What statistical mechanics actually does

David Wallace

June 21, 2013

Abstract

I give a brief account of the way in which thermodynamics and statistical mechanics actually work as contemporary scientific theories, and in particular of what statistical mechanics contributes to thermodynamics over and above any supposed underpinning of the latter's general principles. In doing so, I attempt to illustrate that statistical mechanics should not be thought of wholly or even primarily as itself a foundational project for thermodynamics, and that conceiving of it this way potentially distorts the foundational study of statistical mechanics itself.

1 Introduction

Classical equilibrium thermodynamics is characterised by laws of great generality and scope, but which have no justification within thermodynamics itself. Historically, much of the original impetus of statistical mechanics was to provide a microphysical justification: almost from its outset, the subject had to grapple with the apparent inconsistency between the apparent time-irreversibility of thermodynamics and the apparent time-reversibility of microphysics.

If a 19th-century physicist were transported to the present day and perused the philosophical literature on statistical mechanics, they would be forgiven for thinking that little has changed. Overwhelmingly, the focus of discussion is on the use of statistical-mechanical methods to underpin thermodynamics. For instance, Roman Frigg's recent review states that

> Thermodynamics (TD) correctly 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. ... The fact that many processes in the world are irreversible is enshrined in the so-called Second Law of Thermodynamics ... It is the aim of non-equilibrium [statistical mechanics] to give a precise characterization of irreversibility and to provide a microphysical explanation of why processes in the world are in fact irreversible. (Frigg 2007, pp.99-100)

Craig Callender (2001), slightly more cautiously, observes that "Kinetic theory and statistical mechanics are *in part* attempts to explain the success of thermodynamics in terms of the basic mechanics." (p.540; emphasis mine.) In similar vein, Katinka Ridderbos (2002) notes that "One of the cardinal aims of the theory of statistical mechanics is to underpin thermodynamic regularities by a theory formulated in terms of the dynamical laws governing the motion of the microscopic constituents of a thermodynamic system." (p.66) Examples could easily be multiplied.

Notwithstanding Callender's and Ridderbos' caveats, this essentially *foundational* construal of statistical mechanics is dominant in philosophical discussion. The field is presented as concerned primarily with providing a microscopic underpinning of already-known macroscopic generalities; the point of philosophical concern is whether it does so adequately.

Part of the point of this paper is to suggest that this focus on the foundational role of statistical mechanics is in danger of distorting the discussion. Statistical mechanics is both a huge field in its own right in contemporary physics and a hugely important tool across many (most?) other areas of physics, and only a very small part of the work done under the label of statistical mechanics has anything much to do with the foundations of thermodynamics. Insofar as there are important questions to ask about the conceptual underpinnings of statistical mechanics, it may be misleading to regard statistical mechanics itself as itself wholly or primarily a conceptual underpinning for thermodynamics. I mostly use the neo-Boltzmann approach (advocated recently by, inter alia, Albert (2000), Callender (2001), Goldstein (2001), Lebowitz (2007), North (2002) and Penrose (1994, 1989)) to illustrate where this can be significant.

But the main point of the paper is just to give an overview of what statistical mechanics, as used in contemporary physics, actually does, over and above its supposed foundational role — something that seems to be rather little understood in foundational circles. I make no pretense at conceptual or mathematical rigor: I attempt simply to lay out what the actual methods and (broad-level) techniques of thermodynamics, and of equilibrium and then non-equilibrium statistical mechanics, actually are. I take it that any satisfactory conceptual account of statistical mechanics must succeed not simply at underpinning the general predictions of thermodynamics, but the predictive and explanatory successes of statistical mechanics itself.

2 The content of thermodynamics

What does classical thermodynamics actually tell us about the systems to which it applies? Roughly speaking (that is, with no pretensions to completeness, historical accuracy, conceptual independence or mathematical precision), something like the following:

- That each system, if isolated, relaxes in some reasonable time towards an *equilibrium state* whose "thermodynamic parameters" (roughly speaking,

its macroscopically accessible features) are time-invariant and are determined only by its internal energy U and by whatever external parameters (paradigmatically, its volume V) constrain it. (This equilibration principle is called the 'minus first law of thermodynamics' by Brown and Uffink (2001).)

- That any two such systems can be placed in 'thermal contact', whereby they may be treated as parts of a single system that will reach a joint state of equilibrium.

- That the relation of 'being in equilibrium with', that holds between any two systems at equilibrium which remain unchanged when placed in thermal contact, is an equivalence relation, and that the relation of 'being hotter than', which holds between two systems when energy is transferred from the first to the second when they are placed in thermal contact, is an ordering relation. (The Zeroth Law.)

- That, in part as a consequence of the above, we can define 'empirical temperatures', functions of an equilibrium system's internal energy and external constraints, so that system 1 has a higher temperature than system 2 iff it is hotter.

- That the energy transferred to a system as a result of its transition between two equilibrium states can be consistently divided into 'work', which is energy transferred via variation of the external parameters, and 'heat', which is energy transferred via thermal contact, and that energy is conserved, so that the change in internal energy of a system equals the net work done on the system by varying its parameters plus the net heat flowing into the system from other systems in which it is in thermal contact. (The First Law.)

- That it is possible to speak consistently of arbitrarily small and slow transitions of a system between equilibrium states, so that the infinitesimal change of energy of the system dU in such a transition can be decomposed as

$$dU = \delta Q + \delta W, \tag{1}$$

  where δQ and δW are the infinitesimal work done, and heat transferred, in the transition.

- That the work δW can be decomposed as

$$\delta W = \sum_i P_i dV^i,$$

  where the V^i are the external constraints on the system and the P_i are functions of an equilibrium system's internal energy and external constraints, which can be defined as the rate of change of U with respect to V^i while the system is thermally isolated.

- That there exist functions S ('thermodynamic entropy') and T ('thermodynamic temperature') of an equilibrium system's internal energy and external constraints, such that

$$\delta Q = TdS,$$

  such that T is an empirical temperature, and such that no transition of a thermally isolated system between two equilibrium states can induce a decrease in S. (The Second Law.)

Famously, the above principles (collectively speaking) are *primitives* of thermodynamics: although there are various logical interconnections between them, the system as a whole is merely postulated, and thermodynamics in itself offers no justification for them. But never mind: let us accept them, and take for granted that all of the above is known to hold of, say, a box of gas of known total volume and external energy, and ask: what can be deduced about the the behaviour of the box?

The answer, so far as I can see, is virtually nothing. The box will have *some* equilibrium state, to which it will relax on some unspecified timescale. Increasing or decreasing its volume may (or may not) lead to changes in its internal energy. It will not be possible to use the box to play certain roles in various heat engines: it will not, for instance, be possible to operate on it in a cycle to turn heat into work. It will have some thermodynamical temperature, and if placed in thermal contact with a lower-temperature system, will transfer heat to that system. It will have some entropy, which cannot be induced to decrease in an adiabatic process. But on what the temperature is, or the entropy, for a given volume; on how much work must be done (or will be generated) in contracting the box; on even whether the box is of uniform density...on all these questions, thermodynamics in the abstract is silent. The Second Law, or the First, or the Zeroth, or the Minus First, or all of them together, do not so much as predict that a box of gas initially confined to one half of a box will expand to occupy the whole box.

Nor can it be expected to: the very neutrality of thermodynamics forbids it. Thermodynamics is intended to apply to gases, to liquids and to solids: to boxes filled with plasma, treacle or stone as surely as boxes filled with gas. And it is not a law of nature that a chunk of rock, initially 'confined' to one side of a box by a partition, will expand to fill the whole box.

Thermodynamics only begins to get its bite when its *abstract* principles are supplemented by physical details that pick out the particular system under study. This is normally done via the *equations of state* of the system: the concrete mathematical expressions for the functions P_i and T in terms of the energy U and the constraints V^i. In the case of a gas, for instance, the only salient constraint is the total volume V (note that this already tells us a lot about the gas, for instance that its macrophysics depends on the size of its container but not on the shape) and, in the idealisation of a sufficiently dilute

gas, the equation of state for the parameter $P_1 = P$ is

$$P = \frac{1}{\alpha}\frac{U}{V}, \tag{2}$$

where α is a dimensionless parameter dependent upon the species of gas, and the equation of state for T is

$$T = \frac{1}{\alpha}\frac{m}{MR}U, \tag{3}$$

where M is the mass of the gas, R is an arbitrary scale constant (in fact equal to $\sim 8.3JK^{-1}$), and m is a parameter with the dimensions of mass, dependent on the species of gas (the physical significance of which will be explored shortly, though the reader can probably guess). Given these, we can solve for S, obtaining

$$S = \frac{MR}{m}\left(\ln(V) + \alpha\ln(U)\right) + S_0. \tag{4}$$

Other systems have different equations of state. A box of radiation, for instance, has again volume as its one salient external parameter, and has equations of state

$$P = \frac{1}{4}\frac{U}{V} \tag{5}$$

and

$$T = \left(\frac{c}{4\sigma}\frac{U}{V}\right)^{1/4}, \tag{6}$$

where for the moment c/σ should be thought of as a empirical constant (Blundell and Blundell 2010, p.286). A crystal at very low temperatures has no salient external constraints and can often be treated as having equation of state approximately

$$U = 9\frac{12\pi^4}{5}\frac{MR}{m}T\left(\frac{T}{T_D}\right)^3, \tag{7}$$

where T_D is the *Debye temperature*. In many cases, the equations of state cannot be given in any simple form: the appendices of engineering thermodynamics textbooks (at least those from the pre-Internet era) tend to contain tabulations of the equations of state for the various phases of water.

Without the equations of state of a given system, thermodynamics is largely inert; with them, it has considerable power. But where do the equations of state come from? As far as classical thermodynamics is concerned: from phenomenology, and phenomenology alone.

The equations of state, however, do not by any means tell us *everything* about an equilibrium system — even everything that is macroscopically accessible. Is the average density of an equilibrium system constant? Sometimes — when the system is *extensive* — the equation of state can be manipulated into telling us this, but it is not a law of nature that thermodynamic systems are extensive. What is the frequency distribution of radiation in equilibrium in a box? The particle velocity distribution in an ideal gas? Thermodynamics, even given the equations of state, is silent.

3 Enter statistical mechanics

To summarise the previous section, thermodynamics leaves the following three questions unanswered:

1. Why are the basic postulates of the theory correct?
2. What are the equations of state of any given system?
3. What are the properties of a system at equilibrium, over and above those specified by its equations of state?

Foundational and philosophical work on statistical mechanics tends to cast it as an attempt to answer only the first question: an attempt, what is more, whose success is questionable and controversial. But in fact, statistical mechanics seeks to answer all three questions, and whatver its deficiencies at the first, it has obtained remarkable successes at the second and third. That is: equilibrium statistical mechanics provides us with a calculational method to generate equations of state, and indeed pretty much all other measurable properties of equilibrium systems, given only the microdynamics of those systems. And applying this method has been systematically successful, both in reproducing already-known equations of state and in deriving new ones. Whether or not we understand *why* it works, there can be little doubt *that* it works.

It will be helpful to review how the method is applied. Put aside, for the moment, why any of this is justified; assume, if you like, that physicists follow it unthinkingly and dogmatically. (More seriously, put aside legitimate worries that the method is usually applied in the quantum domain: here, for simplicity, I work purely classically.) The method comes in two flavours: *microcanonical* and *canonical*.

The method of equilibrium statistical mechanics (microcanonical form)

1. Represent the system by a classical phase space Γ and by a Hamiltonian $H[V_1, \ldots V_N]$ that is a functional of the external constraints of the system.
2. Define the *microcanonical ensemble*, for a given energy U and constraint values $V_1, \ldots V_N$, as the probability distribution obtained by restricting the uniform (Liouville) probability distribution to a shell of width $\delta U \ll U$ around the hypersurface $H[V_1, \ldots V_N](x) = U$ and then renormalising.
3. To determine any present (empirically accessible) feature of the system at equilibrium, calculate that feature as if the system's probability of being in a given region of phase space is given by the microcanonical ensemble.
4. Calculate the thermodynamical entropy of the system by taking the logarithm of the volume of the region on which the microcanonical ensemble is defined (and then multiplying by k_B, a scale constant); equivalently, define it as
$$S = -k_B \int_\Gamma \mathrm{d}\mu\,(x)\rho(x)\ln\rho(x) \tag{8}$$

where ρ is the microcanonical distribution and μ is Liouville measure.

5. Calculate the thermodynamic temperature of the system as the rate of change of energy with thermodynamic entropy (keeping the constraints constant).

The method of equilibrium statistical mechanics (canonical form)

1. Represent the system by a classical phase space Γ and by a Hamiltonian $H[V_1, \dots V_N]$ that is a functional of the external constraints of the system.

2. Define the *canonical ensemble*, for a given temperature T and constraint values $V_1, \dots V_N$, as the probability distribution

$$\rho(x) = \frac{1}{Z(T, V_1, \dots V_N)} \exp(-H[V_1, \dots V_N](x)/k_B T) \tag{9}$$

where Z (the *partition function*) is a normalisation factor, given by

$$Z(T, V_1, \dots V_N) = \int_\Gamma \mathrm{d}\mu\,(x) \exp(-H[V_1, \dots V_N](x)/k_B T). \tag{10}$$

3. To determine any present (empirically accessible) feature of the system at equilibrium (including the energy), calculate that feature as if the system's probability of being in a given region of phase space is given by the canonical ensemble.

4. Calculate the thermodynamical entropy of the system via

$$S = -k_B \int_G amma \mathrm{d}\mu\,(x) \rho(x) \ln \rho(x) \tag{11}$$

where ρ is the canonical distribution. (Equivalently, define it as $S = -k_B \ln(Z) + \langle H \rangle / T$, where $\langle H \rangle$ is the expectation value of the Hamiltonian.)

The microcanonical ensemble is intended to be used for systems that are thermally isolated from their environments; the canonical ensemble for systems in thermal contact with a heat bath. In practice (and for reasonably-well-understood mathematical reasons) the two methods give virtually identical results for macroscopic-scale systems, and the choice of which to use is mostly a matter of convenience.

In either case, knowledge of S as a function of U suffices to determine the equation of state. For a dilute gas of N point particles, for instance (in the approximation where we neglect interactions) the phase-space volume occupied by the microcanonical ensemble is

$$\mathrm{Vol} \propto V^N \times U^{3N/2-1} \delta U, \tag{12}$$

so that (when $N \gg 1$) to a very good approximation (and up to a constant term)

$$S(U,V) = k_B(N\ln(V) + \frac{3N}{2}\ln(U)). \tag{13}$$

From this we get

$$dS = k_B(\frac{N}{V}dV + \frac{3N}{2U}dU), \tag{14}$$

and can read off $T = 2U/3Nk_B$ and $P = 2U/3V$. Not only have we recovered the equation of state, we have determined the values of the parameters α and m: the former is $3/2$, the latter is the mass of a single particle. (For a dilute gas whose particles have M internal degrees of freedom, α can again be calculated and (in the classical limit) equals $(3+M)/2$.)

To go beyond the equation of state (again via the microcanonical ensemble), label the $3N$ independent momenta of the particles $x_1, \ldots x_{3N}$, and consider the probability density that the first M have momenta around $p_1, \ldots p_M$. For the ideal gas (cancelling a position integral), this is

$$\begin{aligned}
&\Pr(x_1 = p_1, \ldots x_M = p_M) \\
&= \frac{\int \mathrm{d}x_1 \cdots \mathrm{d}x_{3N}\, \delta(x_1 - p_1)\cdots\delta(x_M - p_M)\delta(\sum_{i=1}^{3N} x_i^2/2m - U)}{\int \mathrm{d}x_1 \cdots \mathrm{d}x_{3N}\, \delta(\sum_{i=1}^{3N} x_i^2/2m - U)} \\
&= \frac{\int \mathrm{d}x_{M+1} \cdots \mathrm{d}x_{3N}\, \delta(\sum_{i=M+1}^{3N} x_i^2/2m - (U - \sum_{i=1}^{M} p_i^2/2m))}{\int \mathrm{d}x_1 \cdots \mathrm{d}x_{3N}\, \delta(\sum_{i=1}^{3N} x_i^2/2m - U)}.
\end{aligned} \tag{15}$$

The numerator and denominator are, respectively, the surface areas of spheres in $3N - M$ and $3N$ dimensional space, so we get

$$\Pr(x_1 = p_1, \ldots x_M = p_M) \propto \frac{(2mU - \sum_{i=1}^{M} p_i^2)^{(3N-M-1)/2}}{(2mU)^{(3N-1/2)}}. \tag{16}$$

Assuming $M \ll N$, this simplifies to

$$\Pr(x_1 = p_1, \ldots x_M = p_M) \propto \left(1 - \sum_{i=1}^{M}(p_i^2/2mU)\right)^{(3N-M-1)/2}, \tag{17}$$

which via the approximation $(1 - x/N)^N \simeq e^{-x}$ for large N (and via the known value for T) simplifies to

$$\Pr(x_1 = p_1, \ldots x_M = p_M) \propto \Pi_{i=1}^{M} \exp(-p_i^2/2mk_BT). \tag{18}$$

That is, the probability distributions over momenta are uncorrelated and the one-particle distribution is the familiar Maxwell-Boltzmann distribution. As such, with extremely high probability the fraction of particles found with a given momentum will be given by this distribution.

Similar results can be obtained for the radiation box (including a calculation of the value of σ), for crystals (including a calculation of the Debye temperature), and for much-less-familiar systems. Indeed, the great bulk of active theoretical research in equilibrium statistical mechanics is concerned with studying the properties of equations of state and microphysical distributions of physical systems, ranging from magnets to plasmas and beyond.

Nor are the successes of statistical mechanics restricted to the expected value of physical parameters: that is, their values averaged over the appropriate ensemble. We can also apply the framework to study *fluctuation phenomena*. On a series of measurements of some physical quantity A, for instance, the expected value of the measurement is given by $\langle A\rangle = \int A\rho$, but we can also measure the *variance* of this measurement, and compare it to its predicted value of $\int (A - \langle A\rangle)^2\rho$. Again, statistical mechanics delivers: fluctuation phenomena are widely observed and conform to the predictions of the equilibrium apparatus.

Finally, nothing in the rules of statistical mechanics prevent it being applied to systems with only a few degrees of freedom, and indeed this is routinely done: systems analysed in this way are of course predicted to have very large levels of fluctuation, and these predictions, too, are routinely confirmed. To take a simple example, the dilute gas can equally well be treated by regarding each particle as in thermal contact with a heat bath comprised of all the other particles. On this basis, the canonical distribution just is the Maxwell-Boltzmann distribution, in agreement with the result we previously derived from the microcanonical treatment of the gas as a whole.

Of course, asserting *that* this machinery correctly predicts the form, and correctly calculuates the coefficients, of equations of state, particle velocity distributions, and fluctuation phenomena, in no way explains *why* it does so. Indeed, this is rather the point: any acceptable account of the *foundations* of statistical mechanics not only has to underpin the general structure of equilibrium thermodynamics (including offering some micro-based function that plays the abstract role of the entropy), it has to reproduce the success of the full statistical-mechanical machinery, including its derivations of the equations of state, and in particular, it has to get the *numerical value of the entropy* right. Thermodynamical entropy is not just some abstractly characterised non-decreasing function: it is (up to a constant) an empirically measureable quantity, and any microphysical analysis of it has to correctly reproduce its actual value.

To summarise: the three questions with which we begin generate three questions for the *foundations* of statistical mechanics:

1. Why are the basic postulates of thermodynamics correct?

2. Why does the (micro)canonical-distribution method correctly determine the equations of state of a system?

3. Why does the (micro)canonical-distribution method correctly determine the properties of a system at equilibrium, over and above those specified by its equations of state?

To this we can now add a fourth (although it could arguably be regarded as part of the third):

4. Why does the (micro)canonical-distribution method correctly determine the fluctuations in the properties of a system at equilibrium?

4 Case study: Boltzmann's characterisation of equilibrium

In Boltzmann's approach to statistical mechanics, it is assumed that:

- A system's phase space is divided into *macrostates*, such that any two phase-space points in a given macrostate have the same or virtually the same macroscopic properties.
- For a sufficiently large system, there is one macrostate (for given values of the energy and the external constraints) whose volume is overwhelmingly larger than all of the others of the same energy and for the same constraints; this macrostate is called the *equilibrium macrostate*.
- The *Boltzmann entropy* of a microstate is k_B times the logarithm of the phase-space volume of the macrostate in which that microstate is located.
- A system is said to be at equilibrium if its microstate lies in the equilibrium macrostate; at equilibrium, the thermodynamic entropy is identified with the Boltzmann entropy.
- Very general features of the dynamics are (it is supposed) heuristically likely to cause the system to evolve from a given macrostate into one of larger volume, so that Boltzmann entropy is in general non-decreasing, and that in particular, the system is very likely to evolve in fairly short order into the equilibrium macrostate.

Note that probability assumptions are heavily suppressed in this framework, although there is at least some appeal to probability tacit in the last assumption.

How does the Boltzmann framework do at answering the four questions posed in the previous section?

(1) Why are the basic postulates of the theory correct?

An explanation for equilibration is at the heart of the Boltzmann framework; if we grant the various dynamical assumptions on which the framework rests, the approach to equilibrium seems to be explained. The Second Law is not so often discussed, but on the prima facie plausible assumption that adiabatic transformations between equilibrium macrostates are supposed to be represented by volume-preserving flows, the Boltzmann entropy has to be non-decreasing under those flows.

On the other hand, this account seems to get no particular leverage on the approach to equilibrium for systems with relatively small numbers of degrees of freedom, where no "macro"state will be wildly larger than another. It might be objected, of course, that such systems cannot reasonably be treated as thermodynamic, so that the concept of "equilibrium" does not apply to them.

(2) Why does the (micro)canonical-distribution method correctly determine the equations of state of a system?

Given the assumption that the equilibrium region occupies the overwhelming majority of the energy hypersurface in phase space, the Boltzmann entropy is extremely close numerically to the logarithm of the volume of the *whole* energy hypersurface: that is, it is extremely close to the entropy as calculated according to the microcanonical method. If we grant that Boltzmann entropy at equilibrium really does play the role of thermodynamic entropy (that is, if we grant the Boltzmannian answer to (1)), then it follows that the microcanonically-calculated entropy is indeed the thermodynamic entropy. From this, the equations of state follow.

(3) Why does the microcanonical- or canonical-distribution method correctly determine the properties of a system at equilibrium, over and above those specified by its equations of state?

The overwhelming majority of the energy hypersurface is (*ex hypothesi*) contained within the equilibrium macrostate. So averaging over the whole hypersurface (the recipe given by the microcanonical method) is to a very close approximation the same as averaging over the equilibrium macrostate. But (again *ex hypothesi*) any macroscopically measurable property of the system is constant, or nearly so, within any given macrostate. So averaging a macroscopic property over the whole macrostate gives the same answer as just picking an arbitrary microstate and evaluating the property on that macrostate.

On the other hand, statistical mechanics can also be applied to systems small enough that the probability distribution is not overwhelmingly concentrated in this fashion. In these cases, the predictions of statistical mechanics become probabilistic, and the Boltzmannian account cannot straightforwardly reproduce them.

(4)Why does the (micro)canonical-distribution method correctly determine the fluctuations in the properties of a system at equilibrium? Here the Boltzmannian method simply fails. Fluctuations by their nature occur only when we cannot idealise the probability of the largest macrostate to be 1. Knowing that the system is "overwhelmingly likely" to have particular values of macroscopic quantities does not in itself give us the machinery to quantify the expected level of deviations from those values.

To summarise: the Boltzmannian approach (once we grant its dynamical assumptions) seems to have the resources to explain the success of statistical mechanics under the approximation that its probabilities are zero or one, and this approximation is fairly reasonable for sufficiently large systems. But in many cases statistical mechanics makes explicitly *probabilistic* predictions, and in this situation the Boltzmannian approach is silent.

It is, I think, relatively straightforward to see how the approach would have to be supplemented to meet this challenge. At present, the Boltzmannian relies on this assumption:

> **Qualitative equilibrium assumption:** after a certain period of time (the 'equilibration timescale'), the system is overwhelmingly likely to be in that macrostate that occupies an overwhelming proportion of the appropriate energy hypersurface.

To properly allow for probabilistic predictions, and in particular for fluctuation phenomena, this has to be strengthened to something like the

> **Quantitative equilibrium assumption:** after a certain period of time (the 'equilibration timescale'), the probability of the system being in any given macrostate within the appropriate energy hypersurface is proportional to the volume of that macrostate.

The quantitative assumption suffices to (a) ground the success of the microcanonical approach (putting aside the canonical approach for the moment) and (b) recover the qualitative assumption as a special case. But the intuition that is generally taken to underpin Boltzmannian statistical mechanics ("basically all the points on the energy hypersurface are in the equilibrium region, so unless the dynamics, or the initial state, are ridiculously special, the system is going to make its way to the equilibrium region and stay there for a ridiculously long time") does not *straightforwardly* serve to underpin the quantitative assumption. The moral, I take it, is that probabilities are not introduced simply as part of the *foundations* of statistical mechanics: that is, as a foundational project intended to ground deterministic macropredictions. Rather, generically speaking the *outputs* of statistical mechanics are probabilistic results, and they reduce to deterministic predictions only for large systems.

Of course, in any case "intuition" should not be the right underpinning for the approach to equilibrium, which is ultimately a dynamical process — and, just as in the equilibrium case, a dynamical process whose foundations are less secure than its calculational outputs.

5 Non-equilibrium statistical mechanics

Reading foundational or philosophical work on non-equilibrium statistical mechanics can give the following impression:

1. What is *known* is that systems evolve to equilibrium over some reasonably-short timescales.

2. What is *needed* is an account of the dynamics of non-equilibrium systems which predicts that they approach equilibrium.

3. The *success conditions* on such an account are (a) that it indeed predicts that systems go to equilibrium; (b) that it is based on conceptually and technically well-motivated assumptions.

(The discussions of non-equilibrium statistical mechanics in the Gibbs approach by Callender (2001), Ridderbos (2002) and Frigg (2007) seem to fit this pattern, for instance.)

On this basis, non-equilibrium statistical mechanics is a foundational project (we already knew *that* systems approach equilibrium, we just seek to find out *why*), and so it is entirely possible to suppose that little or no progress has been made in that project.

As with the equilibrium, this conception of statistical mechanics sharply understates just how much we actually know about the quantitative non-equilibrium processes that physics studies. In fact, we not only know *that* systems approach equilibrium, we know a considerable amount, quantitatively, about *how fast* they approach equilibrium,[1] and we possess a variety of fairly effective calculational algorithms to construct those quantitative predictions from the underlying microphysics. As in the equilibrium case, we may not be sure *why* those algorithms work, but we have abundant evidence *that* they work. And an adequately broad conception of the *foundations* of non-equilibrium statistical mechanics must include in its goals not merely an understanding of the qualitative predictions of the field (i. e. , that systems approach equilibrium) but an underpinning of the quantitative predictions.

Indeed, that description of the task (that underpinning the quantitative part of the subject is an *additional* task for foundational work) understates the case. To see this, and to illustrate the general point, consider again the case of a dilute gas. This provides perhaps the best-known example of an evolution equation for non-equilibrium statistical mechanics: the Boltzmann equation,

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(v) = N\int \mathrm{d}v'\,\mathrm{d}u\,\mathrm{d}u'\,|v-v'|\sigma(uu'\to vv')(\rho(u)\rho(u')-\rho(v)\rho(v')), \qquad (19)$$

where N is the particle number density, $\sigma(uu' \to vv')$ is the the scattering cross-section for particles with velocities u, u' to scatter into a state with velocities v, v', and $\rho(v)$ is either (as Boltzmann originally proposed) the fractional number density of particles with a velocity v, or (as is more usual in contemporary textbooks) the marginal one-particle probability distribution, averaged over particle position.

As is well known, the Boltzmann equation was originally proposed by Boltzmann as a microphysical grounding of the approach of a gas to equilibrium:

[1] And indeed, about the circumstances in which they do *not* approach equilibrium: a box full of a mixture of hydrogen and oxygen at standard temperature and pressure, for instance, does not approach equilibrium on any reasonably quick timescale, in the absence of an externally-generated spark.

systems obeying it will converge on the Maxwell-Boltzmann particle-number distribution. As is equally well known, Boltzmann's own derivation tacitly (via the postulate of the *Stosszahlansatz*) makes assumptions which cannot hold for all microstates of the gas at *any* time (due to reversibility of the dynamics), and which cannot hold for the evolution of any microstate of the gas for *all* times (due to Poincaré recurrence). (For a careful treatment, see Brown, Myrvold, and Uffink (2009).) Partly for these reasons, the Boltzmann equation is frequently[2] referred to as an "early" approach to the foundations of thermodynamics, to be contrasted with later approaches (including the "Boltzmannian" approach described above!) deemed better able to account for the approach to equilibrium.

What is in danger of being lost here is that *the Boltzmann equation works*, not perhaps in the sense of providing a *conceptually* sound understanding of why dilute gasses equilibrate, but in the sense of predicting *quantitatively*, and *accurately*, how dilute gases away from equilibrium actually behave. That is, when the Boltzmann equation says that the rate of change of the one-particle distribution depends in such-and-such a way on the distribution, the overall density, and the scattering cross-section, *that is what is found in nature.*

Nor are its applications limited to Boltzmann's original context. A 1988 survey of the field (Cercignani 1988) notes that the equation "has proved fruitful not only for the study of the classical gases Boltzmann had in mind, but also, properly generalized, for electron transport in nuclear reactors, photon transport in superfluids, and radiative transport in planetary and stellar atmospheres." For an even more exotic application, note that the Boltzmann equation (suitably supplemented to allow for long-range forces) is also central to the kinetic-theory approach to galactic dynamics: to "gases", that is, where the "particles" are stars.

So an approach which succeeds in giving an underpinning to the qualitative fact that dilute gases approach equilibrium, but fails to underpin the Boltmzann equation, has failed to provide a satisfactory account of the statistical mechanics of dilute gases. But conversely, an account that successfully underpins the Boltzmann equation gets the approach to equilibrium as a bonus. For while it may be mysterious why dilute gases in general approach equilibrium, it is a straightforward theorem that dilute gases subject to the Boltzmann equation approach equilibrium.

Furthermore, the Boltzmann equation is by no means the only example of a predictively powerful non-equilibrium evolution equation. (Other examples include the Fokker-Planck equation (discussed in, e. g. , Liboff 2003, p.301 and Le Bellac, Mortessagne, and Batrouni 2004, p.552) and — in quantum theory — the Pauli master equation (discussed in, e. g. , Zwanzig 1966) and the decoherence master equation (discussed in, e. g. , Schlosshauer 2007 and Zurek 2003), as well as indefinitely many variations of the Boltzmann equation). And in each case, these equations are constructed from the underlying mechanics. The algorithms for carrying out this construction are less clearly articulated

[2]See, e. g. , Frigg (2007) or Callender (2001).

and understood in the literature than the algorithms of equilibrium statistical mechanics, but a large fraction of them are constructed via the *method of projections*, which I discuss in the next section.

6 The method of projections

The method of projections, like the canonical- and microcanonical-distribution approaches to equilibrium statistical mechanics, is normally formulated in the language of probability distributions on phase space (or, in the quantum case, of density operators). (If this seems to you simply the wrong way to proceed on conceptual grounds, remember again that I am simply explaining the way statistical physics is done, not claiming that it is on a firm conceptual footing.) The space of such distributions can be denoted $\mathcal{P}$, and — given a Hamiltonian H — Liouville's equation,

$$\dot{\rho} = L_H \rho \equiv \{\rho, H\}\,, \tag{20}$$

gives an evolution equation on $\mathcal{P}$ under the supposed 'true' dynamics of the system. It will be convenient to write $U(t)$ for the time evolution operator generated on $\mathcal{P}$ by Liouville's equation (formally, $U(t) = \exp(\{\cdot, H\}t)$).

The point of the method of projection is to proceed from a dynamical equation for the *full* probability distribution, to an equation for some reduced or restricted version of the distribution, containing only a small part of the information encoded by the original distribution. This is represented by the eponymous *projection*, a map J from $\mathcal{P}$ to itself satisfying $J^2 = J$. The projection is to be thought of as a kind of coarse-graining of a probability distribution, throwing away that part of it which is not relevant to whatever macro-level phenomenology is to be described. In the terminology of Zwanzig Zwanzig (1960, Zwanzig (1966), the projection decomposes ρ into the *relevant* part $\rho_r = J\rho$, and the *irrelevant* part $\rho_{ir} = (1 - J)\rho$, and the objective is to find autonomous dynamical equations for the relevant part. In most (not all) cases, the projection is linear.

Physically significant examples of projections (all of which are linear) include:

Coarse-graining: The phase space is divided into cells (perhaps with each cell corresponding to a macrostate in Boltzmann's sense) and J leaves the probability of each macrostate invariant but smooths the distribution out to be uniform on each macrostate. (Coarse-graining is widely discussed in the foundations of statistical mechanics in the context of the Gibbsian approach.)

Averaging over degrees of freedom: The degrees of freedom are divided into 'relevant' and 'irrelevant' degrees of freedom. For the dilute gas, for instance, the small-scale positional degrees of freedom are designated as irrelevant, and the large-scale positional degrees of freedom, and the momentum degrees of freedom as relevant; for a system interacting with

an environment, the relevant degrees of freedom are the system degrees of freedom and the environment degrees of freedom are designated as irrelevant. The projection then replaces the distribution with one that has the same marginals over relevant degrees of freedom but has some fixed, specified form over irrelevant degrees of freedom; in doing so, any correlations between relevant and irrelevant degrees of freedom are discarded.

Diagonalisation: Only applicable to quantum systems, this projection deletes the off-diagonal elements of the density operator with respect to some specified basis.

In its crudest form, the method of projections works as follows:

1. Evolve the system forward for some reasonably short time Δt under the underlying time evolution operator $U(t)$.
2. Apply the projection operator J. (Sometimes this step is called a *rerandomization posit*.)
3. Iterate.

If J has been appropriately chosen, it will turn out that for some intermediate range of values of Δt, the evolution rule thus determined is insensitive to the exact value of Δt. In this case, we can without ambiguity define the *forward dynamics* associated with J by

$$U_J(N\Delta t)\rho = (JU(\Delta t)\,\rho. \tag{21}$$

Is it conceptually clear what's going on here? Almost certainly not, but that isn't the point. The point is: this is in fact how all manner of equations in contemporary statistical mechanics are constructed (even the Boltzmann equation, in many modern presentations), and the equations thus constructed work — that is, do predict how the relevant part of the probability distribution evolves — so it is incumbent on an adequate foundation of statistical mechanics to understand *why* it works (and not merely to abandon it as unsatisfactory and seek a different explanation for equilibration).

In fact, we can make fairly considerable progress in seeing, qualitatively, what would be involved in the claim that the method works. The *actual* evolution of the relevant part of the distribution, given initial distribution ρ, is given by the underlying dynamics as

$$\rho_r(t) = JU(t)\rho. \tag{22}$$

So for the method of projections to *also* predict correctly how ρ_r evolves, we must have

$$JU(t)\rho = U_J(t)\rho. \tag{23}$$

On pain of violating Poincaré recurrence (at least in the quantum case), this cannot hold for any ρ for all times; on pain of violating time reversal invariance, it cannot hold for all ρ for any times. But there is no a priori problem with

there existing some ρ such, for sub-recurrent positive times, (23) holds. (In the terminology of (Wallace 2010), such ρ are *forward compatible* with J.)

A justification of the method of projections, therefore, requires both some assumptions about the structure of the dynamics, and some condition on the initial distribution. In fact (at least in the case of linear J), it is possible to get rather more precise about this. In more sophisticated applications of the method of projections (following Zwanzig (1960) and references therein), we can differentiate (22) and, after some algebra, obtain the following formally exact equation for ρ_r:

$$\dot{\rho}_r(t) = (JL_HJ)\rho_r(t) + JL_H\mathrm{e}^{(1-J)L_Ht}\rho_{ir}(0) + \int_0^t \mathrm{d}t'\, JL_H\mathrm{e}^{(1-J)L_Ht'}(1-J)L_H\rho_r(t-t'). \tag{24}$$

This expression is (formally) exact, and so makes no time-asymmetric assumptions, but its form is suggestive. The first term is a time-reversible flow term (effectively the projection of the original dynamics onto the relevant subspace). The second term is the only place where $\rho_{ir}(t)$ plays any role: it provides a contribution to the rate of change of ρ_r at time t dependent on the original (time-zero) value of ρ_{ir}. If the second term vanishes (which can be achieved in particular by setting $\rho_{ir}(0) = 0$ as a time-zero boundary condition) then we have a closed integro-differential equation for ρ_r.

Under certain conditions, the integral kernel $JL_H\mathrm{e}^{(1-J)L_Ht}(1-J)$ will drop off rapidly compared to the timescales on which ρ_r evolves.[3] In this case, to a good approximation we can (i) replace the upper limit of the integral with ∞; (ii) replace $\rho_r(t-t')$ with $\rho_r(t)$.

Putting these two results together, we obtain a *differential* equation for ρ_r:

$$\dot{\rho}_r(t) \simeq (JL_HJ)\rho_r(t) + \left(\int_0^\infty \mathrm{d}t'\, JL_H\mathrm{e}^{(1-J)L_Ht'}(1-J)L_H\right)\rho_r(t). \tag{25}$$

This is an explicitly time-asymmetric equation; various particular cases of it include the (empirically verified) equations of decoherence, of radioactive decay, and of diffusion and equilibration in dilute gases.[4]

What is the conceptual significance of these results? Firstly, they show once again how probabilistic concepts are deeply embedded in the contemporary practice of statistical mechanics. One occasionally sees it suggested, in particular in discussions of the Boltzmannian approach, that quantitative probabilistic ideas

[3] A little thought shows that in a *uniformly* recurrent system — that is, one where recurrence occurs for probability distributions rather than just individual phase-space points — this cannot be correct for all times: eventually the kernel must increase again and return to its original value. If this occurs (which is the case in particular in quantum theory, where a *uniform* recurrence theorem holds — see Wallace (2013) for discussion) we must confine application of these methods to times much less than the recurrence timescale. Of course, this is no real restriction.

[4] Strictly speaking, the equation thus derived is the *Prigogine-Brout equation* for the projected dynamics of the full N-particle probability distribution under a projection which smooths out spatial variations; the Boltzmann equation follows from this under further (time-symmetric) assumptions. See Zwanzig (1960) for details.

are ultimately dispensible in statistical mechanics, to be replaced by qualitative ideas about which phase-space regions are much larger than others. We have already seen that this strategy is problematic in the face of fluctuation phenomena; now we find that it is also in tension with the methods of non-equilibrium statistical mechanics.

Secondly, the general method of projections — and, more explicitly, the Zwanzig projection formalism — illustrates a central idea of the Boltzmannian approach: that the applicability of statistical mechanics depends on *time-symmetric* constraints on the dynamics, and on a constraint on the initial state. In the Zwanzig decomposition, the requirement on the initial state is that the second Zwanzig term vanishes; the requirement on the dynamics is that the kernel of the third Zwanzig term falls off sufficiently rapidly.

Thirdly, the method of projections also makes clear the nature of the initial-state constraint: it is a *probabilistic* constraint (a requirement that a certain linear functional of the irrelevant part of the probability distribution vanishes), and it places no constraint whatsoever on those features of the state (the "relevant" features) that we are actually attempting to track using statistical-mechanical methods.

To illustrate this, consider again the Boltzmannian account of the approach to equilibrium. It relies on two assumptions (over and above its general requirements on the form of the dynamics):

1. The initial macrostate of the Universe is some specific macrostate which has a low Boltzmann entropy

2. The probability distribution over current microstates of the Universe is the uniform distribution, conditioned on the present macrostate and on the initial state

Frequently (and, to be fair, more often in popularisations than in technical discussions[5]) it is the first of these conditions — the so-called "Past Hypothesis" — that is described as doing the real work. The probabilistic assumption is generally downplayed.

However, if this account is interpreted in the Zwanzig formalism (using, as our choice of J, coarse-graining over macrostates), the first condition is a constraint on the *relevant* part of the probability distribution, and so is not required to derive time-asymmetric laws. The real work is being done by the second assumption, which is a constraint purely on the irrelevant part of the distribution (specifically, that it vanishes at the initial time). The first condition may be part of an explanation of what the Universe *in fact* has the structure it does, but it is unnecessary as part of an explanation of why, at the emergent level, it obeys the *laws* it does. (I expand upon this point in Wallace (2010).)

[5]See, for example, Penrose (1989).

7 Conclusion

I have attempted to demonstrate that the great bulk of statistical mechanics is concerned little with providing a foundational underpinning for the general principles of thermodynamics, but rather is a collection of techniques used in the modelling of actual systems. These techniques are far from conceptually clear — quite the reverse, in fact — and it is potentially very distorting to look for a philosophically satisfactory underpinning for only the parts of the field most directly concerned with qualitative thermodynamics.

To close with a more general observation, it is perhaps unsurprising that statistical mechanics *in general* seems so little discussed in the foundational literature, despite the burgeoning field of discussions of statistical mechanics *as a foundational endeavour*. For unlike most of the fields studied in philosophy of physics (notably quantum and classical mechanics, classical field theory, and general relativity), statistical mechanics does not lend itself well to a precise, mathematically rigorous formulation, and for good reasons and bad, philosophers seem to find it much easier to engage with the content of a physical theory when a precise formulation is available.

Partly as a consequence, the bulk of philosophical work in statistical mechanics is done from a more or less explicitly historical basis, and a s a consequence is fairly focussed around the preoccupations of the nineteenth and early-twentieth century founders of the subject. Indeed, Jos Uffink's recent "Compendium of the Foundations of Statistical Mechanics" (Uffink 2007) begins by observing that

> Statistical physics ... has not yet developed a set of generally accepted formal axioms, and consequently we have no choice but to dwell on its history.

But — while deep insights can be, and have been, gained into physics both through historical study and through mathematical rigor — they are not, *pace* Uffink, the only ways that philosophers can approach contemporary physics. There is an alternative: look at the actual practice of physics, at least as represented in the papers, research monographs and contemporary texts, and try to make sense of it. Of course, physics understood this way is in general tangled and confused, without the logical clarity that comes from rigorous axiomatisation or the different kind of clarity that comes from following through the historical record. But I take it that the ability to unravel conceptual tangles, and resolve conceptual confusion, is about the one distinctive skill set that philosophers (at least qua philosophers) can bring to the activity of physics. To demand that philosophical attention be confined to those areas of physics already sufficiently well understood that they can be given sharp and precise formulations is to accept a sharply diminished prospect that philosophical study of physics can make any contribution to its development.

References

Albert, D. Z. (2000). *Time and Chance*. Cambridge, MA: Harvard University Press.

Blundell, S. J. and K. M. Blundell (2010). *Concepts in Thermal Physics* (2nd ed.). Oxford: Oxford University Press.

Brown, H. R., W. Myrvold, and J. Uffink (2009). Boltzmann's H-theorem, its discontents, and the birth of statistical mechanics. *Studies in History and Philosophy of Modern Physics 40*, 174–191.

Brown, H. R. and J. Uffink (2001). The origins of time-asymmetry in thermodynamics: The minus first law. *Studies in the History and Philosophy of Modern Physics 32*, 525–538.

Callender, C. (2001). Taking thermodynamics too seriously. *Studies in the History and Philosophy of Modern Physics 32*, 539–553.

Cercignani, C. (1988). *The Boltzmann Equation and its Applications*. New York: Springer-Verlag.

Frigg, R. (2007). A field guide to recent work on the foundations of thermodynamics and statistical mechanics. In D. Rickles (Ed.), *The Ashgate Companion to the New Philosophy of Physics*, pp. 99–196. London: Ashgate.

Goldstein, S. (2001). Boltzmann's approach to statistical mechanics. In J. Bricmont, D. Dürr, M. Galavotti, F. Petruccione, and N. Zanghi (Eds.), *In: Chance in Physics: Foundations and Perspectives*, Berlin, pp. 39. Springer. Available online at http://arxiv.org/abs/cond-mat/0105242.

Le Bellac, M., F. Mortessagne, and G. G. Batrouni (2004). *Equilibrium and Non-Equilibrium Statistical Thermodynamics*. Cambridge: Cambridge.

Lebowitz, J. (2007). From time-symmetric microscopic dynamics to time-asymmetric macroscopic behavior: An overview. Available online at http://arxiv.org/abs/0709.0724.

Liboff, R. L. (2003). *Kinetic Theory: Classical, Quantum, and Relativistic Descriptions* (3rd ed.). New York: Springer-Verlag.

North, J. (2002). What is the problem about the time-asymmetry of thermodynamics? - a reply to Price. *British Journal for the Philosophy of Science 53*, 121–136.

Penrose, R. (1989). *The Emperor's New Mind: concerning computers, brains and the laws of physics*. Oxford: Oxford University Press.

Penrose, R. (1994). On the second law of thermodynamics. *Journal of Statistical Physics 77*, 217–221.

Ridderbos, K. (2002). The coarse-graining approach to statistical mechanics: how blissful is our ignorance? *Studies in the History and Philosophy of Modern Physics 33*, 65–77.

Schlosshauer, M. (2007). *Decoherence and the Quantum-to-Classical Transition*. Berlin: Springer.

Uffink, J. (2007). Compendium of the foundations of classical statistical physics. In J. Butterfield and J. Earman (Eds.), *Handbook for Philosophy of Physics*. Elsevier. Available online at philsci-archive.pitt.edu.

Wallace, D. (2010). The logic of the past hypothesis. Available online at http://users.ox.ac.uk/ mert0130/papers.shtml.

Wallace, D. (2013). Recurrence theorems: a unified account. Available online at http://arxiv.org/abs/1306.3925.

Zurek, W. H. (2003). Decoherence, einselection, and the quantum origins of the classical. *Reviews of Modern Physics 75*, 715.

Zwanzig, R. (1960). Ensemble method in the theory of irreversibility. *Journal of Chemical Physics 33*, 1338–1341.

Zwanzig, R. W. (1966). Statistical mechanics of irreversibility. In P. H. E. Meijer (Ed.), *Quantum Statistical Mechanics*, pp. 139. New York: Gordon and Breach.