

Gibbs versus Boltzmann

Harvey R. Brown

Faculty of Philosophy, University of Oxford, Oxford, U.K.

Michael C. Mackey

*Departments of Physiology, Physics & Mathematics, McGill University, 3655 Promenade
Sir William Osler, Montreal, QC, CANADA,
H3G 1Y6*

1. Introduction

It is instructive to compare the approach used by Boltzmann and Gibbs in their treatments of statistical mechanics. Both started from the assumption that they were dealing with systems of dimension $d = 2s$ whose dynamics were described by s position variables x_i and s momentum variables p_i .

Boltzmann considered the phase space to be a $2s$ dimensional space which is usually called μ **space**. He then considered the evolution of a large number N of identical particles, each with the same dynamics, in μ space. N is large and typically on the order of Avagadro's number, 6×10^{23} . The limiting case of $N \rightarrow \infty$ is the **thermodynamic limit** in which case the Boltzmann approach is equivalent to studying the evolution of a density in μ space.

Gibbs also considered N identical particles operating with these $2s$ dimensional dynamics in a phase space (commonly called the Γ **space**) of dimension $2sN$. He then considered an infinite number of copies of this original system, and gave this construct the name **ensemble**. Thus Gibbs studies the evolution of the ensemble density, and Γ space has proved to be the most useful in statistical mechanics.

The curious, and disturbing, fact is, however, that when we do experiments we do Boltzmann experiments in the sense that we actually make measurements on a single system. We *not* make measurements on an

Email addresses: harvey.brown@philosophy.ox.ac.uk (Harvey R. Brown), michael.mackey@mcgill.ca (Michael C. Mackey)

ensemble! However, the puzzlement is that it is by following the Gibbs approach that we arrive at a sensible definition of *equilibrium entropy* and the calculations using the Gibbs' entropy are the ones that actually agree with what is seen physically.

However, and this is a big however, when it comes to making predictions about non-equilibrium behaviour we then revert to the use of another construct of Boltzmann, namely the Boltzmann equation, and all of the attendant transport equations (for mass, momentum, energy, etc.) that can be derived from it.

So what is the origin of these paradoxes? How can they be resolved? Can they be resolved?

1.1. A short calculation

Consider a gas containing \mathcal{N} molecules per litre (i.e., per 10^3 cc), and each molecule is of mass m and cross sectional area $\pi\sigma^2$. According to (Reichl, 1980, page 459) the collision frequency F_c (units of collisions per unit time) is given by

$$F_c = \frac{\sqrt{2}}{2} \mathcal{N}^2 \sigma^2 \left(\frac{8\pi kT}{m} \right)^{1/2}. \quad (1)$$

Furthermore, from my notes of 1964, the frequency of collisions F_s (units of collisions per unit area per unit time) with a surface in the gas is given by

$$F_s = \frac{\mathcal{N}}{4\pi} \left(\frac{8\pi kT}{m} \right)^{1/2}. \quad (2)$$

Now assume that we divide up our litre of \mathcal{N} molecules (something on the order of 10^{23}) into M tiny little cubes (cubelets), so each one will contain $N = \mathcal{N}/M$ particles in a cube of side L and volume of

$$L^3 = V_\epsilon = \frac{10^3}{M}.$$

Then, the collision frequency within each little volume V_ϵ will be given by

$$F_{c,V_\epsilon} = \frac{\sqrt{2}}{2} \mathcal{N}^2 \sigma^2 L^3 \left(\frac{8\pi kT}{m} \right)^{1/2} \quad (3)$$

$$= \frac{\sqrt{2}}{2} \mathcal{N}^2 \sigma^2 \frac{10^3}{M} \left(\frac{8\pi kT}{m} \right)^{1/2} \quad (4)$$

This value gives us a measure of the frequency of interaction between molecules in each cubelet.

Furthermore, the collision frequency (collisions per unit time) with one of the sides of the cube of cross-sectional area L^2 will be given by

$$\begin{aligned} F_{s,V_\epsilon} &= \frac{\mathcal{N}}{4\pi} L^2 \left(\frac{8\pi kT}{m} \right)^{1/2} \\ &= \frac{10^2 \mathcal{N}}{4\pi M^{2/3}} \left(\frac{8\pi kT}{m} \right)^{1/2}. \end{aligned} \quad (5)$$

However, since there are 6 sides to the cube, the total collision frequency with the sides will be given by

$$F_{s,V_\epsilon,ToT} = \frac{6 \times 10^2 \mathcal{N}}{4\pi M^{2/3}} \left(\frac{8\pi kT}{m} \right)^{1/2}. \quad (6)$$

This “side collision frequency” is, in point of fact, a measure of the frequency with which one little cubelet interacts with its six adjacent neighboring cubelets.

With this preparation in hand, we can now look at the ratio between the interior collision frequency F_{c,V_ϵ} and the side collision frequency $F_{s,V_\epsilon,ToT}$:

$$\begin{aligned} R &= \frac{F_{c,V_\epsilon}}{F_{s,V_\epsilon,ToT}} = \frac{\frac{10\sqrt{2}\mathcal{N}^2}{2M} \sigma^2 \left(\frac{8\pi kT}{m} \right)^{1/2}}{\frac{6 \times 10^2 \mathcal{N}}{4\pi M^{2/3}} \left(\frac{8\pi kT}{m} \right)^{1/2}} \\ &= \frac{10\sqrt{2}\pi\mathcal{N}\sigma^2}{3M^{1/3}} = \frac{10\sqrt{2}\pi}{3} \mathcal{N}\sigma^2 M^{-1/3}. \end{aligned} \quad (7)$$

Thus, we can see that if we have particles with $\sigma = 5 \times 10^{-8}$ cm and take $M \simeq 10^9$ then interactions between the approximately 10^{14} molecules in the interior of the cubelet will occur about 2.25×10^7 times more frequently than will interactions with its six neighboring cubelets. Thus, maybe we can consider each cubelet as a “quasi-particle” and the collection (ensemble) of M quasi-particles is for all practical purposes the ensemble of Gibbs while each of the cubelets is one of Boltzmann’s systems.

This, I think, offers a way in which we can reconcile the fact that we do Boltzmann-like experiments but use a Gibbs approach to compute thermodynamic quantities.

1.2. *Something from statistics to think about*

Walecka (2000) presents the lecture notes of Felix Bloch, and he has a very interesting approach to this entire subject that also avoids the issue of ergodicity. This is a summary of notes that I took from his book while in Cancun on our 25th anniversary, December, 2002. As one can see, it is merely the use of the Chebyshev inequality but disguised!

Consider a system distributed with a density f on a phase space \mathcal{X} , and some functional ϕ on \mathcal{X} , e.g. energy. Then the *mean value* of ϕ is defined by

$$\bar{\phi} = \int_{\mathcal{X}} \phi(x) f(x) dx, \quad (8)$$

and the *deviation* is

$$\Delta\phi(x) \equiv \phi(x) - \bar{\phi} \quad (9)$$

so the *mean square deviation* is given by

$$\begin{aligned} \overline{(\Delta\phi)^2} &= \int_{\mathcal{X}} [\phi(x) - \bar{\phi}]^2 f(x) dx \\ &= \int_{\mathcal{X}} \phi^2(x) f(x) dx - 2 \int_{\mathcal{X}} \phi(x) \bar{\phi} f(x) dx + \int_{\mathcal{X}} \bar{\phi}^2 f(x) dx \\ &= \bar{\phi}^2 - 2\bar{\phi}^2 + \bar{\phi}^2 \\ &\equiv \bar{\phi}^2 - \bar{\phi}^2 \geq 0 \end{aligned} \quad (10)$$

since $[\phi(x) - \bar{\phi}]^2 \geq 0$. Define the RMS (root mean square) deviation by

$$(\Delta\phi)_{rms} \equiv \sqrt{\overline{(\Delta\phi)^2}}. \quad (11)$$

Now pick an arbitrary $\phi^* \equiv \phi(x^*)$ and consider the deviation

$$\Delta\phi^* \equiv \phi(x) - \phi^* \quad (12)$$

The *average deviation*

$$\overline{(\Delta\phi^*)} = \int_{\mathcal{X}} [\phi(x) - \phi^*] f(x) dx \quad (13)$$

vanishes if $\phi^* = \bar{\phi}$. Also, the mean square deviation

$$\begin{aligned} \overline{(\Delta\phi^*)^2} &= \int_{\mathcal{X}} [\phi(x) - \phi^*]^2 f(x) dx \\ &= \int_{\mathcal{X}} \phi^2(x) f(x) dx - 2 \int_{\mathcal{X}} \phi(x) \phi^* f(x) dx + \phi^{*2} \\ &= \bar{\phi}^2 - 2\bar{\phi}\phi^* + \phi^{*2} \end{aligned} \quad (14)$$

is minimized when $\phi^* = \bar{\phi}$.

To see the meaning of the mean $\bar{\phi}$ we want some indication of the deviation $\Delta\phi$ to be expected for a system.¹ Now consider $(\Delta\phi)_{rms} > 0$, and pick a constant $\mathcal{C} > 0$ and $\Omega \subset \mathcal{X}$ such that

$$|\phi(x) - \bar{\phi}| \geq \mathcal{C}(\Delta\phi)_{rms} \quad \text{for all } x \in \Omega. \quad (17)$$

Thus, $\Omega \subset \mathcal{X}$ is the region where $|\phi(x) - \bar{\phi}|$ is \mathcal{C} times greater than $(\Delta\phi)_{rms}$. Square Equation 17, multiply by $f(x)$ and integrate over Ω to give

$$\int_{\Omega} |\phi(x) - \bar{\phi}|^2 f(x) dx \geq \mathcal{C}^2 \overline{(\Delta\phi)^2} \int_{\Omega} f(x) dx. \quad (18)$$

However, by definition

$$\begin{aligned} \overline{(\Delta\phi)^2} &= \int_{\mathcal{X}} [\phi(x) - \bar{\phi}]^2 f(x) dx \\ &= \int_{\mathcal{X} \setminus \Omega} [\phi(x) - \bar{\phi}]^2 f(x) dx + \int_{\Omega} [\phi(x) - \bar{\phi}]^2 f(x) dx \end{aligned} \quad (19)$$

so

$$\begin{aligned} \overline{(\Delta\phi)^2} &\geq \int_{\Omega} [\phi(x) - \bar{\phi}]^2 f(x) dx \\ &\geq \mathcal{C}^2 \overline{(\Delta\phi)^2} \int_{\Omega} f(x) dx. \end{aligned} \quad (20)$$

1

1. Note that the average deviation is

$$\overline{\Delta\phi} \equiv \int_{\mathcal{X}} [\phi(x) - \bar{\phi}] f(x) dx = 0 \quad (15)$$

but the RMS deviation might be a good measure.

2. However, note that

$$(\Delta\phi)_{rms} = \sqrt{\int_{\mathcal{X}} [\phi(x) - \bar{\phi}]^2 f(x) dx} \quad (16)$$

will be identically zero only if $f(x)$ is a delta function on a subset of \mathcal{X} such that $\phi(x) \equiv \bar{\phi}$.

3. Therefore, if $(\Delta\phi)_{rms} \equiv 0 \implies \phi(x) \equiv \bar{\phi}$.

For $\overline{(\Delta\phi)^2}$ finite we have

$$\begin{aligned} 1 &\geq \mathcal{C}^2 \int_{\Omega} f(x) dx \\ &\geq \mathcal{C}^2 \mu(\Omega) \end{aligned} \quad (21)$$

so

$$\mu(\Omega) \leq \frac{1}{\mathcal{C}^2}. \quad (22)$$

That is, the measure of the region Ω in which

$$|\phi(x) - \bar{\phi}| \geq \mathcal{C}(\Delta\phi)_{rms} \quad (23)$$

holds is less than

$$\frac{1}{\mathcal{C}^2}.$$

So what does this all really mean? Simply that for **really large** deviations of $\phi(x)$ from $\bar{\phi}$ (that is, large in absolute value), which would correspond to $\mathcal{C} \gg 1$, then the measure (volume) of phase space for which this will occur is very small (that is, $\leq 1/\mathcal{C}^2$). Therefore, it is highly unlikely to find large deviations of $\phi(x)$ from $\bar{\phi}$.

2. The Gibbs' entropy

With a dynamics S and initial density $f_0(x) = f(0, x)$ of states, the evolution of the density $f(t, x)$ is given by

$$f(t, x) = P_S^t f_0(x), \quad (24)$$

wherein P_S is the *transfer operator* corresponding to S . If the dynamics are described by a system of ordinary differential equations

$$\frac{dx_i}{dt} = F_i(x) \quad i = 1, \dots, d, \quad (25)$$

then the evolution of $f(t, x) \equiv P^t f_0(x)$ is governed by the generalized Liouville equation

$$\frac{\partial f}{\partial t} = - \sum_i \frac{\partial(f F_i)}{\partial x_i}. \quad (26)$$

We define a fixed point f_* of the transfer operator as a *stationary density* with $P^t f_* \equiv f_*$ for all t . A stationary density is important since it corresponds to a state of thermodynamic equilibrium. Note in particular that there is absolutely no requirement for a stationary density to be uniform across the phase space \mathcal{X} . For the system of ordinary differential equations (25), f_* is given by the solution of

$$\sum_i \frac{\partial(f_* F_i)}{\partial x_i} = 0. \quad (27)$$

Note that the uniform density $f_* \equiv 1$ is a stationary density of Equation 26 if and only if

$$\sum_i \frac{\partial F_i}{\partial x_i} = 0. \quad (28)$$

Having postulated that a thermodynamic system has a state characterized by a density f , we are now in a position to develop the physically useful concept of Gibbs' entropy.

First we define an observable $\mathcal{O}(f)$ to be a functional of the thermodynamic state characterizing some aspect of a system, for example the energy, pressure, or temperature. As such, an observable corresponds to a map $\mathcal{O} : \mathcal{D}(\mathcal{X}) \rightarrow R$. The expected, or average, value of the observable $\mathcal{O}(f)$ is given by weighting $\mathcal{O}(f)$ with the system state density f and integrating over the entire phase space:

$$E(\mathcal{O}) = \langle \mathcal{O} \rangle = \int_{\mathcal{X}} \mathcal{O}(f(x)) f(x) dx. \quad (29)$$

In his seminal work Gibbs (1962), assuming the existence of a system state equilibrium density f_* on the phase space \mathcal{X} , introduced the concept of the index of probability given by $\log f_*(x)$ where "log" denotes the natural logarithm. He then identified $-\log f_*$ with the entropy. We identify the entropy H in an equilibrium situation with the average of the index of probability

$$H^G(f_*) = - \int_{\mathcal{X}} f_*(x) \log f_*(x) dx, \quad (30)$$

and call this the Gibbs' entropy. It can be shown that $-\infty < H(f) \leq 0$ for all densities f .

If entropy is to be an extensive quantity (in accord with practical experience) then this definition is unique up to a multiplicative constant (Khinchin,

1949; Skagerstam, 1974). Furthermore, (30) has repeatedly proven to yield correct results when applied to a variety of equilibrium situations. This is why it is the gold standard for equilibrium computations in statistical mechanics and thermodynamics. Thus we identify the equilibrium Gibbs' entropy $H^G(f_*)$ with the thermodynamic entropy S_{TD}^* .

Other contenders for the definition of entropy, such as one version of the Boltzmann entropy (Jaynes, 1965), fail to give proper answers for equilibrium calculations unless all particles are non-interacting.

The uniqueness of the entropy definition (30) (under the assumption that the entropy is an extensive quantity) is so important that it is worthwhile to give the proof. It is short.²

²Consider two systems A and B operating in the phase spaces \mathcal{X}_A and \mathcal{X}_B respectively, and each having the densities of states f_A and f_B . We combine the two systems to form a new system C operating in the product space $\mathcal{X}_C = \mathcal{X}_A \times \mathcal{X}_B$, so system C will have a density of states $f_C(x, y) = f_A(x)f_B(y)$ if A and B do not interact. On experimental grounds we require that when the two systems are combined into a larger system C , then the entropy of system C should equal the sum of the individual entropies of A and B , since entropy is an extensive (additive) system property. We wish to show that the Gibbs' choice for the index of probability is the only choice (up to a multiplicative constant) that will ensure this. Assume that the index of probability is left as an unspecified observable $\mathcal{O}(f)$. If the observable $\mathcal{O}(f)$ is such that it transforms products to sums, $\mathcal{O}(f_C) = \mathcal{O}(f_A f_B) = \mathcal{O}(f_A) + \mathcal{O}(f_B)$, then the relation $H(f_A) + H(f_B) = H(f_C)$ holds. It is clear that picking $\mathcal{O}(w) = d \log w$, where d is any arbitrary non-zero constant, will work but are there any other functions \mathcal{O} with the requisite property? Assume there exists a second continuous observable $\tilde{\mathcal{O}}(f)$ such that

$$\tilde{\mathcal{O}}(f_A f_B) = \tilde{\mathcal{O}}(f_A) + \tilde{\mathcal{O}}(f_B). \quad (31)$$

Define two new functions $v_A(a)$ and $v_B(b)$ through

$$f_A(a) = e^{v_A(a)} \quad \text{and} \quad f_B(b) = e^{v_B(b)}. \quad (32)$$

Then we have

$$\tilde{\mathcal{O}}(e^{v_A+v_B}) = \tilde{\mathcal{O}}(e^{v_A}) + \tilde{\mathcal{O}}(e^{v_B}), \quad (33)$$

or with $h(w) \equiv \tilde{\mathcal{O}}(e^w)$ this becomes

$$h(v_A + v_B) = h(v_A) + h(v_B). \quad (34)$$

This is the Cauchy functional equation that has the unique solution $h(w) = \delta w$ with δ an arbitrary constant (Kuczma, 1985). This implies that $\tilde{\mathcal{O}}(e^w) = \delta w$ so

$$\tilde{\mathcal{O}}(w) = \delta \log w. \quad (35)$$

An examination of the proof of the uniqueness of the Gibbs definition of entropy shows that the proof applies equally well in equilibrium and non-equilibrium situations. It is for this reason that we extend the definition of the equilibrium Gibbs' entropy to non-equilibrium situations and say that the *non-equilibrium Gibbs' entropy* of a density $f(t, x)$ is defined by

$$H_t^G(f) = - \int_{\mathcal{X}} f(t, x) \log f(t, x) dx. \quad (36)$$

We identify $H_t(f)$ with the non-equilibrium entropy $S_{TD}(t)$.

Finally, we extend these notions and define the non-equilibrium conditional entropy of the density $f(t, x)$ with respect to a stationary density $f_*(x)$:

$$H_t^{G,c}(f|f_*) = - \int_{\mathcal{X}} f(t, x) \log \left[\frac{f(t, x)}{f_*(x)} \right] dx. \quad (37)$$

As before it is the case that $-\infty < H_t^c(f|f_*) \leq 0$ for all densities f and f_* . Notice that if the phase space \mathcal{X} is finite and the stationary density is uniform on \mathcal{X} so $f_*(x) = 1/\mu(\mathcal{X})$ for all $x \in \mathcal{X}$ [this is a generalization of the density of the microcanonical ensemble], then (37) reduces to $H_t^{G,c}(f|f_*) = H(f) - \log \mu(\mathcal{X})$. If the space \mathcal{X} is normalized then $f_* = 1$ and $H_t^{G,c}(f|1) = H_t(f)$ as defined in (36). Furthermore, we can write (37) in the form

$$H_t^{G,c}(f|f_*) = H(f) - H(f_*) + \int_{\mathcal{X}} [f(t, x) - f_*(x)] \log f_*(x) dx. \quad (38)$$

Thus it is clear that if there is a convergence $\lim_{t \rightarrow \infty} f(t, x) = f_*(x)$ in some sense, then $\lim_{t \rightarrow \infty} H_t^{G,c}(f|f_*) = 0$. It is for this reason that we identify the convergence of $H_t^{G,c}(f|f_*)$ to zero [$\lim_{t \rightarrow \infty} H_t^c(f|f_*) = 0$] *with the convergence of the entropy difference $\Delta S(t)$ to zero*.

Most importantly for our considerations here, the conditional entropy of any invertible system is constant and uniquely determined by the method of system preparation. In particular, for the system of ordinary differential equations (25) whose density evolves according to the Liouville equation (26)

Thus the observable that gives the requisite additive property for the Gibbs' entropy is the logarithmic function and it is unique up to a multiplicative constant. The question of how a non-equilibrium entropy should be defined has plagued investigators for over a century. However, what is clear is that the definition of non-equilibrium entropy must agree with the Gibbs definition **at equilibrium**.

we can assert that the entropy of the density $P^t f_0$ will be constant for all time and will have the value determined by the initial density f_0 with which the system is prepared. This result can also be proved directly by noting that from the definition of the entropy we may write

$$H_t^{G,c}(f|f_*) = - \int_{R^d} f(x) \left[\log \left(\frac{f}{f_*} \right) + \frac{f_*}{f} - 1 \right] dx \quad (39)$$

when the stationary density is f_* . Differentiating with respect to time gives

$$\frac{dH_t^{G,c}}{dt} = - \int_{R^d} \frac{df}{dt} \log \left[\frac{f}{f_*} \right] dx \quad (40)$$

or, after substituting from (26) for $(\partial f / \partial t)$, and integrating by parts under the assumption that f has compact support,

$$\frac{dH_t^{G,c}}{dt} = \int_{R^d} \frac{f}{f_*} \sum_i \frac{\partial(f_* F_i)}{\partial x_i} dx. \quad (41)$$

However, since f_* is a stationary density of P^t , it is clear from (26) that

$$\frac{dH_t^{G,c}}{dt} = 0, \quad (42)$$

and we conclude that the conditional entropy $H_t^{G,c}(P^t f_0|f_*)$ does not change from its initial value when the dynamics evolve in this manner. This is a specific example of the more general conclusion that entropy can never evolve away from its initial value in systems with invertible (time reversal invariant) dynamics.

3. Coarse graining

To examine the effect of imprecision in our knowledge of dynamical variables on entropy calculations, we introduce the concept of the entropy of a coarse grained density, or more briefly, the **coarse grained entropy**. This concept seems to have been first qualitatively discussed by Gibbs (1962), and quantified by Ehrenfest and Ehrenfest (1959). Denbigh and Denbigh (1985) have considered aspects of the effects of coarse graining on the behaviour of entropy.

Coarse graining is carried out by first partitioning the phase space X (finite and normalized) into discrete cells A_i that satisfy

$$\bigcup_i A_i = X \quad \text{and} \quad A_i \cap_{i \neq j} A_j = \emptyset. \quad (43)$$

Obviously, there is no unique way in which such a partition $\{A_i\}$ may be formed, but we require that the partition is **nontrivial** with respect to the Lebesgue measure μ_L so $0 < \mu_L(A_i) \leq \mu_L(X) = 1$ for all values of i . For every density f , within each cell A_i of this partition we denote the average of f over A_i by $\langle f \rangle_i$,

$$\langle f \rangle_i = \frac{1}{\mu_L(A_i)} \int_{A_i} f(x) dx, \quad (44)$$

so the density f coarse grained with respect to the partition A_i is given by

$$f^{cg}(x) = \sum_i \langle f \rangle_i 1_{A_i}(x). \quad (45)$$

Thus, f^{cg} is constant within each cell A_i . Clearly $\sum_i \langle f \rangle_i \mu_L(A_i) = 1$.

Therefore, given a partition A_i satisfying (45) (nontrivial with respect to Lebesgue measure), a density f , and a coarse grained density f^{cg} defined by (44)-(45), then the Gibbs' entropy of the coarse grained density f^{cg} is given by³

$$H^G(f^{cg}) = - \sum_i \langle f \rangle_i \mu_L(A_i) \log \langle f \rangle_i.$$

³The demonstration is almost trivial since, from (45) and the definition of the entropy, we have

$$\begin{aligned} H^G(f^{cg}) &= - \int_X \left[\sum_i \langle f \rangle_i 1_{A_i}(x) \right] \log \left[\sum_i \langle f \rangle_i 1_{A_i}(x) \right] dx \\ &= - \sum_i \langle f \rangle_i \int_X 1_{A_i}(x) \log \left[\sum_i \langle f \rangle_i 1_{A_i}(x) \right] dx \\ &= - \sum_i \langle f \rangle_i \int_{A_i} \log \left[\sum_i \langle f \rangle_i 1_{A_i}(x) \right] dx \\ &= - \sum_i \langle f \rangle_i \mu_L(A_i) \log \langle f \rangle_i. \end{aligned} \quad (46)$$

It is noteworthy that for any density f , the Gibbs' entropy of the coarse grained density f^{cg} may be greater than the entropy of f , or more specifically:

Theorem 1. *For any density f and any nontrivial partition A_i of the phase space X , $H^G(f) \leq H^G(f^{cg})$.⁴*

Thus, the effect of any error in the estimation of a density f characterizing a system, no matter what the origin, will be to either increase the Gibbs' entropy of the estimated (coarse grained) density $H^G(f^{cg})$ above its actual value $H^G(f)$, or leave it unchanged.

Precisely analogously to the way in which the entropy of the coarse grained density was derived, it is easy to show that the conditional entropy of f^{cg} with respect to a second density g , also coarse grained with respect to the partition A_i , is given by

$$H^{G,c}(f^{cg}|g^{cg}) = - \sum_i \langle f \rangle_i \mu_L(A_i) \log \left[\frac{\langle f \rangle_i}{\langle g \rangle_i} \right].$$

It is equally easy to show that $H^{G,c}(f|g) \leq H^{G,c}(f^{cg}|g^{cg})$ for all densities f and g , and nontrivial partitions A_i of the phase space X .

⁴**Proof** This is quite straightforward to prove using the integrated form of the Gibbs' inequality. First, by use of the indicator function $1_{A_i}(x)$ we may write

$$\begin{aligned} H^G(f) &\equiv - \int_X f(x) \log f(x) dx \\ &= - \sum_i \int_X f(x) 1_{A_i}(x) \log f(x) dx \\ &\leq - \sum_i \int_X f(x) 1_{A_i}(x) \log g(x) dx, \end{aligned} \tag{47}$$

for $g(x)$ an integrable density. Pick $g(x) = \langle f \rangle_i$ so

$$\begin{aligned} H^G(f) &\leq - \sum_i \log \langle f \rangle_i \int_X f(x) 1_{A_i}(x) dx \\ &= - \sum_i \log \langle f \rangle_i \int_{A_i} f(x) dx \\ &= - \sum_i \langle f \rangle_i \mu_L(A_i) \log \langle f \rangle_i \\ &= H(f^{cg}), \end{aligned} \tag{48}$$

and the assertion is proved.

Therefore, in general, *coarse graining of the phase space*, and the consequent coarse graining of a density, *will either increase the Gibbs' entropy or leave it equal to its value before coarse graining*.

Now assume that an initial density f evolves under the action of a Markov operator to give the sequence $\{P^t f\}$. In analogy with (45), the coarse grained $P^t f$ is given by

$$(P^t f(x))^{cg} = \sum_i \langle P^t f \rangle_i 1_{A_i}(x)$$

where

$$\langle P^t f \rangle_i = \frac{1}{\mu_L(A_i)} \int_{A_i} P^t f(x) dx.$$

It is important to realize that we are assuming that the Markov operator operates without any error on the density f , and that the coarse graining arises because of our inability to precisely measure dynamical variables, and consequently densities, for whatever reason.

4. Boltzmann entropy

4.1. Version 1

In any reading of Boltzmann's work, the incredibly good article by Uffink (2004) is really indispensable in trying to follow the twists and turns in Boltzmann's thinking and logic. Other sources that have been helpful include Butterfield (2000), Grandy (1988), Schrödinger (1952), and Tolman (1950).

This part that follows is taken pretty much from Jaynes (1965). If we just think of the Gibbs' entropy in the case that $f_* \equiv 1$ (invariant Lebesgue measure) then

$$H_t^G(f) = - \int_{\mathcal{X}} f(t, x) \log f(t, x) dx. \quad (49)$$

If the system consists of N particles, then $\mathcal{X} = \mathfrak{R}^{6N}$, and $dx = dx_1 \cdots dx_N$.

Now define a *reduced density*

$$\tilde{f}_i(t, x_i) = \int_{X_i} f(t, x) dx_{-i} \quad (50)$$

wherein $dx_{-i} = dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_N$. Then the Boltzmann entropy (first version) is given by

$$H_t^{B,1}(f) = -N \int_{X_i} \tilde{f}_i(t, x_i) \log \tilde{f}_i(t, x_i) dx_i. \quad (51)$$

Theorem 2. $H_t^G \leq H_t^{B,1}$ except when particles do not interact (then $H_t^G \equiv H_t^{B,1}$).⁵

Now Boltzmann originally developed an evolution equation for his version of entropy defined in Equations 51, and this equation is now known as the Boltzmann equation. Using this equations he proved that

$$\frac{dH_t^{B,1}}{dt} \geq 0 \quad (61)$$

and he claimed that this was a dynamical version of the Second Law of thermodynamics.⁶

⁵**Proof.**

$$H_t^{B,1}(f) \equiv -N \int_{X_i} \tilde{f}_i(t, x_i) \log \tilde{f}_i(t, x_i) dx_i \quad (52)$$

$$= -N \int_X f(t, x) \log \tilde{f}_i(t, x_i) dx \quad (53)$$

$$= - \int_X f(t, x) \log \left[\prod_{i=1}^N \tilde{f}_i(t, x_i) \right] dx \quad (54)$$

From the integrated Gibb's inequality

$$H_t^G(f) \equiv - \int_X f(t, x) \log f(t, x) dx \quad (55)$$

$$\leq - \int_X f(t, x) \log g(t, x) dx \quad \forall f, g \in D \quad (56)$$

$$= - \int_X f(t, x) \log \left[\prod_{i=1}^N \tilde{f}_i(t, x_i) \right] dx \quad (57)$$

$$= H_t^{B,1}(f) \quad \text{QED.} \quad (58)$$

Thus

$$H_t^G - H_t^{B,1} = - \int_X f \log f dx + \int_X f \log \prod_{i=1}^N \tilde{f}_i dx \quad (59)$$

$$= - \int_X f \log \left[\frac{f}{\prod_{i=1}^N \tilde{f}_i} \right] dx \quad (60)$$

⁶Additionally, the Boltzmann equation has formed the basis of highly successful treatments of a variety of transport phenomena in gases and plasmas, and extensions of the

However, this was quite a surprising result in light of the constancy of the entropy for the specific example of dynamics described by the time invertible system of ordinary differential equations. The most general conclusion is that *noninvertibility in system dynamics*, as reflected in an evolution of densities via a noninvertible Markov operator, *is necessary for the entropy to increase* as the system evolves (Mackey, 2001). We cannot, however, assert that noninvertibility is sufficient to guarantee this, and indeed it is not the case.

Based on much more specific assumptions, this result concerning the necessity of noninvertibility seems to have been known to Clausius (1879). How, then, did Boltzmann arrive at his conclusion that the entropy would increase to a maximum in a collection of particles moving under the action of (invertible) Hamiltonian dynamics? He managed (probably quite unwittingly) to circumvent this clear problem [the use of invertible (Hamiltonian) dynamics] by his *Stosszahlansatz* (molecular chaos) *postulate* in the derivation of the Boltzmann equation. This reduces, quite simply, to a *postulate of noninvertibility*⁷.

Following the objections of Loschmidt (1876) to Boltzmann’s attempt to justify thermodynamics using (invertible) classical mechanics, a recurrence result was used by Zermelo (1896) as the basis for an attack on Boltzmann’s celebrated “H theorem” concerning the behaviour of the entropy. In what has become known as the *Wiederkehrwand* (objection based on recurrence). Zermelo argued that, because of recurrence, almost all points would constantly revisit the same areas of phase space and thus it would be impossible for the entropy to ever monotonically increase to its maximum.

4.2. Version 2

To understand how Boltzmann thought of his second version of entropy, think of dividing up the μ space into little “cells”, or cubes A_l ($2s$ dimensional), each of volume ω . (For concreteness you could think of each cube as corresponding to a given energy range.) Then the corresponding Γ space for our N microsystems or particles consists of a large $2sN$ dimensional array of these cubes, each with volume or measure ω^N . Label the cells A_l in μ

Boltzmann transport equation have been successfully used in treating transport in fluids. However, this will not be of further concern to us here.

⁷This non-invertibility comes, I believe, because the *Stosszahlansatz* (which assumes a lack of correlation between colliding particles, has to arise because of some mechanism destroying that correlation—which is the way in which non-invertibility is brought in.

space sequentially from $l = 1$ to $l = P$. Call the number of the N systems (particles) that are in each of the cells (no matter which of the μ spaces they are in) n_1, n_2 , up to n_P (clearly, $\sum_{l=1}^P n_l = N$). If we have a state specified by

$$\{n_l\} \equiv \{n_1, n_2, \dots, n_P\} \quad (62)$$

then we will call this a **macrostate**. Boltzmann's insight was to note that there a shit load of ways that you can arrange these microsystems between the cells from $l = 1$ to $l = P$ and still have the same macrostate. Specifically, there are

$$\Omega(\{n_l\}) = \frac{N!}{n_1! \cdots n_P!} \quad (63)$$

ways of doing so. Also, the “thermodynamic probability” (which is no probability) is defined by

$$\mathcal{W} = \Omega(\{n_l\})\omega^N. \quad (64)$$

(I really don't understand why this is done. See Lavis (2004) (spin echo paper and other references above to try to sort out. Also Sears book.) Following this, Boltzmann then *defines* the second version of the Boltzmann entropy by

$$\begin{aligned} H^{B,2} &\equiv \ln \mathcal{W} \\ &= \ln[\Omega(\{n_l\})\omega^N] \\ &= \ln \Omega(\{n_l\}) + N \ln \omega \\ &= \ln N! - \ln[n_1! \cdots n_P!] + N \ln \omega \\ &= \ln N! - \sum_{l=1}^P \ln n_l! + N \ln \omega. \end{aligned} \quad (65)$$

Now Stirling's formula gives us

$$n! \simeq n^n e^{-n} \sqrt{2\pi n}, \quad (66)$$

so

$$\ln n! \simeq n \ln n - n + \frac{1}{2} \ln(2\pi n) \quad (67)$$

and thus we have

$$H^{B,2} \simeq -N \sum_{l=1}^P \frac{n_l}{N} \ln \left(\frac{n_l}{N} \right) + N \ln \omega \quad (68)$$

when N is large *and when all of the n_l are also large* which means, I think, that the n_l are pretty uniformly spread out!

Now what is the connection between this version of the Boltzmann entropy and the Gibbs' entropy? First of note that if all of the microsystems are the same and not interacting with each other, then the density factors in the following way:

$$f(x_1, \dots, x_N, t) = \prod_{l=1}^P f_l(x_l, t) = \prod_{l=1}^P f(x_l, t) \quad (69)$$

Then the Gibbs' entropy just takes the form

$$H_t^G(f) = -N \int_x f(t, x) \log f(t, x) dx. \quad (70)$$

Now the notation above is really screwed up, but the bottom line (which has to be cleaned up later) is that this version of the Boltzmann entropy is, once again going to converge to a coarse grained Gibbs' entropy in the limit of large N *when the distribution of microsystems is close to uniform*. **I have to consult with the Lavis (2004) spin-echo paper, pages 2 through 4, as well as Mackey (1992) to sort out my notation and the argument.**

5. Equilibrium entropies: Gibbs versus Boltzmann

5.1. *My take on the situation: From Oxford, 2002*

5.1.1. *Formulation*

We consider an “ensemble” of M simple systems, and each of these M simple systems consists of N particles. Then, for the j^{th} simple system. the activity at any time is described by the vector

$$x^j = (x_1^j, \dots, x_N^j), \quad j = 1, \dots, M. \quad (71)$$

Suppose, for the sake of argument, that the state space of each of the N elements of the simple system is X and is finite. In the terminology of Boltzmann X is μ space. Then the state space of one of the simple systems in our ensemble is just given by

$$\mathcal{X} = X^N, \quad (72)$$

which, as for Gibbs, would be our Γ space.

5.1.2. *A Histogram of State Variables in One Simple System*

Form a partition $\{A_l\}_{l=1}^P$ on X (remember that this is μ space) such that

$$\cup_{l=1}^P A_l = X \quad A_l \cap A_{l'} = \emptyset, \quad l \neq l'. \quad (73)$$

Now note that the fraction of the N particles of the j^{th} simple system in A_l is given by

$$F_{P,l}^j = \frac{1}{N} \sum_{i=1}^N 1_{A_l}(x_i^j). \quad (74)$$

As a check notice that

$$\begin{aligned} \sum_{l=1}^P F_{P,l}^j &= \sum_{l=1}^P \frac{1}{N} \sum_{i=1}^N 1_{A_l}(x_i^j) \\ &= \frac{1}{N} \sum_{i=1}^N \left(\sum_{l=1}^P 1_{A_l}(x_i^j) \right) \\ &= \frac{1}{N} \sum_{i=1}^N (1) \\ &= \frac{1}{N} \times N \\ &= 1. \end{aligned} \quad (75)$$

Using this idea and notation, we could write a histogram approximation to the “collapsed density” of the j^{th} simple system as

$$\mathcal{F}_P^j(x) = \sum_{l=1}^P F_{P,l}^j \frac{1_{A_l}(x)}{\mu(A_l)} \quad (76)$$

$$= \sum_{l=1}^P \left\{ \frac{1}{N} \sum_{i=1}^N 1_{A_l}(x_i^j) \right\} \frac{1_{A_l}(x)}{\mu(A_l)}, \quad x \in X, \quad (77)$$

where $\mu(A_l)$ is the measure of A_l . Again, as a check, note that

$$\int_X \mathcal{F}_P^j(x) dx = \sum_{l=1}^P F_{P,l}^j \frac{1}{\mu(A_l)} \int_X 1_{A_l}(x) dx \quad (78)$$

$$= \sum_{l=1}^P \left\{ \frac{1}{N} \sum_{i=1}^N 1_{A_l}(x_i^j) \right\} \frac{1}{\mu(A_l)} \mu(A_l) = 1. \quad (79)$$

By Bernoulli's theorem⁸, we know that $F_{P,l}^j$ is a random variable and $\lim_{N \rightarrow \infty} F_{P,l}^j$ converges in mean to the probability $\tilde{\xi}_l$ of being in A_l , or

$$\lim_{N \rightarrow \infty} F_{P,l}^j = \tilde{\xi}_l. \quad (80)$$

Thus we have for large N :

$$\lim_{N \rightarrow \infty} \mathcal{F}_P^j(x) = \sum_{l=1}^P \tilde{\xi}_l \frac{1_{A_l}(x)}{\mu(A_l)}, \quad x \in X. \quad (81)$$

5.1.3. Histogram of State Variables in an Ensemble of Simple Systems

Now let's consider M copies of this single system, $j = 1, \dots, M$, and we assume that each of these copies is independent of all others. Form a second partition $\{B_q\}_{q=1}^Q$ on \mathcal{X} (remember that this is Γ space) such that

$$\cup_{q=1}^Q B_q = \mathcal{X} = X^N \quad B_q \cap B_{q'} = \emptyset, \quad q \neq q'. \quad (82)$$

Now note that the fraction of the M copies of the single system that are in B_q is given by

$$F_{Q,q} = \frac{1}{M} \sum_{j=1}^M 1_{B_q}(x^j). \quad (83)$$

As a check notice that

$$\sum_{q=1}^Q F_{Q,q} = \sum_{q=1}^Q \frac{1}{M} \sum_{j=1}^M 1_{B_q}(x^j) = \frac{1}{M} \sum_{i=1}^M \left(\sum_{q=1}^Q 1_{B_q}(x^i) \right) = \frac{1}{M} \sum_{i=1}^M (1) = \frac{1}{M} \times M = 1. \quad (84)$$

I will use the argument $y \in \mathcal{X} = X^N$ to avoid confusion, and construct a histogram approximation to the coarse grained density of the M copies of the original system as:

$$\mathcal{F}_Q^j(y) = \sum_{q=1}^Q F_{Q,q}^j \frac{1_{B_q}(y)}{\mu(B_q)} \quad (85)$$

$$= \sum_{q=1}^Q \left\{ \frac{1}{M} \sum_{i=1}^M 1_{B_q}(x^i) \right\} \frac{1_{B_q}(y)}{\mu(B_q)}, \quad y \in \mathcal{X} \equiv X^N, \quad (86)$$

⁸The relative frequency $h[E] = n_E/n$ of realizing the event E in n independent trials is a random variable which converges to $P[E]$ in mean, and thus also in probability, as $n \rightarrow \infty$.

where $\mu(B_q)$ is the Lebesgue measure of B_q . Again, as a check, note that

$$\int_{\mathcal{X}} \mathcal{F}_Q^j(y) dy = \sum_{q=1}^Q F_{Q,q}^j \frac{1}{\mu(B_q)} \int_{\mathcal{X}} 1_{B_q}(y) dy \quad (87)$$

$$= \sum_{q=1}^Q \left\{ \frac{1}{M} \sum_{i=1}^M 1_{B_q}(x^i) \right\} \frac{1}{\mu(B_q)} \mu(B_q) = 1. \quad (88)$$

By Bernoulli's theorem, we know that $F_{Q,q}^j$ is a random variable and $\lim_{M \rightarrow \infty} F_{Q,q}^j$ converges in mean to the probability $\hat{\xi}_q$ of being in B_q , or

$$\lim_{M \rightarrow \infty} F_{Q,q}^j = \hat{\xi}_q. \quad (89)$$

Thus we have for large M (the ensemble limit):

$$\lim_{M \rightarrow \infty} \mathcal{F}_Q^j(y) = \sum_{q=1}^Q \hat{\xi}_q \frac{1_{B_q}(y)}{\mu(B_q)}, \quad y \in \mathcal{X} \equiv X^N. \quad (90)$$

5.1.4. "Collapsed" Density from an Ensemble Density

Suppose we know that the **ensemble** density for our system is given by $\hat{f}(x_1, \dots, x_N)$. Then we could define a **collapsed density** by

$$\hat{f}_{\text{collapsed}}(x) = \int_X \cdots \int_X \hat{f}(x_1, \dots, x_N) \prod_{i=1}^N \delta(x_i - x) dx_i \quad (91)$$

as in Mackey and Milton and Kaneko. Suppose we apply this idea to our ensemble histogram (86) to give

$$\int_{\mathcal{X}} \lim_{M \rightarrow \infty} \mathcal{F}_Q^j(y) dy = \int_X \cdots \int_X \mathcal{F}_Q^j(x_1, \dots, x_N) \prod_{i=1}^N \delta(x_i - x) dx_i \quad (92)$$

$$= \sum_{q=1}^Q \hat{\xi}_q \int_X \cdots \int_X \frac{1_{B_q}(x_1, \dots, x_N)}{\mu(B_q)} \prod_{i=1}^N \delta(x_i - x) dx_i \quad (93)$$

$$= \sum_{l=1}^P \tilde{\xi}_l \frac{[\mu(A_l)]^{N-1} 1_{A_l}(x)}{[\mu(A_l)]^N} \quad (94)$$

$$= \lim_{N \rightarrow \infty} \mathcal{F}_P^j \quad (95)$$

$$= \sum_{l=1}^P \tilde{\xi}_l \frac{1_{A_l}(x)}{\mu(A_l)} \quad (96)$$

$$= \lim_{N \rightarrow \infty} \mathcal{F}_P^j(x) \quad (97)$$

where we have taken $Q = P$ and $q = l$ so $\mu(B_q) = [\mu(A_l)]^N$ and $\hat{\xi}_q = \tilde{\xi}_l$ (since the probability of being in B_q is just the same as the probability of being in A_l after collapse).

This, it would seem, offers the justification for understanding why we can make a measurement on a single system (equivalent to looking at quantities averaged with respect to $\tilde{f}(x)$, e.g. Equation 81) but use Gibbsian ensemble techniques to make calculations (Equation 97, equivalent to looking at quantities averaged with respect to $\hat{f}(x)$).

5.1.5. Passage to Continuous Densities

Up to now we have been representing $\lim_{M \rightarrow \infty} \mathcal{F}_Q^M(x)$ and $\lim_{N \rightarrow \infty} \mathcal{F}_P^j(x)$ as simple functions. However, by the convergence of simple functions and the Lebesgue dominated convergence theorem we know that in the limit of Q or $P \rightarrow \infty$ we can recover integrable functions (densities) from these:

$$\tilde{f}(x) = \lim_{P \rightarrow \infty} \lim_{N \rightarrow \infty} \mathcal{F}_P^j(x) \quad x \in X \quad (98)$$

and

$$\hat{f}(y) = \lim_{Q \rightarrow \infty} \lim_{M \rightarrow \infty} \mathcal{F}_Q^M(y) \quad y \in \mathcal{X} \equiv X^N. \quad (99)$$

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council (NSERC, Canada).

Butterfield, J., 2000. Philosophy of thermal physics: Ht2000. Unpublished lecture notes, University of Oxford.

Clausius, R., 1879. The Mechanical Theory of Heat. Macmillan, London.

Denbigh, K., Denbigh, J., 1985. Entropy in Relation to Incomplete Knowledge. Cambridge University Press, Cambridge.

Ehrenfest, P., Ehrenfest, T., 1959. The Conceptual Foundations of the Statistical Approach in Mechanics. Cornell University Press, Ithaca.

Gibbs, J., 1962. Elementary Principles in Statistical Mechanics. Dover, New York.

Grandy, W., 1988. Foundations of Statistical Mechanics: I & II. Reidel, Dordrecht.

Jaynes, E., 1965. Gibbs vs Boltzmann entropies. Amer. J. Phys. 33, 391–398.

Khinchin, A., 1949. Mathematical Foundations of Statistical Mechanics. Dover, New York, NY.

Kuczma, M., 1985. An Introduction to the Theory of Functional Equations and Inequalities. Silesian University Press, Katowice (Poland).

Lavis, D., 2004. The spin-echo system reconsidered. Foundations of Physics 34 (4), 669–688.

Loschmidt, J., 1876. Über den Zustand des Wärmegleichgewichtes eines Systeme von Körpern mit Rücksicht auf die Schwerkraft. Wien. Ber. 73, 128–142.

Mackey, M. C., 1992. Time's Arrow: The Origins of Thermodynamic Behaviour. Springer-Verlag, Berlin, New York, Heidelberg.

- Mackey, M. C., 2001. Microscopic dynamics and the second law of thermodynamics. In: Mugnai, C., Ranfagni, A., Schulman, L. (Eds.), *Time's Arrows, Quantum Measurement and Superluminal Behavior*. Consiglio Nazionale Delle Ricerche, Roma, pp. 49–65.
- Reichl, L., 1980. *A Modern Course in Statistical Physics*. University of Texas Press, Austin.
- Schrödinger, E., 1952. *Statistical Thermodynamics*. Cambridge University Press, Cambridge.
- Skagerstam, B., 1974. On the mathematical definition of entropy. *Z. Naturforsch.* 29A, 1239–1243.
- Tolman, R., 1950. *The Principles of Statistical Mechanics*. Oxford University Press, Oxford.
- Uffink, J., 2004. SEHR CUM GRANO: Boltzmann's work in statistical physics, preprint supplied by email to M. C. Mackey.
- Walecka, J., 2000. *Fundamentals of Statistical Mechanics: Manuscript and Notes of Felix Bloch*. Imperial College Press and World Scientific, London, Singapore.
- Zermelo, E., 1896. Über einen Satz der Dynamik und der Mechanischen Wärmetheorie. *Ann. der Phys.* 57, 485–494.