WHY ERGODIC THEORY DOES NOT EXPLAIN THE SUCCESS OF EQUILIBRIUM STATISTICAL MECHANICS

John Earman
Department of History and Philosophy of Science
University of Pittsburgh
Miklós Rédei
Center for Philosophy of Science
University of Pittsburgh†

The British Journal for the Philosophy of Science **47** (1996) 63-78

ABSTRACT

We argue that, contrary to some analyses in the philosophy of science literature, ergodic theory falls short in explaining the success of classical equilibrium statistical mechanics. Our claim is based on the observations that dynamical systems for which statistical mechanics works are most likely not ergodic, and that ergodicity is both too strong and too weak a condition for the required explanation: one needs only ergodic-like behavior for the finite set of observables that matter, but the behavior must ensure that the approach to equilibrium for these obsersvables is on the appropriate time scale.

1 Introduction

Ever since Boltzmann [1871] introduced the concept of ergodicity, its role in statistical mechanics has been controversial.[1] The controversy erupts sporadically in the philosophical literature, e.g. Sklar [1973], Friedman [1976], Lavis [1977], Quay [1977], Malament and Zabell [1980], Railton [1981], Leeds [1989], Clark [1987], Butterfield [1987] and Batterman [1990, 1992]. The survey of the current situation given in Sklar's [1993] *Physics and Chance* leaves the impression that the relevance of ergodic theory to explaining the success of equilibrium statistical mechanics is still open to debate. On the contrary, we feel that Sklar's (otherwise) laudable book contains all of the information needed to show that ergodic theory is not explanatorily relevant. In preparation for giving our argument for this claim, we briefly review some of the relevant aspects of ergodic theory in section 2 and then summarize the debate on the explanatory relevance in sections 3 and 5. Our argument is contained mainly in section 4. Sections 6 and 7 provide some concluding remarks.

2 Basic notions and results of ergodic theory

The setting for modern ergodic theory is a dynamical system (X, ϕ_t, μ). X, the *state space* or *phase space*, is a topological space which in most intended applications is compact and metrizable. The *flow* $\phi_t \colon X \to X$, $(t \in \mathbb{R})$, is a one parameter family of homomorphisms with the group properties $\phi_0 = \mathrm{id}$, $\phi_{t_1} \circ \phi_{t_2} = \phi_{t_1+t_2}$, and $\phi_{-t} = \phi_t^{-1}$.[2] μ is a normed measure on X and is invariant under the flow, i.e. for any measurable set $A \subseteq X$, $\mu(\phi_t(A)) = \mu(A)$ for all t. There are many equivalent ways to characterize ergodicity. Perhaps the closest to Boltzmann's original intentions is given in

Def. 1 The dynamical system (X, ϕ_t, μ) is *ergodic* iff for any measurable set $A \subseteq X$ such that $\mu(A) \neq 0$ and for almost every (a.e.) $x \in X$ it holds that $\{\phi_t(x)\} \cap A \neq \emptyset$ for some t.[3]

Somewhat more useful for certain purposes is an equivalent definition:

Def. 2 The dynamical system (X, ϕ_t, μ) is said to be *decomposable* iff X can be partitioned into two (or more) invariant regions of non-zero measure, i.e. there are $A, B \subset X$ such that $A \cap B = \emptyset$, $A \cup B = X$, $\mu(A) \neq 0 \neq \mu(B)$, and $\phi_t(A) \subseteq A$ and $\phi_t(B) \subseteq B$ for all t. A dynamical system is said to be *ergodic* iff it is indecomposable.

The central result in ergodic theory follows from a theorem in functional analysis due to Birkhoff [1931]. Before stating the result and the theorem, we need to define two kinds of averages for phase functions $f \colon X \to \mathbb{R}$. The *phase average* $\langle f \rangle$ of f is given by

$$\langle f \rangle = \int_X f(x) \mathrm{d}\mu(x) \tag{1}$$

The *time average* $f^*(x)$ of $f(x)$ is given by

$$f^*(x) = \lim_{T \to \infty} \frac{1}{T} \int_{t_0}^{t_0+T} f(\phi_t(x)) \mathrm{d}t \tag{2}$$

Theorem 1 (Birkhoff). Let (X, ϕ_t, μ) be a dynamical system, and let f be an integrable function on X. Then the time average $f^*(x)$ exists for a.e. $x \in X$, is integrable, and is independent of the initial time t_0.

The hard work goes into establishing Theorem 1. It is then comparatively easy to establish

Theorem 2 (Ergodic Theorem). Let (X, ϕ_t, μ) be a dynamical system and let f be an integrable phase function. Then $f^*(x) = \langle f \rangle$ for a.e. $x \in X$ iff the system is ergodic.

The equality of time averages and phase averages has traditionally been the starting point for those who want to claim that ergodic theory explains why phase space averages work (see section 3).

Three corollaries of the Ergodic Theorem are worth mentioning. The first connects probability in the sense of μ-measure to probability in the sense of the relative time the system spends in a given phase space region.

Cor. 1 The dynamical system (X, ϕ_t, μ) is ergodic iff for a.e. $x \in X$, the limit of the relative time that the orbit $\phi_t(x)$ spends in a measurable set $A \subseteq X$ is $\mu(A)$.

If the dynamical system is discrete, then Corollary 1 says that $\mu(A)$ can be viewed as the limit of the relative *frequency* of how many times the phase point visits the set A. Thus Corollary 1 is a "continuum version" of the relative frequency interpretation of probability.

The second corollary demonstrates the uniqueness of an ergodic measure relative to a continuity requirement. Recall that μ' is said to be *absolutely continuous* (a.c.) with respect to μ iff for any measurable $A \subseteq X$ the condition $\mu(A) = 0$ implies $\mu'(A) = 0$.

Cor. 2 Suppose that the system (X, ϕ_t, μ) is ergodic. Let μ' be a ϕ_t-invariant measure that is absolutely continuous with respect to μ. Then $\mu' = \mu$.

The third corollary, an easy consequence of Cor. 2, locates the ergodic measure μ in the convex set S of all ϕ_t-invariant states. Before stating the corollary recall that an element μ in S is said to be *extremal* in S if it can not be written as a non-trivial convex sum of two distinct elements in S i.e. μ is extremal if $\mu \neq \lambda\mu_1 + (1-\lambda)\mu_2$ with $0 < \lambda < 1$ and $\mu_1, \mu_2 \in S$ such that $\mu_1 \neq \mu_2$.

Cor. 3 If the system (X, ϕ_t, μ) is ergodic then μ is *extremal* in the set S of all ϕ_t-invariant states.

The application of Cor. 2 to Hamiltonian dynamical systems is the focus of Malament and Zabell's [1980] defense of the explanatory relevance of ergodic theory. For a system with n degrees of freedom, take $X = \mathbb{R}^{2n}$ with coordinates $q_1, q_2, \ldots, q_n$ (generalized position coordinates), and $p_1, p_2, \ldots, p_n$ (generalized momenta). The dynamics is generated by the specification of the *Hamiltonian* $H(q,p)$. *Hamiltonian's equations* are

$$\dot{q_i} = \frac{\mathrm{d}q_i}{\mathrm{d}t} = \frac{\partial H}{\partial p_i} \tag{3}$$

$$\dot{p_i} = \frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial H}{\partial q_i} \tag{4}$$

The solutions to these equations define a flow ϕ_t on $\mathbb{R}^{2n}$. If H does not depend on t, then $H(q,p) = E$ =constant. So the motion is confined to a $2n-1$ dimensional subspace S_E (the energy surface). Assume that S_E is a smooth submanifold of $\mathbb{R}^{2n}$ and that it is closed and bounded. Then there is a ϕ_t-invariant measure μ on S_E such that $\mu(S_E) = 1$. This *microcanonical measure* (a.k.a. *microcanonical ensemble*) is given by

$$\mu(A) = \int_A \frac{\mathrm{d}x_{2n-1}}{\mid \mathrm{grad}H \mid_{H=E}} \tag{5}$$

where $\mathrm{d}x_{2n-1}$ denotes the $2n-1$ dimensional Lebesgue measure. Thus, μ and Lebesgue measure are mutually absolutely continuous. It follows from Cor. 2 that for an ergodic Hamiltonian system, μ is the only invariant measure on S_E a.c. with respect to Lebesgue measure on S_E.[4]

For future reference we introduce another ergodic property that is stronger than ergodicity.

Def. 3 (X, ϕ_t, μ) is *mixing* iff for any measurable $A, B \subseteq X$ it holds that

$$\lim_{|t|\to\infty} \mu(\phi_t(A) \cap B) = \mu(A)\mu(B)$$

The relevance of mixing to present concerns is due to

Theorem 3 Let (X, ϕ_t, μ) be a dynamical system and let ρ be a measure that is a.c. with respect to μ. Define $\rho_t(A) \equiv \rho(\phi_t(A))$ for measurable $A \subseteq X$. If the system is mixing, then $\rho_t \to \mu$ as $|t| \to \infty$ in the sense that

$$\lim_{|t|\to\infty} \int f \mathrm{d}\rho_t = \int f \mathrm{d}\mu$$

for any bounded measurable f.

For Hamiltonian dynamical systems it follows that any measure on S_E that is a.c. with respect to Lebesgue measure converges to the microcanonical measure.

3 The debate over the explanatory relevance of ergodic theory

Consider a statistical mechanical system, such as a box of gas. Choose a phase function corresponding to some macroscopically measurable quantity, and compute its phase average using the microcanonical measure. The value thus obtained reliably provides an accurate prediction of the actually measured value of the quantity when the system is in equilibrium. Why should this be so? The answer often given by the proponents of the ergodic approach goes something like this. Suppose that the system is ergodic. Then by the Ergodic Theorem the phase average of the function in question will equal its time average. The macroscopic measurement of the corresponding quantity will take some time, and, thus, the value reported will be a time average. The time interval involved may be short on a typical macroscopic scale but will be long on a microscale, and, consequently, the finite time average will closely approximate the infinite time average appearing in the Ergodic Theorem. Ergo, the measured value will be closely approximated by the phase average.[5] Two objections to this explanation have been repeatedly raised. First, it has been noted that the time taken to complete a macroscopic measurement may not be long on some appropriate microscale; after all,

macroscopic measurements are capable of revealing that statistical mechanical systems are not in equilibrium. Second, because of the "a.e." qualification in the Ergodic Theorem, the explanation shows at best why using phase averages works with "measure one." But why should μ-measure one be equated with probabilistic certainty in the physically relevant sense? Cor. 1 which guarantees that for an ergodic system μ-measure is interpretable as limiting relative "frequency" might be taken to provide the start of an answer. But it is an answer that is fraught with a version of the "single case problem" of the relative frequency interpretation of probability, now transmuted into a problem about the infinite time limit. Given that the infinite limit of the relative time spent by a phase orbit in a region $A \subseteq X$ is $\mu(A)$, how does this fact bear on the question of how probable it is to find the phase point of the system in A *now*?

To these complaints Sklar [1973] added two further ones. First, ergodicity cannot be a sufficient explanation for the success of using phase averages to predict equilibrium values. For systems with a small number N of particles can be ergodic – e.g. two hard spheres in a cubical box or one hard sphere in a stadium.[6] But for such systems there is not even a relevant notion of equilibrium. Large N must then enter the story somewhere. Second, ergodicity is not necessary to explain why phase space averages work, for (Sklar argues) there is a full and correct explanation that is independent of ergodicity.

> It goes like this: How a gas behaves over time depends on (1) its microscopic constitution; (2) the laws governing the interaction of its micro-constituents; (3) the constraints placed upon it; and (4) *the initial conditions characterizing the microstate of the gas at a given time* ... Clause (4) is crucial. It is the matter-of-fact distribution of such initial conditions among gas samples in the world which is responsible for many of the most important macroscopic features of the gas ... The actual distribution of initial states is such that calculations done by the Gibbs method [i.e. by using phase averages calculated from the microcanonical measure] ... 'works.' This is a matter of fact, not of law. The 'facts' explain the success of the Gibbs method. In a clear sense they are the only legitimate explanation of its success. (1973, p. 210)

Malament and Zabell [1980] agree that the fudge on finite and infinite time averages is unacceptable. But they argue that ergodicity does have an important though limited role to play within an explanatory scheme for why microcanonical phase averages work. It would seem that any candidate for a measure to represent equilibrium probabilities should be stationary or invariant.[7] The microcanonical measure passes that test. And, furthermore, we have seen that if the system is Hamiltonian and ergodic, then the microcanonical measure is the only invariant measure that is a.c. with respect to Lebesgue measure on the energy surface S_E. Malament and Zabell go on to motivate the condition of a.c. by showing that it is equivalent to a condition of translation continuity, which requires that for two measurable sets on S_E such that one is a small displacement of the other, "the probability of finding the exact microstate of the system in the one set should be close to that of finding it in the other" [1980, p. 346].[8] Thus it would seem that if physical probabilities obey the Malament-Zabell continuity requirement, ergodicity is the basis of a convincing justification for adopting the microcanonical measure to compute probabilities for the equilibrium state. This, however, is not the end of their story, for they agree with Sklar that, even with the addition of their continuity requirement, ergodicity is not sufficient. This is where large N enters. For systems with a large number of particles, theorems by Khinchin [1949] and Lanford [1973] show that if a phase function f corresponding to a macroscopic quantity satisfies some symmetry requirements, then the dispersion about the expected value $\langle f \rangle$ is small. Putting these various pieces together, we finally arrive at a satisfactory explanation of why it is overwhelmingly probable that using microcanonical phase averages works to predict equilibrium values for macroscopic quantities, or at least those quantities subject to the Khinchin-Lanford dispersion theorems. There has been no attempt to respond to Sklar's final and perhaps most telling criticism. We will return to it in section 6.

4 The explanatory irrelevance of ergodic theory

Two types of criticism can be brought to bear on the Malament-Zabell strategy. The first concedes that the strategy has a valid thrust but denies that the presuppositions of the strategy apply to systems of interest. The second denies the efficacy of the strategy even should its presuppositions hold. Leeds' [1989] criticism belong to the latter category; we will take it up in the following section. In this section we will consider a criticism of the first kind. The criticism is part of a general skepticism about the explanatory relevance of ergodic theory.

Our skepticism about the explanatory relevance of ergodic theory starts from the observation that typical systems treated in classical statistical mechanics are very likely to be not ergodic (see Wightman [1985]. As mentioned above, systems such as perfectly hard spheres in a cubical box and one perfectly hard sphere in a stadium have been proven to be ergodic.[9] But, of course, real molecules are not perfectly hard spheres. Geodesic motion on a manifold of negative curvature is also ergodic.[10] But while this case is relevant to the cosmological context (see Ellis and Tavakol [1994]

it is irrelevant to the bread-and-butter systems studied in classical statistical mechanics. In sum, the evidence for the applicability of ergodicity where it is required is non-existent. Furthermore, the evidence against the applicability is strong. The KAM Theorem leads one to expect that for systems where the interactions among the molecules are non-singular, the phase space will contain islands of stability where the flow is non-ergodic.[11] Thus all of the debate about how ergodicty would explain why phase averages work is purely academic since most the systems in question are not ergodic, or so all of the available evidence suggests. We now want to consider a series of responses that would attempt to salvage some explanatory role for ergodic theory.

Reaction 1 The proponents of the explanatory relevance of ergodic theory could still hope to show that some statistical mechanical systems are ergodic, and if this hope is fulfilled, they can tell their explanatory story for this restricted set of cases. But not only does this retreat take ergodic theorists off the main part of the playing field, it also takes them onto unfirm ground. If the typical system in classical statistical mechanics is non-ergodic, and yet using phase averages calculated from the microcanonical measure works, then the explanation of why it works will have to invoke non-ergodic mechanisms and properties. It is a reasonable hypothesis that these non-ergodic mechanisms and properties are responsible for the success of equilibrium statistical mechanics even in those cases where the system is ergodic.

Reaction 2 If retreat turns to rout, perhaps attack is the better strategy. The proponent of ergodic theory could try to show that ergodicity explains the success of equilibrium statistical mechanics in the so-called thermodynamic limit where $N \to \infty$ and $V \to \infty$ while N/V stays finite. One idea would be to show that in this limit the relative volume of the phase space in which the flow is non-ergodic approaches zero. The evidence regarding this idea is mixed. Numerical simulations on simple model systems have confirmed that the relative volume occupied by the invariant KAM tori (on which the flow is non-ergodic) decreases as the number of degrees of freedom increases. However, the transition to effective gross ergodicity has not been confirmed (see, for example, Hurd et al. [1994]. Another idea would be to show that while the dynamical (X^V, ϕ_t^V, μ^V) system describing a physical system confined to the finite region V might not be ergodic, there is a dynamical system $(X^\infty, \phi_t^\infty, \mu^\infty)$ describing the physical system in thermodynamic limit, which means in particular that $\mu^\infty = \lim_{V\to\infty} \mu^V$ in some appropriate sense of the limit, and $(X^\infty, \phi_t^\infty, \mu^\infty)$ is already ergodic. The problem with this second defense is that infinite systems can exhibit the phenomenon of phase transition: an infinte system can possess more than one equilibrium state, i.e. the set S of ϕ_t^∞-invariant states can contain more than one time invariant measure. On the other hand, by ergodicity one can explain, at best, one single state, the μ^∞ measure only, which is one of the extremal states in S (Cor. 3). Thus whenever equilibrium statistical mechanics proves in some way or another the existence of more than one single equilibrium state, and to the extent these different equilibrium states are viewed as physically relevant, ergodicity proves to be just insufficient to even account for them, much less to explain why/how equilibrium statistical mechanics works. A common problem of both proposals is that the relevance of the ideal thermodynamical limit to explaining the behavior of actual systems where N and V are finite is far from apparent – regardless whether the relative volume of non-ergodic regions does vanish in the thermodynamic limit, or whether the infinite system is ergodic. The fudge between large N and V on one hand and infinite N and V on the other seems just as suspect as the original fudge between finite and infinite time averages.

Reaction 3 Sklar [1993], taking the part of the ergodic theorist he so trenchantly criticized twenty years previously, tries to put the best face on matters.

> There is good reason to think, however, that large finite systems will have small regions of stable trajectories [as suggested by KAM], and that the overwhelmingly largest part of the available phase space, once more with sizes measured in the standard way, will be at least ergodic-like. (p. 175)

But what is the relevance of the fact – if indeed it is a fact – that for typical large finite systems, the flow on the overwhelmingly largest part of the phase space is ergodic, at least with "largest" as judged by the microcanonical μ? The only plausible story we have so far of the explanatory relevance of ergodic theory is that it is the key ingredient in the result that singles out the microcanonical μ as the unique measure having certain desirable properties. But even if the system is just a "little bit" non-ergodic, the uniqueness result fails – and it fails entirely, not just a little bit. Hence, from the point of view of the problem of explanatory role of ergodicity, there is no middle ground between ergodicity and non-ergodicity. Of course, one could secure the conclusion that it is very likely that typical large finite systems will exhibit ergodic behavior by adding the postulate that these systems are likely to be found in states belonging to regions of phase space with large μ measure. But this postulate is just what ergodic theory was supposed to justify.

Reaction 4 It might be that for large N and V, the macroscopic observables we care about are "insensitive to the non-ergodic portion of the flow even if its relative phase volume does not go to zero" (Wightman [1985], p. 20). What

would such an insensitivity mean? According to Def. 2, the failure of ergodicity means that X can be partitioned into two (or more) invariant subspaces A and B of non-zero measure. Suppose that the flow is ergodic on A but not on B. Then, as measured in the microcanonical μ, the insensitivity of the macro-observable O to B would mean that $\langle f_O\rangle \approx \langle f_O\rangle|_A \equiv \int_A f_O \mathrm{d}\mu$, where f_O is the phase function that represents O. The trouble is that under the present assumptions there are lots of other measures besides μ that are ϕ_t-invariant and are a.c. with respect to Lebesgue measure. Since the flow is assumed to be ergodic on A, the reduced measure $\mu_A(\bullet) = \mu(\bullet \cap A)/\mu(A)$ is the unique normed, invariant, and a.c. measure concentrated on A. Let μ'_B be any normed, invariant, and a.c. measure concentrated on B. Then $\mu_\epsilon = \epsilon\mu_A + (1-\epsilon)\mu'_B$ $(0 \leq \epsilon \leq 1)$ is a normed, invariant, and a.c. measure on X. Computing the phase average of f_O using μ_ϵ gives

$$\langle f_O\rangle_\epsilon = \frac{\epsilon}{\mu(A)}\langle f_O\rangle|_A + (1-\epsilon)\langle f_O\rangle'_B \tag{6}$$

By choosing the value of ϵ appropriately, one can make $\langle f_O\rangle_\epsilon$ as insentive to either the ergodic or the non-ergodic portion of the flow as desired. Of course, one expects that when ϵ is close to 0 (and thus $\langle f_O\rangle_\epsilon$ is insensitive to the ergodic portion of the flow) $\langle f_O\rangle_\epsilon$ will give a poor prediction of the measured equilibrium value of O. But ergodic theory does not explain why this is so, or least the explanation does not follow any of the lines explored so far. Neverteless, Wightman's suggestion of focusing on a restricted set of macroscopic observables is a valuable one. We will return to it in section 6 where we will give our own twist to it.

Reaction 5 Boltzmann wrote:

> The great irregularity of thermal motion and the manifold forces affecting bodies from the outside make it probable that the atoms of the warm body, through the motion we call heat, run through all the positions and velocities compatible with the equation of kinetic energy, so that we can use the equations developed above [the equality of phase measure of a region and the average relative time spent by the phase point in the region] ... (1871, p. 284)[12]

A plausible construction to put on this passage goes as follows. Considered as closed systems, typical systems of statistical mechanics are non-ergodic. But actual physical systems are not closed. And the effect of the perturbation of outside forces may be to break up the stable tori of KAM and to make the system ergodic in the sense that phase trajectories will pass arbitrarily close to any given point of the phase space (pace Def. 1). The trouble, however, with moving to open systems is that the underpinnings of ergodic theory are kicked out. A closed Hamiltonian system is at least a dynamical system. But if the system is subject to perturbations from outside, there is no reason to think that the definition of a dynamical system is satisfied. *A fortiori*, ergodicity doesn't even make sense. Of course, the open system in question may be a subsystem of a larger system that is closed and that does satisfy the definition of a dynamical system. But once again, is there any reason to think that this larger system is ergodic? We now start back at the beginning of the cycle that lead to this juncture. The present reaction contains, however, the germ of a potentially valuable idea that will be developed in section 6.

5 Leeds' criticism of the Malament-Zabell strategy

The complex argumentation of Leeds' [1989] defies capsule summary. But the nub of his concern about the Malament-Zabell strategy of bypassing limit theorems and seeing the key role of ergodicity in the proof of uniqueness of the equilibrium measure focuses on their assumption that a system in equilibrium should be characterized by a measure that is stationary in the technical sense of being invariant under the flow. When one asks for an explanation of the equilibrium behavior of, say, a cup of coffee to which cream has been added, what one is referring to are properties exhibited by the system after some characteristic macroscopic time when the coffee and cream have stopped sloshing around and have "settled down". Why should an equilibrium system in this sense be described, exactly or to some good approximation, by a stationary measure in the technical sense?

Leeds himself suggests an answer that does not have the form proposed by Malament and Zabell but does appeal to the ergodic hierarchy. As seen in section 2 above, the mixing property guarantees that if the measure that characterizes the pre-equilibrium situation is a.c. with respect to Lebesgue measure on S_E, then it will converge to a stationary measure and, indeed, to the microcanonical measure. Now neither mixing nor any other property in the standard ergodic hierarchy guarantees that the rate convergence occurs on a time scale on which we actually observe the system of interest to settle down to "equilibrium." So ergodic properties are not sufficient to explain why equilibrium statistical mechanic works. But on the present construal they are playing an important role in the explanation.

But now comes our complaint again. Most of the systems we are interested in are very probably not even ergodic much less mixing. There are various ways to try to salvage an explanatory role for mixing, but we feel that they will run into analogues of the difficulties already discussed in the preceeding section.

6 Why then does equilibrium statistical mechanics work?

Let (S_E, ϕ_t, μ) be a Hamiltonian system with S_E an energy surface and μ the microcanonical measure on S_E. Say that the system (S_E, ϕ_t, μ) is *mixing with respect to the finite set of observables* $\{O_1, O_2, \ldots, O_n\}$ iff for any measure ρ a.c. with respect to μ,

$$\lim_{|t|\to\infty} \int f_{O_i} \mathrm{d}\rho_t = \int f_{O_i} \mathrm{d}\mu \qquad (i = 1, 2, \ldots n) \tag{7}$$

where $\rho_t(A) \equiv \rho(\phi_t(A))$. In fact, we can only measure a small handful of macroscopic observables. To explain why equilibrium statistical mechanics works in predicting values of these observables it is not necessary to appeal to full ergodicity or mixing but only to finite mixing with respect to those observables that matter. Our suggestion is that if one wants to preserve some explanatory role for ergodic theory, one should investigate whether or not typical systems studied in classical statistical mechanics do have the finite mixing property with respect to those observables that matter. This is, of course, much more demanding than proving one clean mathematical theorem, especially since the relevant set of finite observables may differ from system to system. But who said that life had to be simple? Alas, even if the systems of interest do have the finite mixing property with respect to the relevant set of observables, that is not enough to explain why equilibrium statistical mechanics works for these systems. It also needs to be shown that (a) the convergence in (7) is sufficiently rapid and (b) the dispersion about the microcanonical average of f_{O_i} is small for large N. As for (a), however, it is surely the case that the convergence in (7) is *not* rapid for every choice of initial distribution ρ: there will always be mathematically possible ρ's such that for finite t the expectation value $\int f_{O_i} \mathrm{d}\rho_t$ of some observable O_i we care about will exhibit antithermodynamical behavior, and this is so even if the system is fully ergodic or mixing.[13]. It thus seems, as Sklar has urged (see section 3), that the explanation we seek cannot avoid reference to the matter-of-fact initial conditions for the systems we observe. But these conditions may be such that the equilibrium behavior we seek will emerge whether or not the system is ergodic or mixing.

A different line of investigation harkens back to Reaction 5 of section 4. Although a dead end from the point of view of ergodic theory, it promises to open a different route to explaining the success of equilibrium statistical mechanics. The leading idea starts from the recognition that the actual statistical mechanical systems we deal with are open in the sense of being subject to perturbations from the outside. The form of the perturbation can either be postulated or possibly derived if the system of interest is a subsystem of a larger system that is itself a closed dynamical system.[14] In either case the hope is that the perturbation will act as a kind of "stirring" mechanism which will rapidly drive the observed values of macroscopic quantities for the systems of interest to those predicted by the standard method and that this will be so regardless of whether the dynamics of the system, considered in isolation from the perturbations from without, is ergodic. It is even possible that the stirring mechanism is so effective that assumptions about the matter-of-fact initial distribution of states can be avoided.

There is no general agreement about the prospects of the above approaches and of others that have been explored in the literature. But it seems fair to say that ergodic theory in its traditional form is unlikely to play more than a cameo role in whatever the final explanation of the success of equilibrium statistical mechanics turns out to be.

7 Conclusion

Our discussion has been narrowly focused on some technical issues in the foundations of statistical mechanics. It would be wrong to think, however, that the topic at issue does not have any implications for more general issues in the philosophy of science. To give one example, Batterman [1992] uses classical statistical mechanics to argue that there are important modes of statistical explanation that fit neither Hempel's [1965] Inductive-Statistical model nor Railton's [1978, 1981] Deductive-Nomological-Probabilistic model. Roughly the idea is that a nearly pervasive patterns of behavior can be explained by showing that with measure one the systems in question will display this type of behavior. The explanations Batterman surveys appeal to ergodicity and also to stronger properties higher up in the ergodic hierarchy. (In strictly increasing strength, one has ergodic systems, weakly mixing systems, mixing systems, K-systems, and Bernoulli systems.[15]) In the concluding section of his paper, Batterman adds a final footnote, which reads:

> There is in fact a theorem, known as the KAM theorem, which says roughly that for most systems there will exist regions of stability in the phase space such that states initially within these regions will remain there as time goes on. In such cases the system cannot be ergodic, in which case we lose the straightforward justification for the microcanonical distribution. [1992, fn. 14]

We would add that not only is the straightforward justification of the microcanonical measure lost, but so is any obvious justification. What shape explanations in statistical mechanics will take once this loss is fully digested remains to be seen. This is a matter of some importance not only for the foundations of statistical mechanics but for the philosophy of scientific methodology.

Notes

† We are grateful to David Malament and Larry Sklar for helpful comments on an earlier draft of this paper. M. Rédei thanks for the financial support provided by the Center for Philosophy of Science, by the Fulbright and Széchenyi Foundations and by OTKA (contract numbers T 015606 and T 013853).

1. For historical accounts of Boltzmann's use of ergodicity, see Brush [1976] and von Plato [1991, 1994].

2. In some applications X will be a differentiable manifold. ϕ_t: $X \to X$ will then be required to be a diffeomorphism.

3. "Almost every $x \in X$" means for all $x \in X$ except for a set of measure 0.

4. It is not true, as claimed in Batterman [1990, p. 401], that (given ergodicity) the microcanonical measure is the only invariant measure on S_E.

5. Here is a slightly different twist given by Lebowitz and Penrose: "The physical importance of ergodicity is that it can be used to justify the use of the microcanonical ensemble for calculating equilibrium values and fluctuations. Suppose f is some macroscopic observable and the system is started at time zero from a dynamical state x, for which $f(x)$ has a value that is very far from its equilibrium value. As time proceeds, we expect that the current value of f, which is $f[\phi_t(x)]$, will approach and mostly stay very close to an equilibrium value with only very rare large fluctuations away from this value. This equilibrium value should therefore be equal to the time average because the initial period during which equilibrium is established constitutes only negligibly to the formula defining $f^*(x)$. The [ergodic] theorem tells us that this equilibrium value is almost always equal to $\langle f \rangle$, the average value of f in the microcanonical ensemble, provided the system is ergodic." [1973, p. 25]

6. A stadium shape is composed of two half circles joined by straight line segments.

7. But see the discussion in section 5 where this assumption comes into question.

8. In fact they do not show this. The theorem they present and prove in the Appendix of their paper states that if a μ measure on $\mathbb{R}^n$ is continuous with respect to displacement then μ is a.c. with respect to the Lebesgue measure on $\mathbb{R}^n$. But this is not what one needs. As Leeds [1989] already points out, what is needed is a similar theorem on the constant energy surface, and the needed theorem is non-trivial, since continuity with respect to displacement does not even make (global) sense: displacements take the regions off the energy surface. Subsequently, Malament and Zabell (private communication) have shown how to extend their original theorem from $\mathbb{R}^n$ to an arbitrary n-dimensional manifold.

9. Actually the full proof of the ergodicity of perfectly hard spheres in a box has never been published; see Wightman's [1985] remark.

10. Indeed, it is fully chaotic in the sense of being Bernoulli; see below.

11. For an account of the K(olmogorov)A(rnould)M(oser) Theorem, see Lichtenberg and Lieberman [1983].

12. Translation from von Plato [1991, p. 77].

13. We are indebted to Larry Sklar for making us appreciate this point.

14. See Lanford [1979] for a review of various attempts to implement these approaches.

15. For an account of the ergodic hierarchy, see Wightman [1985] or Arnold and Avez [1968].

References

Arnold, V. and Avez A. [1968]: *Ergodic Problems of Classical Mechanics.* New York, Benjamin.

Batterman, R.W. [1990]: 'Irreversibility and Statistical Mechanics: A New Approach?', *Philosophy of Science*, **57**, pp. 395-419.

Batterman, R.W. [1992]: 'Explanatory Stability', *Noûs*, **26**, pp. 325-348.

Birkhoff, G.D. [1931]: 'Proof of the Ergodic Theorem', *Proceedings of the National Academy of Sciences*, **17**, pp. 656-660.

Boltzmann, L. [1871]: 'Einige allgemeine Sätze über Wärmegleichgewicht', in *Wissentschaftliche Abhandlungen*, Vol. 1, pp. 259-287.

Brush, S. [1976]: *The Kind of Motion We Call Heat.* 2 Vols. Amsterdam, North Holland.

Butterfield, J. [1987]: 'Probability and Disturbing Measurement', *Proceedings of the Aristotle Society, Suppl. Volume LXI.*, pp. 211-243.

Clark, P. [1987]: 'Determinism and Probability in Physics' *Proceedings of the Aristotle Society, Suppl. Volume LXI.*, pp. 185-210.

Ellis, G. and Tavakol, R. [1994]: 'Mixing Properties of Compact $K = -1$ FLRW Models', in Hobhill et al. (eds), *Deterministic Chaos in General Relativity*, New York, Plenum Press pp. 237-250.

Friedman, K.S. [1976]: 'A Partial Vindication of Ergodic Theory', *Philosophy of Science*, **43**, pp. 151-162.

Hempel, C.G. [1965]: 'Inductive-Statistical Explantion', in *Aspects of Scientific Explanation*, New York, Free Press, pp. 381ff.

Hurd, L., Grebogi, C. and Ott, E. [1994]: 'On the Tendency Towards Ergodicity with Increasing Number of Degrees of Freedom in Hamiltonian Systems', in J. Seimens (ed), *Hamiltonian Mechanics: Integrability and Chaotic Behavior*, New York, Plenum Press, pp. 128-138.

Khinchin, A.I. [1949]: *Mathematical Foundations of Statistical Mechanics*, New York, Dover.

Lavis, D. [1977]: 'The Role of Statistical Mechanics in Classical Physics', *British Journal for the Philosophy of Science*, **28**, pp. 255-279.

Lanford, O. [1973]: 'Entropy and Equilibrium States in Classical Statistical Mechanics', in A. Lenard (ed), *Statistical Mechanics and mathematical Problems*, Berlin, Springer Verlag, pp. 1-113..

Lebowitz, L. and Penrose, O. [1973]: 'Modern Ergodic Theory', *Physics Today*, **26**, (no. 2), pp. 23-29.

Leeds, S. [1989] 'Malament and Zabell on Gibbs Phase Averaging', *Philosophy of Science*, **56**, pp. 325-340.

Lichtenberg, A. J. and Lieberman, M. A. [1983]: *Regular and Stochastic Motion*, Berlin, Springer Verlag.

Malament, D. and Zabell, S.L. [1980]: 'Why Gibbs Phase Space Averages Work – The Role of Ergodic Theory', *Philosophy of Science*, **47**, pp. 339-349.

Quay, P. [1978]: 'A Philosophical Explanation of the Explanatory Function of Ergodic Theory', *Philosophy of Science*, **45**, pp. 47-59.

Railton, P. [1978]: 'A Deductive-Nomological Model of Probabilistic Explanation', *Philosophy of Science*, **45**, pp. 206-226.

Railton, P. [1981]: 'Probability, Explanation, and Information', *Synthese*, **48**, pp. 233-256.

Sklar, L. [1973]: 'Statistical Explanation and Ergodic Theory', *Philosophy of Science*, **40**, pp. 194-212.

Sklar, L. [1993]: *Physics and Chance*, Cambridge, Cambridge University Press.

von Plato, J. [1991]: 'Boltzmann's Ergodic Hypothesis', *Archive for the History of Exact Sciences*, **42**, 71-89.

von Plato, J. [1994]: *Creating Modern Probability Theory*, Cambridge, Cambridge University Press.

Wightman, A.S. [1985]: 'Regular and Chaotic Motions in Dynamical Systems. Introduction to the Problems', in G. Velo and A.S. Wightman (eds), *Regular and Chaotic Motions in Dynamical Systems*, New York, Plenum Press, pp. 1-26.