A Simple Derivation of the Landauer Bound

Wayne C. Myrvold
Department of Philosophy
The University of Western Ontario

August 8, 2024

Contents

1 Introduction

Consider a typical case of relaxation of a system, initially out of thermal equilibrium, to a state of thermal equilibrium. It might, for instance, involve two objects, initially at different temperatures, placed in thermal contact, which, after a while, come to have the same temperature. I'd like to highlight two things about processes of this sort. First, they represent a lost opportunity to obtain work from the system; one could have used the two objects as reservoirs for a heat engine, extracting heat from the warmer one, converting some to work and discarding the rest into the cooler one, until their temperatures are equal. The flip side of this coin is: once the temperatures have equalized, one has to do work or expend some resource that could be used to do work, in order to restore the initial state. We call processes like this *dissipatory*. Spontaneous equilibration involves dissipation.

Second, they involve "forgetting" of the past (at least at the macroscopic level); distinct initial states can lead to the same final state. Once two objects, initially at different temperatures, have reached thermal equilibrium with each other, you won't be able to tell, *from examination of those objects*, which was initially the hotter and which, the cooler (though, of course, there might be a record of the initial states somewhere else).

Equilibration involves forgetting of the past, and it involves dissipation. The question addressed by the literature on what has come to be called *Landauer's principle* is: does *every* process of forgetting (that is, loss of distinguishability of states) likewise involve dissipation?

The literature on Landauer's principle is, as John Norton in particular has persuasively argued, a bit of a mess (see Norton 2005, 2011, 2013). Nonetheless, there is something right in what is said, or, at least, something in the neighbourhood of what is said is right. In my *BJPS* paper (Myrvold, 2024) I present a proof of a bound on dissipation associated with processes that take distinguishable initial states to the same state, with care to lay out exactly what is and isn't being assumed. The proof, in its full generality, applies to situations in which there is less than complete distinguishability of the initial states, and associates dissipation with any loss of distinguishability. It applies to any number of input states, and to situations involving arbitrary numbers of heat reservoirs. In what follows, after some preliminaries, I present in §4 a proof of a special case, which sacrifices generality for ease of reading. Though less general than the version given in the paper, the conceptual underpinnings are the same. The proof goes through in both classical

and quantum mechanics; for the sake of brevity, we'll treat of the classical case. The quantum version goes through almost word-for-word, with mixed states replacing probability distributions on classical phase space.

2   Illustration: An ideal gas

Although the scope of the proof will by no means be restricted to any particular sort of example, it can be helpful to keep an example in mind, as an illustration of what is to be proven. We consider an ideal gas consisting of N molecules, confined to a cylindrical chamber, in thermal contact with a heat bath at temperature T. There's a partition that can be inserted, dividing the chamber into two parts, or removed. We assume that the partition can be inserted at any distance along the length of the chamber, so that the two parts into which it divides the chamber need not be of the same size. In addition, the partition can be used as a piston to expand or compress the gas in one sub-chamber or the other. We also assume that, if the gas is confined to some region, we have the means to slide it over into any other region of the same volume, without doing any work (think, if you will, of a pair of moveable pistons).

Let L be a state in which the partition divides the container into two equal parts, left and right, and the gas is in the left half, with temperature T. Let R be the same, except the gas is in the right half. And let F (for *full*) be a state in which the partition has been removed, and the gas fills the chamber, again at temperature T.

If, starting from either L or R, we slowly expand the gas until it occupies a fraction p (greater than $1/2$) of the container, some quantity Q of heat will pass from the heat reservoir to the gas. If it is possible to compress the gas back to its original volume, while transferring the same quantity Q of heat from the gas into the reservoir, we will say that the original expansion is *thermodynamically reversible*. Otherwise, the process is *dissipatory*.[1]

Given the set-up as we have described it, one process that is dissipatory

---

[1]In this section, we entertain the fiction that uncertainties about the result of a process can be disregarded, even at the molecular level, and apply standard thermodynamic calculations to calculate heat flows in and out of a gas, even one consisting of a small number of molecules. This fiction will be removed in §4, in which probabilities over the state of the system will be employed, and we will couch our discussion in terms of *expectation values* of heat exchanges.

is the process that starts with the gas in state L, and proceeds by removing the partition, and allowing the gas to diffuse into the newly available volume. Initial and final states of this process are at the same temperature. Since it's an ideal gas, total internal energy is a function only of temperature, and so the internal energy is unchanged, and there's no net transfer of heat between the gas and the reservoir.

This is a dissipatory process: no heat passes from the reservoir into the gas in the course of the process, but it is not possible to restore the initial state without doing work and passing heat into the reservoir. We want a quantitative measure of the amount of dissipation. As long as all heat transfers in and out of the system involve reservoirs at the same temperature T, we can take, as a measure of the amount of dissipation associated with a given procedure M that takes a system from a state a to a state b, the difference between the minimum amount of heat transferred out of the system, among all processes that restore the initial state, and the quantity of heat that enters the system in the course of the process.[2] Call this quantity $D_M^T(a \rightarrow b)$.

If the initial state is one that can be restored from the final state with no heat exchange whatsoever between the system and its environment (which is true of the cases often considered in the literature on Landauer's bound), this is conceptually simpler. In such a case, the dissipation associated with a process is just the negative of the heat that enters the system in the course of the process, or, to say the same thing, it's the quantity of heat that is transferred *out* of the system in the course of the process. For that reason, dissipation is often equated, in the Landauer literature, with heat transferred to the environment.

So, for example, suppose that the initial state is either L or R, and the final state is required to be L. For each of these initial states, it's possible to produce the state L with no heat exchange between the system and the environment. If the initial state is L, do nothing. If the initial state is R, then slide the gas over, at constant volume, into the left side of the container. These are processes without dissipation. Note that different manipulations are required to produce L without dissipation, depending on whether we start with L or R.

Suppose, now, we want a single manipulation that yields a final state L from either initial state, L or R. Here's one way to do it. From initial state

---

[2]If the process M involves a net transfer of heat *from* the gas to the reservoir, take *heat entering the system* to be a negative quantity.

L or R, remove the partition and allow the gas to expand into the container, that, is produce the state F. Then slowly compress the gas to L, doing work on it and expelling heat. The full process is dissipatory, and the amount of dissipation is the amount of heat expelled. The dissipation occurs in the first step, though it's in the second step that the heat is expelled. We'll call this manipulation $M_{1/2}$; it is a member of a family of processes $\{M_p\}$, which we will define shortly.

Dissipation so defined will do as a measure of dissipation in the special case in which all transfers of heat in and out of the system involve reservoirs at the same temperature T. Suppose, now, that we have available to us heat reservoirs at two different temperatures, a hotter temperature T_1 and a cooler one T_2. Suppose that some quantity of heat Q is absorbed from the hotter reservoir, via a thermodynamically reversible process. The initial state can be restored by depositing the same amount of heat back into the hotter reservoir, or else we could run a Carnot cycle and restore the initial state while depositing a *smaller* amount of heat Q' into the cooler reservoir, where

$$Q' = \left(\frac{T_2}{T_1}\right) Q, \tag{1}$$

or,

$$\frac{Q'}{T_2} = \frac{Q}{T_1}. \tag{2}$$

Following Clausius (1854), we will call the quantity Q/T the *equivalence-value* of a quantity Q of heat transferred reversibly to or from a reservoir at temperature T. The rational for this terminology is: if we want to restore the initial state, we can either expel an amount Q at temperature T_1, or an amount Q' at temperature T_2; a quantity Q at temperature T_2 is equivalent, as far as its usefulness in restoring the initial state, to a quantity Q' at temperature T_1; they have the same equivalence-value. The mark of a thermodynamically reversible cycle is that the sum of equivalence-values of heat transferred out of a system be equal in magnitude to the sum of equivalence-values of heat transferred into the system. The general definition of dissipation incurred in the course of a process that takes a state a to a state b is: the difference between the minimum value of the sum of equivalence-values of heat transferred out of the system, among all processes that take b back to a, and the sum of equivalence-values of heat transfers into the system during the course of M. Call this quantity $\delta_M(a \to b)$. For the case in which reservoirs have the same temperature T, it doesn't really matter whether we use $D_M^T(a \to b)$

5

or $\delta_M(a \to b)$ as our measure of quantity of dissipation, since the two are related simply by,

$$D_M^T(a \to b) = T\,\delta_M(a \to b). \tag{3}$$

Returning to our example: if the gas can be compressed reversibly back to its original volume, by a standard calculation, this involves transfer of a quantity of heat from the gas to the reservoir equal to

$$Q_{rev}(F \to L) = NkT \log 2. \tag{4}$$

Therefore, since the process of removal of the partition, from an initial state of either L or R, involves no heat passing into the system, we subtract zero from this quantity, to get a dissipation equal to

$$D_{M_{1/2}}^T(L \to L) = D_{M_{1/2}}^T(R \to L) = NkT \log 2 - 0 = NkT \log 2. \tag{5}$$

So, we've considered one process that produces a final state L from either initial state L or R, and it's dissipatory; we have dissipation $NkT \log 2$, whether the initial state is L or R.

There are other processes that produce L from initial state L or R, a whole continuum of them. For any p, $0 < p < 1$, consider a process, which we will call M_p, in which the partition is moved slowly from its initial position at the middle of the container to a position in which the left and right parts have volumes that are fractions p and $1 - p$, respectively, of the whole. For p greater than $1/2$, this involves expansion of the gas, if the initial state is L, and compression, if it is R. We then remove the partition, and allow the gas to fill the whole of the container, producing state F. As a final step, we reversibly compress the gas to L.

The dissipation associated with each of these processes is just the net heat expelled into the environment. This is $NkT \log 2$ for the final stage. For the first stage, if p is greater than $1/2$, then heat passed *into* the system if the initial state is L, and out of the system if the initial state is R. Thus, for $p > 1/2$, the net heat expelled from the system is greater than $NkT \log 2$ if the initial state is L, and less than $NkT \log 2$ if the initial state is R. These are reversed for $p < 1/2$.

We can calculate how much greater or less these quantities are than $NkT \log 2$.

The dissipations associated with such a process, for the two initial states,

6

are

$$D^T_{M_p}(L \rightarrow L) = NkT \log 2 - NkT \log(2p) = -NkT \log p$$
$$D^T_{M_p}(R \rightarrow L) = NkT \log 2 - NkT \log(2(1-p)) = -NkT \log(1-p). \tag{6}$$

Note that, by choosing p close enough to one, we can make the dissipation associated with the process, for initial state L, as close to zero as we like. That, is, reversibly expanding the gas until it occupies almost all of the container, then removing the partition, is close to being a reversible process. But this small dissipation, if the initial state is L, comes at the cost of a large dissipation if the initial state is R. When the initial state is R, the process involves compressing the gas into a very small volume, and then removing the partition.

We have here a tradeoff; by choosing p, we can make the dissipation associated with one of the initial states as close to zero as we want, at the expense of making the dissipation associated with the other initial state high. Here's the way of expressing the tradeoff that I find clearest. Using D_L and D_R as abbreviations for these two quantities, we have, for process M_p,

$$e^{-D_L/kT} + e^{-D_R/kT} = p^N + (1-p)^N \leq 1. \tag{7}$$

(We have used the fact that $N \geq 1$, since a gas can't have less than one molecule in it.) For a one-molecule gas, we have equality,

$$e^{-D_L/kT} + e^{-D_R/kT} = p + (1-p) = 1. \tag{8}$$

To sum up: we have considered a set of processes $\{M_p\}$, each of which takes the two initial states, L and R, to the same final state L. All the processes in this set satisfy,

$$e^{-D_L/kT} + e^{-D_R/kT} \leq 1, \tag{9}$$

where D_L and D_R are the dissipations incurred if the gas is initial in state L and R, respectively. For the one-molecule gas, we have equality in (9), for each M_p.

The question we now ask is: Could we, for any of these processes M_p, cook up a process that takes both L and R to the same final state, that assuredly has lower dissipation than incurred by M_p, for both initial states? The fact that we haven't thought of one yet is, of course, no proof that it

can't be done. But, as we will prove in section 4.3, for the one-molecule gas, the answer is negative. Each of the processes M_p is in a certain sense optimal, in that there is no process that has lower dissipation than M_p for *both* of the initial states L and R. What will be proven is that, in general, for any process that takes both of two distinguishable states to the same final state, with associated dissipations D_1 and D_2, the pair (D_1, D_2) must satisfy,

$$e^{-D_1/kT} + e^{-D_2/kT} \le 1. \tag{10}$$

The set of values of the pair (D_1, D_2) permitted by (10) is graphed in Figure 1. This allows us to visualize the tradeoff already mentioned. If one wants to make one of the dissipations small, this comes at the price of making the other large. For the special case in which the same dissipation is incurred, regardless of initial state, the bound says that this common dissipation must be at least $kT \log 2$.
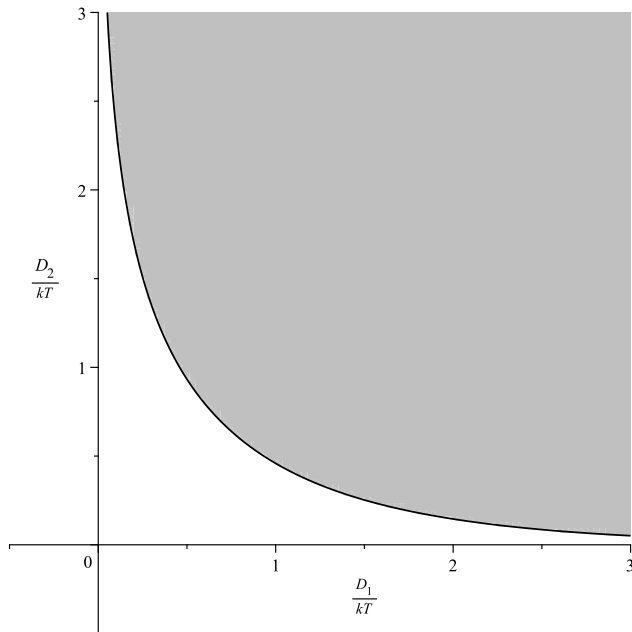

Figure 1: Values of (D_1, D_2) permitted by the Landauer bound.

8

For the one-molecule gas, each of the procedures we have discussed has a pair of dissipations (D_L, D_R) that is on the boundary of the permitted region. This means that, for any of these procedures, there is no procedure that will reduce dissipation for *both* initial states.

As Szilard (1929) pointed out, the bound (10) can be expressed, equivalently, as,

For all p in $(0, 1)$,

$$p\, D_1 + (1 - p)\, D_2 \;\geq\; -kT \left[p \log p + (1 - p) \log(1 - p) \right]. \quad (11)$$

This is the form in which the bound is usually expressed, in the literature on Landauer's principle. For proof of the equivalence, see Appendix B of Myrvold (2024).

Here's how to think about this equivalence. For any value of p, the set of pairs (D_1, D_2) satisfying the equation

$$p\, D_1 + (1 - p)\, D_2 \;=\; -kT \left[p \log p + (1 - p) \log(1 - p) \right]. \quad (12)$$

is a straight line in (D_1, D_2) space tangent to the boundary of the shaded area in Figure 1 at the point $(-kT \log p, -kT \log(1 - p))$. The inequality (11) says that, for each p, the point (D_1, D_2) is on or above the line defined by equation (12). The set of all such lines is the envelope of the shaded area in Figure 1, the set of all (D_1, D_2) satisfying the Landauer bound, inequality (10).

3   Introducing probabilities

Maxwell, writing in 1878, defined thermodynamics as "the investigation of the dynamical and thermal properties of bodies, deduced entirely from what are called the First and Second laws of Thermodynamics, without any hypotheses as to the molecular constitution of bodies. . . ."(Maxwell 1878, p. 258; Niven 1890, pp. 664–665).

As long as we confine our attention to the macroscopic level, we can ignore the molecular structure of matter, and treat the results of our manipulations as predictable, free of the uncertainties induced by molecular fluctuations. This predictability is implicit in thermodynamics as it was developed in the 19th century, in that one speaks of *the* quantity of heat that *will* be exchanged

9

in a given process. If we take a container of gas, and insert a partition halfway along the length, then, for a gas with a macroscopic number of molecules, we expect, with a high degree of confidence, that the quantities of gas now trapped in each of the two halves will be pretty much the same. For a gas consisting of only a few molecules, we have to take into account the non-negligible chance that the quantities of gas trapped in the two halves will be appreciably different, and, for the one-molecule gas, the gas will, with certainty, be trapped entirely in either the left or the right half. The results of subsequent operations will depend on which half it is in. If we want to extend thermodynamics down to the molecular scale, therefore, we will have to live with limitations on predictability. If all goes well, however, we will be able to attach probabilities to the results of our operations.

What we can expect, in the realm of probabilistic processes, is a bound on the expectation values of dissipations.[3] As Szilard put it,

> if we want to use the fluctuation phenomena in order to gain energy at the expense of heat, we are in the same position as playing a game of chance, in which we may win certain amounts now and then, although the expectation value of the winnings is zero or negative (Szilard 1964, p. 302; Leff and Rex 2003, p. 111, from Szilard 1929, p. 841).

It is customary, in statistical thermodynamics, to associate with a system that is in thermal contact with a heat bath, a probability distribution known as a *canonical distribution*, or, sometimes, a *thermal distribution*. If $H(x)$ is

---

[3] *Expectation values*, not *expected values*. The phrase "expected value" is a solecism that seems to have arisen only in the 20th century; the *OED* gives 1915 as its earliest attestation. An *expectation value*, in the original meaning of the phrase, is the *value of an expectation*. If you have a ticket that entitles you to a reward tomorrow if some event E happens in the meantime, you have an expectation: an expectation of receiving the reward if the condition is met. If E is not certain to occur, the value of the expectation of receiving a reward if E happens is less than the value of receiving the same reward unconditionally, and diminishes as the probability of E decreases. This sort of terminology has its origin in Huygens' treatise, *Van Rekeningh in Spelen van Geluck* (pub. 1660), which is all about calculating the values of various expectations.

My guess is that the use of "expected value" stems from over-emphasis on probability distributions that are strongly peaked near their expectation values. For those special cases, with high probability, the actual value will be close to the expectation value. But one can also have cases in which it's *impossible* for the actual value to be close to the expectation value.

the Hamiltonian of the system, the canonical distribution at temperature T is the probability distribution that has a density function, with respect to Liouville measure, given by,

$$\rho(x) = Z^{-1} e^{-H(x)/kT}, \tag{13}$$

where Z is a normalization constant chosen to make the integral of this function over the available region of phase space equal to one. In what follows it is assumed that probabilities of heat exchanges calculated using distributions of this form are correct. This assumption, which plays a central role in statistical thermodynamics, is what I have elsewhere (Myrvold, 2021) called the *Canonical Postulate*.

4 The proof

4.1 The framework

The framework used in the proof will be the more-or-less standard one for statistical thermodynamics, going back to Gibbs (1902).[4] We have a system of interest, A, that can interact with a heat reservoir B. We can, of course, consider cases that involve two or more heat reservoirs; the generalization is straightforward, but for simplicity we will consider the case of only one.

We assume that there is a probability distribution that gives probabilities for the state of the composite system AB at time t_0. What does a good part of the work in the proof is that this is assumed to take a particular form: the probability distribution for the reservoir B is a canonical distribution at temperature T, and the system A (whose distribution is *not* assumed to be canonical) is uncorrelated with B at time t_0.

Between time t_0 and time t_1 the system AB undergoes Hamiltonian evolution, according to some, possibly time-varying, Hamiltonian $H_{AB}(t)$. The course of $H_{AB}(t)$ is assumed to be pre-programmed, independent of the state of the system AB. We assume that at times t_0 and t_1, there is no term in the total Hamiltonian corresponding to an interaction between A and B, and hence that at these times the total energy of the composite system AB is just the sum of the energies A and B. We also assume that the Hamiltonian

---

[4]Of course, it doesn't follow from this that we are thereby committed to any of the more outlandish things that you may have been told "Gibbsians" believe.

for B, H_B, is the same at t_0 and t_1.[5] We don't assume that the Hamiltonian for A is the same at both times. In between times t_0 and t_1, there might be coupling between A and B, allowing the two to exchange energy. In addition, the internal energy of A might change as a result of changes in H_A. Changes in the internal energy of A due to exchanges of energy with B are to be counted as *heat*, changes due to changes in the Hamiltonian, as *work*.

The Hamiltonian evolution consists of a dynamical flow on the phase space of the system AB, a mapping $\mathcal{T}$ from phase-points at time t_0 to phase-points at time t_1. The flow induces a mapping from a probability distribution over microstates of the system at time t_0 to a corresponding probability distribution over microstates at time t_1: the probability that, at time t_1, the phase point is in a given region Δ of phase space is the probability that, at time t_0, the phase point is among those that get mapped by $\mathcal{T}$ into a point in Δ. The only property of the Hamiltonian flow on phase space that we will be invoking is that there is a measure on phase space, the Liouville measure, λ, that is conserved by this flow. That is, if we apply our recipe for evolving probability distributions to the Liouville measure λ, it turns out that any region Δ assigned a probability by λ at time t_0 is assigned the same probability by the result of applying the Hamiltonian flow to λ to produce a probability distribution over states of affairs at time t_1.

For the purposes of the proof, we need not say much about the status of these probability distributions, or about the notion of probability invoked. The theorem relates expectation values of heat exchanges, calculated with respect to some probability distributions, to other quantities defined in terms of those same probability distributions. The only caveat is: one has to accept that it makes sense to talk about the probability that the state of the system is in a given subset of the phase space at a given time, and to apply the dynamical evolution of the system to relate probabilities of states time t_0 to probabilities at time t_1. A point of view that held that probability-talk applied to physical systems makes sense only for systems in equilibrium, and identified the probability ascribed to a region with the long-term average of the fraction of time spent in that region, would not permit this. This is an indication of a limitation of that point of view, a signal that, though it might be of use in some applications, it is not adequate for all applications of probability in physics.

---

[5]That is, we assume that the Hamiltonian is the same function of the phase point at these two times; we don't assume that the internal energy of B is the same at both times.

A probability distribution over conditions at time t_0, together with the dynamical map $\mathcal{T}$, yields a probability distribution over the amount of heat exchanged between A and B in the interim between t_0 and t_1. This probability distribution, in turn, yields an expectation value, $\langle Q \rangle$, for the heat exchange. We take heat going from the reservoir B to the system A as positive, and as negative if it goes in the opposite direction. The fundamental theorem of statistical thermodynamics, proved in the next section, will be about such expectation values.

It should be stressed that it is *not* assumed that the probability distribution for Q is tightly focussed near its expectation value, or that the actual value is close to the expectation value. The theorems we will prove will be entirely general, and remain valid when the spread of the distribution for Q is large.

We first prove, in subsection 4.2, what I have called the *Fundamental Theorem of Statistical Thermodynamics*, which is an analogue, within statistical mechanics, of the second law of thermodynamics. We then apply this, in subsection 4.3, to get the Landauer bound.

4.2   The fundamental theorem of statistical thermodynamics

We will be concerned with probability distributions on the phase space of our system, with the restriction that they can be represented by a density function with respect to the Liouville measure. We define a functional on the set of such probability distributions, which (unfortunately, in my opinion) has come to be known as the "Gibbs entropy." We will use the notation $\mathcal{S}[\rho]$ for this functional. The letter S is, of course, typically used for entropy, but don't let the notation fool you. *No connection between $\mathcal{S}$ and thermodynamic entropy is presumed in the proof.* In particular, we are not using changes in $\mathcal{S}$ as a measure of quantity of dissipation, and so arguments that it is unsuitable for that role *have no relevance whatsoever to our proof.*

If ρ is a probability density function with respect to the Liouville measure λ, we define $\mathcal{S}[\rho]$ as,

$$\mathcal{S}[\rho] = -k\langle \log \rho \rangle_\rho = -k \int \rho(x) \log \rho(x) \, d\lambda. \qquad (14)$$

Given a composite system AB, with a probability distribution represented by a density function ρ_{AB}, marginal densities ρ_A and ρ_B are formed by inte-

grating out the degrees of freedom of the other system. We will, of course, be considering time-evolving probability distributions. For brevity, we will write $\mathcal{S}_{AB}(t)$ for $\mathcal{S}[\rho_{AB}(t)]$, and similarly for $\mathcal{S}_A(t)$ and $\mathcal{S}_B(t)$.

What we will prove is the following.[6]

> **The fundamental theorem of statistical thermodynamics.** Assume that, at time t_0, the probability distribution over the composite system AB is such that the distribution of B is a canonical distribution, at temperature T, and the distributions of A and B are probabilistically independent. Between times t_0 and t_1 the composite system AB undergoes Hamiltonian evolution, according to a Hamiltonian that may vary with time; the only restrictions are that at the times t_0 and t_1 the total Hamiltonian H_{AB} is the sum of subsystem Hamiltonians H_A and H_B, and that $H_B(t_1) = H_B(t_0)$. We define $\langle Q \rangle$ as the expectation value of energy obtained by A from B:
>
> $$\langle Q \rangle = - \left(\langle H_B(t_1) \rangle - \langle H_B(t_0) \rangle \right).$$
>
> Then
>
> $$\langle Q \rangle \leq T \left(\mathcal{S}_A[t_1] - \mathcal{S}_A[t_0] \right). \tag{15}$$

This is a mathematical result about probability distributions evolving via Hamiltonian evolution; it holds for any probability distribution that has a density with respect to Liouville measure, whether or not that probability distribution is held to have any physical significance.

The features of $\mathcal{S}$ that will be used in the proof are the following.

Lemma 1. *For any Hamiltonian H, and any $T > 0$, the canonical distribution at temperature T minimizes*

$$\langle H \rangle_\rho - T\mathcal{S}[\rho].$$

Lemma 2 (Subadditivity). *For a composite system AB,*

$$\mathcal{S}[\rho_{AB}] \leq \mathcal{S}[\rho_A] + \mathcal{S}[\rho_B],$$

with equality if and only if the subsystems are probabilistically independent.

---

[6]This is not a new theorem. The classical version of it is found in Gibbs (1902, pp. 160–164), and the quantum version, in Tolman (1938, §128–130). Nonetheless, it is not as well-known in the philosophical literature on statistical mechanics and thermodynamics as it should be. Maroney (2009) refers to it as a *generalized Landauer principle*.

Lemma 3 (Conservation of $\mathcal{S}$). $\mathcal{S}[\rho]$ *is conserved under Hamiltonian evolution.*

From there, the fundamental theorem is pretty simple. A and B are assumed to be uncorrelated at t_0, but might be correlated at t_1. Subadditivity gives us,

$$\mathcal{S}_A(t_0) + \mathcal{S}_B(t_0) = \mathcal{S}_{AB}(t_0); \tag{16}$$

$$\mathcal{S}_A(t_1) + \mathcal{S}_B(t_1) \geq \mathcal{S}_{AB}(t_1). \tag{17}$$

Because the evolution of AB is Hamiltonian in the interim between t_0 and t_1,

$$\mathcal{S}_{AB}(t_0) = \mathcal{S}_{AB}(t_1). \tag{18}$$

Combining these gives,

$$\mathcal{S}_A(t_1) + \mathcal{S}_B(t_1) \geq \mathcal{S}_A(t_0) + \mathcal{S}_B(t_0) \tag{19}$$

or, more succinctly,

$$\Delta \mathcal{S}_A + \Delta \mathcal{S}_B \geq 0. \tag{20}$$

The distribution of B at time t_0 is assumed to be canonical at temperature T; this is not assumed for t_1. From this, Lemma 1 gives us,

$$\langle H_B(t_1) \rangle - T\mathcal{S}_B(t_1) \geq \langle H_B(t_0) \rangle - T\mathcal{S}_B(t_0), \tag{21}$$

or,

$$\langle \Delta H_B \rangle \geq T\Delta \mathcal{S}_B. \tag{22}$$

Combining (22) and (20), and using the fact that $T > 0$, gives,

$$T\Delta \mathcal{S}_A + \langle \Delta H_B \rangle \geq T\Delta \mathcal{S}_A + T\Delta \mathcal{S}_B \geq 0. \tag{23}$$

That is,

$$\langle Q \rangle = -\langle \Delta H_B \rangle \leq T\Delta \mathcal{S}_A, \tag{24}$$

which is what was to be proved.

To some readers, it might look like we have done the impossible, and derived a time-asymmetric conclusion from time-symmetric premises. Those readers might wonder whether some implicit time-asymmetric assumption has been smuggled in. The answer to this is that nothing has been smuggled in that has not been explicitly stated. The theorem, taken on its own, exhibits no time asymmetry; nothing in its statement or its proof requires that t_0 be

to the past of t_1. The two times t_0 and t_1 do not, however, enter into the statement of the theorem symmetrically; absence of correlation between A and B is assumed at t_0 but not at t_1. In application to physical systems, we will assume that the process of equilibration undergone by the heat reservoir B has effectively effaced any correlations it might have with the rest of the world, and for this reason presume absence of correlation between A and B prior to any interactions they may undergo in the course of the manipulation at hand.

4.3 The Landauer bound

4.3.1 Example: our ideal gas

As a warm-up to the more general theorem, let's apply these considerations to our gas of §2.

The strategy is as follows. Let a_L and a_R be probability distributions associated with states L and R, respectively. They might, for example, be canonical distributions with support confined to molecular positions in L and R, respectively. Suppose we have some manipulation M that, applied to a_L or a_R, yields a_L at the end. This manipulation M could, for example, be any one of the procedures M_p we discussed in §2, which take the gas to the state F, followed by reversible compression to L. It could also be something else that we haven't thought of, so long as it falls within the framework outlined in §4.1. For simplicity, in this section we assume that $\mathcal{S}[a_L] = \mathcal{S}[a_R]$, and that the final state is one of the initial states. However, the proof in its full generality doesn't require these assumptions, which will be dropped in the next section.

Let $\langle Q_L \rangle_M$ be the expectation value of quantities of heat deposited *into* the reservoir over the course of manipulation M when the initial state is L, and similarly for $\langle Q_R \rangle_M$.

The strategy of the proof is as follows. Given the probability distributions a_L and a_R, for any number p in $(0,1)$, there is another probability distribution that is a mixture of the two, a weighted average with weights p and $1 - p$. Call this mixture $\bar{a}_p$. If you like, you can imagine a situation in which a randomizing device is employed to choose whether to prepare a_L or a_R, but this is not necessary. We are employing the mixture to find out something about the quantities $\langle Q_L \rangle_M$ and $\langle Q_R \rangle_M$, and for that purpose it

is not necessary to associate any physical set-up with this mixture.[7]

The key idea is that, since the manipulation M takes both of a_L and a_R to L, it also takes a mixture $\bar{a}_p$ to L, for any value of p. The associated expectation value of heat exchanges will be the corresponding weighted average of heat exchanges associated with the two states a_L and a_R. That is, the expectation value of heat exchanged, when M is applied to the mixture $\bar{a}_p$, is,

$$\langle Q_{\bar{a}_p} \rangle_M = p \langle Q_L \rangle_M + (1 - p)\langle Q_R \rangle_M. \tag{25}$$

Furthermore, since the Fundamental Theorem applies to any probability distribution, we can employ it to get a relation between the these quantities. Recalling that $\langle Q_L \rangle_M$ and $\langle Q_R \rangle_M$ are expectation values of heat transferred *out of* the gas, an application of the Fundamental Theorem gives us,

$$- \left(p\langle Q_L \rangle_M + (1 - p)\langle Q_R \rangle_M \right) \leq T \left(\mathcal{S}[a_L] - \mathcal{S}[\bar{a}_p] \right), \tag{26}$$

or,

$$p\langle Q_L \rangle_M + (1 - p)\langle Q_R \rangle_M \geq T \left(\mathcal{S}[\bar{a}_p] - \mathcal{S}[a_L] \right). \tag{27}$$

The distributions a_L and a_R have non-overlapping supports. For such distributions there is a simple relation between the value that the functional $\mathcal{S}$ assigns to the mixture $\bar{a}_p$, and the values it assigns to a_L and a_R.

$$\mathcal{S}[\bar{a}_p] = p\,\mathcal{S}[a_L] + (1 - p)\mathcal{S}[a_R] - k\left[p \log p + (1 - p) \log(1 - p) \right]. \tag{28}$$

And, since $\mathcal{S}[a_R] = \mathcal{S}[a_L]$,

$$\mathcal{S}[\bar{a}_p] = \mathcal{S}[a_L] - k\left[p \log p + (1 - p) \log(1 - p) \right]. \tag{29}$$

Combining (27) and (29) gives,

$$p \langle Q_L \rangle_M + (1 - p) \langle Q_R \rangle_M \geq -kT\left[p \log p + (1 - p) \log(1 - p) \right]. \tag{30}$$

---

[7]In particular, if someone were to insist that the *only* legitimate probabilities in statistical physics come from canonical distributions, we could accept that for the sake of argument, with no effect whatsoever on the proof. Let a_L and a_R be two canonical distributions at temperature T, with support in the left and right sides of the box, respectively. For any manipulation, these yield probability distributions over heat exchanges, which yield expectation values that we can use to define $\langle Q_L \rangle_M$ and $\langle Q_R \rangle_M$. As a purely mathematical fact any weighted average of probability distributions is a probability distribution, and we can ask what the dynamical evolution associated with our manipulation will do to that distribution, and from that gain information about the quantities $\langle Q_L \rangle_M$ and $\langle Q_R \rangle_M$.

Since p can be any number in $(0, 1)$, $\langle Q_L \rangle_M$ and $\langle Q_R \rangle_M$ have to be such that (30) holds for *all* p in $(0, 1)$. This, as already mentioned, is equivalent to

$$e^{-\langle Q_L \rangle_M / kT} + e^{-\langle Q_R \rangle_M / kT} \leq 1. \tag{31}$$

Note that neither the weights p, nor the functional $\mathcal{S}$, appear in this result. It is a result about expectation values of heat exchanges for any manipulation that takes both L and R to a final state L; consideration of mixtures, and the quantity $\mathcal{S}$, are used as auxiliaries in a derivation of a relation that is not about them.

If you've followed this derivation, then you've grasped the conceptual apparatus needed for the more general result proved in the next section.

4.3.2 The theorem generalized

In the previous section we considered a special case in which the value of $\mathcal{S}$ is the same for the final state and the two initial states, and in which the initial states can be restored from the final with no heat transfer. In this section we lift those restrictions.

We consider the case in which there are two initial states of system A, a_1, a_2, which are both taken, by some manipulation M, to the same final state b, and a single heat reservoir. These states are not microstates of the system, but consist of a specification of values of external parameters (such as the position of the piston), represented in the Hamiltonian of the system, and a probability distribution over microstates of the system. The manipulation M is represented by a time-dependent Hamiltonian. Generalization to a larger number of initial states, and more than one heat reservoir, is straightforward.

As the specifications of a_1 and a_2 include specifications of probability distributions over the state of AB, and the manipulation M defines a dynamical flow on the state space of the composite system AB, there will be associated probability distributions over heat exchanges. Let $\mathcal{Q}_M^T(a_i \to b)$ be the expectation value of heat received by A from the reservoir B, in the course of the manipulation M, when the initial state is a_i.

We assume that there are processes that take b back to a_1, and other processes that take b back to a_2. Among processes that take b to a_1 there will be a minimum expectation value of the quantity of heat that is transferred from A to the reservoir. Recall that, in section 4.2, we took heat *into* A as positive, heat flowing *out* of it, as negative. So a minimum value of heat out means a maximum value of the (negative) quantity Q. We define $\hat{\mathcal{Q}}^T(b \to a_1)$

18

as the maximum expectation value of heat transfer from A to B among all processes that take b to a_1. We similarly define $\hat{\mathcal{Q}}^T(b \to a_2)$.

If $\mathcal{Q}^T_M(a_i \to b)$ is equal in absolute magnitude to $\hat{\mathcal{Q}}^T(b \to a_i)$, then the manipulation M taking a_i to b will be called *reversible*. In general, restoring the state a_i will require (on average) more heat to be expelled to the reservoir than was obtained from the reservoir during the transition from a_i to b. For that reason, we will have,[8]

$$\hat{\mathcal{Q}}^T(b \to a_i) + \mathcal{Q}^T_M(a_i \to b) \leq 0. \tag{32}$$

In this context, we define the *dissipation* associated with the process M, when started on state a_i, to be the absolute value of this quantity (that is, how far below zero it is).

$$\mathcal{D}^T_M(a_i \to b) = -\left(\hat{\mathcal{Q}}^T(b \to a_i) + \mathcal{Q}^T_M(a_i \to b) \right). \tag{33}$$

So, given the manipulation M, we have two quantities, $\mathcal{D}^T_M(a_1 \to b)$ and $\mathcal{D}^T_M(a_2 \to b)$, that we want to know about. In what follows, we will abbreviate these as D_1 and D_2. What we will show is that, for all p in the interval $(0, 1)$,

$$p\, D_1 + (1 - p)\, D_2 \geq -kT \left[p \log p + (1 - p) \log(1 - p) \right]. \tag{34}$$

As already remarked, this is equivalent to

$$e^{-D_1/kT} + e^{-D_2/kT} \leq 1. \tag{35}$$

We employ the same strategy as before. We consider a mixture $\bar{a}_p$, of a_1 and a_2, with weights p and $1 - p$, respectively. As before, the expectation value of heat exchanges associated with $\bar{a}_p$ will be the corresponding weighted average of heat exchanges associated with the two states a_1 and a_2. That is,

$$\mathcal{Q}^T_M(\bar{a}_p \to b) = p\, \mathcal{Q}^T_M(a_1 \to b) + (1 - p)\mathcal{Q}^T_M(a_2 \to b). \tag{36}$$

We will also use, for the case in which a_1 and a_2 are distinguishable (that is, they have non-overlapping supports), the fact that

$$\mathcal{S}[\bar{a}_p] = p\, \mathcal{S}[a_1] + (1 - p)\mathcal{S}[a_2] - k \left[p \log p + (1 - p) \log(1 - p) \right]. \tag{37}$$

---

[8]This follows from the Fundamental Theorem.

The proof involves three applications of the Fundamental Theorem. The first two give us,

$$\hat{\mathcal{Q}}^T(b \to a_1) \leq T\left(\mathcal{S}[a_1] - \mathcal{S}[b]\right)$$

$$\hat{\mathcal{Q}}^T(b \to a_2) \leq T\left(\mathcal{S}[a_2] - \mathcal{S}[b]\right) \tag{38}$$

The third application is

$$\mathcal{Q}_M^T(\bar{a}_p \to b) \leq T\left(\mathcal{S}[b] - \mathcal{S}[\bar{a}_p]\right). \tag{39}$$

What we want to prove is (34). We recall the definition of D_1 and D_2.

$$D_i = -(\hat{\mathcal{Q}}^T(b \to a_i) + \mathcal{Q}_M^T(a_i \to b)). \tag{40}$$

Therefore,

$$p\, D_1 + (1-p)\, D_2 = -\left(p\, \hat{\mathcal{Q}}^T(b \to a_1) + (1-p)\hat{\mathcal{Q}}^T(b \to a_2) \right.$$
$$\left. + p\, \mathcal{Q}_M^T(a_1 \to b) + (1-p)\mathcal{Q}_M^T(a_2 \to b) \right). \tag{41}$$

Using (36), this is,

$$p\, D_1 + (1-p)\, D_2 =$$
$$-\left(p\, \hat{\mathcal{Q}}^T(b \to a_1) + (1-p)\hat{\mathcal{Q}}^T(b \to a_2) + \mathcal{Q}_M^T(\bar{a}_p \to b) \right). \tag{42}$$

Employing (38) and (39), we have,

$$p\, \hat{\mathcal{Q}}^T(b \to a_1) + (1-p)\hat{\mathcal{Q}}^T(b \to a_2) + \mathcal{Q}_M^T(\bar{a}_p \to b)$$
$$\leq T\left(p\, S[a_1] + (1-p)S[a_2] - S[\bar{a}_p]\right). \tag{43}$$

Plugging this into (42) and doing a bit of rearranging gives us,

$$p\, D_1 + (1-p)D_2 \geq T\left(S[\bar{a}_p] - p\, S[a_1] - (1-p)S[a_2]\right). \tag{44}$$

Inequality (44) holds whether or not the distributions a_1 and a_2 are distinguishable. But, in the case in which they are, (37) applies; plugging this into the right-hand-side of (44) gives us, finally,

$$p\, D_1 + (1-p)D_2 \geq -kT\left[p\log p + (1-p)\log(1-p)\right], \tag{45}$$

which is what was to be proved.

5   On the internal/external division

In our proof, we assumed that the composite system, consisting of the system A and any heat reservoirs it interacts with, undergoes Hamiltonian evolution. We didn't assume that it's isolated; the Hamiltonian may be changing with time, and hence energy may be exchanged between AB and the systems driving the changes. All that we require is that the course of the Hamiltonian be independent of the state of the system AB.

That means that there will typically be, in addition to the dissipation within the system A, also dissipation associated with the systems driving the changes in the Hamiltonian. Also, in cases in which the work done on or by the system A differs, depending on the initial state of A, the external world might contain a record of the initial state of A, even though A itself does not. *We have obtained a result about dissipation associated with a manipulation of A, in terms of heat transfer between A and heat reservoirs, associated with any process that takes two or more distinguishable states of A to the same state of A.*

This set-up might strike some as odd. Shouldn't we consider the sum total of dissipations in *all systems involved*? And, if we're concerned about erasure, shouldn't we ensure that the external world contains no trace of the initial state of A?

The first thing to say about this is: if we can get a result about lower bounds on dissipation within a system, associated with loss of distinguishability of states of that system, this is a stronger result than one that merely deals with total dissipation in all systems involved. And if we have a theorem that applies to any situation in which distinguishable states of A are brought to the same state, whether or not there are traces of its initial state in the external world, this is a stronger result than one confined to circumstances in which all systems, including the ones we have externalized, are brought to the same state. The more restricted scenarios can be obtained as special cases.

But another reason for this way of proceeding is that we are seeking to extend, as far as possible, thermodynamical concepts to statistical mechanics. This, it might seem, would motivate taking all dissipations in all systems involved in a process. It might be thought that, in fact, this is *required* by the second law of thermodynamics, which is often glossed as saying that the total entropy of all systems involved in a process must be non-decreasing. Sometimes one hears it said that the second law, strictly speaking, applies

only to isolated systems.

This isn't right. To see why, let's think about the second law for a bit.

In one of its forms, the second law says that, if a system undergoes a cyclic process (that is, a process that returns it to its initial thermodynamic state), exchanging heats Q_i with heat reservoirs at temperatures T_i, then

$$\sum_i \frac{Q_i}{T_i} \leq 0. \tag{46}$$

If the process is thermodynamically reversible, it can go in the opposite direction, reversing the signs of the heats Q_i, and so we must have equality in (46).

Another formulation of the second law is,

There exists a state function S such that, for any process that takes a system from state a to state b,

$$\sum_i \frac{Q_i}{T_i} \leq S(b) - S(a). \tag{47}$$

If we add the further assumption, often tacitly assumed in thermodynamics, that any two states can be connected reversibly, then any two state functions satisfying (47) differ at most by an constant, and we have the familiar thermodynamic entropy, defined up to an arbitrary additive constant.

The second law has the consequence that:

(*) If a system does not exchange heat with any other system, its entropy does not decrease.

This, in turn, has the further consequence,

(**) The entropy of an energetically isolated system—that is, a system not exchanging any energy at all, either as heat or as work, with any other system—does not decrease.

The statement (**) is strictly weaker than (*), and hence, strictly weaker than the second law of thermodynamics. If one could lower the entropy of a system by extracting work from it, without any heat transfer from the system to its environment, then (*) would be refuted, but, if the total entropy of all systems involved increased, (**) would not be. Nonetheless, one sometimes

sees (**) as a statement of the second law. It's not. It's a *consequence* of the second law, but not equivalent to it.[9]

If one could manipulate a system A in such a way that a quantity of heat is absorbed from a reservoir (increasing the energy of the system) and then entirely extracted from the system as work (restoring its energy to its initial value), and, at the end of the process, the system is restored to its initial state, then the second of law thermodynamics, in the formulation (46), would be refuted. This is true even if the process inevitably involved compensatory dissipation in the system that system A does work on.

6   Erasure with dissipation but without increase of Boltzmann entropy

In the minds of some, the somewhat rusty, old-fashioned concept of "entropy" as Clausius defined it, more than a century and a half ago, has been replaced by a newer, more scientifically respectable concept, namely, *Boltzmann entropy*. Someone might with some justice wonder why I have not framed the issue of dissipation during erasure in terms of increase of Boltzmann entropy.

The answer is that, though the concept of Boltzmann entropy has its uses, it is not suited to the task of assessing dissipation associated with chancy processes.

Consider, once again, a one-molecule gas. Suppose the initial state is L. The partition is lifted, and the molecule is allowed to roam freely throughout the container. After a decent interval of time, long compared to the characteristic time-scale of thermalization for the gas, the partition is re-inserted.

The new state is either L or R. These two states have the same Boltzmann entropy, the same Boltzmann entropy as the initial state.

The process is a dissipatory process. As we have shown, it would not be possible to devise a procedure that is guaranteed to restore the initial state that does involves no expulsion of heat from the gas to the reservoir, whether the new state is L or R.

---

[9]Yes, I know that there are physics textbooks that give this as a statement of the second law. And, yes, Clausius himself closed one of his papers (Clausius, 1865) with a playful remark that the two laws of thermodynamics could be formulated as 1) The energy of the universe is constant. 2) The entropy of the world strives towards a maximum. Don't take this too seriously; it's not one of his official forms of the second law.

It is also a process that produces erasure. Removal and then, after a decent interval of time, re-insertion of the partition yields, when applied to L, a state that is either L or R, with equal probability. The same is true if it is applied to R. One can learn nothing about the initial state from examination of the final state.

If one were to try to use increase of Boltzmann entropy as a measure of dissipation, one would erroneously conclude that this is a procedure that effects erasure with no dissipation. Of course, it's not; this illustrates the fact that, in cases in which the final macrostate is not predictable from the initial macrostate, dissipation need not involve an increase in Boltzmann entropy.

The fact that erasure can be effected without increase of Boltzmann entropy has been discussed by Hemmo and Shenker (2012, Ch. 12). They take this to be a counter-example to Landauer's thesis. It would be, if what was claimed was that there could be no erasure without increase of Boltzmann entropy. I have no interest, here, in engaging in textual exegesis of the proponents of Landauer's thesis (*e.g.* Bennett 2003); let me just say that I am skeptical that the relevant texts can be made to support that interpretation.

What this illustrates, I think, is the danger of burdening the poor word "entropy" with so many different meanings. Someone might make a claim, using one meaning, and someone else might, unknowingly, substitute a different meaning, and end up rebutting a claim different from the one that was made.

7 Entropy and Information

The astute reader will have noticed that our proof did not rely on the identification of any of the quantities involved with thermodynamic entropy, and, indeed, proceeded without using the word "entropy" at all. In particular, nothing about the functional $\mathcal{S}$ was used that is not a consequence of its definition. But some have called $\mathcal{S}$ an entropy, intended to be thought of as such, not only for canonical distributions, but also for epistemic mixtures of them.

Whether we do so is, of course, purely a matter of terminology. Nothing about how physical systems behave, or what we can do with them, rides on our choice of how to use the word "entropy." But there is a choice to be made. Clausius, who coined the term *Entropie*, was working within a framework in which it was tacitly assumed that the results of our manipulations are

predictable. His definition doesn't tell us how to apply it to situations in which chance plays a role.

To help us think about this matter, consider two families of processes by which the gas discussed in section 2 may be prepared. One is the *deterministic family*, which has two members, $\mathbf{Det}(L)$ and $\mathbf{Det}(R)$, which prepare states L, and R, respectively, with certainty. The other is the *chancy family*: for any number p in the interval $(0, 1)$, the process $\mathbf{Chance}(p)$, using some randomizing device, prepares state L with probability p, and state R with probability $1 - p$.

If, having employed some preparation $\mathbf{Chance}(p)$ from the chancy family, we now want to prepare state L, there is, as we have seen, a thermodynamic cost to doing so. Any process that is guaranteed to produce L from the output of $\mathbf{Chance}(p)$ must incur some dissipation D_L or D_R, depending on which of these states has been chosen, and the expectation value of the dissipation must satisfy,

$$\langle D \rangle_p = p\, D_L + (1 - p) D_R \geq -kT \left[p \log p + (1 - p) \log(1 - p) \right]. \qquad (48)$$

Should we now recognize this thermodynamic disadvantage of the preparation process $\mathbf{Chance}(p)$, compared to the deterministic processes, by associating with $\mathbf{Chance}(p)$ an entropy that is greater than that of the state L, by an amount equal to $-k \left[p \log p + (1 - p) \log(1 - p) \right]$?

There are considerations for and against this extension of the term "entropy." On the nay side, one might say (and this is the sort of thing that people *do* say),

> Entropy should be a physical property of physical systems, not an attribute of a state of knowledge. The physical state resulting from $\mathbf{Chance}(p)$ is one of two low-entropy states, L or R; the fact that we don't know which doesn't change the physical state. If there were some oracle that could ascertain which of these two states the system is in without disturbing it, and convey that information to you, this would not change the state of the system, and hence not change its entropy.

On the yea side, one might say,

> Entropy is connected to the availability of energy for performing useful work. If a system is in one of two states, L or R, but it

is not known which, this is less useful for obtaining work than if it is known. This motivates calling the quantity that tracks this reduction in usefulness, *entropy*.

Note that this sort of reply would not warrant associating an entropy with *every* state of ignorance in which epistemic probabilities can be given numerical values. If I flip a coin, and afterwards assign epistemic probability of one-half to each of the alternatives *Heads* and *Tails*, this is not automatically a state that differs thermodynamically from a state in which the coin is known to have landed *Heads*. It is only when I can use the knowledge that the coin is *Heads* to obtain more work from the coin, or obtain it more reliably, than I could if I were ignorant, that I should assign a lower entropy to the state of more complete knowledge. That is, on the view that motivates the *yea* answer to our terminological question, entropy is relative both to a state of knowledge and to a class of available manipulations.[10]

These two sorts of replies stem from two conceptions of the nature of thermodynamics, which I have elsewhere referred to as *Planckian* and *Maxwellian* (Myrvold, 2022). On the Maxwellian view, which was the conception of the founders of thermodynamics, thermodynamics is what physicists these days would call a *resource theory*, a theory about how agents with access to specified physical and informational resources and the ability to perform specified operations can use those resources to accomplish specified tasks. On the Planckian view, which came to dominate the twentieth-century textbook tradition, thermodynamics is the study of the properties of macroscopic systems in equilibrium. It cannot be emphasized too much that these are both perfectly legitimate fields of investigation, and are in no way to be thought of rivals for the title of The One And Only Thermodynamics.

I myself don't particularly care which choice of terminology is made, in connection with extension of the use of the term "entropy" to chancy setups. If I had my way, use of the term *entropy* would have been restricted to the context in which Clausius originally defined it, and novel terms would have been coined for the allied concepts in statistical mechanics, the so-

---

[10]It would be a mistake to take differences in the functional $\mathcal{S}$ corresponding to differences in a state of information about a system to correspond, in every case, to differences in thermodynamic entropy, so conceived, without consideration of the class of available manipulations. For a detailed discussion of the conditions under which a difference in the functional $\mathcal{S}$ *does* indicate a difference in thermodynamic entropy, see §6 of Myrvold (2020).

called Boltzmann entropy and the so-called Gibbs entropies. Likewise for the Shannon entropy in information theory.

8 Conclusion

I believe that there is nothing in the above that the critics of the literature on Landauer's principle will find objectionable. I haven't made any attempt to provide detailed references to the literature, but I think you will find that, if you have qualms about Landauer's principle that stem from something you've read, you will find that the argument presented here addresses those worries.

9 Acknowledgments

I am grateful to John Norton, Jacob Barandes, and the members of the University of Western Ontario philosophy of physics reading group for helpful discussions. I thank, in particular, John Norton, for urging me to write up this simpler proof.

References

Bennett, C. H. (2003). Notes on Landauer's principle, reversible computation, and Maxwell's demon. *Studies in History and Philosophy of Modern Physics 34*, 501–510.

Clausius, R. (1854). Ueber eine veränderte Form des zweiten Hauptsatzes der mechanischen Wärmetheorie. *Annalen der Physik 93*, 481–506. Reprinted in Clausius (1864, 127–154); English translation in Clausius (1856), and in Clausius (1867b).

Clausius, R. (1856). On a modified form of the second fundamental theorem in the mechanical theory of heat. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 12*, 81–88. English translation of Clausius (1854).

Clausius, R. (1864). *Abhandlungen über die mechanische Wärmetheorie*, Volume 1. Braunschweig: Friedrich Vieweg und Sohn.

Clausius, R. (1865). Ueber verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie. *Annalen der Physik 125*, 353–400. Reprinted in Clausius (1867a, 1–44). English translation in Clausius (1867b).

Clausius, R. (1867a). *Abhandlungen über die mechanische Wärmetheorie*, Volume 2. Braunschweig: Friedrich Vieweg und Sohn.

Clausius, R. (1867b). *The Mechanical Theory of Heat, with its Applications to the Steam Engine, and to the Physical Properties of Bodies.* London: John van Voorst. English translation, with one additional paper, of Clausius (1864).

Gibbs, J. W. (1902). *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics.* New York: Charles Scribner's Sons.

Hemmo, M. and O. R. Shenker (2012). *The Road to Maxwell's Demon: Conceptual Foundations of Statistical Mechanics.* Cambridge: Cambridge University Press.

Leff, H. S. and A. F. Rex (Eds.) (2003). *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*. Bristol and Philadelphia: Institute of Physics Publishing.

Maroney, O. (2009). Generalizing Landauer's principle. *Physical Review E 79*, 031105.

Maxwell, J. C. (1878). Tait's "Thermodynamics". *Nature 17*, 257–259, 278–280. Reprinted in Niven (1890, 660–671).

Myrvold, W. C. (2020). The science of $\Theta\Delta^{\mathrm{cs}}$. *Foundations of Physics 50*, 1219–1251.

Myrvold, W. C. (2021). On the relation of the laws of thermodynamics to statistical mechanics. Available at `https://philsci-archive.pitt.edu/23539/`.

Myrvold, W. C. (2022). Philosophical issues in thermal physics. *Oxford Research Encyclopedia of Physics*. `https://oxfordre.com/physics/display/10.1093/acrefore/9780190871994.001.0001/acrefore-9780190871994-e-42`.

Myrvold, W. C. (2024). Shakin' all over: Proving Landauer's principle without neglect of fluctuations. *The British Journal for the Philosophy of Science 75*.

Niven, W. D. (Ed.) (1890). *The Scientific Papers of James Clerk Maxwell*, Volume Two. Cambridge: Cambridge University Press.

Norton, J. D. (2005). Eaters of the lotus: Landauer's principle and the return of Maxwell's demon. *Studies in History and Philosophy of Modern Physics 36*, 375–411.

Norton, J. D. (2011). Waiting for Landauer. *Studies in History and Philosophy of Modern Physics 42*, 184–198.

Norton, J. D. (2013). All shook up: Fluctuations, Maxwell's demon and the thermodynamics of computation. *Entropy 15*, 4432–4483.

Szilard, L. (1929). Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Zeitschrift für Physik 53*,

840–856. Reprinted in Szilard (1972, 103–119). English translation in Szilard (1972, 120–129), and in Leff and Rex (2003, 110–119).

Szilard, L. (1964). On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Behavioral Science 9*, 301–310. English translation of Szilard (1929). Reprinted in Szilard (1972, 120–129), and in Leff and Rex (2003, 110–119).

Szilard, L. (1972). *The Collected Works of Leo Szilard: Scientific Papers.* Cambridge, MA: The MIT Press.

Tolman, R. C. (1938). *The Principles of Statistical Mechanics.* Oxford: Clarendon Press.