What Is Gibbs's Canonical Distribution?

Kevin Davey†‡

Although the canonical distribution is one of the central tools of statistical mechanics, the reason for its effectiveness is poorly understood. This is due in part to the fact that there is no clear consensus on what it *means* to use the canonical distribution to describe a system in equilibrium with a heat bath. I examine some traditional views as to what sort of thing we should take the canonical distribution to represent. I argue that a less explored alternative, according to which the canonical distribution represents a time ensemble of sorts, has a number of advantages that rival interpretations lack.

1. Introduction. One striking thing about the machinery of modern statistical mechanics is how effective it is. It is able to explain and predict a great number of facts about systems in equilibrium and even some facts about systems out of equilibrium. More striking, however, is the fact that the *reason* for the effectiveness of this machinery is so poorly understood, despite so much effort by physicists, mathematicians, and philosophers on this very issue.

Let us focus for now on the use of the canonical distribution to calculate quantities associated with a system in equilibrium with a heat bath. What justifies this procedure? Answering this question requires carefully untangling several distinct issues. First, we must ask what it *means* to describe the state of a system using the canonical distribution. Armed with the answer to this question, we must argue that we are *right* to describe systems in equilibrium with heat baths using the canonical distribution. Finally, we must argue that when we do correctly describe a system using the canonical ensemble, we are *justified* in expecting the value of a macroscopic observable on the system to be that given by integrating the observable over phase space using the canonical measure.

Some small progress can be made toward justifying the use of the canonical ensemble by carefully distinguishing these questions and treat-

†To contact the author, please write to: Department of Philosophy, University of Chicago, 1115 E. 58th St., Chicago, IL 60637; e-mail: kjdavey@uchicago.edu.

‡Thanks to Brandon Fogel, Nick Huggett, and John Norton for their feedback and suggestions.

Philosophy of Science, 76 (December 2009) pp. 970–983. 0031-8248/2009/7605-0006$10.00
Copyright 2009 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/605793 Published online by Cambridge University Press

ing them one by one. In what follows, I provide a sketch of how to begin this task, focusing mainly on the question of what it means to describe the state of a system using the canonical distribution.

2. What Is the Canonical Ensemble?

2.1. Equilibrium. We want to ask what it means to describe a system in equilibrium with a heat bath using the canonical distribution. It will be useful to begin by asking what it means to be *in equilibrium* with a heat bath. (At this point, we take the concept of a heat bath as unproblematic.)

The concept of equilibrium has a disparate set of meanings, and we need to fix on one. Sometimes when we talk about a system being in equilibrium, we mean that the system is currently in a particular macrostate—typically, one of the macrostates described by the formalism of classical equilibrium thermodynamics. I call this a static notion of equilibrium, insofar as it is used to determine whether a system is in equilibrium at a given time. This is not the only way in which we can think of equilibrium. In fact, when we talk about a gas being in equilibrium with a heat bath, we generally do not have a static conception of equilibrium in mind. The gas will exchange energy with the heat bath in such a way that it will occasionally find itself out of static equilibrium. So the following rough definition will suffice for our purposes: a system is in equilibrium with a heat bath if, as it evolves over time, it spends most of its time in the macrostate (or macrostates) associated with static equilibrium, only occasionally fluctuating into nonequilibrium macrostates. This is a temporally extended notion of equilibrium, insofar as saying that a system is in equilibrium with a heat bath in this sense is now to make a claim about the behavior of the system over time.

Because heat baths are supposed to be infinite, it may well turn out that any finite system placed in contact with a heat bath will spend much of its life in states of static equilibrium, if considered over a sufficiently large timescale. Thus, it may well turn out that any finite system that remains in contact with a heat bath indefinitely will be in equilibrium with the heat bath in the sense just defined. This should cause no confusion, so long as it is understood that there are different concepts of equilibrium, and a system in one sort of equilibrium need not be in another type of equilibrium.

2.2. Traditional Interpretations. Let us return to the canonical distribution. According to Gibbs, the state of a system in equilibrium with a

https://doi.org/10.1086/605793 Published online by Cambridge University Press

heat bath at temperature T is described by the *canonical distribution*:

$$\rho(X) = \frac{e^{-\beta H(X)}}{\mathcal{Z}}, \tag{1}$$

where X is a point in phase space, H is the Hamiltonian, $\mathcal{Z}$ is the partition function, and $\beta = 1/kT$, where k is Boltzmann's constant. For simplicity, we suppose that the Hamiltonian is time independent and that our system is constrained to lie in some finite volume of physical space. But what does it *mean* to say of a system that it may be described by the canonical distribution (1)? I shall take it for granted that describing systems in equilibrium with heat baths using the canonical distribution involves making some sort of probabilistic or statistical claim about such systems. But precisely what sort of probabilistic or statistical claim could this be?

One approach is to think of the canonical distribution as giving us *epistemic* probabilities. In its simplest form, the claim here is that when we know a system is in equilibrium with a heat bath, it is rational to assign the degree of belief $\rho(X)|dX|$ to the proposition that the system is in a small region dX at time t, where ρ is the canonical distribution (1). But a claim like this is surely implausible. I might have a justified belief that the number of particles in a gas lies in a certain range (say, between 1.1×10^{23} and 1.2×10^{23} particles). In ordinary situations, for each integer in this range I will assign a very low probability to that integer being the *exact* number of particles in the system. The canonical distribution, however, will assign probability 1 to the claim that the gas has exactly N particles, where N is the actual number of particles described in the Hamiltonian. In this case, the canonical distribution gives a probability of 1 where my epistemic probability is almost 0. A more sophisticated proponent of the epistemic interpretation of statistical mechanics might have something to say about this sort of worry, although whether related worries arise is a question I am happy to leave open.[1] For the purpose of this article, I will focus instead on the idea that describing a system with the canonical distribution involves making an objective, nonepistemic probabilistic or statistical claim about the system in question.

More specifically, according to the approaches that I do want to consider, using the canonical distribution to describe a system in equilibrium

1. One might try to argue that our epistemic probabilities are given by *weighted sums* of canonical distributions, each involving a slightly different Hamiltonian, and that because the observables that matter have (roughly) the same expected value when calculated with this weighted sum as when calculated with the canonical distribution involving the *actual* Hamiltonian, we may take the canonical distribution to represent our epistemic state well enough for the purpose of calculating the expected macrostate of the system.

https://doi.org/10.1086/605793 Published online by Cambridge University Press

with a heat bath at temperature T involves making two observations: first, that the system is a member of a particular *ensemble* (or set of systems) and, second, that it is appropriate (in some sense to be specified) to describe this ensemble using the canonical distribution (1). To pursue this approach, we must first construct an ensemble of systems and then explain in what sense it is appropriate to describe this ensemble with the canonical distribution (1).

There are a few ways to carve up possible ways of proceeding. One thing we might do is try to construct our ensemble out of actually occurring systems with the given Hamiltonian H in equilibrium with a heat bath at a given temperature T. We call this an *actualist* approach.

> **Actualist Approach**. The ensemble should consist of some appropriately chosen set of actual (i.e., physically instantiated) systems with the Hamiltonian H in equilibrium with a heat bath at temperature T.

Given an ensemble constructed in accordance with this requirement, we might then claim that the statistics of the ensemble are (approximately) given by the canonical distribution. More precisely, we might claim that the proportion of systems in the ensemble occupying a fixed small region dX of phase space is (approximately) equal to $\rho(X)|dX|$, where ρ is the canonical distribution. We call this an *(approximate) identity* requirement.

> **(Approximate) Identity Requirement**. The proportion of systems in the ensemble that lie in a fixed small region dX of phase space is (approximately) equal to $\rho(X)|dX|$, where ρ is the canonical distribution.

But the combination of an actualist approach with an approximate identity requirement is not promising at all. Presumably, there are systems that have found themselves in contact with heat baths only a small number of times in the history of the universe. Only a tiny region of their phase space will ever have been instantiated, and perhaps only a tiny region of their phase space will ever be instantiated. For such a system, the approximate identity requirement will typically not hold of any ensemble constructed using the actualist approach.

In response to this, let us try to keep the actualist approach but relax the approximate identity requirement. To motivate our replacement for the approximate identity requirement, consider the case of an ordinary coin that we can toss as much as we like. Take the following hypothesis.

> *Hyp*: If the coin were to be thrown continually, it would come up heads about half the time and tails about half the time.

Suppose now that the coin has been thrown 1,000 times and that it comes

https://doi.org/10.1086/605793 Published online by Cambridge University Press

up heads two times and tails 998 times. In such a case, we would be inclined to say that *Hyp* has been strongly disconfirmed. (By *strong disconfirmation* I mean a type of disconfirmation sufficiently severe to render belief in the given hypothesis to be unjustified.) If the coin came up heads 400 times and tails 600 times, we might say that the hypothesis has been disconfirmed but not strongly so. If the coin came up heads 480 times and tails 520 times, we would probably say that *Hyp* has not been disconfirmed at all. Exactly how statistical confirmation or disconfirmation works will not be important to us. The important point is that claims about the distribution of an ensemble can sometimes be strongly disconfirmed by observation of a sample.

Just as one might have some sort of procedure for tossing coins, or discovering tossed coins, one might also have some sort of procedure for creating or discovering systems with a prespecified Hamiltonian H in contact with a heat bath at temperature T. In general, we call such a procedure a *sampling procedure*. Note that a possible sampling procedure for tossed coins might be to take note of *all* coins in the universe that have just been tossed and list them sequentially. (I am supposing that the set of tossed coins at any moment is finite.) Thus, we do not require that a sampling procedure involves the rejection of otherwise eligible data points, nor do we require that it be a procedure that any human being could actually perform. Note that when a sampling procedure is applied over time, it generates an ensemble. At any moment of time, the ensemble generated thus far is assumed to be finite, even though the ensemble generated over all time may well be infinite. Suppose then that we have a sampling procedure for systems with Hamiltonian H in contact with a heat bath at temperature T. Consider now the hypothesis.

> *Hyp*: The ensemble generated over all time by this sampling procedure has the property that the proportion of systems in the ensemble in a fixed small region dX of phase space is (approximately) equal to $\rho(X)|dX|$, where ρ is the canonical distribution.

This hypothesis is capable of being disconfirmed—and even strongly disconfirmed—by the finite set of systems generated by our sampling procedure thus far. And so, instead of the approximate identity requirement, we impose the following requirement.

> **Nondisconfirmation Requirement**. *Hyp* is not strongly disconfirmed by the finite set of systems generated by our sampling procedure thus far.

The nondisconfirmation requirement can handle the case of uninstantiated or rarely instantiated systems much better, insofar as we typically do not

https://doi.org/10.1086/605793 Published online by Cambridge University Press

take hypotheses about distributions to be strongly disconfirmed by very small samples.

The nondisconfirmation requirement is very modest. It may well be too modest to be of use. It is therefore interesting to note that even a requirement this modest gets us into trouble. To see how, let us consider some of the possible sampling procedures we might use to generate our ensemble of systems with Hamiltonian H in contact with a heat bath at temperature T.

Suppose that our sampling procedure tells us to include in our ensemble the microstate of any system with Hamiltonian H in contact with a heat bath at temperature T, but only at the moment the system is first placed in contact with the heat bath. The problem here is that many systems are way out of static equilibrium when first placed in contact with a heat bath. Because of this, the method of sampling just described may consistently produce a disproportionate number of systems in microstates that are extraordinarily rare according to the canonical distribution (1). If this is so, I take it that *Hyp* will be strongly disconfirmed.

Perhaps it is unfair to consider systems only at the moment they are placed in contact with a heat bath. A problem, however, is that if our sampling procedure tells us to take note of *all* of the microstates of the system whenever it is in contact with a heat bath, the out-of-static-equilibrium microstates may still be disproportionately represented. (Imagine, e.g., a system with a very long relaxation time, such that up until now, such systems in contact with heat baths at a given temperature have spent almost all of their lives in regions of phase space that are extremely rare according to the canonical distribution.) It will also not do to sample the microstate the moment the system reaches static equilibrium, as the canonical distribution predicts energy fluctuations that will not be present if we sample microstates only with fixed energy. The absence of such energy fluctuations will count as strong disconfirmation of *Hyp*, particularly in cases in which the occurrence of such fluctuations is moderately probable.

The most promising strategy is to sample the microstate of the system in contact with the heat bath only after some prespecified amount of time—some sort of 'relaxation time'—appropriate to the system at hand. But I am skeptical as to whether this sort of approach can work. Highly contingent facts about the world, including facts about the way in which we choose to prepare systems, have a way of affecting the composition of any such ensemble. For instance, consider a system that, at a given temperature, has multiple macroscopically distinguishable equilibrium states. Any system with potentially variable anisotropic structure, such as a crystal, will do. For a more artificial example, consider a one-dimensional Ising ring with no external magnetic field, that is, a system of N particles

https://doi.org/10.1086/605793 Published online by Cambridge University Press

with Hamiltonian

$$H = -J\sum_{i=1}^{N}\sigma_i\sigma_{i+1},$$

in which each σ_i takes on the value of plus or minus 1, and $\sigma_{N+1} = \sigma_1$. At low temperatures, this system organizes itself into large, macroscopic blocks of ones or minus ones. Because the value of the Hamiltonian is unchanged under the transformation $\sigma \rightarrow -\sigma$, for each equilibrium macrostate at a given temperature there is a distinct equilibrium macrostate at the same temperature in which each macroscopic chunk is replaced by one of the opposite sign. At sufficiently low temperatures, these equilibrium macrostates are clearly distinguishable. To make the argument that follows as clean as possible, let us suppose that systems with the given Hamiltonian do not occur naturally but are always manufactured by us. It is perfectly possible for us, having manufactured such a system, to place it in contact with a heat bath at a certain extremely cold temperature T (much lower than that of the environment in which it was produced) if and only if the system is in a *particular* macrostate—for instance, if and only if the largest macroscopic block consists of ones rather than minus ones.

When placed in contact with the heat bath, such a system is very likely to relax into a particular sort of macrostate—specifically, a macrostate in which it remains the case that the largest macroscopic block consists of ones rather than minus ones. Such a system is very unlikely to fluctuate out of this state if the temperature is very low. As a result, insofar as we sample systems only after this relaxation time, our sample will consist primarily of microstates in a very specific region of phase space—namely, the region of phase space corresponding to this particular sort of macrostate. This ensemble will look extraordinarily unlikely, given the canonical probability distribution. If we place 1,000 such systems in contact with heat baths, and essentially all of them end up with the largest macroscopic block consisting of ones rather than minus ones, it will be as if a fair coin thrown 1,000 times came up heads more or less every time. This is a situation with extraordinarily low probability, and so we would surely have to say that in such a situation *Hyp* has been strongly disconfirmed.[2]

2. This is not to deny that if such a system remained in contact with the heat bath long enough, there would eventually be large energy fluctuations after which the system might find itself in one of the other possible macrostates at the given temperature. But this will typically occur only after an astronomically long time. So if we try to get around the problem by insisting that we sample systems only after sufficiently large energy fluctuations have occurred, essentially *all* our ensembles will be empty, and our description of such ensembles with the canonical distribution starts to border on vacuity.

https://doi.org/10.1086/605793 Published online by Cambridge University Press

I think that difficulties like these show that the entire actualist approach is in error. Insofar as our interpretation of the canonical distribution involves identifying an ensemble, I suspect that the ensemble will have to be populated not just with actual physical systems but also with counterfactual systems. This is in agreement with Gibbs, who denied that his ensembles had any sort of *objective existence*, saying instead that they were *creations of the imagination* (see Gibbs 1981, 188). But what counterfactual ensemble should we pick?

The traditional view, championed by Gibbs, was to simply *posit* an imaginary collection of systems distributed in accordance with the canonical distribution. According to this approach, when one describes a system using the canonical distribution, one is making a claim about an imaginary ensemble of which the particular system is a member. There are many well-known problems with this approach. For instance, the actual system is a member of all sorts of imaginary ensembles—so why privilege the canonical ensemble when it comes to making predictions?[3] I think the fact that problems like this are so difficult suggests that this way of thinking of the canonical distribution does not get things right and that it makes sense to look for more 'naturally occurring' counterfactual ensembles with which to work.

2.3. The Time Ensemble. One proposal seems particularly natural. The rough idea is this—take a system in contact with a heat bath and imagine allowing it to stay in contact with the heat bath indefinitely. To describe this system using the canonical distribution is just to say that the proportion of time the system spends in any small region of phase space dX containing the point X is approximately

$$\frac{e^{-\beta H(X)}}{\mathcal{Z}}\,|dX|.$$

In other words, the ensemble we consider is the *time ensemble* consisting of all the states through which the system would pass were it left in contact with the heat bath indefinitely. This is obviously a counterfactual ensemble, insofar as it involves microstates that may not have been (and may

Whether obstacles like this can be satisfactorily surmounted is something I am happy to leave open—at any rate, these obstacles are challenging enough that it is worth considering alternative points of view.

3. For extremely thorough discussions of this problem, although mainly with respect to the microcanonical rather than the canonical ensemble, see Earman and Redei 1996 and Chapter 5 of Sklar 1993.

https://doi.org/10.1086/605793 Published online by Cambridge University Press

never be) realized.[4] Although we will suggest some slight modifications of this idea shortly, let us focus on the idea in this simpler form first.

The main idea must be formulated with a little care. The most natural way to define a probability density ϕ on the set of microstates through which our system passes is as follows: focus at first on the set of microstates through which the system passes in a fixed interval of time from $t = 0$ to $t = T$. We suppose that initially the system and the heat bath are in some fixed, prespecified state. Let dX be a very small region of phase space—for simplicity, we assume it is a ball centered around a point X of phase space—and let $\tau(dX)$ be the amount of time that the system spends in the region dX of phase space in the interval of time $[0, T]$. Our goal is to define a probability distribution ϕ_T that describes the states of the system in the time interval from zero to T. We would like it to be the case that

$$\phi_T(X)|dX| \approx \frac{\tau(dX)}{T},$$

and so it is natural to define

$$\phi_T(X) = \lim_{|dX|\to 0} \frac{\tau(dX)}{T|dX|},$$

where we imagine the radius of the open ball dX going to zero.

The problem is that thus defined, $\phi_T(X) = 0$ almost everywhere.[5] This is not a useful distribution with which to work. This sort of situation is not unfamiliar. In his *Lectures on Gas Theory*, Boltzmann (1964) considers the case of a homogeneous gas in a vessel. He defines a function $f(\vec{v}, t)$—the so-called density of velocities—by requiring $f(\vec{v}, t)|d\vec{v}|$ to be the number of particles in a (fixed) region of space whose velocities lie in the region

4. John Norton has pointed out that Einstein sketched a similar idea in a 1903 article (Einstein 1903). That Einstein did not flesh this idea out is perhaps due to his (incorrect) supposition that Gibbs sorted all this out in his book (1981).

5. Under reasonable assumptions, in a finite amount of time the path our system traces through phase space will consist of finitely many continuous curves, each of finite length. It may be shown to follow from this that the set of points through which the system does not pass in this interval of time is an open set, possibly plus some finite set of points, and that the set of points through which the system does pass is of measure zero. If we take a point X through which the system does not pass, it follows that, with finitely many exceptions, there will be some open ball centered around that point through which the system does not pass. And so when dX is a sufficiently small ball centered around X, we will have $\tau(dX) = 0$. Thus, for sufficiently small dX, $\tau(dX)/(T|dX|) = 0$, and so it follows from our definition that $\phi_T(X) = 0$. So with at most finitely many exceptions, $\phi_T(X) = 0$ whenever the system does not pass through X. Thus, $\phi_T(X) = 0$ almost everywhere.

https://doi.org/10.1086/605793 Published online by Cambridge University Press

$d\vec{v}$ at time t. Let us suppose we are dealing with a gas consisting of finitely many point particles, and let us suppose that the region $d\vec{v}$ is a sphere centered around a fixed velocity $\vec{v}$. Suppose $N(d\vec{v})$ is the number of particles in the given spatial region with velocity in the region $d\vec{v}$ (we suppress the argument t for simplicity). Then we would like

$$f(\vec{v}, t)|d\vec{v}| \approx N(d\vec{v}),$$

and so it is natural to define

$$f(\vec{v}, t) = \lim_{|d\vec{v}|\to 0} \frac{N(d\vec{v})}{|d\vec{v}|}.$$

But the problem is that for almost all velocities $\vec{v}$, once $d\vec{v}$ is sufficiently small $N(d\vec{v}) = 0$, and so $f(\vec{v}, t) = 0$ almost everywhere.

Boltzmann's solution was to think of $d\vec{v}$ as a tiny *fixed* cell of velocity space centered around $\vec{v}$. Because $d\vec{v}$ is fixed, it is not the sort of thing that can be sent to zero. The equation

$$f(\vec{v}, t)|d\vec{v}| = N(d\vec{v})$$

is then taken to define the value of $f(\vec{v}, t)$. Boltzmann's idea was that we should choose our cells $d\vec{v}$ in such a way that the resulting distribution of velocities $f(\vec{v}, t)$ is more or less continuous—so that when we move from a cell of velocity space to an 'adjacent' cell, $f(\vec{v}, t)$ does not change too much. That it is possible to choose cells' $d\vec{v}$ in this way is a substantive assumption Boltzmann was simply happy to make. So long as this assumption was made, the quantity $f(\vec{v}, t)$ was well defined enough for Boltzmann's purposes.

We mimic this move. We divide phase space into small cells. Let dX be the cell containing X. We then define

$$\phi_T(X) = \frac{\tau(dX)}{T|dX|}. \tag{2}$$

The idea here is to choose the cells in such a way that ϕ_T is more or less continuous—so that when we move from one of these cells to an adjacent one, $\phi_T(X)$ does not change too much. That it is possible to do this is a substantive assumption that we simply take for granted in what follows. So long as this assumption is made, the quantity $\phi_T(X)$ is well defined enough to work with.

Given all this, there are various ways in which one might formulate the idea that the canonical distribution describes the time ensemble of states through which the system would evolve were it left in contact with the heat bath. For instance, consider what follows.

https://doi.org/10.1086/605793 Published online by Cambridge University Press

Definition. For a system in contact with a heat bath to be describable by the canonical distribution is for it to be the case that, for all points X in phase space, $\lim_{T\to\infty} \phi_T(X) \approx e^{-\beta H(X)}/\mathcal{Z}$.

Here, we suppose that the system and the heat bath are in fixed, prespecified states at $T = 0$. (We leave open the possibility that different coarse grainings may be required at different times T.)

All sorts of variations on this main idea are possible. One of the more important variations involves thinking of the interaction between the heat bath and the system indeterministically, in much the same way that we think of the interactions between a Brownian particle and its environment indeterministically. In this case, the system will not trace out a determinate path through phase space in the interval of time $[0, T]$—instead, we will have to describe what happens to the system in the interval $[0, T]$ using a probability measure over paths through phase space.

In this indeterministic setting, there are a number of ways in which one might try to capture the idea that describing a system with the canonical distribution involves making a claim about the set of states through which the system evolves. One way follows. Given a small region of dX of phase space, let $E_T[\tau(dX)]$ be the expected value of the amount of time the system spends in the region dX in the interval of time $[0, T]$. (We suppose that the system is in a fixed, prespecified state at $T = 0$.) We then define

$$\phi_T^*(X) = \frac{E_T[\tau(dX)]}{T|dX|}. \tag{3}$$

For such a system in a particular initial state to be describable by the canonical distribution is for it to be the case that for all points X in phase space,

$$\lim_{T\to\infty} \phi_T^*(X) \approx \frac{e^{-\beta H(X)}}{\mathcal{Z}}.$$

A different approach is to work with the definition of ϕ_T given in the deterministic setting and insist instead that

$$\lim_{T\to\infty} \Pr\left[\phi_T^*(X) \approx \frac{e^{-\beta H(X)}}{\mathcal{Z}}\right] \approx 1.$$

Even this later equation admits of variations. My goal here is not to commit to one particular approach but rather to indicate some possible ways of thinking of the canonical distribution in which time ensembles are central.

Even without committing to a specific definition, some advantages of the general approach sketched here are already visible. According to this

https://doi.org/10.1086/605793 Published online by Cambridge University Press

approach, to say that a system can be described by the canonical distribution is to make a statistical claim about the time ensemble of microstates through which our system passes. Note that one advantage of working with these definitions is that the focus is on time averages right from the start. If time averages are so important—as the huge amount of literature directed toward understanding the relationship between time and ensemble averages suggests—it seems much more natural to just *define* the canonical ensemble as a time ensemble of sorts. This is not to suggest that by redefining words we can avoid any of the major philosophical problems with equilibrium statistical mechanics. All the old problems still exist, although many of them will need to be formulated in different language. The hope, however, is that better definitions can give us a better perspective on what these problems are and how deep they run.

Also note that according to our definition, to describe a system with the canonical distribution is to make a claim about the (very) long-term behavior of the system. As such, the values of observables calculated using the canonical distribution are to be thought of as time-averaged expectation values. The time-averaged expectation value of an observable can be quite different from the value of the observable at any moment of time. However, supposing we can show that the variance of the observable over time is small, we can at least assert that the observed value will *probably* be approximately equal to the time-averaged expectation value.[6] The observed value probably being approximately equal to the time-averaged expectation value is compatible with the observed value being radically different from the time-averaged expectation value. But so long as we realize that the canonical distribution gives us only the probable value of an observable, there is no real problem with the fact that the actual value of the observable might be very different, unless we insist on demanding from the canonical ensemble something that it is simply not designed to give.[7]

3. Concluding Remarks. In this article, I have focused on what it means to describe the state of a system using the canonical distribution. Quite separate from this is the question of whether we are usually *right* when we describe systems in equilibrium with heat baths using the canonical distribution.

6. An analogous move is made in Malament and Zabell 1980, albeit with respect to ensemble averages and the microcanonical distribution.

7. Here, we are talking about probability derived from the time ensemble in the way described. There may be other notions of probability at play, due to the fact that the system in question will be a part of other ensembles or collectives. The canonical ensemble makes no claims about those other ensembles.

https://doi.org/10.1086/605793 Published online by Cambridge University Press

On the view I have been urging, describing a system with the canonical distribution involves making a highly nontrivial assertion about the system. It is not obvious that it is ever correct to describe any system, let alone the multitude of systems to which the canonical distribution is actually applied, with the canonical distribution thus defined. Much of the technical literature on this problem has focused on deterministic models of heat baths and has been somewhat inconclusive. (In its modern form, this literature largely springs from the seminal article by Ford, Kac, and Mazur [1965], as well as work by Huerta and Robertson [1969, 1971]. For a more modern discussion, see Tegmark and Yeh 1994.) Much less work has been done on the long-term behavior of a Hamiltonian system immersed in an indeterministic heat bath. (An exception is the simple case of Brownian motion, which is very well understood.) Whether it is any easier to calculate the long-term behavior of a system in an indeterministic rather than a deterministic heat bath is not known, although there are reasons for thinking that this might be so.

Even once we have settled the problem of determining whether we are right to describe a system in contact with a heat bath using the canonical distribution, other difficult issues then require attention. For instance, why it is possible to treat a physical system in equilibrium with a large, finite reservoir of heat as if it were in contact with an *infinite* reservoir of heat? And why it is permissible to use the canonical distribution to describe systems that are not in contact with any sort of reservoir of heat, as is often done? (The fact that the microcanonical and canonical distributions approach each other in the thermodynamic limit is often used to justify the use of the canonical distribution for closed systems, but it is not clear whether this argument is completely convincing.) These questions remain unanswered. Nevertheless, thinking about what sort of thing the canonical distribution could be is of great help in clarifying the mathematical and conceptual questions that must be addressed in order to justify its use. This is at least a modest sort of progress.

REFERENCES

Boltzmann, L. (1964), *Lectures on Gas Theory*. Berkeley: University of California Press.

Earman, J., and M. Redei (1996), "Why Ergodic Theory Does Not Explain the Success of Statistical Mechanics", *British Journal for the Philosophy of Science* 47: 63–78.

Einstein, A. (1903), "A Theory of the Foundations of Thermodynamics", *Annalen der Physik* 11: 170–187.

Ford, G., M. Kac, and P. Mazur (1965), "Statistical Mechanics of Assemblies of Coupled Oscillators", *Journal of Mathematical Physics* 6 (4): 504–515.

Gibbs, J. Willard (1981), *Elementary Principles in Statistical Mechanics*. Woodbridge, CT: Ox Bow.

Huerta, M., and H. Robertson (1969), "Entropy, Information Theory, and the Approach to Equilibrium of Coupled Harmonic Oscillator Systems", *Journal of Statistical Physics* 1 (3): 393–414.

https://doi.org/10.1086/605793 Published online by Cambridge University Press

——— (1971), "Approach to Equilibrium of Coupled Harmonic Oscillator Systems II", *Journal of Statistical Physics* 3 (2): 171–189.
Malament, D., and S. Zabell (1980), "Why Gibbs Phase Averages Work—the Role of Ergodic Theory", *Philosophy of Science* 47: 339–349.
Sklar, L. (1993), *Physics and Chance*. Cambridge: Cambridge University Press.
Tegmark, M., and L. Yeh (1994), "Steady States of Harmonic Oscillator Chains and Shortcomings of Harmonic Heat Baths", *Physica A* 202: 342–262.

https://doi.org/10.1086/605793 Published online by Cambridge University Press