

§12. P.S. More About Entropy.

A. Secheorichiu  
Stat Mech - HT13  
Oxford.

I have tried in the foregoing to take us as quickly as possible from the ~~the~~ (perhaps somewhat murky) conceptual underpinnings of the S.M. to a state of operational clarity as to how we would compute things. More things will be computed in PART IV.

Fortunately (or sadly), you do not need to really understand why maximising  $-\sum p_x \ln p_x$  works or what it really means - you can simply embrace the straightforward Gibbs prescription:

- 1) Compute  $Z$  from knowledge of  $\{d\}$  and  $\{E_d\}$
- 2) Compute  $F = -k_B T \ln Z$
- 3) Compute  $P, S, U$  from that (and  $C_V$ , usually)
- 4) Move on to the next problem (in S.M. or in life)

From this utilitarian viewpoint, my task of introducing the "fundamentals of S.M." is complete. Nevertheless, in this section, I wish to discuss the notion of entropy and the meaning of  $\{p_x\}$  a little more and also to survey some alternative schemes for setting up Stat. Mech. (they all eventually lead to the same practical prescriptions).

This is for those of you who wish to make sense of the formalism and be convinced that we are on a firm ground, intellectually (or are we?)

## 12.1 Shannon's Theorem

In 8.1, I argued ~~that~~ that, in order to achieve the "fairest" and most unbiased assignment of probabilities to microstates, one must maximize

$$S_G = - \sum_{\alpha} p_{\alpha} \ln p_{\alpha} \quad (1)$$

(Gibbs entropy, "information entropy", measure of uncertainty etc.)

I did this by presenting a reasonable <sup>practical</sup> scheme for assigning probabilities, which I asked you to agree was the fairest imaginable.

In the spirit of formalistic nit-picking, you might be tempted to ask whether the result (1) is in any sense unique - could we have invented other "fair games" leading to different definitions of entropy? Here's an argument that addresses this.

~~Given~~ Faced with some set of probabilities  $\{p_{\alpha}\}$ , let us seek to define a function

$$H(p_1, p_2, \dots, p_r)$$

that will characterize the uncertainty associated with this set of probabilities.

In order to be a suitable such measure,  $H$  must satisfy certain basic properties:

- 1)  $H$  should be a continuous function of  $p_\alpha$ 's  
(changing  $p_\alpha$ 's a little should not dramatically change the measure of uncertainty associated with them)
- 2)  $H$  should be symmetric wrt permutations of  $\{p_\alpha\}$   
(it should not matter in what order we list the microstates)
- 3)  $\forall \{p_1 \dots p_\Omega\}$ ,  $H(p_1, \dots, p_\Omega) < H(\frac{1}{\Omega}, \dots, \frac{1}{\Omega}) \equiv H_\Omega$  (2)  
(situation with equal probabilities corresponds to maximum uncertainty)
- 4) If  $\Omega' > \Omega$ ,  $H_{\Omega'} > H_\Omega$  (3)  
(more <sup>equiprobable</sup> microstates  $\Rightarrow$  more uncertainty)
- 5)  $H$  should be additive and independent of how we count the microstates, in the foll. sense.  
If the choice of a microstate is broken down into two successive choices, <sup>(first a group, then the state)</sup> the total  $H$  should be a weighted sum of individual values of  $H$  associated with each subgroup:

Split the microstates into groups:

$$\underbrace{\alpha = 1, \dots, m_1}_{\text{Group } i=1} \quad \underbrace{m_1+1, \dots, m_1+m_2, \dots}_{\text{Group } i=2} \quad \dots \quad \underbrace{\dots}_{\text{Group } i=M} \quad \sum_{i=1}^M m_i = \Omega$$

$i = 1 \quad 2 \quad \dots \quad M$   
 $w_1 \quad w_2 \quad \dots \quad w_M$  probabilities of these groups.

Within each subgroup, let us assign (conditional) probabilities to microstates:

$$P_{\alpha}^{(i)} = \text{probability to pick microstate } \alpha \text{ within subgroup } (i) = \frac{P_{\alpha}}{W_i}$$

$\leftarrow$  unconditional probability of  $\alpha$   
 $\uparrow$  probability to pick group  $i$

So we want  $H$  to satisfy

$$H(p_1, \dots, p_{\Omega}) = H(w_1, \dots, w_M) + w_1 H(p_1^{(1)} \dots p_{m_1}^{(1)}) + w_2 H(p_{m_1+1}^{(2)} \dots p_{m_1+m_2}^{(2)}) + \dots = \quad (2)$$

$\uparrow$  total uncertainty  
 $\uparrow$  uncertainty in the distribution of groups  
 $\leftarrow$  probability of a particular group  
 $\leftarrow$  uncertainty within a group

$$= H(w_1, \dots, w_M) + w_1 H\left(\frac{p_1}{w_1}, \dots, \frac{p_{m_1}}{w_1}\right) + w_2 H\left(\frac{p_{m_1+1}}{w_2}, \dots, \frac{p_{m_1+m_2}}{w_2}\right) + \dots$$

Theorem. The only  $H$  satisfying these assumptions is

$$H = -k \sum_{\alpha} p_{\alpha} \ln p_{\alpha},$$

where  $k > 0$  is a constant

Proof. Consider a special case: all  $p_{\alpha} = \frac{1}{\Omega}$ .

Then  $w_i = \frac{m_i}{\Omega}$  and  $p_{\alpha}^{(i)} = \frac{p_{\alpha}}{w_i} = \frac{1}{m_i}$

Then from (4),

$$\begin{aligned}
 H(p_1 \dots p_n) &= H\left(\frac{1}{\Omega}, \dots, \frac{1}{\Omega}\right) = H_{\Omega} = \\
 &= H(w_1, \dots, w_M) + \sum_{i=1}^M w_i \underbrace{H\left(\frac{1}{m_i}, \dots, \frac{1}{m_i}\right)}_{H_{m_i}}
 \end{aligned}$$

So we have

$$H(w_1, \dots, w_M) = H_{\Omega} - \sum_{i=1}^M w_i H_{m_i} \quad (5)$$

Now consider a special case of this formula for the situation when all  $m_i = m$  are the same.

Then  $\Omega = mM$  and  $w_i = \frac{m}{\Omega} = \frac{1}{M}$ .

From (5),

$$H_M = H_{mM} - H_m \quad (6)$$

This is a functional equation: if  $H_m = f(m)$

then  $f(mn) = f(m) + f(n) \quad (7)$

Lemma. The only monotonically increasing function that satisfies (7) is  $f(m) = k \ln m$

[proof below]

If this is true, then

$$\boxed{H_{\Omega} = k \ln \Omega} \quad (8)$$

("Boltzmann entropy", see below)

must be because  $H_{\Omega_1} > H_{\Omega_2}$  for  $\Omega_1 > \Omega_2$  condition (3)

$\uparrow$  positive const.



Proof of the Lemma (p. 115).

$$f(mn) = f(m) + f(n) \\ f \text{ monotonically increasing} \quad \Rightarrow \quad f(m) = k \ln m$$

For  $\forall$  integers  $m, n > 1$ , we can always find integers  $r$  and (arb. large)  $s$  such that

$$\frac{r}{s} < \frac{\ln m}{\ln n} < \frac{r+1}{s} \quad \Rightarrow \quad n^r < m^s < n^{r+1} \quad (11)$$

As  $f$  is monotonically increasing,

$$f(n^r) < f(m^s) < f(n^{r+1})$$

But  $f(mn) = f(m) + f(n) \Rightarrow f(n^r) = r f(n)$  etc., so

$$r f(n) < s f(m) < (r+1) f(n)$$

$$\text{So, } \frac{r}{s} < \frac{f(m)}{f(n)} < \frac{r+1}{s} \quad (12)$$

$$\text{From (11) \& (12), } \left| \frac{f(m)}{f(n)} - \frac{\ln m}{\ln n} \right| < \frac{1}{s}$$

$$\text{or } \left| \frac{f(m)}{\ln m} - \frac{f(n)}{\ln n} \right| < \frac{1}{s} \frac{f(n)}{\ln m} \rightarrow 0$$

$$\text{So } \frac{f(m)}{\ln(m)} = \frac{f(n)}{\ln(n)} = \text{const} = k \quad \text{q.e.d.} \\ \text{because } s \text{ arb. large.}$$

Of course, all this is obviated by the fact that eq. (1) "works" but it's nice to know its generality can be proven in this way.

Read Shannon's paper on connection to info. theory.  
L. 5 ended here.

## 12.2 Boltzmann Entropy & the Canonical Ensemble

Ludwig Boltzmann's tombstone has a famous formula carved into it:

$$S = k \log W \quad (12)$$

— "Boltzmann entropy", where  $k$  is a constant (technically speaking arbitrary  $k > 0$ , traditionally  $k_B$ , introduced, btw, by Planck, not Boltzmann) and  $W$  is "the number of complexions".

What does this mean and how does it relate to what we have discussed so far?

A "complexion" is the same thing that I have so far referred to as a "state" ( $\alpha$ ). Boltzmann's formula introduces entropy for a system where all states are equiprobable, so it is actually the same expression as we ~~found~~ found for the case of all  $p_\alpha = \text{each other} = \frac{1}{\Omega}$  and  $W = \Omega$ : eq. (12) is the same as

$$S = k_B \ln \Omega \quad (13) \quad \left[ \begin{array}{l} \text{see eq. (7) p. 74} \\ \text{eq. (19) p. 77} \\ \text{eq. (8) p. 115} \end{array} \right]$$

~~It was introduced (by Boltzmann) on similar grounds as the Shannon-Gibbs entropy in 12.1:~~

It was introduced (by Boltzmann) on similar grounds as the Shannon-Gibbs entropy in 12.1:



- 1) It must be larger if  $\Omega$  is larger
- 2) It must be additive for several systems when they are put together:

$$\text{Since } \Omega_{12} = \Omega_1 \Omega_2, \quad S_{12} = S_1 + S_2 \quad (14)$$

As such, Boltzmann's entropy simply appears to be a particular case of the Shannon-Gibbs entropy for isolated systems (where all states are equiprobable).

In fact, it is possible <sup>to</sup> turn the argument around and get the Gibbs entropy from the Boltzmann entropy. Actually, I pretty much did that in 8.1 (pp 71-73): I assigned  $N$  "quanta of probability" to  $\Omega$  microstates in a "fair and balanced" fashion and found that the # of ways in which any particular set of probabilities  $\{p_\alpha = \frac{N_\alpha}{N}\}$  could be obtained was [eq (4) p. 72]

$$W = \frac{N!}{N_1! \dots N_\Omega!} \quad (15)$$

Entropy then was simply ~~the Boltzmann entropy~~

$$S = \frac{k_B \ln W}{N} = -k_B \sum_\alpha p_\alpha \ln p_\alpha \quad \left( \text{all } N_\alpha \rightarrow \infty, N \rightarrow \infty, \right. \\ \left. N_\alpha/N \rightarrow p_\alpha \right) \quad (16)$$

[eq. (6) p. 73]

This was justified as a counting scheme: larger  $W$  gave the more probable assignment of  $\{p_\alpha\}$  and it was convenient to take log to make  $S$  additive.

At that stage, I presented this scheme simply as a "reasonable" procedure for assigning probabilities based on available knowledge. In 12.1, I removed the need for this by showing that the Gibbs ~~formula~~<sup>expression</sup> for  $S$  was in some sense the uniquely suitable choice.

This information/uncertainty-based view is in fact not how the Gibbs construction has traditionally been thought about (e.g. by Gibbs).

- Rather, one makes  $N$  mental copies of the system one is interested in, ~~and~~<sup>and</sup> joins them together into one isolated über-system <sup>("ensemble")</sup>. The states ("complexions") of this über-system are characterized by

$$\begin{array}{cccccccccccc}
 N_1 & & & & & & & & & & & & & 1 \\
 \vdots & & & & & & & & & & & & & \\
 N_\Omega & & " & & " & & " & & " & & " & & " & & " & & " & & " & & \Omega
 \end{array}$$

and so the # of these über-states is  $W$ , given by (15).

- Since the über-system is isolated, all these states are equiprobable and the entropy is Boltzmann entropy,  $S_N = k_B \ln W$ , which, if maximised, will give the most probable über-state (maximising entropy per system,  $S = S_N/N$  as in eq. (16), is equivalent) — this is the ~~the~~ equilibrium state of the über-system.

- If  $\{N_1 \dots N_\alpha\}$  is the most probable ~~state~~ über-state, then  $p_\alpha = \frac{N_\alpha}{N}$  are the <sup>desired</sup> probabilities of the microstates any of the copies of the original system might be in if picked from the über-system (ensemble).

- To complete this construction, one proves that fluctuations around the most probable über-state vanish as  $N \rightarrow \infty$  [see, e.g., Schrödinger Ch. V, VI]  
 $\uparrow$  always OK because  $N$  is arbitrary

- Recall that to get the canonical (Gibbs) distribution, we maximized  $\frac{S_N}{N} = -k_B \sum_\alpha p_\alpha \ln p_\alpha$

Subject to fixed mean energy

$$\sum_\alpha p_\alpha E_\alpha = U \quad (\Rightarrow p_\alpha = \frac{e^{-\beta E_\alpha}}{Z} \text{ etc.})$$

Since  $p_\alpha = N_\alpha/N$ , this is the same as

$$\sum_\alpha N_\alpha E_\alpha = NU = \mathcal{E} \quad (17)$$

So the (isolated) über-system has the exact total energy  $\mathcal{E} = NU$ .

} Thus, seeking the equilibrium of a system at fixed mean energy  $U$  (@ fixed temperature) is the same as seeking the most likely way exact energy  $NU$  would distribute itself between very many ( $N$ ) copies of the system, isolated from the world.

- Clearly, this describes a system in contact with a heat bath (see 10.3) — in this case the heat bath is made of  $N-1$  copies of the system itself, but that does not matter because the nature of the heat bath does not matter, ~~and~~ only the value of the temperature that it sets for the system!

### 12.3 Microcanonical Ensemble

This is an opportune moment to mention yet another way in which S.M. and thermodynamics can be constructed (and is, in the majority of textbooks). Effectively, this is an attempt to impart a veneer of "objective reality" to the foundations of the subject, which, in the way I presented them, were perhaps uncomfortably dependent on such subjective notions as the observer's information about the system (or the "existence" of  $N$  mental copies of it). So, under this scheme, we start by considering a completely isolated system and postulate that all its microstates are equiprobable.

Since the system is isolated, its energy is exactly conserved, so those microstates are in fact not all possible ones, but only those with

$$E_{\alpha} = E \quad (18)$$

$$\text{So, } \left\{ \begin{array}{l} p_\alpha = \frac{1}{\Omega(E)} \quad \text{if } E_\alpha = E \\ p_\alpha = 0 \quad \text{if } E_\alpha \neq E \end{array} \right. \quad (19)$$

$\uparrow$  # of microstates with  $E_\alpha = E$

Microcanonical distribution

[Note: It is possible to show that in a steady state, so in equilibrium, the distribution can only be a function of conserved quantities; since it is usually possible to find a frame where the system is at rest, only  $E$  matters.]

The entropy of this distribution is

$$S = k_B \ln \Omega(E). \quad (20)$$

To get the canonical (Gibbs) distribution from (19), pick a small part of the system ~~with energy  $E$~~  and ask what is the probability for it to have energy  $\epsilon \ll E$ ? # of microstates of the rest ("reservoir", heat bath) with energy  $E - \epsilon$

$$p(\epsilon) = \frac{\Omega_{\text{part}}(\epsilon) \Omega_R(E - \epsilon)}{\Omega(E)} = \frac{\Omega_{\text{part}}(\epsilon)}{\Omega(E)} e^{\frac{1}{k_B} S_R(E - \epsilon)} \approx$$

$\equiv \frac{1}{T}$  temperature of the reservoir

# of microstates of the part that have energy  $\epsilon$

$$\approx \frac{\Omega_{\text{part}}(\epsilon)}{\Omega(E)} e^{\frac{1}{k_B} \left[ S_R(E) - \epsilon \left( \frac{\partial S_R}{\partial E} \right) + \dots \right]}$$

$$= \frac{e^{S_R(E)/k_B}}{\Omega(E)} \Omega_{\text{part}}(\epsilon) e^{-\frac{\epsilon}{k_B T}} \quad (21)$$

$\uparrow$  constant indep. of  $\epsilon$ , fixed by normalization



$\Omega$  by generalising Boltzmann's entropy (only valid for isolated systems) along the same lines as the requirement of additivity/independence of the state-counting on p.114 (eq.(4)):

$$S = k_B \ln \Omega(E) = \underbrace{S_{\text{part}}}_{\substack{\uparrow \\ \text{entropy of} \\ \text{its small} \\ \text{part}}} + \underbrace{\langle S_R(E-\epsilon) \rangle}_{\substack{\uparrow \\ \text{avg. entropy of the} \\ \text{reservoir (overall } \epsilon \text{)}}$$

the big isolated system

Then

$$S_{\text{part}} = k_B \ln \Omega(E) - \sum_{\epsilon} p(\epsilon) \underbrace{k_B \ln \Omega_R(E-\epsilon)}_{S_R(E-\epsilon)} =$$

$$= -k_B \sum_{\epsilon} \underbrace{p(\epsilon)}_{\substack{\parallel \\ \Omega(\epsilon) p_{\alpha}}} \ln \underbrace{\frac{\Omega_R(E-\epsilon)}{\Omega(E)}}_{\substack{\parallel \leftarrow \text{see eq. (22)} \\ \frac{p(\epsilon)}{\Omega_{\text{part}}(\epsilon)} = p_{\alpha}}} = -k_B \sum_{\alpha} p_{\alpha} \ln p_{\alpha}$$

Gibbs entropy  
(24)

So, everything is consistent.

Note: The canonical ensemble of Gibbs, when thought of in terms of one über-system containing  $N$  copies of the original system (pp. 120-121) with exact total energy  $NU$  is basically a case of microcanonical distribution being applied to this imaginary assemblage.

12.4 Gibbs vs. Boltzmann / Meaning of Probabilities

Let us summarise the 3 main schemes for the construction of statistical mechanics and thermodynamics

"BOLTZMANN"

(& a lot of other people)  
e.g. Landau, Pauli, Ehrenfest, ...

- Consider an isolated system with fixed E
- Assume equal probabilities ~~for all its states~~ for all its states



Microcanonical ensemble

- Consider a small subsystem



"GIBBS"

(& Schrödinger)

- Imagine an ensemble of  $N$  identical copies of a system
- Distribute energy  $NU$  between them



Canonical ensemble

- Maximise Gibbs entropy subject to fixed mean energy  $U$



Get Gibbs distribution



Calculate everything



Get testable results

} only this part of theory falsifiable (by experiment)

"SHANNON"

(or Jaynes)

- Admit nearly total ignorance
- Seek to make a statistical inference about likelihoods of states subject to a few scraps of knowledge and no other bias



- Maximise Shannon (=Gibbs) entropy subject to constraints imposed by knowledge, e.g.  $U$





The "Gibbs" and "Shannon" schemes really are versions of one another, just the language is a bit different - and one might even argue that the "Shannon" construction is what Gibbs really had in mind. So I will refer to this entire strand as Gibbsian (perhaps "Gibbsian heresy")

"Boltzmann" ("Boltzmannite orthodoxy") is different: we are invited to think of every step in the construction as describing some form of objective reality, while under the Gibbs approach, we are effectively just trying to come up with the best possible guess, given limited information.

The reality of the Boltzmannite construction is somewhat illusory:

1) An isolated system with a fixed energy is a fiction:

- it is practically impossible to set up
- if set up, it is inaccessible to measurement (because it is isolated!)

so it is in fact as imaginary as, say, Gibbs's ensemble.

2) What is the basis for assuming equal probabilities?

The usual view within this strand of thought is as follows:

As the isolated system in question evolves in time, it samples (repeatedly) its entire phase space - i.e., it visits all possible microstates consistent with its conservation laws ( $E = \text{const}$ ). Thus, the probability for it to be in any given microstate or set of microstates is simply the fraction of time that it spends there:

$$p_\alpha = f_\alpha \text{ (frequency of visit } \alpha \text{)} \quad (25)$$

In other words, time averages of any quantities of interest are equal to the statistical averages (i.e., ~~ensemble~~ averages over <sup>all</sup> microstates)

This last statement is known as the ergodic hypothesis.

[in fact the assumption about probabilities is stronger:

$$\left. \begin{array}{l} \text{time spent in} \\ \text{a subset of} \\ \text{microstates} \\ \text{(subvolume of} \\ \text{phase space)} \end{array} \right\} = \frac{\text{\# of microstates} \\ \text{in this subset}}{\text{total \# of} \\ \text{microstates}} \quad (26)$$

So, we do all our practical calculations via statistical averages (with  $p_\alpha = 1/\Omega$  etc.), but the physical justification for that is that the system is time-averaging over them (we can't compute time averages because we cannot calculate precise dynamics).

~~Microstates~~

~~unconvincing~~ The objection to this view that I find most compelling is ~~unconvincing~~ simply that the size of the phase space (# of microstates) of any ~~macroscopic~~ macroscopic system is so enormous that it is in fact not possible for the system to visit all of it over any reasonable times [pointed out by Jaynes].

Note. The key divide here is rooted in the old argument about meaning of probabilities:

- probabilities as frequencies: "objective"

vs.

- probabilities as (a priori) likelihoods: [Laplace, Bayes, Keynes, Jeffreys, Jaynes]  
"subjective" - measuring our (lack of) knowledge.

NB: I am not, in choosing to go with the 2nd view and putting up all these objections to the 1st, suggesting that one is right and the other wrong. Remember that the falsifiable <sup>(and examinable)</sup> content of the theory is the same either way, so the issue is ~~which~~ which of the logical constructions lead up to it makes more sense to me and to you - and I urge you to explore the literature on your own and decide for yourselves whether you are a Gibbsian or a Boltzmannite (either way, you are in good company) - or, indeed, whether you wish to invent a third way!

[This is a bit like the thorny matter of interpretations of Q.M. Everyone agrees on the results but not on why it all works!] L.6 ended here

To preempt some of the confusions about the "subjective" nature of maximising uncertainty (whose uncertainty?) let me deal with the common objection that surely, if two observers have different amounts of info about the same system and so arrive at two different entropy-maximising sets of  $p_i$ 's, we would be in trouble if those different sets gave different testable predictions about the system!  
 (Heat capacity of a room filled with air can't depend on who is looking!)

= If Mr. B has more constraints (more knowledge) than Mr. A, but his additional constraints are in fact desirable from Mr. A's, they will in fact get the same distribution because ~~the~~ Mr. B's additional Lagrange multipliers will turn out to be <sup>arbitrary</sup> ~~arbitrary~~ and so can be set = 0 (easy to see if you work through some example you fancy) See PS-5

= If Mr. B's <sup>additional</sup> constraints are incompatible with Mr. A's, the method of Lagrange multipliers will produce a set of equations for  $\lambda$ 's that will have no real solutions - telling us the system of constraints is logically contradictory and so no theory exists (basically means one of them got their constraints wrong)

- Finally, if Mr. B's additional constraints are neither incompatible with nor derivable from Mr. A's, that means he ~~is~~ has discovered new physics.

In fact, if experiment failed to agree with our max. entropy predictions in a systematic fashion, we would be forced to conclude that our system of constraints is in some essential way incomplete and to search for new ones!

---

## 12.6 Second Law

So far, we have not involved the idea of true in our considerations: we were always interested in some eventual equilibrium and the way to calculate it was to maximise  $S_G = - \sum_{\alpha} p_{\alpha} \ln p_{\alpha}$  subject to constraints representing some measurable properties of this equilibrium.

This maximisation of  $S_G$  is not the same thing as the 2<sup>nd</sup> law of thermodynamics, which states, effectively, that the thermodynamic entropy  $S$  of the world (or a closed system) must not decrease in any process — and so in time.

This statement is famously replete with hard metaphysical questions (although it's quite straightforward when it comes to calculating entropy changes in mundane situations) — so it is perhaps useful to see how it emerges within the conceptual framework I am advocating for you.

The following proof is what I believe to be an acceptable vulgarisation of an argument <sup>due</sup> to Jaynes (1965).

Time t: Consider a closed system (the world) in equilibrium, subject to some set of its properties having just been measured and no other info available.

Then our best guess as to its state at this time t

is obtained by

$S_G \rightarrow \max$  with constraints obtained at  
time  $t$

$\Downarrow$

$\{p_\alpha\}$

In this state, the obtained max. value of  $S_G$  is equal to the thermodynamic entropy (see proof on p. 81):

Time  $t'$ :  $k_B S_{G, \max}(t) = S(t) \quad (27)$

Now consider the evolution of this system from time  $t$  to time  $t' > t$ , starting from the  $p_\alpha$ 's and the set of states that we inferred at time  $t$  and using Hamilton's equations (if the system is classical) or the time-dependent Schrödinger equation (if it is quantum, as it always really is). During this evolution the

~~entropy,  $S_G(t')$ , is constant~~

Gibbs entropy,  $S_G(t') = -\sum_\alpha p_\alpha \ln p_\alpha = S_G(t)$ , stays constant: indeed, the Sch. Equ evolves the states  $d$ , but if the system was in some state  $d$  at time  $t$  with probability  $p_\alpha$ , it will be in the descendant of this state at time  $t'$  with exactly the same probability (so this is like changing the labels in the expression for  $S_G$ , while  $p_\alpha$ 's stay the same — and so does  $S_G$ ).

Thus,  $S_G(t') = S_G(t) = S_{G, \max}(t) = \frac{1}{k_B} S(t) \quad (28)$

see

§13.4

Now forget all previous info and make a new set of measurements at time  $t'$  and work out a new set of  $p_i$ 's at  $t'$ , subject only to these new constraints, by maximising Gibbs entropy and from it infer the new thermodynamic (equilibrium) entropy:

$$S'(t') = k_B S_{G, \max}(t') \geq k_B S_G(t') = k_B S_G(t) = S(t) \quad (2a)$$

Thus,

$$\boxed{S(t') \geq S(t)}$$

second law.

↑  
the "real"  
 $S_G$ , evolved  
from time  $t$

The meaning of this is that the increase of  $S$  reflects our insistence to forget most of the detailed knowledge we possess as a result of evolving in time any earlier state (even if based on an earlier guess) and to reapply at every later time the rules of stat. inference based on the very little knowledge we can obtain in later measurements.

If you are sufficiently steeped in quantum ways of thinking by now, you will pounce and ask who is doing all these measurements?

If it is an external observer or ~~scientific~~ apparatus, then the system isn't really closed and, in particular, the measurement at the later time  $t'$  will potentially destroy ~~all~~ the identification of all those



microstates with their progenitors at time  $t$ , so  $S_G(t') \neq S_G(t)$ . [in a classical system, this would not be a problem because you can make measurements w/o altering the system, but in QM you cannot!]

Moreover, what if your measurements at  $t'$  are so much better than at the technologically less advanced time  $t$ ? ~~surely~~ You might imagine an extreme case in which you determine the state of the system at  $t'$  precisely and so  $S_G(t') = 0$ !

- Clearly, the observer is in fact not external, but is inside the system
- And as he/she/it does their measurement, ~~then~~ not just the entropy of the object of measurement (which is part of the system) but also of the observer and his apparatus changes.

The argument above implies that a very precise ~~entropy-decrease~~ measurement must massively increase the entropy of the observer and the observer's kit to compensate and ensure  $S(t') \geq S(t)$ .

~~Entropy-decrease~~

[Note that this was the basis of Szilard's (1929) exorcism of Maxwell's demon]