What Gibbsian Statistical Mechanics Says: in defense of bare probabilism

David Wallace*

February 2, 2025

Abstract

I expound and defend the "bare probabilism" reading of Gibbsian (i.e. mainstream) statistical mechanics, responding to Frigg and Werndl's recent (*BJPS* 72 (2021), 105-129) plea: "can somebody please say what Gibbsian statistical mechanics says?"

1 Introduction

Mainstream statistical physics proceeds by assigning probability functions to classical systems, and mixed quantum states to quantum systems, and then calculating synchronic and diachronic properties of those functions. Contemporary philosophy of physics refers to this mainstream approach as "Gibbsian statistical mechanics" (henceforth GSM) and contrasts it, usually unfavorably, to (so-called) "Boltzmannian statistical mechanics", in which the role of probability is lessened and in some versions eliminated altogether.

Recently, however, there has been increasing interest in engaging with rather than simply dismissing GSM (see, e. g. , Luczak 2016, Robertson 2020, 2022, Myrvold 2021, Wallace 2015, 2020), and this has led to controversy about its actual content. The issue has been raised in a recent (and characteristically clear) paper by Frigg and Werndl (2021) (henceforth FW), titled "Can somebody please say what Gibbsian statistical mechanics says?" In their introduction, they say

*Department of History and Philosophy of Science / Department of Philosophy, University of Pittsburgh, PA 15260, USA; email david.wallace@pitt.edu

> GSM is widely used and considered by many to be the most important formalism of statistical mechanics. Yet a closer look at GSM reveals that it is unclear what the theory says and how it bears on experimental practice. ... Hence our plea: can somebody please says what GSM says? (p.105)

This paper is an attempt to answer their plea (building on a previous, less explicit answer in (Wallace 2015)), as well as a critique of the alternative answer that FW themselves develop. I defend the claim that GSM should be interpreted — and is interpreted by working physicists — via what FW call *bare probabilism*, the view that there is *nothing* to GSM beyond the probabilistic claims that it makes about systems. FW regard this as unacceptable in large part because it fails to provide an adequate basis for thermodynamics; I will argue that this follows only because FW fail to recognize that in modern physics, thermodynamics is interpreted statistically just as statistical mechanics is.

In section 2 I review the formalism of GSM, and in section 3 I summarize FW's criticisms and their eventual position. In section 4 I defend the view that contemporary thermodynamics should itself be understood statistically, and in sections 5–6 I discuss fluctuations in GSM, which FW regard as a major issue for its interpretation. Section 7 is the conclusion.

A disclaimer about terminology: the philosophy literature can easily give the impression that (1) the "Gibbs-vs-Boltzmann" debate is a genuine schism within modern physics, and (2) what is called "Gibbsian" or "Boltzmannian" statistical mechanics is an accurate description of the historical views of (respectively) Gibbs and Boltzmann. I reject both claims. As to the first: I agree with FW that "Gibbsian statistical mechanics" is simply what modern physicists call statistical mechanics; "Boltzmannian statistical mechanics" can be understood varyingly as a historically held position, as the view of a heterodox minority (Lebowitz 1993; Goldstein 2001), or as a special case of mainstream statistical mechanics (Wallace 2020). As to the second: Wayne Myrvold (2021, ch.7) persuasively argues that the historical Gibbs and Boltzmann held significantly more nuanced, and more mutually compatible, views than the contemporary literature might suggest. I have stuck with the standard terminology for ease of communication, but to be clear: by "GSM" I mean no more or less than "statistical mechanics as it has in fact been practiced in mainstream physics over the last seventy years or so".

2 The Gibbsian formalism

The *formalism* of GSM (leaving aside its interpretation for now) is as follows:

1. A *system* is characterized either classically by a phase space and a collection of functions on that space representing the dynamical variables (including the Hamiltonian H, which generates the system's dynamics) or quantum-mechanically by a Hilbert space and a collection of operators on that space representing the dynamical variables (again including the Hamiltonian H).

2. The *statistical state* ρ_t of the system at time t (often I just write ρ, leaving any time dependence tacit) is defined either classically by a probability measure over the phase space (normally expressed as a function on phase space expressing the measure in terms of the Liouville measure), or quantum-mechanically by a quantum state (pure or mixed) on the Hilbert space. (Notice an important quantum/classical disanalogy: in classical mechanics there is a clear distinction between the statistical state and the *microstates* of the system, which are represented by *points* in the phase space; in quantum mechanics the distinction is elided).

3. The statistical state of an undisturbed system evolves over time under Liouville's equation, which classically is

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_t = \{H, \rho_t\}_{PB} \tag{1}$$

with $\{,\}_{PB}$ the Poisson bracket, and quantum-mechanically as

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_t = -\frac{i}{\hbar}[H, \rho_t] \tag{2}$$

with $[,]$ the commutator.

4. Given a dynamical variable X, its *expectation value* with respect to statistical state ρ is given classically by

$$\langle X \rangle_\rho = \int \mathrm{d}\mu\,(x)\rho(x)X(x) \tag{3}$$

(where μ is the Liouville measure) and quantum-mechanically by

$$\langle X \rangle_\rho = \mathsf{Tr}(\rho X). \tag{4}$$

A system is in *exact statistical equilibrium* if its statistical state is invariant under time, $\mathrm{d}\rho/\mathrm{d}t = 0$. One important class of states in exact statistical equilibrium are the *canonical states*, which have form

$$\rho_\beta = \frac{1}{Z(\beta)} \mathrm{e}^{-\beta H} \tag{5}$$

(interpreted either classically or quantum-mechanically; $Z(\beta)$ is in either case a normalization factor), and a system is in *exact canonical equilibrium* if its statistical state is one of the canonical states. Exact canonical equilibrium is thus a special case of exact statistical equilibrium. The notion of canonical state generalizes naturally to include dependence on other conserved quantities such as N, the number of particles; for this case, we have

$$\rho_{\beta,\mu} = \frac{1}{Z(\beta,\mu)} \mathrm{e}^{-\beta(H+\mu N)}. \tag{6}$$

(States like this are sometimes called *grand canonical states.*)

It is common[1] to introduce a weaker notion of statistical equilibrium. We fix some set of *macrovariables*, which are a small subset of the degrees of freedom intended to represent collective, large-scale degrees of freedom and/or degrees of freedom which are realistically measurable. Common choices include: for a spatially extended system the microvariables averaged over some lengthscale L large compared to intermolecular spacing but small compared to the system scale (or, more practically, the Fourier components of the dynamical variables with wavelengths $> L$), and for a system of N particles, the complete set of M-particle dynamical variables for fixed $M \ll N$. (These roughly correspond to the choices of macrovariables generally made in the two main paradigms for non-equilibrium statistical mechanics: the Langevin and BBGKY paradigms (see, e. g. , (Calzetta and Hu 2008, ch.2) or (Wallace 2016). In the former paradigm the macrovariables are sometimes called *relevant* variables.) A system is then in *statistical macroequilibrium* at t if the expectation value $\langle M(t)\rangle \equiv \langle M\rangle_{\rho_t}$ is time-invariant at t for all macrovariables M. And it is in *canonical macroequilibrium* if the expectation values of all macrovariables match their values on some canonical state.

[1]In the philosophy literature FW attribute this definition to van Lith (1999). It is very widespread in the physics literature: notably, it is the central notion used by the various recent programs to explain equilibration in quantum statistical mechanics (cf section 6).

This formalism is related to thermodynamics by an *averaging principle* which I will call AP*: thermodynamic quantities like energy or particle number are identified with the expectation values of the corresponding statistical system, *calculated at canonical equilibrium*. To spell this out a little:[2] consider a fluid, which in phenomenological thermodynamics is characterized by its energy U, the number of particles n, and its volume V; the fluid's thermodynamic behavior is described by its equation of state $S = S(U, n, V)$, where S is the system's thermodynamic entropy. In Gibbsian statistical mechanics we identify $S(U, n, V)$ with the Gibbs entropy of the unique canonical state that has expected energy $\langle H(V) \rangle = U$ and expected number of particles $\langle N \rangle = n$, where the system volume V is interpreted as an external parameter for the Hamiltonian $H(V)$ and where N is the classical function or quantum operator representing particle number. In (Wallace 2023, section 3) I call this fuller recipe (which includes AP* as a part of it) the *canonical recipe*: using it, in those situations where it is calculationally tractable to do so, reliably recovers the measured thermodynamic equation of state.

3 Frigg and Werndl on the interpretation of GSM

FW take GSM as committed to a somewhat different averaging principle which they call AP:

> when measuring the property f on a system in [thermodynamic equilibrium], the observed equilibrium value of the property is the ensemble average $\langle f \rangle$ of an ensemble in [statistical equilibrium].

Then they raise two, related, questions about GSM and AP:

1. How is the formalism to be interpreted; that is, what if anything are we saying *about a physical system* when we assign a statistical state to it?

2. How is AP to be justified, given some interpretation of GSM?

[2]For a more detailed description, see (Wallace 2023, section 3), or any good statistical mechanics textbook.

(I have stated (2) as secondary to (1), as I think this is faithful to FW, but I would myself be inclined to break it loose from any question of interpretation and ask: why and to what extent does AP (and, more generally, the canonical recipe) in fact give *empirically correct predictions* for thermodynamic systems?)

They begin by considering, and fairly swiftly rejecting, two possibilities: *quietism* (the view that AP is simply a posit in no need of further justification) and the *time average approach*, which equates the Gibbsian probability measure to a long-term time average. In both cases I agree with their assessment and will not discuss these approaches further.[3]

FW's main focus is *probabilism*: the view that GSM is to be interpreted in the obvious way as making probabilistic statements about systems, so that in classical mechanics, when we say 'the statistical state of this system is ρ', we mean, the probability density of the system having microstate x is $\rho(x)$. FW do not discuss quantum mechanics explicitly, but the obvious quantum version of probabilism is just that a system's statistical state is its actual quantum state. (What that means, and in particular whether it is itself a probabilistic or categorical statement, of course depends on one's preferred solution to the measurement problem.)

FW consider several versions of probabilism (as well as a digression to discuss a recent proposal by McCoy (2020), which they ultimately reject; I will not discuss that proposal here). The first is *bare probabilism*: the view that *all there is* to Gibbsian statistical mechanics is the probabilistic interpretation. As FW say, "[o]n such a view GSM really is just a study of the statistical properties of [the statistical state] with nothing else added."[4]

To FW, the problem with bare probabilism is that — as they interpret it — it abandons any hope of a statistical-mechanical grounding of thermodynamics. ("Any attempt to read more into GSM, in particular any attempt to read a notion of thermodynamic equilibrium into it, is misguided and should

[3]Interestingly, FW describe the time average approach as "the 'standard' textbook approach"; if so, I think that reflects badly on the writers of textbooks (or at least, reflects those writers' impatience to get through foundations as quickly as possible so as to discuss applications), since severe criticisms of the approach go back at least as far as (Tolman 1938, pp.65-70). Penrose (1979, p.1944) described this approach as "out of fashion now" over forty years ago.

[4]They cite my (2015) when they present bare probabilism, but say (fn. 16) that they have been unable to find an explicit statement of bare probabilism in print. I had intended to give one; obviously I was insufficiently clear!

be resisted.") They go on to spell out the concern in more detail:

> The fact that thermodynamic equilibrium and statistical equilibrium are both equilibria does not mean that they are somehow similar, or that statistical equilibrium can serve as a stand-in for thermodynamic equilibrium when the latter is excised. In fact, statistical and thermodynamic equilibrium are not only conceptually different, the two notions also do not have the same extension. An ensemble in statistical equilibrium not only contains systems at thermodynamic equilibrium; it can also contain systems that are not at thermodynamic equilibrium. (FW p.114)

Here, FW are making substantive assumptions about thermodynamics that are worth spelling out. Specifically, they are committed to the view (standard in Boltzmannian statistical mechanics) that thermodynamic properties like energy, entropy and being at equilibrium are properties of *microstates*. For equilibrium in particular, and given again a collection of macrovariables we can define a microstate of a classical system as being at *Boltzmannian equilibrium* if the value of the macrovariables is time-invariant, perhaps up to some small fluctuation term; FW (section 2) define thermodynamic equilibrium as Boltzmannian equilibrium.[5]

Under these assumptions about thermodynamics, FW are surely correct that thermodynamics does not reduce to GSM given Bare Probabilism, that statistical equilibrium is not thermodynamic equilibrium (it is certainly not Boltzmannian equilibrium), and that AP does not follow. Indeed, Bare Probabilism entails that the *expected* value of a measurement of a quantity will be its average, but says nothing categorical about the *actual* value: only in the

[5]As a matter of exegesis: FW's formal definition of when a system is in thermodynamic equilibrium is that 'none of its thermodynamic properties are changing with time'. That in itself is universally held, but the key question is whether those 'thermodynamic properties' are to be understood as the actually-possessed values of the macrovariables for a given system (this is the Boltzmannian definition) or as a statistical average of those macrovariables (as, I will argue in section 4, is assumed in GSW). It is clear from FW's discussion (for instance, from the above quote, or from their account of fluctuations on pp.117-8) that they endorse the former view, so that thermodynamic equilibrium is a *non-probabilistic* property, holding of phase-space points rather than phase-space distributions. Of course, the distinction is usually irrelevant in most large systems, where fluctuations tend to be negligible, and so it is often elided, especially in textbook discussions of phenomenological thermodynamics.

Thanks to an anonymous referee for pressing this point.

limiting case of a dispersion-free probability distribution will the actually-measured value be guaranteed to match the expected value.

Given these issues, FW regard bare probabilism as unsatisfactory. They consider two alternatives, both of which have as a starting point a consideration of statistical fluctuations. Staying within classical mechanics, they define the *fluctuation* of a system with microstate $x(t)$ at time t (with respect to some macrovariable f) as

$$\Delta(t) = f(x(t)) - \langle f \rangle_\rho \tag{7}$$

where ρ is the equilibrium distribution (presumably the canonical or microcanonical distribution, though FW are not explicit here). Then they suggest that AP would be justified as a good approximation for any system where the probability of large $\Delta(t)$ is low (a condition that they call *thermodynamic fluctuations*). How we could generalize this to quantum mechanics is unclear, but FW restrict their attention to classical mechanics.

FW then suggest two modifications of probabilism. In *qualified probabilism*, AP is added as a requirement of inter-theoretic reduction: 'bare probabilism must ensure that AP holds whenever GSM is used in tandem with [thermodynamics]'. In *fluctuation probabilism*, AP is a restriction on GSM itself: '[f]luctuation probabilism has to restrict dynamical laws and observables that are allowable in GSM to those that produce thermodynamic fluctuations'.

FW claim that it is an 'aesthetic matter' which of qualified or fluctuation probabilism we accept, but it's not clear why, unless we regard GSM as purely a foundation for thermodynamics: according to qualified probabilism we can use GSM outside the regimes where AP holds, so long as we don't try to relate it to thermodynamics; according to fluctuation probabilism we cannot use GSM at all unless AP holds. But in any case, FW regard all forms of probabilism, including bare probabilism, as committed to a position they call 'ρ-universalism' — literally, the view that ρ at all times does after all give the correct probability distribution at least for macroscopic observables — which they go on to criticize.

Their criticism is based on a consideration of fluctuations, for which they consider two possible interpretations. The first is that the fluctuations represent probabilistic expectations of how *the same* system's macrovariables will fluctuate over time. As they correctly point out, this is impossible if there are conserved quantities such that the distribution ρ assigns different probabilities to different values of those quantities: in this situation, we would predict that some conserved quantity C has nonzero probability of having value c at time t, and also nonzero probability of having value $c' \neq c$ at time t', but (say FW) these are incompatible claims.

On the second interpretation, the fluctuations represent probabilistic expectations of how identically-prepared systems will differ in the measured values of macrovariables. FW concede that in this case, analytically ρ-universalism holds, but object that this is not the empirical situation we find ourselves in: 'in laboratory measurements we observe the *same* system at consecutive times rather than drawing different systems out of an urn at random' (p.122; emphasis theirs).

Exploring how probabilism can be saved from these objections, FW formulate two technical conditions (both expressed with respect to a set of macrovariables). A system satisfies *masking* if the fluctuations for each macrovariable are the same for any time-invariant statistical state ρ (so that unrecognized conserved quantities do not affect macrovariable expectation values). A system satisfies *independence* (tacitly, with respect to a given time-invariant statistical state ρ) if for sufficiently large times t and for arbitrary initial statistical state ρ' the expectation value of any macrovariable with respect to the time-evolved ρ' and to ρ coincide to any desired degree of accuracy. FW's conclusion is that probabilism is only justified when one or other of these conditions holds, and that it is non-trivial to establish that they do.

4 Statistical thermodynamics

The core claim of this paper is fairly simple: bare probabilism by itself suffices as a reductive base for thermodynamics, because thermodynamics itself ought to be interpreted as a statistical theory. That is: where a quantity in thermodynamics has a direct microphysical interpretation, like energy, work, or particle number, the numerical value of that quantity used in the equations and inequalities of equilibrium thermodynamics should be interpreted as the expectation value of that quantity; other state-dependent quantities, like thermodynamic entropy, should be interpreted as other (non-expectation) functions of the probability distribution characterizing a thermodynamic system. (And quantities corresponding to externally controlled parameters, such as the volume of a container of gas, continue to have a non-statistical interpretation.[6]) So the statement that a system is in thermodynamic equilibrium, and hence that the numerical values of its quantities are time-independent, should thus be interpreted as a probabilistic statement about the system: that the expected value of any macrovariable, and the values of other functionals of the probability distribution, are time-invariant.

In more detail, and more carefully: we can distinguish *statistical thermodynamics*, whose parameters are interpreted probabilistically in this way, from *categorical thermodynamics*, whose parameters are interpreted as describing an actual system. My claim is that statistical thermodynamics is derivable from GSM under the bare probabilism interpretation of the latter. Categorical thermodynamics then follows, approximately and with high probability, from statistical thermodynamics for certain sufficiently large systems, basically via the law of large numbers. (And in non-historical contexts, it's not obvious that we benefit from treating categorical thermodynamics as a *theory* rather than just the Law of Large Numbers applied to statistical thermodynamics.)

The details of any such derivation are beyond the scope of this paper (I discuss the technical details of such a derivation in (Wallace 2023)) but the important point here is that there can be no barrier of principle in deriving statistical thermodynamics from the bare probabilism reading of GSM. If the claims of thermodynamics are understood as categorical then the equating of thermodynamic quantities with statstical averages becomes a substantive extra assumption, in need of justification or explicit posit. But if thermodynamics is itself statistical, there is no conceptual difficulty in supposing (e.g.) that a claim about the *average* work extractable from a system by cyclical processes might be calculable from a probabilistic description of that system. Similarly, if a system is at thermodynamic equilibrium when the *expected* values of its macrovariables are time-invariant, there is no obstacle to identifying thermodynamic equilibrium with statistical macroequilibrium; indeed, that identification is practically analytic.

Why think that statistical thermodynamics is how modern physics interprets thermodynamics? I find the question somewhat difficult to answer: to me, it seems transparent in practically all the modern literature, at least as far back as (Tolman 1938):

[6]This tripartite distinction between types of thermodynamic variables has a long history, going back at least to (Tolman 1938, pp.524-540). (Thanks to an anonymous referee for drawing my attention to this.)

> [I]t is to be emphasized, in accordance with the viewpoint here chosen, that the proposed methods are to be regarded as *really statistical* in character, and that the results which they provide are to be regarded as true *on the average* for the systems in an appropriately chosen ensemble, rather than as necessarily precisely true in any individual case. (63-4, emphasis in original[7]).

FW read the physics literature differently, though: they survey thirty textbooks and find most or all committed to AP — that is, to the position that the actually measured value, not just the expectation of that value, equals the phase-space average — and hence, in their reading, to something beyond Bare Probabilism. I am not sure how substantive the difference in our readings is: of course the averaging principle AP* is true, indeed analytic, on the statistical reading of thermodynamics, so in at least some cases the issue is not whether a textbook asserts *an* averaging principle but whether it is committed to the categorical reading of thermodynamics, so that the averaging principle is the stronger AP. And in at least some cases they clearly are not: (Chandler 1987, p.58), for instance (cited by FW) writes that

> The primary assumption of statistical mechanics —- that the observed value of a property corresponds to the ensemble average of that property — seems reasonable, therefore, if the observation is carried out over a very long time **or if the observation is actually the average over very many independent observations**

(emphasis mine). But to explore this in all cases would be tedious and in any case not decisive (a pretty clear lesson of 20th century philosophy of science is that textbooks are not definitive guides to scientific practice), so I will proceed differently, by identifying three major research programs in contemporary thermodynamics which clearly presume the statistical interpretation of thermodynamics, and indeed would be unintelligible without it.

1. The protection of the Second Law from Maxwell's demon via considerations of the thermodynamic cost of erasure ("Landauer's Principle"), as inaugurated by Bennett (1982). The main focus of this literature

[7] I take Tolman's reference to an 'ensemble' as being an artifact of a then-prevalent frequentist reading of statistical probability, rather than as essential in the argument.

has been on very small thermodynamic systems (like the infamous 'one-molecule gas'), for which of course there can be no prospect of a version of the Second Law that holds for each individual case: fluctuations can always lead to spontaneous generation of work. The form of the Second Law being defended in this literature is very clearly and explicitly the expected-value form: even with access to a Maxwell demon, one cannot *in expectation* transform work to heat once erasure costs are considered. And this literature equally clearly relies on the identification of thermodynamic entropy with Gibbs entropy, which as Bennett notes (p.936) 'is an inherently statistical concept' that cannot even be expressed as the expected value of some microphysically stateable quantity.

2. The revolution in classical thermodynamics begun by Jarzynski's (1997) celebrated inequality and Crooks' (1998) closely related fluctuation theorem. The Jarzynski equality takes as its starting point a *statistical* interpretation of the Second Law for a system in contact with a heat reservoir: that the *expected* work performed on a system is greater or equal to its change in free energy,

$$\langle W \rangle \geq \Delta F. \tag{8}$$

   Jarzynski then strengthens this result to an *equality* (i. e., equation), again statistical in nature:

$$\langle \mathrm{e}^{-W/k_B T} \rangle = \mathrm{e}^{-\Delta F/k_B T} \tag{9}$$

   from which the classical inequality follows.

3. In quantum thermodynamics, the 'one-shot' program begun by Brandão *et al* (2015) again starts with the interpretation of the Second Law as a statistical *average*, and asks what we can say about work extraction if we instead consider *individual* approaches: a variety of inequalities have been established that strengthen the Second Law in the quantum context.

These are not niche programs in modern physics. (Bennett 1982) has over 2500 citations[8], and the Bennett-Landauer approach to Maxwell's demon is (disapprovingly) acknowledged as the current orthodoxy in Earman and

[8]All citation counts from Google Scholar (scholar.google.com), accessed 6/14/2024.

Norton's (1999) historical review of Maxwell's Demon. (Jarzynski 1997) has over 5800 citations; (Crooks 1998) has over 2900; the Nobel Prize committee[9] lists 'the first experimental test of Jarzynski's equality' as one of the major applications of optical tweezers, for which Arthur Ashkin shared the 2018 Physics prize. The single-shot program is much newer; still, (Brandão *et al* 2015) has over 750 citations.

None of these programs make any sense without a statistical reading of thermodynamics: in each case they quite explicitly make use of it. And in none of the sources I cite above — or indeed, in any of the other work I'm familiar with in any of these programs — is there any suggestion that the statistical reading is in need of defending or indeed that it admits any alternative; in each case, it is simply assumed without question. The authors of these sources cannot be regarded as fringe figures; indeed, some of them are acknowledged leaders of their field.[10] It is reasonable to take their views as expressing orthodoxy in at least some large part of the physics community.

To be sure, in *philosophy* of physics the statistical reading of thermodynamics has been severely criticized; it is varyingly claimed to rely on an incoherent notion of objective classical probability, to make objective thermodynamic facts an epistemic and subject-relative matter, to improperly move our focus from how actual individual systems behave, and more. (Variants of these criticisms are made by, e. g. , Albert (2000), Goldstein (2001), Callender (1999, 2001). FW cite (Callender 1999) approvingly.)

I find these complaints unpersuasive. Objective probability, however philosophically confusing it might be, is manifestly required for the empirical application of statistical mechanics and in any case looks much better once quantum mechanics is taken into account. The epistemic reading of Gibbsian statistical mechanics is in large part optional and in any case thermodynamics is concerned with our capacities, which might reasonably be influenced by our epistemic situation. And there is no difficulty understanding thermodynamics as making claims about individual systems, provided those claims are understood as probabilistic — something we are in any case used to from quantum physics. (I expand on this defence in (Wallace 2020, 2023); for

[9] Advanced Information, NobelPrize.org. Nobel Prize Outreach AB 2024. https://www.nobelprize.org/prizes/physics/2018/advanced-information/ Accessed 6/14/2024.

[10] For instance: Bennett received the Wolf Prize in 2018 and the Breakthrough Prize in 2023; Jarzynski received the Lars Onsager prize in 2019. These are among the highest honors in physics.

other defenses, see (Maroney 2007, Myrvold 2021, Robertson 2022).)

But in any case, whether the statistical reading of thermodynamics is correct or even defensible is beside the point. The question that FW ask, and that I am concerned with here, is 'can somebody please say what GSM says?' They, and I, are not concerned with the further question: 'is what GSM says correct?' I claim that there is overwhelming evidence that in contemporary physics practice, thermodynamics is understood statistically in the first instance, with categorical claims following, insofar as they do follow, only via large-number statistics, and so if we want to understand how GSM is interpreted in contemporary physics, our understanding needs to take this into account. When it is taken into account, there is no difficulty in bare probabilism for the reductive project of deriving thermodynamics from statistical mechanics.

5 Fluctuations revisited

Let's consider again FW's question of how to interpret fluctuations in GSM. Suppose that some system is initially at exact statistical equilibrium — that is, its statistical state ρ is time-invariant. (The interpretation of the claim that the system's statistical state is ρ will depend on our account of statistical-mechanical probabilities — it might mean that ρ is a quantum state, or a classical distribution understood as the classical limit of a quantum state (Wallace 2016), or ρ might be an 'epistemic chance' in the sense of (Myrvold 2021), or it might describe the relative frequencies of microstates in a large collection of systems from which our system has been randomly selected. Nothing in this section hinges on this interpretative question.)

Now suppose that O is a macrovariable for the system, and we measure O at some time t_1, the outcome will be somewhat random, with the probability of a given outcome determined by ρ, and following FW we can express the outcome as a fluctuation: that is, as a difference between the actually-obtained value and the mean value. If we want to *determine* this probability distribution, how should we proceed? Two strategies naturally come to mind. We might take a single system and measure O at a long sequence of times $t_1, t_2, \ldots t_N$. Alternately, we might take a large class of N identically prepared systems and measure O on each system once.

Whichever we do, the *probability measure* over values of O for each of the N measurements will be identical and will be determined by ρ. In the first case this follows from the fact that ρ is time-invariant; in the latter it is an analytic consequence of our assumption that the system is prepared in statistical state ρ. Does it follow, however, that the *statistics* of the N measurements can be expected to approximate the probability measure, if N is large enough?

Standard probability theory gives us an answer: yes, we can expect this, for large enough N, *provided* that the measurement outcomes are *uncorrelated*. In the case where we measure identically prepared systems, this is true by assumption, and so this method is reliable as a means of determining the distribution (or verifying a theoretical prediction of its value). But there is in general no reason why the outcomes of measurements at different times might not be correlated, and indeed there are perfectly realistic cases where we would expect just this. FW give some examples of this (p.124) but their discussion might give the impression that only exotic systems have this feature. In fact it can readily be found in systems of interest to mainstream statistical mechanics. I will give two examples.

The first is a single harmonic oscillator, previously in thermal contact with a heat bath (so that its statistical state is initially canonical) but now isolated from that bath. The canonical distribution will correctly describe the probability distribution over (say) joint measurements of its position and momentum, at any single time — but of course any such joint measurement at one time will completely fix the values of any measurement at any other time, so that a sequence of such measurements will not appear as if drawn randomly from an ensemble described by ρ but will trace out a (randomly initialized) harmonic trajectory. (And, since the oscillator is not a chaotic system, this will continue to be true approximately even if the initial measurement has finite precision.)

The second is a macroscopically, but finitely, large ferromagnet, initially heated to well above its ferromagnetic temperature and then allowed to cool. The canonical distribution will predict that the magnetic field at some specified point in the magnet has a non-zero magnitude and a random direction. But the direction at one time will almost always coincide with the direction at another time: the multi-time correlation between magnetic field values at any point in the magnet is almost perfect.

In each case, the existence of multi-time correlations is not something additional to GSM, but something that can be, and routinely is, calculated within GSM. Given a macrovariable O, let $O(t)$ be the macrovariable describing the value of O after the system has evolved under its Hamiltonian dynamics for time t. Formally,

$$O(t) = \exp(\{\cdot, H\}t)O \tag{10}$$

where $\{,\}$ is the Poisson bracket or $-i/\hbar$ times the commutator, as appropriate. (In QM this can be rewritten as

$$O(t) = \mathrm{e}^{iHt/\hbar} O \mathrm{e}^{-iHt/\hbar} \tag{11}$$

and we can recognize it as the Heisenberg-picture version of O.) Then the two-time correlation function

$$\langle O(t)O(0)\rangle_\rho - \langle O(t)\rangle_\rho \langle O\rangle_\rho \tag{12}$$

calculated on the equilibrium distribution ρ is just one more equilibrium expectation value. If it decreases to ~ 0 in a time τ, then measurements of O separated by time τ can be expected to be uncorrelated and their statistical distribution will match the single-time probability distribution of O. If it remains large at time τ, those measurements will be correlated and their statistics will not match that distribution.

Alternately (at least classically), we can state this in terms of events and probabilities. Given a subset Δ of phase space, define Π_Δ as a macrovariable taking value 1 on Δ and 0 otherwise. Now consider some sequence $\Delta_1, \ldots \Delta_N$, and ask what is the probability of the system being in Δ_1 at time 1, in Δ_2 and time 2, etc. We have

$$\Pr(\Delta_1, t_1; \Delta_2, t_2; \ldots \Delta_N, t_N) = \langle \Pi_{\Delta_n}(t_n) \cdots \Pi_{\Delta_1}(t_1)\rangle_\rho. \tag{13}$$

(A formally similar expression can be written down in quantum mechanics, but its interpretation as a probability requires *decoherence*; further discussion lies beyond the scope of this paper, though see (Wallace 2012, ch.3) for more.)

There is another, equivalent, way to look at this. Suppose we do measure O at time 0 and get result o. The correct probability distribution to use to describe the system with respect to subsequent measurements is no longer ρ but ρ conditionalized on $O = o$: that is, we should update the distribution to

$$\rho'(x) = \mathcal{N}\rho(x)\delta(O(x) - o) \tag{14}$$

with $\mathcal{N}$ a normalization constant. This will in general not be an equilibrium distribution; that is, conditionalizing on the result of the measurement has moved the system out of statistical equilibrium. Only if the system relaxes to equilibrium (or at least to macroequilibrium) before the next measurement will the probability distribution over that next measurement be independent of the result obtained. However, the *un*conditional distribution over results of that later measurement is unaffected by the mere presence of the measurement (unless it is done in such a way as to mechanically disturb the system): from an external point of view, there is a correlation between the probability distribution over system state and measurement outcome, but the marginal probability distribution over the system alone is unaffected.

At this point it will be helpful to reconsider FW's principle of ρ-universalism, which is

> the assertion that ρ is universally valid in that it gives the correct probabilities for all events (recognized by the theory) to occur at all times. (p.113)

We can now see that there is a subtle ambiguity in this principle. It could be interpreted (specializing for simplicity as any of

1. ρ gives the correct probability for any single event (identified with a phase-space region Δ), at any time t:
$$\Pr(\Delta, t) = \int_\Delta \rho. \tag{15}$$

2. ρ gives the correct probability for any sequence of events (identified with phase space regions $\Delta_1, \ldots \Delta_N$) at any times $t_1, \ldots t_n$, calculated via formula (13). Or equivalently: using ρ to predict the probability of an event conditional on a previous event requires us to conditionalize ρ itself on that previous event.

3. ρ gives the correct probability for any sequence of events (identified with phase space regions $\Delta_1, \ldots \Delta_N$) at any times $t_1, \ldots t_n$, calculated via
$$\Pr(\Delta_1, t_1; \Delta_2, t_2; \ldots \Delta_N, t_N) = \int_{\Delta_1} \rho \times \cdots \int_{\Delta_N} \rho. \tag{16}$$
Or equivalently: ρ gives the correct probability for any single event via formula (15), *irrespective* of what previous measurements have been made and what outcomes they have received.

Of these, (1) and (2) are direct consequences of Bare Probabilism; indeed, (2) is basically equivalent to bare probabilism plus a choice of Hamiltonian ((1) is strictly weaker, since it does not contain multi-time information and so does not fully constrain the Hamiltonian). But (3) follows from Bare Probabilism only if multi-time correlations decay quickly compared to the timescale on which repeated measurements are made — and so (3) is correct in GSM only when this further assumption is made.

In any case: when correlations do not decay sufficiently quickly, the probability distribution over a macrovariable is not measurable by repeated measurements of that same macrovariable on a single system. To measure that distribution, we instead need to use the alternative method: carry out single measurements on a number of identically prepared systems.

FW have concerns about this 'alternative method'. For one thing, they regard it as mismatched to experimental practice:

> In laboratory experiments we observe the same system at consecutive times rather than drawing different systems out of an urn at random. This difference is significant because successive observations on the same system are generally not independent in the way in which draws from an urn are, and so it remains unclear what we can infer about laboratory experiments from GSM under an 'urn interpretation'. (p.122)

But experimental statistical mechanics is more pluralistic than FW recognize. In modern nanoscale thermodynamics, for instance, microscopic systems — individual strands of DNA, say — are allowed to equilibrate with a thermal environment; then the experimenter intervenes on them, and the statistics of their responses are measured and compared with the probabilistic predictions. This has become a major area of experimental thermodynamics in the 21st century, driven by theoretical advances like the Jarzynski equality and technological developments like optical tweezers.

Still, FW are certainly correct that in at least an important class of applications of GSM, it *is* assumed that multi-time correlations can be neglected, so that conditionalization on previous measurement results is not needed and ρ can be interpreted directly as giving the statistics of measurements of a single system. And they have technical concerns both that this can in fact be justified in appropriate systems, and that in the case of identically-prepared systems we are in fact entitled to assume that the preparation method prepares systems in the appropriate (statistical equilibrium) state. I turn now to these concerns.

6 Gibbsian equilibration

In FW's discussion of repeated measurements of the same system, they worry that there are no general reasons to expect those repeated measurements to give statistics matching the synchronic probability distribution. We have seen that this worry is well-founded for some physically-realistic systems, including microscopic and broken-symmetry systems. Still, we can ask (and FW do ask): why expect *any* system to display these statistics on repeated measurement? As they correctly note: any system with conserved quantities other than energy will display multi-time correlations in repeated measurements of those quantities; if any macrovariable's probability distribution conditional on one value of the conserved quantity differs from its distribution conditional on another variable, then inevitably there will be persistent multi-time correlations in measurements of that quantity.

One class of such conserved quantities can be set aside as harmless. If there are *macrovariables* other than energy that are conserved, it is well known that (as I noted in my presentation of GSM in section 2) the appropriate choice of equilibrium distribution must take that into account. Particle number, for instance, is normally conserved; allowing for this fact leads us to use either a generalization of the canonical ensemble which is determined by the expectation values of both energy and particle number, or a generalized microcanonical ensemble that is uniform on a thin shell of states around a given energy and particle number. In general, the space of systems at equilibrium is parameterized by energy *and any other conserved macrovariables*, both in statistical mechanics and in thermodynamics.

But what if there is a conserved quantity that is *not* a macrovariable, and yet which is such that some macrovariables' values are correlated with it? Then, by definition, the system will not equilibrate: its statistical state will not evolve into a macroequilibrium state. In other words, the problem FW identify is just the problem of establishing that systems which we think 'ought' to equilibrate actually do. The properties of masking and independence which they discuss are just aspects of the definition of macroequilibration.

More precisely: in GSM, and given the statistical interpretation of thermodynamics, a given system *macro-equilibrates* with a relaxation time τ iff, given any physically reasonable way of preparing the system at an initial time t_0, the system is in statistical macroequilibrium for later times t with $t - t_0 \gg \tau$. And it *canonically macroequilibrates*, with respect to some (possibly empty) set of conserved quantities, if in addition the distribution it evolves to determines the same probability distribution over macrovariables as the (unique) canonical distribution determined by the initial expected values of the energy and those conserved quantities. (The reference to 'physically reasonable ways of preparing a system' is somewhat vague, but can be sharpened in the context of repeated measurements: there we are interested in preparing a system by beginning with one already in statistical equilibrium and then sequentially measuring some macrovariables.)

If a system canonically macroequilibrates to a distribution ρ with equilibration time τ, then repeated measurements made at intervals $\gg \tau$ will be statistically independent (and determined by ρ at each time separately). And this is exactly what is required to justify any match of the statistics obtained in such a sequence of measurements with the probability distribution determined by ρ.

FW are also concerned (and here they follow earlier observations by Leeds (1989) and Werndl (2013)) that we cannot assume that the preparation methods used to prepare 'identically prepared systems' after all succeed in placing the system in statistical macroequilibrium, let alone in the canonical distribution or a state macroequivalent to it. But when one looks at those preparation methods in detail, nearly always they involve waiting long enough for some system to come to equilibrium, and then extracting the system to be studied as a subsystem of that system. In nanoscale thermodynamics, for instance, the DNA strand or other system being studied is initially in thermal contact with a heat bath, and remains so for a time much longer than its relaxation time prior to the start of the experiment.[11] So again, the problem of establishing a system is properly prepared is in most cases just an aspect of the general problem of establishing that a system equilibrates, and does so on an appropriate timescale.

(What about systems that are *not* so prepared? Then it is necessary to check the actual preparation method and see if it prepares the system in

[11]In general, in fact, it remains in thermal contact with the heat bath throughout the experiment, though typically the measurement timescales in the experiment are *not* long compared to the relaxation time.

statistical equilibrium. If it does not, then equilibrium statistical mechanics cannot be used to study it — but whoever thought otherwise? When systems equilibrate quickly, of course more or less any preparation process will do, but if they don't, we need to pay attention to the way the system is prepared, to make sure it actually has been prepared in statistical equilibrium. In the two cases I described in section 5 (the single particle and the ferromagnet) I was careful to describe their preparation to ensure this: the single particle was extracted from an equilibrated collection of many particles; the ferromagnet was cooled from a temperature at which it would have quickly equilibrated. Alternative preparation processes might not produce systems in statistical equilibrium and in this case the equilibrium distribution will not give the right answers. If I prepared a collection of ferromagnets by magnetizing them all and then giving them to a toddler to play with, probably this will not produce a collection of systems at statistical equilibrium — maybe she likes lining things up).

Of course, rigorously proving equilibration is notoriously difficult, and for a century the standard approach has been to provide heuristic arguments as to why equilibration should be expected and then to proceed with statistical mechanics as if it were correct. One could dispute that strategy, but that dispute does not seem particularly relevant to the interpretation of GSM. (Nor is a Boltzmannian approach to statistical mechanics any better off here. On the Boltzmannian conception of equilibration, a "macrostate" is a region of phase space where the macrovariables are approximately constant, and the "equilibrium macrostate" is the largest such macrostate. Microscopic conserved quantities, if there are any, can cause a system to get stuck in the "wrong" macrostate just as easily as they can cause a system to display the "wrong" statistics.)

That said, the last 30 years have seen very substantial progress in our understanding of equilibration, mostly in the quantum rather than classical regime. A full review would be far beyond the scope of this paper (see, e. g. , (d'Alessio *et al* 2016) and references therein), but to give one important example: the "eigenstate thermalization hypothesis" (Deutsch 1991; Srednicki 1994) relates equilibration to certain statistical features of the distribution of a large system's energy eigenstates — statistical features which provably characterize 'typical' Hamiltonians (for an appropriate definition of 'typical') and which can be checked for specific systems via numerical simulation. Equilibration is a difficult problem, but it is not an intractable one.

In any case, the task at hand is (once again) to establish what GSM says, not whether what it says is correct. Perhaps physicists are mistaken in thinking that equilibration ought to be characterized in terms of a probability distribution evolving towards one macroscopically equivalent to the canonical distribution; perhaps they are overconfident that the non-rigorous calculational tools of theoretical statistical mechanics both point towards an understanding of equilibration and give us means to calculate quantitatively what relaxation times are.[12] But all that is required to understand the content of GSM is to recognize that at least a large fraction of physicists hold these views about equilibration.

7 Conclusion

Gibbsian statistical mechanics assigns a probability distribution ρ_t to a system at time t, and its interpretation is nothing more or less than that this probability distribution is, indeed, the statistical state of the system: that is, it determines the probability distribution over, and so the expectation value of, any variables for the system, including multi-time variables. But a great deal follows from this. GSM determines the multi-time correlation functions that describe both fluctuations at equilibrium and the approach to equilibrium. It determines the control theory of interventions on a statistical-mechanical system that is statistical thermodynamics: a theory which establishes limits on the expectation values of the work extractable from a system through various control operations. It provides the foundation for the remarkable recent work in thermodynamics that goes beyond these limits and actually provides sharp predictions of expectation values in thermodynamic change. And, through the law of large numbers, it provides a microphysical grounding for the categorical thermodynamics of large systems.

That's what Gibbsian statistical mechanics says.

Acknowledgements

Thanks to Roman Frigg, Charlotte Werndl, and Wayne Myrvold for helpful discussions.

[12]For the avoidance of doubt: I reject both suggestions: to me both the arguments and the empirical evidence seem overwhelming that the Gibbsian characterization of equilibration is broadly correct, and that the methods of modern non-equilibrium statistical mechanics give deep insight into it, qualitatively and quantitatively.

References

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

Bennett, C. H. (1982). The thermodynamics of computation — a review. *International Journal of Theoretical Physics 21*, 905–940.

Brandão, F., M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner (2015). The second laws of quantum thermodynamics. *PNAS 112*, 3275–3279.

Callender, C. (1999). Reducing thermodynamics to statistical mechanics: the case of entropy. *Journal of Philosophy 96*, 348–373.

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

Calzetta, E. A. and B.-L. B. Hu (2008). *Nonequilibrium Quantum Field Theory*. Cambridge: Cambridge University Press.

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

Crooks, G. E. (1998). Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *Journal of Statistical Physics 90*, 1481.

D'Alessio, L., Y. Kafri, A. Polkovnikov, and M. Rigol (2016). From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics. *Advances in Physics 65*, 239–362.

Deutsch, J. (1991). Quantum statistical mechanics in a closed system. *Physical Review A 43*, 2046.

Earman, J. and J. Norton (1999). EXORCIST XIV: The wrath of Maxwell's demon. part II. from Szilard to Landauer and beyond. *Studies in the History and Philosophy of Modern Physics 30*, 1–40.

Frigg, R. and C. Werndl (2021). Can someone please say what Gibbsian statistical mechanics says? *British Journal for the Philosophy of Science 72*, 105–129.

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

Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters 78*, 2690.

Lebowitz, J. L. (1993). Boltzmann's entropy and time's arrow. *Physics Today 46*, 32–38.

Leeds, S. (1989). Malament and Zabell on Gibbs phase averaging. *Philosophy of Science 56*, 325–340.

Luczak, J. (2016). On how to approach the approach to equilibrium. *Philosophy of Science 83*, 393–411.

Maroney, O. (2007). The physical basis of the Gibbs-von Neumann entropy. https://arxiv.org/abs/quant-ph/0701127.

McCoy, C. (2020). An alternative interpretation of statistical mechanics. *Erkenntnis 85*, 1–21.

Myrvold, W. C. (2021). *Beyond Chance and Credence*. Oxford. Oxford University Press.

Penrose, O. (1979). Foundations of statistical mechanics. *Reports on Progress in Physics 42*, 1937–2006.

Robertson, K. (2020). Asymmetry, abstraction, and autonomy: Justifying coarse-graining in statistical mechanics. *British Journal for the Philosophy of Science 71*, 547–579.

Robertson, K. (2022). In search of the holy grail: How to reduce the second law of thermodynamics. *British Journal for the Philosophy of Science 73*, 987–1020.

Srednicki, M. (1994). Chaos and quantum thermalization. *Physical Review E 50*, 888.

Tolman, R. C. (1938). *The principles of statistical mechanics*. Oxford University P.

van Lith, J. (1999). Reconsidering the concept of equilibrium in classical statistical mechanics. *Philosophy of Science 66*, S107–S118.

Wallace, D. (2012). *The Emergent Multiverse: Quantum Theory according to the Everett Interpretation*. Oxford: Oxford University Press.

Wallace, D. (2015). The quantitative content of statistical mechanics. *Studies in the History and Philosophy of modern Physics 52*, 285–293. Originally published online under the title "What statistical mechanics actually does".

Wallace, D. (2016). Probability and irreversibility in modern statistical mechanics: Classical and quantum. To appear in D. Bedingham, O. Maroney and C. Timpson (eds.), *Quantum Foundations of Statistical Mechanics* (Oxford University Press, forthcoming). Preprint at https://arxiv.org/abs/2104.11223.

Wallace, D. (2020). The necessity of Gibbsian statistical mechanics. In V. Allori (Ed.), *Statistical Mechanics and Scientific Explanation: Determinism, Indeterminism and Laws of Nature*. World Scientific.

Wallace, D. (2023). Thermodynamics with and without irreversibility. To appear in Olimpia Lombardi and Cristian Lopez (eds.), *The Arrow of Time: From Local Systems to the Whole Universe* (Cambridge University Press, forthcoming). Preprint at https://philsci-archive.pitt.edu/22273/.

Werndl, C. (2013). Justifying typicality measures of Boltzmannian statistical mechanics and dynamical systems. *Studies in History and Philosophy of Modern Physics 44*, 470–479.