

# What is Statistical Mechanics?

**Roman Frigg**

*Department of Philosophy, Logic and Scientific Method, London School of Economics, UK*

**Forthcoming** in Carlos Galles, Pablo Lorenzano, Eduardo Ortiz, and Hans-Jörg Rheinberger (eds.): *History and Philosophy of Science and Technology*, Encyclopedia of Life Support Systems Volume 4, Isle of Man: Eolss.

**Keywords:** Statistical Mechanics, Entropy, Second Law of Thermodynamics, Equilibrium.

## Contents

1. Introduction
  2. Classical Mechanics
  3. The Boltzmann Approach
  4. The Gibbs Approach
  5. Conclusion
- Bibliography

## Summary

Thermodynamics describes a large class of phenomena we observe in macroscopic systems. The aim of statistical mechanics is to account for this behaviour in terms of the dynamical laws governing the microscopic constituents of macroscopic systems and probabilistic assumptions. This article provides a survey of the discussion about the foundation of statistical mechanics by introducing the basic approaches and discussing their merits as well as their problems. After a brief review of classical mechanics, which provides the background against which statistical mechanics is formulated, we discuss the two main theoretical approaches to statistical mechanics, one of which can be associated with Boltzmann and the other with Gibbs. We end with a discussion of remaining issues and open questions.

## 1. Introduction

Let us begin with a characteristic example. Consider a gas that is confined to the left half of a box. Now we remove the barrier separating the two halves of the box. As a result, the gas quickly disperses, and it continues to do so until it homogeneously fills the entire box. This is illustrated in Figure 1.



Figure 1: The gas spreading after the opening of the shutter

The gas has approached equilibrium. This process has one important feature: it is *unidirectional*. We see the gas spread – I.e. we see it evolve towards equilibrium – but we never observe gases spontaneously reverting to the left half of a box – i.e. we never see them move away from equilibrium when left alone. And this is not a specific feature of our gas. We see ice cubes melting, coffee getting cold when left alone, and milk mix with tea; but we never observe the opposite happening. Ice cubes don't suddenly emerge from lukewarm water, cold coffee doesn't spontaneously heat up, and white tea doesn't un-mix, leaving a spoonful of milk at the top of a cup otherwise filled with black tea. In fact, all systems, irrespective of their specific makeup, behave in this way! This fact is enshrined in the so-called Second Law of thermodynamics (TD), which, roughly, states that transitions from equilibrium to non-equilibrium states cannot occur in isolated systems. Thermodynamics describes a system in terms of macroscopic quantities such as pressure, volume and temperature. Its laws are formulated solely in terms of these and it makes no reference to a system's microscopic constitution. For this reason TD is a 'macro theory'. In order to give a precise formulation of the Second Law, TD introduces a quantity called entropy (the precise definition of which need not occupy us here). The Second Law then says that entropy in closed systems (such as our gas) cannot decrease, and in fact processes like the spreading of the gas are characterised by an increase in entropy.

But there is an altogether different way of looking at that same gas. The gas consists of a large number of gas molecules (a vessel on a laboratory table contains something like  $10^{23}$  molecules).

These molecules bounce around under the influence of the forces exerted onto them when they crash into the walls of the vessel and when they collide with each other. The motion of each molecule under these forces is governed by laws of mechanics, which, in what follows, we assume to be the laws of classical mechanics (CM). We now know that quantum mechanics and not classical mechanics is the fundamental theory of matter. In the current context we nevertheless stick with classical mechanics because the central problems in connection with statistical mechanics remain, *mutatis mutandis*, the same if we replace classical mechanics with quantum mechanics and is easier to discuss these if we do not at the same time also have to deal with all the conceptual problems raised by quantum mechanics.

In other words, the gas is a large mechanical system and everything that happens in it is determined by the laws of mechanics. Since classical mechanics in this context governs the behaviour of the micro constituents of a system it is referred to as the 'micro theory'.

This raises the question of how the two ways of looking at the gas fit together. Since neither the thermodynamic nor the mechanical approach is in any way privileged, both have to lead to the same conclusions. In particular, it has to follow from the mechanical description that the Second Law is valid (since this law is a well-confirmed empirical fact). Statistical mechanics (SM) is the discipline that addresses this task. Its two most fundamental questions are the following. First, how can we characterise equilibrium from a mechanical point of view? This is the task addressed in equilibrium SM. Second, what is it about molecules and their motions that leads them to spread out and assume a new equilibrium state when the shutter is removed? And crucially, what accounts for the fact that the reverse process won't happen? This is the central problem of non-equilibrium SM.

Hence, from a more abstract point of view we can say that SM is the study of the connection between micro-physics and macro-physics: it aims to account for this behaviour in terms of the dynamical laws governing the microscopic constituents of macroscopic systems. The term 'statistical' in its name is owed to the fact that, as we will see, a mechanical explanation can only be given if we also introduce probabilistic elements into the theory. The aim of this chapter is to lay out the main tenets of SM, and to explain the foundational problems that it faces.

Such a project faces an immediate difficulty. Foundational debates in many other fields of physics can take as their point of departure a generally accepted formalism. Not so in SM. Unlike quantum mechanics and relativity theory, SM has not yet found a generally accepted theoretical framework,

let alone a canonical formulation. What we find in SM is a plethora of different approaches and schools, each with its own programme and mathematical apparatus.

All these schools use (slight variants) of either of two theoretical frameworks, one of which can be associated with Boltzmann's 1877 landmark paper and the other with Gibbs' seminal 1902 book, and can thereby be classified as either 'Boltzmannian' or 'Gibbsian'. For this reason I divide my presentation of SM into a Boltzmannian and a Gibbsian part. It is important, at least initially, to keep these two frameworks apart because they give rise to markedly different characterisations both of equilibrium and of non-equilibrium, and accordingly the problems that beset accounts formulated within either framework are peculiar to one framework and often do not have a counterpart in the other. But before embarking on a discussion of these approaches, we need to say more about CM.

## 2. Classical Mechanics

CM can be presented in various more or less but not entirely equivalent formulations: Newtonian mechanics, Lagrangean mechanics, Hamiltonian mechanics and Hamilton-Jacobi theory. Hamiltonian Mechanics is best suited to the purposes of SM, and therefore this section focuses entirely on this version of CM.

CM describes the world as consisting of point-particles, which are located at a particular point in space and have a particular momentum (where a particle's momentum essentially is its velocity times its mass). A system's state is fully determined by a specification of each particle's position and momentum – that is, if you know the positions and the momenta of all particles in the system you know everything that there is to know about the system's state from a mechanical point of view. Conjoining the space and momentum dimension of all particles of a system in one vector space yields the so-called phase space  $\Gamma$  of the system. For a particle moving around in the three dimensional space of our every-day experience, the phase space basically consists of all points  $X = (x, y, z, p_x, p_y, p_z)$ , where  $x$ ,  $y$ , and  $z$  are the three directions in space, and  $p_x$ ,  $p_y$ , and  $p_z$  are the momenta in the  $x$ ,  $y$ , and  $z$  directions. So the phase space of one particle has six (mathematical) dimensions. The phase space of a system consisting of two particles is the collection of all points  $X = (x_1, y_1, z_1, x_2, y_2, z_2, p_{x_1}, p_{y_1}, p_{z_1}, p_{x_2}, p_{y_2}, p_{z_2})$ , where  $x_1$ ,  $y_1$ , and  $z_1$  are the spatial locations of the first particle,  $x_2$ ,  $y_2$ , and  $z_2$  the one of the second particle, and  $p_{x_1}$ , ..., are the momenta in the respective directions. Hence, the phase space of such a system is twelve-

dimensional. The generalisation of this to a system of  $n$  particles – which is what SM studies – is now straightforward: it is a  $6n$  dimensional abstract mathematical space. If  $X$  is the state of an  $n$  particle gas, it is also referred to as the system's *micro-state*.

An important feature of  $\Gamma$  is that it is endowed with a so-called Lebesgue measure  $\mu$ . Although  $\Gamma$  is an abstract mathematical space, the leading idea of a measure is exactly the same as the one of a volume in the three dimensional space of every-day experience: it is a device to attribute sizes to parts of space. We say that a certain collection of points of this space (for instance the ones that lie inside a bottle) have a certain volume (for instance one litre), and in the same way can we say that a certain set of points in  $\Gamma$  has a certain  $\mu$ -measure. If  $A$  is a set of points in  $\Gamma$ , we write  $\mu(A)$  to denote the  $\mu$ -measure of this set. At first it may seem counterintuitive to have measures ('volumes') in spaces of more than three dimensions (as the above  $X$  show, the space of a one particle system has six and the one of a two particle system twelve dimensions). However, the idea of a higher dimensional measure becomes rather natural when we recall that the moves we make when introducing higher dimensional measures are the same as when we generalise one-dimensional length, which *is* the Lebesgue measure in one dimension, to two dimensions, where the usual surface is the Lebesgue measure, and then to three dimensions, where volume is the Lebesgue measure.

The system's state usually changes in time; for instance, it might change from  $X = (0,0,0,1,3,2)$  to  $X = (7,5,8,3,2,6)$  over the course of five seconds. This change does not happen in an arbitrary way; in fact it is governed by the so-called *Hamiltonian equations of motion*. The precise form and character of these equations need not occupy us here, but two points are important. First in their general form the equations of motions are an 'equation schema' since they contain a blank, so to speak. This blank has to be filled with a function  $H$  which specifies how the energy of the system depends on its state. This dependence takes different forms in different systems. It is at this point that the specific physical properties of a system under investigation come into play, and different Hamiltonians give raise to different kinds of motions. Second, if the Hamiltonian that we plug into the general equations has certain nice properties (which it does in the cases we are interested in here), then the equations have unique solutions in the following sense: pick a particular instance of time (you can pick any instant you like, but you can pick only one!) and call it  $t_0$ ; then prepare the system in a particular state, the so-called 'initial condition', which you are also free to choose as you like; if you have done this, then the equations of motion (which now contain the particular Hamiltonian of your system) uniquely determine the state of the system at *any* other time  $t$ . To make this more vivid consider the following thought experiment. You have 1000 copies of the same

system (in the sense that they consist of the same number of particles and are governed by the same Hamiltonian). Then you pick a particular instance of time  $t_0$  and make sure that at that time all 1000 systems are in the same state. Now you let the evolution in each system take its course and look at the systems again one hour later. You will then find that also one hour later all 1000 systems are in the same state – they have evolved in exactly the same way! For instance, if all systems were in state  $X = (0,0,0,1,3,2)$  at  $t_0$  and the first system has evolved into  $X = (7,5,8,3,2,6)$  after one hour, then all other systems have evolved into  $X = (7,5,8,3,2,6)$  as well. And this will be so if you have 10 000 systems or 100 000 or any number you like; and it will be true after one hour, after two hours, after three hours ... after any amount of time you like. There will never be a system that deviates from what the others do.

The function that tells us what the system's state at some later point will be is called a 'phase flow' and we denote it with the letter  $\phi$ . We write  $\phi_t(X)$  to denote the state into which  $X$  evolves under the dynamics of the system if time  $t$  (e.g. one hour) elapses, and similarly we write  $\phi_t(A)$  to denote the image of a set  $A$  (of states) under the dynamics of the system. The 'line' that  $\phi_t(X)$  traces through the phase space is called a trajectory.

Let us illustrate this with a simple example. Consider a pendulum of the kind we know from grandmothers' clocks: a bob of mass  $m$  is fixed to a string of length  $l$  and oscillates back and forth. To facilitate the calculations we assume that the string is massless, there is neither friction nor air resistance, and the only force acting on the pendulum bob is gravity. It is then easy to write down the pendulum's Hamiltonian and solve the equations. The solutions, it turns out, are ellipses in phase space – so the system's trajectory is an ellipse. Figure 2a shows the leftmost and the rightmost snapshot of the motion and the pendulum's trajectory in phase space. Depending on the initial condition (e.g. how far to the left you move the bob before you let go), the bob moves on a different ellipse. This is shown in the upper half of Figure 2b.

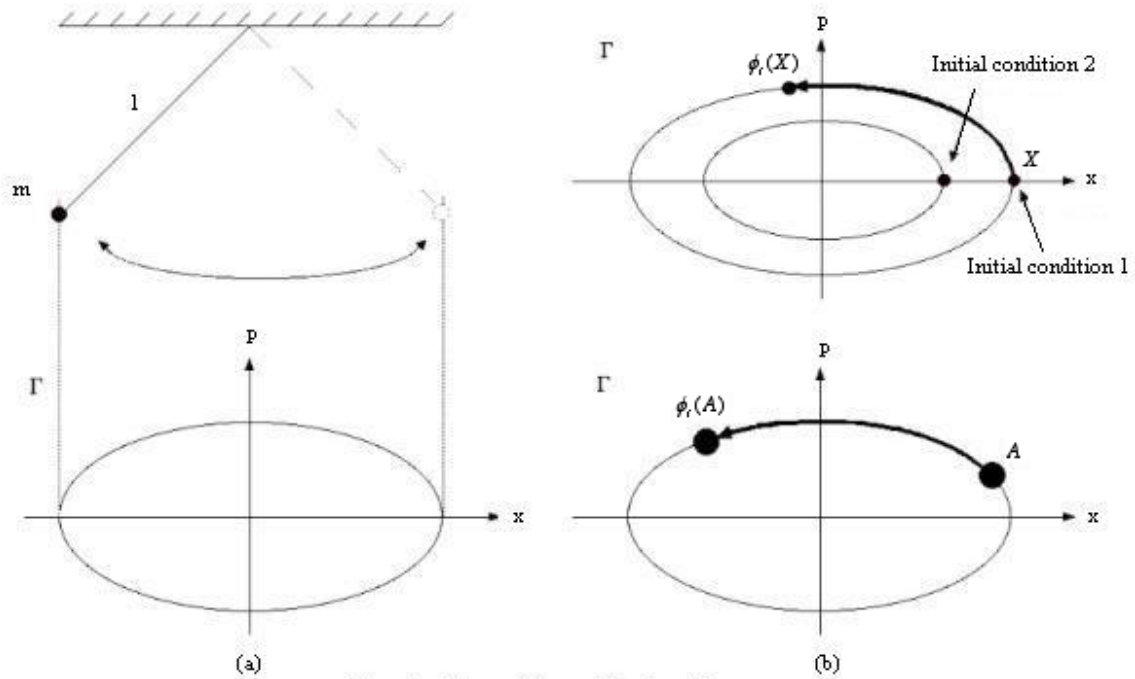


Figure 2: The pendulum and its phase flow

The Hamiltonian dynamics of systems has three distinctive features. The first one, known as *Liouville's theorem*, says that the measure  $\mu$  is invariant under the dynamics of the system:  $\mu(A) = \mu(\phi_t(A))$  for all  $A$  and all  $t$ . This means that the measure of a set does not change in time: if you measure a set now and you measure it tomorrow you are bound to find the same value. This is illustrated in the lower half of figure 2b, where a circular set  $A$  moves around under the dynamics of the pendulum and does not change its surface.

The second distinctive feature is *Poincaré's recurrence theorem*. Roughly speaking this theorem says that a system will sooner or later return arbitrarily close to its initial state. The time that it takes the system to return close to its initial state is called 'Poincaré recurrence time'. Recurrence can also easily be seen in the above example: the system returns to the exact same state after every full oscillation. In this simple example Poincaré recurrence is obvious; the surprising thing is that we find this kind of recurrence in *every* system, no matter how complicated its phase flow.

The third is so-called *time reversal invariance*. The Hamiltonian equations of motion in a sense perform the function of a censor: they say which time evolutions are allowed by the theory and which ones are not. Now consider a ball moving from left to right and record this process on videotape. Intuitively, time reversal amounts to playing the tape backwards, which makes us see a ball moving from right to left. So we can ask the question: if the first process (motion from left to

right) is allowed by the theory, is the reverse of this process (motion from right to left) allowed too? If the answer to this question is 'yes' in all cases, then the theory is said to be time reversal invariant. It turns out that the Hamiltonian equations of motion have this property (and this is non-trivial: not all equations of motion are time reversal invariant). Again, time reversal invariance is easy to see in our example: both the original motion and its reverse are possible.

### 3. The Boltzmann Approach

Over the years Boltzmann developed a multitude of different approaches to SM. However, contemporary Boltzmannians, take the account introduced by Boltzmann in his seminal 1877 paper as their starting point. For this reason we concentrate on this approach.

Let us start with macro-states. We assume that every system has a certain number of macro states  $M_1, \dots, M_k$  (where  $k$  is a natural number that depends on the specifics of the system), which are characterised by the values of macroscopic variables, in the case of a gas pressure, temperature, and volume. In the introductory example one macro-state corresponds to the gas being confined to the left half, another one to it being spread out. In fact, these two states have special status: the former is the gas' initial state, which, for reasons that will become clear later, we call the *past state* and label by  $M_p$ ; the latter is the gas' equilibrium state, which we label  $M_{eq}$ .

What is the relation between micro-states and macro-states? It is one of the fundamental posits of the Boltzmann approach that the former determine the latter. More specifically, the posit is that a system's macro-state supervene on its micro-state, meaning that a change in the macro-state must be accompanied by a change in the micro-state  $X$ : if  $M$  changes then  $X$  has to change too. For instance, it is not possible to change the pressure of a system and at the same time keep its micro-state constant. Hence, to every given micro-state  $X \in \Gamma$  there corresponds *exactly one* macro-state. Let us refer to this macro-state as  $M(X)$ . This determination relation is not one-to-one; in fact many different  $X$  can correspond to the same macro-state. We now group together all micro-states  $X$  that correspond to the same macro-state, which yields a partitioning of the phase space in non-overlapping regions that each correspond to a macro-state. For this reason we use the same letters,  $M_1, \dots, M_k$ , to refer to macro-states and the corresponding regions in phase space. This is illustrated in Figure 3a.



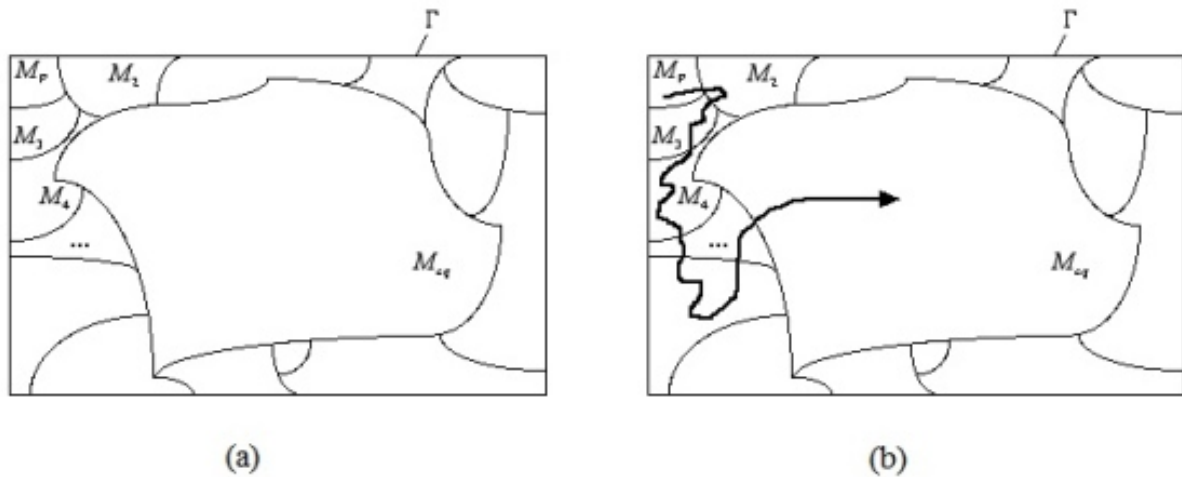


Figure 3: Partitioning and approach to equilibrium

We are now in a position to introduce the Boltzmann entropy. To this end recall that we have a measure on  $\Gamma$  that assigns to every set a particular volume, hence *a fortiori* also to macro-states. With this in mind, we define the Boltzmann entropy of a macro-state  $M_i$ ,  $S_B(M_i)$ , as a constant – the so-called Boltzmann constant  $k_B$  – times the logarithm of the measure of the macro-state:  $S_B(M_i) = k_B \log[\mu(M_i)]$ . The important feature of the logarithm is that it is a monotonic function: the larger  $\mu(M_i)$ , the larger its logarithm. From this it follows that the largest macro-state also has the highest entropy!

One can show that, at least in the case of dilute gases, the Boltzmann entropy coincides with the thermodynamic entropy (in the sense that both have the same functional dependence on the basic state variables), and so it is plausible to say that the equilibrium state is the macro-state for which the Boltzmann entropy is maximal (since TD posits that entropy be maximal for equilibrium states). By assumption the system begins in a low entropy state, the past state  $M_p$ . The problem of explaining the approach to equilibrium then amounts to answering the question: why does a system originally in  $M_p$  eventually move into  $M_{eq}$ ? This is illustrated in Figure 3b.

Or more precisely, to underwrite the Second Law one would have to show that *all* trajectories starting in  $M_p$  must end up in  $M_{eq}$ . Unfortunately it is clear that this would be aiming too high – it is generally accepted that the best we can hope for is to get a justification of something a bit weaker than the strict Second Law, namely a 'probabilistic version' of it which I call *Boltzmann's Law* (BL):

Consider an arbitrary instant of time  $t = t_1$  and assume that the Boltzmann entropy of the system at that time,  $S_B(t_1)$ , is far below its maximum value. It is then highly probable that at any later time  $t_2 > t_1$  we have  $S_B(t_2) \geq S_B(t_1)$ .

The central problem now is to elucidate the notion of an entropy increase being 'highly likely'. This problem has two aspects, one conceptual and one formal. The conceptual problem consists in explaining what notion of probability is at play here: what do we mean when we talk about it being likely that the entropy increases? The formal problem is to provide a justification that the claim made in BL is indeed true, which essentially depends on the system's dynamics.

There are two prominent ways of introducing probabilities into the Boltzmannian apparatus developed so far. The first, which is associated with Boltzmann himself, interprets probabilities as time averages. More specifically, the view is that the probability of a macro-state is the proportion of time that the system actually spends in that state in the long run. For instance, if the system spends 10% of the time in macro-state  $M_1$ , then the probability of this state is 0.1.

Already Boltzmann realised that a strong dynamical assumption is needed to make this suggestion fly: the system has to be ergodic. Roughly speaking, a system is ergodic if, on average, the time it spends in a subset of the phase space is proportional to the portion of the phase space occupied by that set. So if, for instance, set  $A$  occupies one quarter of the phase space, then an ergodic system spends one quarter of its time in  $A$ . It follows immediately that the most likely macro-state is the equilibrium state, as we would expect.

The proper mathematical formulation of ergodicity was a formidable problem that was solved satisfactorily only half a century after Boltzmann had proposed the idea. But even this was not the end of difficulties. On the one hand it was soon realised that there are systems showing the right sort of behaviour (i.e. they approach equilibrium) while they fail to be ergodic and hence ergodicity does not seem to be the key ingredient in an explanation of thermodynamic behaviour. On the other hand there are technical problems with the account that cast doubt on its workability.

The second approach focusses on the internal structure of macro-states and assigns probabilities using the so-called *statistical postulate* (SP), the posit that given the system is in *macro-state*  $M$ , the probability of finding the system's *micro-state* in a certain subset  $A$  of  $M$  is proportional to that set's size:  $p(A) = \mu(A)/\mu(M)$ . This is illustrated in Figure 4a.

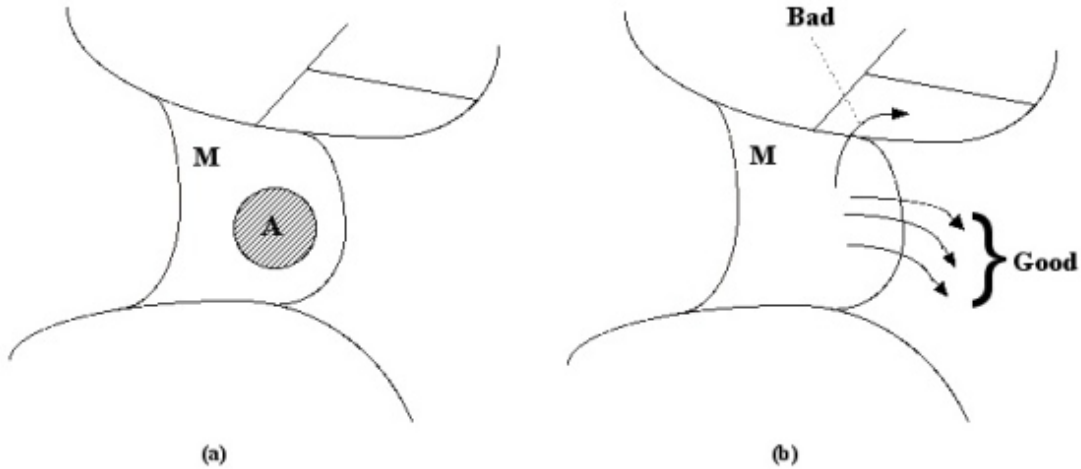


Figure 4: The statistical postulate.

How is this useful to explain BL? The answer to this question lies in recalling that the phase flow  $\phi_t$  completely determines the future for every point  $X$  in phase space, and hence *a fortiori* for every point in  $M$ . Given this, we can sort the points in  $M$  in 'good' and 'bad', where the good ones are those that move into a macro-state of higher entropy when the trajectory on which they lie leave  $M$ ; the bad ones are those that move towards macro-states of lower entropy. This is illustrated in Figure 4b. Now take  $A$  to be the set of good points. Then the probability for an entropy increase is  $\mu(A)/\mu(M)$ . This is the probability that BL talks about, and requires that it be high.

So the crucial question then is: what reasons are there to believe that  $\mu(A)/\mu(M)$  is high for all macro-states (except the equilibrium state itself). Whether or not this is the case depends on the system's phase flow  $\phi_t$ , which, in turn, depends on the system's Hamiltonian (its energy function). It is clear that not all Hamiltonians give rise to phase flows that make SP true. So there is a substantive question, first, about which class of Hamiltonians does, and, second, whether the actual system under scrutiny belongs to this class. Although this question is of central importance, it has, somewhat surprisingly, received relatively little attention in the recent literature on Boltzmann, and the same is true of the question of why a more complex version of SP to which we turn below holds true. The most promising approach to this problem seems to be one employing typicality arguments originally proposed by Goldstein.

But even if this question is answered satisfactorily, there are more problems to come. As Loschmidt pointed out in a controversy with Boltzmann in the 1870s, trying to explain unidirectional

behaviour by appeal to dynamical features of a system is highly problematic because there is no such unidirectionality at the mechanical level. In fact, as we have seen above, Hamiltonian mechanics is time reversal invariant and so everything that can happen in one direction can also happen in the other. More specifically, if the transition from a certain low entropy state to a higher entropy state is permitted by the underlying dynamics (which is what we want), then the reverse transition from the high to the low entropy state is permitted as well (which is what we don't want!). This point is known as Loschmidt's reversibility objection.

One might now try to mitigate the force of this argument by pointing to the fact that BL is a probabilistic and not a universal law (and hence allows for some unwanted transitions), and then arguing that the unwanted transitions are unlikely. Unfortunately this hope is shattered as soon as we try to make good on this suggestion. Calculations show that *if* the system, in macro-state  $M$ , is very likely to evolve towards a macro-state of higher entropy in the future (which we want to be the case), *then*, because of the time reversal invariance of the underlying dynamics, the system is also very likely to have evolved into the current macro-state  $M$  from another macro-state  $M'$  of *higher* entropy than  $M$ . So whenever the system is very likely to have a high entropy future it is also very likely to have a high entropy past. This stands in stark contradiction with both common sense experience and BL itself. If we have a lukewarm cup of coffee on the desk, SP makes the radically wrong retrodiction that is overwhelmingly likely that 5 minutes ago the coffee was cold (and the air in the room warmer), but then fluctuated away from equilibrium to become lukewarm and five minutes from now will be cold again. However, in fact the coffee was hot five minutes ago, cooled down a bit and will have further cooled down five minutes from now.

Before addressing this problem, let us add another difficulty, now known as Zermelo's Recurrence Objection. As we have seen above, Poincaré's recurrence theorem says, roughly, that almost every point in the systems phase space lies on a trajectory that will, after some *finite* time (the Poincaré recurrence time), return arbitrarily close to that point. As Zermelo pointed out in 1896, this has the unwelcome consequence that entropy cannot keep increasing all the time; sooner or later there will be a period of time during which the entropy of the system decreases. For instance, if we consider again the initial example of the gas (Figure 1), it follows from Poincaré's recurrence theorem that there some time in the future the gas will return to the left half of the container *all by itself*. This is not what we expect.

In response to the first problem (Loschmidt's objection) it has been pointed out that it is no surprise that an approach trying to underwrite BL *solely* by appeal to the laws of dynamics fails because a

system's actual behaviour is determined by its dynamical laws *and* its initial condition. Hence there need not be a contradiction between time reversal invariant laws and the fact that high to low entropy transitions do only very rarely occur. All we have to do is factor in that the system have low entropy initial conditions. The problem with wrong claims about the past can then be solved by explicitly conditionalising on the system's initial state,  $M_p$ . This amounts to replacing SP, which makes no reference to the system's past, by a rule that does. Such a rule can be constructed by considering a different class of states when attributing probabilities to increasing entropy. SP considers *all* states in  $M$  and then asks what proportion of them have a higher entropy future. But, so the argument goes, this is the wrong consideration. We should only consider those states in  $M$  which have the right past; i.e. those that have started off in  $M_p$ . So the right question to ask is not what portion of micro-states in  $M$ , but rather what portion of microstates in  $R_t = M \cap \phi_t(M_p)$  has a higher entropy future, where  $\phi_t(M_p)$  is the image of the initial state under the dynamics of the system since the process started. We then have to replace SP by SP\*:  $p(A) = \mu(A \cap R_t) / \mu(R_t)$ . This is illustrated in Figure 5.

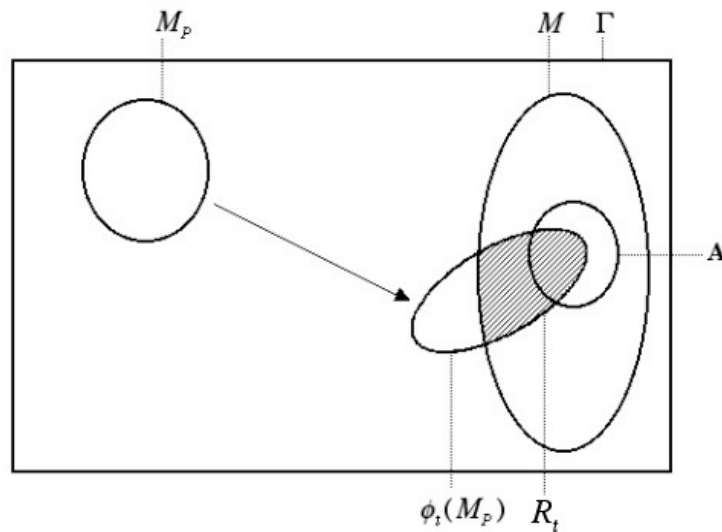


Figure 5: Conditionalising on  $M_p$ .

By construction, those fine-grained micro-states in  $M$  having the wrong past have been ruled out, which is what we need. Given this, we can formulate a condition for BL to be true: it has to be the case that if we choose  $A$  to be the set of those states that have a higher entropy future, then the probabilities given by SP\* for a high entropy future have to come out high. As indicated above, it is a substantial question for which class of Hamiltonians this is true. Unfortunately we do not see much discussion of this problem in the literature, and probably the most promising – but as yet still

underexplored – suggestion might be an approach based on typicality.

There is controversy over what exactly counts as  $M_p$ . The issue is at what point in time the relevant low entropy initial condition is assumed to hold. A natural answer would be that the beginning of an experiment is the relevant instant; we prepare the gas such that it sits in the left half of the container before we open the shutter and this is the low entropy initial condition that we need. This is how we have been talking about the problem so far.

Many physicists and philosophers think that this is wrong because the original problem (explaining why entropy increases towards the future) recurs if we think about how the low entropy state at the beginning of the experiment came about in the first place. Our gas is part of a larger system consisting of the laboratory and even the person who prepared the gas, and this system already existed prior to the beginning of the experiment. Since this larger system is also governed by the laws of CM we are forced to say that the system as a whole is highly likely to have come into the state it is in at the beginning of the experiment from one of higher entropy. And this argument can be repeated for every instance you choose. The problem is obvious by now: whichever point in time we chose to be the point for the low entropy initial condition to hold, it follows that the overwhelming majority of trajectories compatible with this state are such that their entropy was higher in the past. An infinite regress looms large. This regress can be undercut by assuming that there is an instant that simply has no past, in which case it simply does not make sense to say that the system has evolved into that state from another state. In other words, we have to assume that the low entropy condition holds at the beginning of the universe. And this is indeed that move many are willing to make:  $M_p$  is the state of the universe just after the big bang. When understood in this way, the claim that the system started off in a low entropy state is called the *Past Hypothesis*, and  $M_p$  is referred to as the *Past State*. Cosmology is then taken to provide evidence for the truth of the Past Hypothesis, since modern cosmology informs us that the universe was created in the big bang a long but finite time ago and that it then was in a low entropy state.

But the Past Hypothesis has not gone unchallenged. Earman argues that it is 'not even false', since the Boltzmann entropy is undefinable in the relevant (relativistic) cosmological models. Another, more philosophical, worry is that the need to introduce the Past Hypothesis to begin with only arises if one has a particular view of laws of nature. We started with a pledge to explain the behaviour of homely systems like a vessel full of gas and ended up talking about the universe as a whole due to the above regress argument. But this argument relies on the assumption that laws are universal in the sense of being valid all the time and everywhere: the whole world – not only the gas, but also its

laboratory environment and even the person preparing the system for the experiment – are governed by the deterministic laws of classical mechanics. Only under this assumption there is a problem about the system's high entropy past prior to the beginning of the experiment. However, the universal validity of laws, and in particular the laws of mechanics, is not uncontroversial. Some believe that these laws are valid only locally and claiming universal validity is simply a mistake. But if one denies that the large system consisting of the gas, the stuff in the laboratory, the physicist doing the experiment, and ultimately the entire universe is one big mechanical systems, then we can't use mechanics to predict that the system prior to the beginning of the experiment is very likely to have been in a state of higher entropy and the need for a cosmological Past Hypothesis evaporates.

Those who hold such a more 'local' view of laws pursue what is known as a 'branch systems approach'. The leading idea is that the isolated systems relevant to SM have neither been in existence forever, nor continue to exist forever after the thermodynamic processes took place. Rather, they separate off from the environment at some point (they 'branch') then exist as energetically isolated systems for a while and then usually merge again with the environment. Such systems are referred to as 'branch systems'. For instance, the system consisting of a glass and an ice cube comes into existence when someone puts the ice cube into the water, and it ceases to exist when someone pours it into the sink. So the question becomes why a branch system like the water with the ice cube behaves in the way it does. An explanation can be given along the lines of the past hypothesis, with the essential difference that the initial low entropy state has to be postulated not for the beginning of the universe but only for the state of the system immediately after the branching. Since the system, by stipulation, did not exist before that moment, there is also no question of whether the system has evolved into the current state from a higher entropy state. This way of looking at things is in line with how working physicists think about these matters for the simple reason that low entropy states are routinely prepared in laboratories.

Irrespective of how this issue is resolved, there are three further issues that need to be addressed. The first is the interpretation of the probabilities in SP\*. So far we have not said anything about how those probabilities should be interpreted. And in fact this is not an easy question. The most plausible interpretation is to interpret these probabilities as Humean chances in Lewis' sense.

The second is Zermelo's recurrence objection, which, roughly, says that entropy cannot always increase because every mechanical system returns arbitrarily close to its initial state after some *finite* time. If directed at the full Second Law this objection is indeed fatal. However, there is no

logical contradiction between recurrence and BL since BL does not require that entropy *always* increase. However, it could still be the case that recurrence is so prevalent at times that the probabilities for entropy increase are no longer high. While there is no in principle reason to rule this out, the standard response to the objection points out that this is something we never experience: according to the Past Hypothesis, the universe is still today in a low entropy state far away from equilibrium and recurrence will therefore presumably not occur within all relevant observation times. This, of course, is compatible with there being periods of decreasing entropy at some later point in the history of the universe. Hence, we simply should not view BL as valid at all times. And this reply also works for those who adopt a branch systems approach since even for a small system like a gas in box the Poincaré recurrence time is larger than the age of the universe (Boltzmann himself estimated that the time needed for a recurrence to occur for a system consisting of a cubic centimeter of air was about  $10^{10^{19}}$  seconds). In sum, we get around Zermelo's objection by giving up not only the strict Second Law, but also the universal validity of BL. This is a high price to pay, but it is the only way to reconcile entropy increase with Poincaré's recurrence.

The third issue is reductionism. We have so far made various reductionist assumptions. We have assumed that the gas really *is* just a collection of molecules, and, more controversially, we have assumed that the Second Law, or some close cousin of it, has to be derivable from the mechanical laws governing the motion of the gas molecules. In philosophical parlance this amounts to saying that the aim of SM is to *reduce* TD to mechanics plus probabilistic assumptions.

What does such a reduction involve? Over the past decades the issue of reductionism has attracted the attention of many philosophers and a vast body of literature on the topic has grown. This enthusiasm did not resonate with those writing on the foundations of SM and the philosophical debates over the nature (and even desirability) of reduction had rather little impact on work done on the foundations of SM (this is true for both the Boltzmannian and Gibbsian traditions). This led to a curious mismatch between the two debates. A look at how reductionism is dealt with in the literature on SM shows that, by and large, there is agreement that the aim of SM is to derive the laws of TD (or something very much like it) from the underlying micro theory. This has a familiar ring to it for those who know the philosophical debates over reductionism. In fact, it is precisely what Nagel declared to be the aim of reduction. So one can say that the Nagelian model of reduction is the (usually unquestioned and unacknowledged) 'background philosophy' of SM.

However, Nagel's theory of reduction is widely claimed to be seriously flawed and therefore untenable. But this puts us into an uneasy situation: here we have a respectable physical theory, but



this theory is based on a conception of reduction that is generally regarded as unacceptable by philosophers. This cannot be. Either the criticisms put forward against Nagel's model of reduction have no bite, at least within the context of SM, or there must be another, better, notion of reduction which can account for the practices of SM. Somewhat surprisingly, this problem has not been recognised in the debate and so we don't know what notion of reductionism is at work in SM.

#### 4. The Gibbs Approach

At the beginning of the Gibbs approach stands a radical rupture with the Boltzmann programme. The object of study for the Boltzmannians is an individual system, consisting of a large but finite number of micro constituents. By contrast, within the Gibbs framework the object of study is a so-called *ensemble*, an uncountably infinite collection of independent systems that are all governed by the same Hamiltonian but distributed over different states. Gibbs introduces the concept as follows:

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

Ensembles are fictions, or mental copies of the one system under consideration; they do not interact with each other, each system has its own dynamics, and they are not located in space and time. Hence, it is important not to confuse ensembles with collections of micro-objects such as the molecules of a gas. The ensemble corresponding to a gas made up of  $n$  molecules, say, consists of an infinite number of copies of the *entire* gas.

Now consider an ensemble of systems. The instantaneous state of one system of the ensemble is specified by one point in its phase space. The state of the ensemble *as a whole* is therefore specified by a density function  $\rho$  on the system's phase space.  $\rho$  is then regarded as a probability density, reflecting the probability of finding the state of a system chosen at random from the entire ensemble in region  $R$  of  $\Gamma$ :  $p(R) = \int_R \rho d\Gamma$ . To make this more intuitive consider the following simple example. You play a special kind of darts: you fix a plank to the wall, which serves as your dart board. For some reason you know that the probability of your dart landing at a particular place on

the board is given by the curve shown in Figure 6. You are then asked what the probability is that your next dart lands in the left half of the board. The answer is  $1/2$  since one half of the surface underneath the curve  $\rho$  is above the left side (and the integral of  $\rho$  over a certain region is just the surface that the curve encloses over that region). In SM  $R$  plays the role of a particular part of the board (in the example here the left half), and  $\rho$  is the probability, but not for a dart landing but for finding a system there.

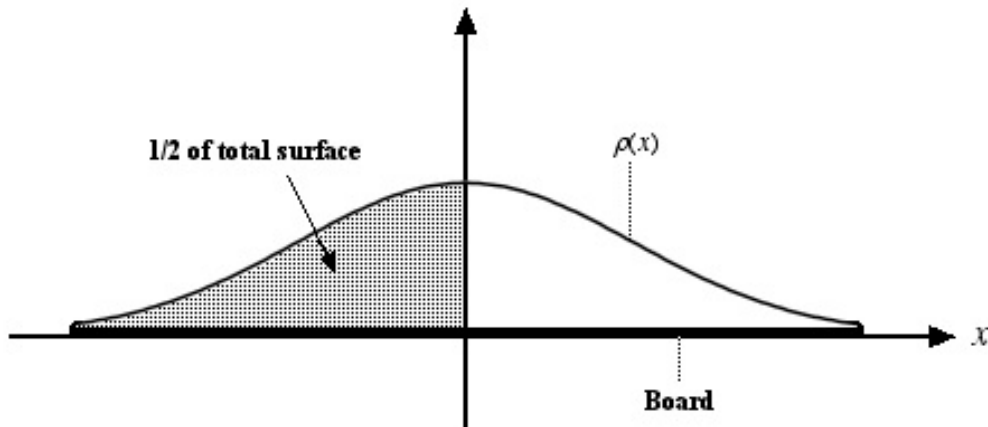


Figure 6: Probability of a dart landing at a particular place on the board.

The importance of this is that it allows us to calculate expectation values. Assume that the game is such that you get one Pound if the dart hits the left half and three Pounds if it lands on the right half. What is your average gain? The answer is  $1/2 \times 1 \text{ Pound} + 1/2 \times 3 \text{ Pounds} = 2 \text{ Pounds}$ . This is the expectation value. The same idea is at work in SM in general. Physical magnitudes like, for instance, pressure, are associated with functions  $f$  on  $\Gamma$  and then we calculate the expectation value, which, in general is given by  $\bar{f} = \int_{\Gamma} f \rho d\Gamma$ . These expectation values, in the context of SM also referred to as *phase averages* or *ensemble averages* are of central importance because it is one of the central posits of Gibbsian SM that these values are what we observe in experiments! So if you want to use the formalism to make predictions, you first have to figure out what the probability distribution  $\rho$  is, then find the function  $f$  corresponding to the physical quantity you are interested in, and then calculate the phase average. Neither of these steps is easy in practice and working physicists spend most of their time doing these calculations. However, these difficulties need not occupy us here.

Given that observable quantities are associated with phase averages and that equilibrium is defined in terms of the constancy of the macroscopic parameters characterising the system, it is natural to regard the stationarity of the distribution as a necessary condition for equilibrium because stationary distributions yield constant averages. For this reason Gibbs refers to stationarity as the 'condition of statistical equilibrium'.

Among all stationary distributions those satisfying a further requirement, the *Gibbsian maximum entropy* principle, play a special role. The Gibbs entropy (sometimes also 'ensemble entropy') is defined as  $S_G(\rho) = -k_B \int_{\Gamma} \rho \log(\rho) d\Gamma$ , where  $\rho$  is the above probability density and  $k_B$  the Boltzmann constant. The Gibbsian maximum entropy principle then requires that  $S_G(\rho)$  be maximal, given the constraints that are imposed on the system.

The last clause is essential because different constraints single out different distributions. A common choice is to keep both the energy and the particle number in the system fixed:  $E = \text{const}$  and  $n = \text{const}$  (while also assuming that the spatial extension of the system is finite). One can prove that under these circumstances  $S_G(\rho)$  is maximal for the so-called the 'microcanonical distribution' (or 'microcanonical ensemble'). If we choose to hold the number of particles constant while allowing for energy fluctuations around a given mean value we obtain the so-called canonical distribution; if we also allow the particle number to fluctuate around a given mean value we find the so-called grand-canonical distribution.

This formalism is enormously successful in that correct predictions can be derived for a vast class of systems. But the success of this formalism is rather puzzling. The first and most obvious question concerns the relation of systems and ensembles. The probability distribution in the Gibbs approach is defined over an ensemble, the formalism provides ensemble averages, and equilibrium is regarded as a property of an ensemble. But what we are really interested in is the behaviour of a single system. What can the properties of an ensemble, a fictional entity consisting of infinitely many copies of a system, tell us about the one real system on the laboratory table? And more specifically, why do averages over an ensemble coincide with the values found in measurements performed on an actual physical system in equilibrium? There is no obvious reason why this should be so.

Common textbook wisdom justifies the use of phase averages as follows. As we have seen the Gibbs formalism associates physical quantities with functions  $f$  on the system's phase space.

Making an experiment to measure one of these quantities takes time. So what measurement devices register is not the instantaneous value of the function in question, but rather its time average over the duration of the measurement; hence time averages are what is empirically accessible. Then, so the argument continues, although measurements take an amount of time that is short by human standards, it is long compared to microscopic time scales on which typical molecular processes take place. For this reason the actually measured value is approximately equal to the *infinite* time average of the measured function. We now assume that the system is ergodic. In this case time averages equal phase averages, and the latter can easily be obtained from the formalism. Hence we have found the sought-after connection: the Gibbs formalism provides phase averages which, by ergodicity, are equal to infinite time averages, and these are, to a good approximation, equal to the finite time averages obtained from measurements.

This argument is problematic for at least two reasons. First, from the fact that measurements take some time it does not follow that what is actually measured are time averages. So we would need an argument for the conclusion that measurements produce time averages. Second, even if we take it for granted that measurements do produce finite time averages, then equating these with infinite time averages is problematic. Even if the duration of the measurement is very long (which is often not the case as actual measurement may not take that much time), finite and infinite averages may assume very different values. And the infinity is crucial: if we replace infinite time averages by finite ones (no matter how long the relevant period is taken to be), then the ergodic theorem does not hold any more and the explanation is false.

These criticisms seem decisive and call for a different strategy. Three suggestions stand out. The first response, due to Malament and Zabell, tackles this challenge by suggesting a way of explaining the success of equilibrium theory that still invokes ergodicity, but avoids altogether appeal to time averages. This avoids the above mentioned problems, but suffers from the difficulty that many systems that are successfully dealt with by the formalism of SM are not ergodic. To circumvent this difficulty Vranas has suggested replacing ergodicity with what he calls  $\varepsilon$ -ergodicity. Intuitively a system is  $\varepsilon$ -ergodic if it is ergodic not on the entire phase space, but on a very large part of it. The leading idea behind his approach is to challenge the commonly held belief that even if a system is just a 'little bit' non-ergodic, then it behaves in a completely 'un-ergodic' way. Vranas points out that there is a middle ground and then argues that this middle ground actually provides us with everything we need. This is a promising proposal, but it faces three challenges. First, it needs to be shown that all relevant systems really are  $\varepsilon$ -ergodic. Second, the argument so far has only been developed for the microcanonical ensemble, but one would like to know whether, and if so how, it

works for the canonical and the grandcanonical ensemble. Third, it is still based on the assumption that equilibrium is characterised by a stationary distribution, which, as we will see below, is an obstacle when it comes to formulating a workable Gibbsian non-equilibrium theory.

The second response begins with Khinchin's work, who pointed out that the problems of the ergodic programme are due to the fact that it focusses on too general a class of systems. Rather than studying dynamical systems at a general level, we should focus on those cases that are relevant in statistical mechanics. This involves two restrictions. First, we only have to consider systems with a large number of degrees of freedom; second, we only need to take into account a special class of phase functions, so-called sum functions, basically functions that are a sum of one-particle functions. Under these assumption Khinchin could prove that as  $n$  becomes larger, the measure of those regions on the energy hypersurface where the time and the space means differ by more than a small amount tends towards zero. Roughly speaking, this result says that for large  $n$  the system behaves, for all practical purposes, as if it was ergododic.

The problem with this result is that it is valid only for sum functions, and in particular only if the Hamiltonian itself is a sum function, which usually is not the case. So the question is how this result can be generalised to more realistic cases. This problem stands at the starting point of a research programme now known as the 'thermodynamic limit'. Its leading question is whether one can still prove 'Khinchin-like' results in the case of Hamiltonians *with* interaction terms. Results of this kind can be proven in the limit for  $n \rightarrow \infty$ , if also the volume  $V$  of the system tends towards infinity in such a way that the numberdensity  $n/V$  remains constant.

Both programmes discussed so far remain silent about the interpretation of probability. So how could probabilities in these theories be understood? There are two obvious choices. The first is some sort of frequentism. A common way of looking at ensembles, suggested by Gibbs himself, is to think about them in analogy with urns, but rather than containing balls of different colours they contain systems in different micro-states. The distribution  $\rho$  then gives the frequency with which we get a system in a certain state when drawing a system at random from the ensemble. First appearances notwithstanding, this is problematic. Ensembles just aren't urns from which one can draw systems at random! They are imaginary constructs and it is unclear at best what sense to make of the notion of drawing a system from an ensemble.

The other way to interpret probabilities are time averages. This is a workable suggestion provided the system is ergodic. Its main problem is that it undercuts an extension of the approach to non-

equilibrium situations (to which we turn below). Interpreting probabilities as infinite time averages yields stationary probabilities. As a result, phase averages are constant. This is what we expect in equilibrium, but it is at odds with the fact that we witness change and observe systems approaching equilibrium departing from a non-equilibrium state. This evolution has to be reflected in a change of the probability distribution, which is impossible if it is stationary by definition.

Discontent with these approaches to probability is the point of departure for the third suggestion, which urges us to adopt an epistemic interpretation of probability. This view can be traced back to Tolman and has been developed into an all-encompassing approach to SM by Jaynes. At the heart of Jaynes' approach to SM lies a radical reconceptualisation of what SM is. On his view, SM is about our knowledge of the world, not about the world itself. The probability distribution represents our state of knowledge about the system at hand and not matters of fact about the system itself. More specifically, the distribution represents our lack of knowledge about a system's micro-state given its macro condition; and, in particular, entropy becomes a measure of how much knowledge we lack.

To put this suggestion onto secure footing, Jaynes uses Shannon's notion of entropy together with the apparatus of information theory in which it is embedded. The leading idea then is that we should always choose the distribution that corresponds to a maximal amount of uncertainty, i.e. is maximally non-committal with respect to the missing information. Since in the continuous case the Shannon entropy has the same mathematical form as the Gibbs entropy, this immediately leads to the prescription to choose the distribution that maximises the Gibbs entropy, which is exactly what the formalism instructs us to do!

This is striking, yet the very notion, as well as the application to SM, of the information theoretic entropy are fraught with controversy, which centres around the question of why it is rational to choose a high entropy distribution in the absence of relevant information. Moreover, interpreting the entropy as an expression of our knowledge rather than a property of the system has the counterintuitive consequence that both entropy and equilibrium are no longer properties of the *system* but rather of our epistemic situation.

So far we have dealt with equilibrium. Let us now turn to the question of how the approach to equilibrium can be explained from a Gibbsian point of view. Unfortunately there are formidable obstacles. The first is that it is a consequence of the formalism that the Gibbs entropy is a constant! This precludes a characterisation of the approach to equilibrium in terms of increasing Gibbs

entropy, which is what one would expect if we were to treat the Gibbs entropy as the SM counterpart of the thermodynamic entropy.

The second problem is the characterisation of equilibrium in terms of a stationary distribution. It is a mathematical matter of fact that the Hamiltonian equations of motion, which govern the system, preclude an evolution from a non-stationary to a stationary distribution: if, at some point in time, the distribution is non-stationary, then it will remain non-stationary for all times and, conversely, if it is stationary at some time, then it must have been stationary all along. For this reason a characterisation of equilibrium in terms of stationary distributions contradicts the fact that an approach to equilibrium takes place in systems that are not initially in equilibrium. Clearly, this is a *reductio* of a characterisation of equilibrium in terms of stationary distributions.

Hence the main challenge for Gibbsian non-equilibrium theory is to find a way to get the Gibbs entropy moving, and to characterise equilibrium in a way that does not preclude change in the system. This can be done in different ways, and there is indeed a plethora of approaches offering distinct solutions. Coarse graining, interventionism, stochastic dynamics, the Brussels School, and the BBGKY hierarchy. These theories are beyond the scope of this introduction.

Let me end this section with some remarks about reductionism. The conceptual problems in connection with reductionism mentioned above remain also crop up in the Gibbs framework. But on top of these, there are some issues that are specific to this framework.

Boltzmann took over from TD the notion that entropy and equilibrium are properties of an individual system and sacrificed the idea that equilibrium (and the associated entropy values) are stationary. Gibbs, on the contrary, retains the stationarity of equilibrium, but at the price of making entropy and equilibrium properties of an ensemble rather than an individual system. This is because both equilibrium and entropy are defined in terms of the probability distribution  $\rho$ , which is a distribution over an ensemble and not over an individual system. Since a particular system can be a member of many different ensembles one can no longer assert that an individual system is in equilibrium. This 'ensemble character' carries over to other physical quantities, most notably temperature, which are also properties of an ensemble and not of an individual system.

This is problematic because the state of an individual system can change considerably as time evolves while the ensemble average does not change at all; so we cannot infer from the behaviour of an ensemble to the behaviour of an individual system. However, what we are dealing with in

experimental contexts are individual systems; and so the shift to ensembles has been deemed inadequate by many.

It is worth observing, however, that Gibbs himself never claimed to have reduced TD to SM and only spoke about 'thermodynamic analogies' when discussing the relation between TD and SM. The notion of analogy is weaker than that of reduction, but it is at least an open question whether this is an advantage. If the analogy is based on purely algebraic properties of certain variables then it is not clear what, if anything, SM contributes to our understanding of thermal phenomena; if the analogy is more than a merely formal one, then at least some of the problems that we have been discussing in connection with reduction are bound to surface again.

## **5. Conclusion**

Even if all the inherent problems of the Boltzmannian and the Gibbsian approach could be solved, there would remain one big bad bug: the very existence of two different frameworks. One of the foremost problems of the foundation of SM is the lack of a generally accepted and universally used formalism, which leads to a kind of schizophrenia in the field. The Gibbs formalism has a wider range of application and is therefore the practitioner's workhorse. In fact, virtually all practical applications of SM are based on the Gibbsian machinery. The weight of successful applications notwithstanding, a consensus has emerged over the last decade and a half that the Gibbs formalism cannot explain why SM works and that when it comes to foundational issues the Boltzmannian approach is the only viable option. Hence, whenever the question arises of why SM is so successful, an explanation is given in Boltzmannian terms.

So we are in the odd situation that we have one formalism to answer foundational questions, and another one for applications. This would not be particularly worrisome if the two formalisms were intertranslatable or equivalent in some other sense (like, for instance, the Schrödinger and the Heisenberg picture in quantum mechanics). However, as we have seen above, this is not the case. The two frameworks disagree fundamentally over what the object of study is, the definition of equilibrium, and the nature of entropy to mention just a few. So even if all the internal difficulties of either of these approaches were to find a satisfactory solution, we would still be left with the question of how the two relate.



A suggestion of how these two frameworks could be reconciled has recently been presented by Lavis. His approach involves the radical suggestion to give up the notion of equilibrium, which is binary in that systems either are or not in equilibrium, and to replace it by the continuous property of 'commonness'. Whether this move is justified and whether it solves the problem is a question that needs to be discussed in the future.

## Acknowledgements

I would like to thank Soazig Le Bihan, Martin Frigg, and Stefani Thrasyvoulou and Radin Dardashti for helpful comments on earlier drafts; special thanks to Radin for drawing the figures and preparing the manuscript for publication.

## Glossary

**Entropy:** Physical property of a system which reaches its maximum if the system is in equilibrium.

**Equilibrium:** The state of a system in which all its macro-properties (such as temperature, pressure and volume) assume constant values (i.e. do not change over time). In statistical mechanics this state is characterized as having maximum entropy.

**Second Law of Thermodynamics:** The proposition that The that entropy in closed systems cannot decrease.

**Statistical Mechanics:** The study of the connection between micro-physics and macro-physics, which aims to account for system's macroscopic behaviour in terms of the dynamical laws governing its microscopic constituents and probabilistic assumptions.

## Nomenclature

**CM:** Classical Mechanics

$\bar{f}$  : Phase or ensemble average of function  $f$

$\phi$  : Phase flow

$\Gamma$  : Phase space

$k_B$  : Boltzmann constant

$M_i$  : Macro states

$M_{eq}$  : Equilibrium macro state

$\mu$  : Lebesgue-measure

$S_B$  : Boltzmann Entropy

$S_G$  : Gibbs Entropy

**SM:** Statistical Mechanics

**SP:** Statistical Postulate

**TD:** Thermodynamics

$X$  :State of the system

## Bibliography

Albert, David (2000). *Time and Chance*. Cambridge/MA and London: Harvard University Press. [Provides a contemporary formulation of Boltzmannian SM and introduces the Past Hypothesis.]

Boltzmann, Ludwig (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. Reprinted in F. Hasenöhl (ed.): *Wissenschaftliche Abhandlungen*. Leipzig: J. A. Barth 1909, Vol. 2, 164-223. [The *locus classicus* of what is now known as Boltzmannian SM.]

Callender, Craig (1999). Reducing Thermodynamics to Statistical Mechanics: The Case of Entropy. *Journal of Philosophy* 96, 348-373. [Provides an in-depth discussion of the problems occurring when reducing TD to SM with a special focus on entropy.]

Dizadji-Bahmani, Foad, Roman Frigg and Stephan Hartmann (2010). Who's Afraid of Nagelian Reduction? *Erkenntnis* 73, 393-412. [Gives a contemporary formulation of Nagelian reduction and defends it against criticisms.]

Earman, John (2006). The "Past Hypothesis": Not Even False. *Studies in History and Philosophy of Modern Physics* 37, 399-430. [Offers a sustained criticism of the Past Hypothesis.]

Earman, John & Miklós Rédei (1996). Why Ergodic Theory Does Not Explain the Success of Equilibrium Statistical Mechanics. *British Journal for the Philosophy of Science* 47, 63-78. [Critically discusses the ergodic programme and argues that ergodicity does not explain the success of SM.]

Emch, Gérard (2007). Quantum Statistical Physics. In: Jeremy Butterfield & John Earman (eds.): *Philosophy of Physics*. Amsterdam: North Holland, 1075-1182. [Provides a detailed survey of quantum SM.]

Frigg, Roman (2008). A Field Guide to Recent Work on the Foundations of Statistical Mechanics. In: Dean Rickles (ed.): *The Ashgate Companion to Contemporary Philosophy of Physics*. London: Ashgate, 99-196. [Gives a systematic survey of problems and issues in the foundation of SM.]

Frigg, Roman (2009). Typicality and the Approach to Equilibrium in Boltzmannian Statistical Mechanics. *Philosophy of Science* (Supplement) 76, 2009, 997–1008. [Provides a careful formulation of the typicality approach to SM and outlines the problems that it faces.]

Gibbs, J. Willard (1902). *Elementary Principles in Statistical Mechanics*. Woodbridge: Ox Bow Press 1981. [The *locus classicus* of what is now known as Gibbsian SM]

Goldstein, Sheldon (2001). Boltzmann's Approach to Statistical Mechanics. In: Detlef Dürr, Maria Carla Galavotti, GianCarlo Ghirardi, Francesco Petruccione & Nino Zanghí (eds.) (2001): *Chance in Physics: Foundations and Perspectives*. Berlin and New York: Springer, 39-54. [Presents a contemporary formulation of SM with a special focus on typicality.]

Jaynes, Edwin T. (1983). *Papers on Probability, Statistics, and Statistical Physics*. Ed. by R. D. Rosenkrantz. Dordrecht: Reidel. [Reformulates SM in information-theoretic terms and makes the radical claim that SM should be considered a branch of statistics.]

Khinchin, Alexander I. (1949). *Mathematical Foundations of Statistical Mechanics*. Mineola/NY: Dover Publications 1960. [Lays the foundation for the programme that is now known as the thermodynamic limit.]

Lavis, David (2005). Boltzmann and Gibbs: An Attempted Reconciliation. *Studies in History and Philosophy of Modern Physics* 36, 245-273. [Presents a reconciliation of the Boltzmannian and the Gibbsian approach to SM.]

Lewis, David (1994). Humean Supervenience Debugged. *Mind* 103, 473-90. [Gives the canonical formulation of the interpretation of probability that is now known as Humean objective chance.]

Malament, David & Sandy L. Zabell (1980). Why Gibbs Phase Averages Work. *Philosophy of Science* 47, 339-349. [Present an explanation of Gibbsian equilibrium SM based on ergodicity.]

Sklar, Lawrence (1993). *Physics and Chance. Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press. [Provides a comprehensive review of the different approaches in the foundation of SM.]

Tolman, Richard C. (1938). *The Principles of Statistical Mechanics*. Mineola/New York: Dover 1979. [A classical text on SM and the *locus classicus* for the epistemic interpretation of probabilities in SM.]

Uffink, Jos (2007). Compendium of the Foundations of Classical Statistical Physics. In: Jeremy Butterfield & John Earman (eds.): *Philosophy of Physics*. Amsterdam: North Holland, 923-1047. [A detailed and in-depth review of different schools of thought in the foundation of SM from a historical point of view.]

Vranas, Peter B. M. (1998). Epsilon-Ergodicity and the Success of Equilibrium Statistical Mechanics. *Philosophy of Science* 65, 688-708. [Introduces the notion of  $\epsilon$ -ergodicity into the discussion of SM and shows how this notion can be used to solve some problems that arise in Malament & Zabell's theory.]