibrium values come apart, the Boltzmannian values are correct in sense that $F_B = F_T$ (which means that BEP is true). There are two reasons for this. First, the examples in the previous section show that in cases in which Gibbsian phase averages and Boltzmannian equilibrium disagree, there are good reasons not use the phase averages. For example, when observing the polarisation on a physical system that is accurately represented by the six vertex model, the measurements do not return the phase average (0, 0) as a measurement outcome, because most of the time one measures the Boltzmaniian equilibrium values, i.e. either $(1, 1)$ or $(-1, -1)$. Similarly, for the magnetisation, observations on a physical system that is accurately represented by the Ising model will not return the phase average $(0, 0)$ as a measurement result; instead one measures the Boltzmannian equilibrium values, i.e. either $-sN$ or sN (cf. Lavis and Bell 1999, 299). Second, in recent paper Frigg and Werndl (2019) argue that BSM is a fundamental theory while GSM is an effective theory. If this is correct (and we find their arguments convincing) then it follows that BSM provides the correct results.

Irrespective of how this question is resolved, we hope that this paper has shed some light on the relation between BSM and GSM, and that the results presented here will be useful in future investigations of the relation between the two theories.

Acknowledgements

We would like to thank Wayne Myrvold for helpful discussions and constructive comments on earlier versions oft the paper, as well as for shepherding the paper through a lengthy refereeing process. Thanks also to the anonymous referees for helpful comments. In the process of researching this paper we benefited from discussions with Jeffrey Barrett, Jeremy Butterfield, David Lavis, Stephan

Hartmann, Patricia Palacios, Jos Uffink and Giovanni Valente, and we want to thank them for sharing their insights with us.

Appendix: Proof of the Cancelling Out-Theorem

The proof is stated for the deterministic case (it is obvious how it carries over to the stochastic case). Relabel the macro-state in such a way that the Boltzmannian equilibrium macro-state is M_{equ} and the other macro-states are $M_{k,l}$ ($l = 1$ or 2 and $1 \leq k \leq q/2$) such that (i) $\mu_Z(Z_{M_{k,1}}) = \mu_Z(Z_{M_{k,2}})$ and (ii) $V_{M_{k,1}} + V_{M_{k,2}} = 2V_{M_{equ}}$ for all k. By assumption, the Boltzmannian equilibrium value is $V_{M_{equ}}$. The Gibbsian phase average is given by

$$\sum_{k=1}^{q/2} \mu_Z(Z_{M_{k,1}})V_{M_{k,1}} + \mu_Z(Z_{M_{k,2}})V_{M_{k,2}} + \mu_Z(M_{equ})V_{M_{equ}}. \tag{35}$$

This equals (applying (i)):

$$\sum_{k=1}^{q/2} \mu_Z(Z_{M_{k,1}})[V_{M_{k,1}} + V_{M_{k,2}}] + \mu_Z(M_{equ})V_{M_{equ}}. \tag{36}$$

And this equals (applying (ii)):

$$\sum_{k=1}^{q/2} 2\mu_Z(Z_{M_{k,1}})V_{M_{equ}} + \mu_Z(M_{equ})V_{M_{equ}}. \tag{37}$$

Applying (i) once again then yields:

$$\sum_{k=1}^{q/2} [\mu_Z(Z_{M_{k,1}}) + \mu_Z(Z_{M_{k,2}})]V_{M_{equ}} + \mu_Z(M_{equ})V_{M_{equ}} = V_{M_{equ}}. \tag{38}$$

References

Adler, M. (2016). *Monte-Carlo-Simulation of the Ising Model*. Anchor Academic Publishing.

Allison, D., and Reshetikin N. (2005). Numerical Study of the Six-Vertex Model with Domain-Wall Boundary Conditions. *Annales De L'Institut Fourier* 55, 1847-1869.

Badino, M. (2006). The Foundational Role of Ergodic Theory. *Foundations of Science* 11, 323-347.

Barkema, E. D. and Newman, M. E. J. (1998). Monte Carlo Simulation of Ice Models. *Physical Review* E 57, 1155.

Batterman, R. (1998). Why Equilibrium Statistical Mechanics Works: Universality and the Renormalization Group. *Philosophy of Science* 65, 183-208.

Baxter, R. J. (1982). *Exactly Solved Models in Statistical Mechanics*. Academic Press.

Berger, A. (2001). *Chaos and Chance: an Introduction to Stochastic Aspects of Dynamics*. Springer.

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.

Bricmont, J. (2001). Bayes, Boltzmann and Bohm: Probabilities in Physics. In *Chance in Physics: Foundations and Perspectives*, eds. J. Bricmont, D. Dürr, M. C. Galvotti, G. Ghirardi, F. Petruccione and N. Zanghì, 3-21. Springer.

Chandler D. (1987). *Introduction to Modern Statistical Mechanics*. Oxford University Press.

Cipra, A. (1987). An Introduction to the Ising Model. *American Mathematical Monthly* 94, 937-954.

Cassandro, M., Galavotti, G., Lebowitz, J. L. and Monroe J. L. (1973). Existence and Uniqueness of Equilibrium States for Some Spin and Continuum Systems. *Communication in Mathematical Physics* 32, 153-165.

Davey, K. (2009). What Is Gibbs' Canonical Distribution? *Philosophy of Science* 76, 970-983.

Doob, J. L. (1953). *Stochastic Processes*. John Wiley and Sons.

Ehrenfest, P. and Ehrenfest-Afanassjewa, T. (1959). *The Conceptual Foundations of the Statistical Approach in Mechanics*. Cornell University Press.

Fermi, E. (2000): *Thermodynamics*, Mineola/NY: Dover Publications.

Ferrenberg, A. M. and Landau, D. P. (1991). Critical Behavior of the Three-Dimensional Ising model: A High-Resolution Monte Carlo Study. *Physical Review B* 44, 5081.

Frigg, R. (2008). A Field Guide to Recent Work on the Foundations of Statistical Mechanics. In *The Ashgate Companion to Contemporary Philosophy of Physics*, ed. D. Rickles, 99-196. Ashgate.

Frigg, R. and Nguyen, J. (2017). Models and Representation. In *Springer Handbook of Model-Based Science*, eds. L. Magnani and T. Bertolotti, 49-102, Springer.

Frigg, R. and Werndl, C. (2011). Explaining Thermodynamic-Like Behaviour in Terms of Epsilon-Ergodicity. *Philosophy of Science* 78 (4), 628-652.

Frigg, R. and Werndl C. (2019). Statistical Mechanics: A Tale of Two Theories. *The Monist* 102, 424-438.

Frigg, R. and Werndl C. (forthcoming-a). Can Somebody Please Say What Gibbsian Statistical Mechanics Says?, *The British Journal for the Philosophy of Science*.

Frigg, Roman and Charlotte Werndl (forthcoming-b): Equilibrium in Gibbsian Statistical Mechanics, forthcoming in *Routledge Companion to Philosophy of Physics*, eds. E. Knox and A. Wilson. Routledge.

Gibbs, J. W. (1902). *Elementary Principles in Statistical Mechanics*. New York: Scribner's Sons.

Gonsalves, R. J. (2007). *Statistical Mechanics, Phase Transitions, and the Ising Model*. Lecturing notes at the University of Buffalo. Available at:
http://www.physics.buffalo.edu/phy411-506-2008/chapter8/index.html

Guggenheim, E. A. (1967). *Thermodynamics. An Advanced Treatment for Chemists and Physicists*. North-Holland.

Hill, T. L. (1987). *Statistical Mechanics: Principles and Selected Applications*. Dover.

Hill, T. L. (1986). *An Introduction to Statistical Thermodynamics*. Dover.

Isihara, A. (1971). *Statistical Physics*. Academic Press.

Jeans, J. H. (1916). *The Dynamical Theory of Gases*. Cambridge University Press.

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

Kittel, C. (1958). *Elementary Statistical Mechanics*. Dover.

Landau, L. D. and Lifshitz, E. M. (1980): *Statistical Physics*. Elsevier.

Lavis, D. (2005). Boltzmann and Gibbs: An Attempted Reconciliation. *Studies in History and Philosophy of Modern Physics* 36, 245-273.

Lavis, D. (2008). Boltzmann, Gibbs and the Concept of Equilibrium. *Philosophy of Science* 75, 682-696.

Lavis, D. and Bell, G.M. (1999). *Statistical Mechanics of Lattice Systems, Volume 1: Closed Form and Exact Solutions*. Ellis Horwood.

Lawden, D. F. (1987). *Principles of Thermodynamics and Statistical Mechanics*. Dover.

Levis, D. (2012). *Two-dimensional Spin-Ice and the Sixteen Vertex Model*. Ph.D. thesis, Université de Pierre et Marrie Currie.

Lieb, E. H. and Yngvason, J. (1999). The Physics and Mathematics of the Second Law of Thermodynamics. *Physics Reports* 310, 1-96.

Malament, D. B. and Zabell, S. L. (1980). Why Gibbs Phase Averages Work? The Role of Ergodic Theory. *Philosophy of Science* 56, 339-349.

Maroney, O. J. E. (2008). The Physical Basis of the Gibbs-von Neumann entropy. Unpublished Manuscript. Available on https://arxiv.org/abs/quant-ph/0701127

Meester, R. (2003). *A Natural Introduction to Probability Theory*. Springer.

Myrvold, W. C. (2016). Probabilities in Statistical Mechanics. In *The Oxford Handbook of Probability and Philosophy*, eds. A. Hajek and C. Hitchcock, 573-600. Oxford University Press.

Pathria, R. K. and Beale, P. D. (2011). *Statistical Mechanics*, 3rd ed. Elsevier.

Penrose, O. (1970). *Foundations of Statistical Mechanics: A Deductive Treatment*. Dover, 2005.

Reiss, H. (1996): *Methods of Thermodynamics*, Mineola, NY: Dover.

Ruelle, D. (1969): *Statistical Mechanics: Rigorous Results*, London: Imperial College Press.

Schrödinger, E. (1989). Statistical Thermodynamics. Dover Publications

Sekular, P. (Unpublished). Monte-Carlo simulation of small 2D Ising lattice with Metropolis dynamics. Available at: http://secular.me.uk/physics/ising-model.pdf

Sklar, L. (1973). Physics and Chance: Philosophical Issues In The Foundations Of Statistical Mechanics. Cambridge University Press.

Slater, J. C. (1941). Theory of the Transition in KH2PO4. *Journal of Chemical Physics* 9, 16-33.

Swendsen, R. H. (2012). An Introduction to Statistical Mechanics and Thermodynamics. Oxford University Press.

Syljuasen, O. F. and Zvonarev, M. B. (2004). Monte Carlo Simulations of Vertex Models. *Physical Review E* 70, 016118.

Szilard, L. (1925). On the extension of phenomenological thermodynamics to fluctuation phenomena. *Zeitschrift fur Physik* 32, 753-788.

Tolman, R. J. (1938). *Principles of Statistical Mechanics*. Dover.

Uffink, J. (2007). Compendium of the Foundations of Classical Statistical Physics. In *Philosophy of Physics*, eds. J. Butterfield and J. Earman, 923-1047. North Holland.

van Enter, A. C. D. and van Hemmen, J. L. (1984). Statistical-Mechanical Formalism for Spin Glass. *Physical Review* 29, 355-365.

van Lith, J. (2001). *Stir in Stillness: A Study in the Foundations of Equilibrium Statistical Mechanics*. PhD Thesis, University of Utrecht. Available at: http://www.library.uu.nl/digiarchief/dip/diss/1957294/inhoud.htm

Thompson, C. J. (1972). Mathematical Statistical Physics. Princeton University Press.

Venaille, A. and Bouchet, F. (2009). Statistical Ensemble Inequivalence and Bicritical Points for Two-Dimensional Flows and Geophysical Flows. *Physical Review Letters* 104501.

Vranas, P. B. (1998). Epsilon-ergodicity and the Success of Equilibrium Statistical Mechanics. *Philosophy of Science* 65, 688-708.

Wallace, D. (2015). The quantitative content of statistical mechanics. *Studies in History and Philosophy of Modern Physics* 52, 285-293.

Wallace, D. (Unpublished). What Statistical Mechanics Actually Does. PhilSci archive.

Werndl, C. (2013). Justifying Typicality Measures of Boltzmannian Statistical Mechanics and Dynamical Systems. *Studies in History and Philosophy of Modern Physics* 44 (4), 470-479.

Werndl, C. and Frigg, R. (2015a). Rethinking Boltzmannian Equilibrium. *Philosophy of Science* 82, 1224-1235.

Werndl, C. and Frigg, R. (2015b). Reconceptualising Equilibrium in Boltzmannian Statistical Mechanics and Characterising its Existence. *Studies in History and Philosophy of Modern Physics* 49, 19-31.

Werndl, C. and Frigg, R. (2017a). Mind the Gap: Boltzmannian versus Gibbsian Equilibrium. Philosophy of Science 84 (5), 1289-1302.

Werndl, C. and Frigg, R. (2017b). Stochastic Boltzmann Equilibrium. In *D üsseldorf 2015: The European Philosophy of Science Association Proceedings*, eds. M. Massimi and J.-W. van Romejin. Springer.

Werndl, C. and Frigg, R. (forthcoming-a). When Does a Boltzmannian Equilibrium Exist?. In *Quantum Foundations of Statistical Mechanics*, eds. D. Bedingham, O. Maroney and C. Timpson. Oxford University Press.