Shakin' All Over: Proving Landauer's principle without neglect of fluctuations

Wayne C. Myrvold
Department of Philosophy
The University of Western Ontario
wmyrvold@uwo.ca

Abstract

Landauer's principle is, roughly, the principle that logically irreversible operations cannot be performed without dissipation of energy, with a specified lower bound on that dissipation. Though widely accepted in the literature on the thermodynamics of computation, it has been the subject of considerable dispute in the philosophical literature. Proofs of the principle have been questioned on the grounds of insufficient generality and on the grounds of the assumption, used in the proofs, of the availability of reversible processes at the microscale. The relevance of the principle, should it be true, has also been questioned, as it has been argued that microscale fluctuations entail dissipation that always greatly exceeds the Landauer bound. In this article Landauer's principle is treated within statistical mechanics, and a proof of the principle is given that neither relies on neglect of fluctuations nor assumes the availability of thermodynamically reversible processes. In addition, it is argued that microscale fluctuations are no obstacle to approximating thermodynamic reversibility, in the appropriate sense, as closely as one would like.

1 Introduction

The statement that has come to be known as *Landauer's principle* is, roughly, that there are specifiable bounds on how far one can reduce the dissipation of energy associated with implementation of a logically irreversible operation, that is, an operation whose input state cannot be recovered from its output state. See section 2 for a precise statement of the principle. It is widely accepted in the literature on the thermodynamics of computation; see Leff and Rex ([2003]) for a sampling of the relevant literature and an extensive bibliography. Nonetheless, it has been the subject of considerable controversy in the philosophical literature (Earman and Norton [1999]; Norton [2005]; Ladyman, Presnell, Short, and Groisman [2007]; Ladyman, Presnell, and Short [2008]; Norton [2011]; Hemmo and Shenker [2012, 2013, 2019]; Ladyman and Robertson [2013]; Norton [2013a, 2013b, 2013c]; Ladyman and Robertson [2014]; Ladyman [2018]; Norton [2018]).

Ladyman, Presnell, Short, and Groisman ([2007]), hereinafter referred to as LPSG, presented a proof of Landauer's principle. The proof, like any proof, rests on assumptions. The operative assumptions of the proof are that a probabilistic version of the second law of thermodynamics holds, and that certain processes can be performed reversibly. These processes include, crucially, expansion of a single-molecule gas. Norton ([2011, 2013b, 2013c]) has argued that inevitable fluctuations at the molecular level invalidate the assumption of even approximate thermodynamic reversibility of processes at the microscale, and also that any process involves dissipation in excess of the bounds required by Landauer's principle, rendering the principle moot. This is regarded by Norton as a 'no-go' result, invalidating the basic framework within which most of the work on the thermodynamics of computation has been carried out.

Ladyman and Robertson ([2014]) addressed the purported no-go result, arguing that the conclusion has not been established. They acknowledged, however, a concern about the assumption, ubiquitous in the literature on thermodynamics of computation, of molecular-scale processes carried out with negligible dissipation.

In this article, the subject of Landauer's principle is addressed from the point of view of statistical mechanics, and a derivation within statistical mechanics of the Landauer principle is given that neither relies on neglect of fluctuations nor assumes the availability of thermodynamically reversible processes. In this context, one cannot expect the second law of thermodynamics, as originally conceived, to hold. Because of statistical fluctuations, a heat engine operating between two reservoirs might on occasion extract more work than the Carnot bound on efficiencies allows. The version of the second law relevant to this context is a probabilistic one, which we will call the *statistical second law*, involving *expectation values* of energy exchanges. It is provable within statistical mechanics, in two versions, classical and quantum. It is therefore not required as an independent assumption. The basic terminology and concepts needed for the proof are presented in section 2, and the proof itself in section 3. As Norton has rightly emphasized, a theorem of this sort is moot if the processes involved depart sufficiently far from thermodynamic reversibility. This is explicit in the theorem we prove. Unless there are processes available that approximate reversibility sufficiently closely, the theorem places no bounds on *extra* dissipation associated with logical irreversibility. This point is addressed in section 4, where it is argued that, given the notion of thermodynamic reversibility relevant to the context at hand, fluctuations, even ones that are large on the scale at which the processes are taking place, pose no threat to the assumption that processes can take place that approximate thermodynamic reversibility as closely as one would like. In section 5 these considerations are applied to the stock example of a one-particle gas. With this in place, one can show that worries about the LPSG proof that stem from the assumption that a one-particle gas can be expanded reversibly can be put to rest (section 6).

It has been claimed that Landauer's principle can be shown to be false, and that explicit counterexamples can be given. These supposed counterexamples have to do, not with the theorem proven in this paper, but with a different claim, which has to do with Boltzmann entropy. This matter will be discussed in section 7.

2 The set-up

As is usual in thermodynamics, the thermodynamic state of a system A is defined with respect to some set of manipulable variables $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, which may represent, for example, the positions of the walls of a container the system is constrained to be in, or the value of applied fields. We thus consider a family of Hamiltonians $\{H_{\boldsymbol{\lambda}}\}$. The variables $\boldsymbol{\lambda}$ are treated as exogenous, meaning that we do not include in our physical description the systems that are the sources of these applied fields, and we do not consider the influence of the system A on those systems. They may also be freely specified, independently of the state of the system. We consider some set $\mathcal{M}$ of manipulations of the system, where each manipulation consists of some specification of $\boldsymbol{\lambda}(t)$ through some interval $t_0 \leq t \leq t_1$. In addition, when needed it is assumed that there are available one or more heat reservoirs $\{B_i\}$ at temperatures T_i, with which the system can exchange heat. The system A may be coupled and decoupled from these heat reservoirs during the course of its evolution. That is, the interaction terms in the total Hamiltonian consisting of the system A and the reservoirs $\{B_i\}$ are also treated as manipulable variables.

Since the time of Maxwell ([1871, 1878]), it has been recognized that the kinetic theory of heat entails that the second law of thermodynamics, as originally formulated, cannot hold strictly. When compressing a gas with a piston, we might find on some occasion that, due to a fluctuation in the force exerted by the gas on the piston, less work is needed to compress the gas than one would expect on average, and so in a given cycle of a heat engine we might obtain more net work than allowed by the second law from a given quantity of heat extracted. By the same token we might obtain less net work than expected. We do not, however, expect that we will be able to *consistently and reliably* violate the Carnot limit on efficiency of a heat engine. The original version of the second law should be replaced by a probabilistic one. The second law will then be, to employ Szilard's vivid analogy, like a theorem about the impossibility of a gambling system intended to beat the odds set by a casino.

> Consider somebody playing a thermodynamical gamble with the help of cyclic processes and with the intention of decreasing the entropy of the heat reservoirs. Nature will deal with him like a well established casino, in which it is possible to make an occasional win but for which no system exists ensuring the gambler a profit (Szilard [1972], p. 73, from Szilard [1925], p. 757).

On a macroscopic scale, we expect fluctuations to be negligible. At the microscale on which in-principle limitations on the thermal cost of computation are investigated, fluctuations are far from negligible. Accordingly, we will invoke probabilistic considerations, and treat of the evolution of probability distributions over the state of a system subjected to various manipulations. In connection with the amount of work needed to perform an operation or the amount of heat exchanged in the course of the evolution of the system, the quantities we will consider are the *expectation values* of work and heat exchanges, calculated with respect to those probability distributions.

We assume it makes sense to associate a probability distribution with a preparation procedure, and to compute on its basis probabilities for outcomes of subsequent manipulations. We need not enquire into the status of these probabilities, so long as they serve this purpose.

A *caveat* is in order, however. We will be considering probabilistic mixtures of preparations. A probabilistic mixture of preparations involves choosing from some set of preparations, with specified probabilities as to which preparation is performed. The probability distribution associated with a mixture of preparations is a weighted average of the distributions corresponding to the component preparations. This means that we are *not* identifying the probability assigned to a given region of state space with the long-run fraction of time spent by the system in that region. We may consider, for example, a set-up in which a particle is confined either to the left or to the right of a partition dividing a container, with equal probability for each. Then, for each side of the container, the probability that the particle is on that side is one-half, even though, for each outcome of the preparation, all of the particle's time is spent on one side or the other.

We will treat of "states" $a = (\rho_a, H_a)$, consisting, in the classical context, of a probability distribution over the phase space of the system A, represented by a density function ρ_a, and a Hamiltonian H_a, which, as already noted, may depend on exogenous, manipulable variables. In the quantum context, the density function is replaced by a density operator on the system's Hilbert space. We consider the effects on those states of manipulations in some class $\mathcal{M}$ of manipulations.

As is usual in statistical mechanics, the distributions associated with the heat reservoirs B_i will be canonical distributions, uncorrelated with the system A (see Maroney ([2007]) for discussion of the justification for this use of canonical distributions). In the classical context, a canonical distribution is a distribution that has density, with respect to Liouville measure,

$$\rho_\beta = Z^{-1} e^{-\beta H}, \tag{2.1}$$

where β is the inverse temperature $1/kT$, and Z is the normalization constant required to make the integral of this density over all phase space unity. This depends both on the Hamiltonian H and on β, and is called the *partition function*. In the quantum context, a canonical state is represented by density operator

$$\hat{\rho}_\beta = Z^{-1} e^{-\beta \hat{H}}, \tag{2.2}$$

where, again, Z is the constant required to normalize the state.

The manipulations of a system A we will be considering will be ones of the following form:

- At time t_0, the system has some probability distribution ρ_a, and the Hamiltonian of the system A is H_a.
- At time t_0, the heat reservoirs B_i have canonical distributions at temperatures T_i, uncorrelated with A, and are not interacting with A.
- During the time interval $[t_0, t_1]$, the composite system consisting of A and the reservoirs $\{B_i\}$ undergoes Hamiltonian evolution, governed by a time-dependent Hamiltonian $H(t)$, which may include successive couplings between A and the heat reservoirs $\{B_i\}$.
- The internal Hamiltonians of the reservoirs $\{B_i\}$ do not change.
- At time t_1, the Hamiltonian of the system A is H_b, and, as a result of Hamiltonian evolution of the composite system, the marginal probability distribution of A is ρ_b.

This is a manipulation that takes a state $a = (\rho_a, H_a)$ to state $b = (\rho_b, H_b)$.

It should be noted that we are *not* considering manipulations that consist of a measurement performed on the system A followed by a manipulation of the exogenous variables whose choice depends on the outcome of the measurement. Controlled operations are allowed, but the control mechanism must be internalized, that is, included in the system under study. The system A could consist of two parts A_1 and A_2, which interact in such a way that the state of A_1 affects what happens to A_2, which subsequently affects what happens to A_1. But all of this must be encoded in the Hamiltonian $H(t)$, which may be time-varying but which undergoes a preprogrammed evolution that is *not* dependent on the state of the system A. Otherwise, there may be dissipation associated with the operation of the control mechanism that gets left out of the analysis.

We count energy exchanges with the reservoirs as heat (to be counted as positive if A gains energy from the reservoir, negative if A loses energy), and energy changes to A due to changes in the external potentials as work (again, counted as positive if A gains energy, negative if it loses energy).

There is no restriction whatsoever on the Hamiltonian $H(t)$, as the only fact about Hamiltonian evolution that will be invoked is conservation of Liouville measure (in the classical context), or conservation of inner product (in the quantum context). Thus, the theorems we will prove will apply even to hypothetical cases involving more fine-grained control of the system's evolution than would be feasible in practice. It is also not assumed that the heat reservoirs have a canonical distribution or any other equilibrium distribution *after* they have interacted with the system A, though it is assumed that, if the system A is to interact again with a reservoir after having once interacted with it, enough time has elapsed for thermalization to occur in the reservoir, so that it may be treated as canonically distributed. We are taking this condition as a necessary condition for exchanges of energy between A and the reservoir to count as exchanges of *heat*. (Without some distinction between heat exchange and work, neither the second law nor any other law of thermodynamics can be formulated.)

Dropping the assumption of the availability of reversible processes requires revision of the familiar framework of thermodynamics, as it means dropping the assumption of the availability of an entropy function. In its place we will define quantities $S_{\mathcal{M}}(a \to b)$, defined relative to a class of available manipulations $\mathcal{M}$, to be thought of as analogues, in the current context, of entropy differences between states a and b. These will be representable as differences in the values of some state function only in the limiting case in which all states can be connected reversibly.

For any manipulation M that takes a state a to a state b, we define $\langle Q_i(a \to b)\rangle_M$ as the expectation value of the heat obtained by A from reservoir B_i. We use these to define,

$$\sigma_M(a \to b) = \sum_i \frac{\langle Q_i(a \to b)\rangle_M}{T_i}. \tag{2.3}$$

Let $\mathcal{M}(a \to b)$ be the set of manipulations in $\mathcal{M}$ that take a to b, and define, as an analogue of the entropy difference between a and b,[1]

$$S_{\mathcal{M}}(a \to b) = \text{l.u.b.}\{\, \sigma_M(a \to b),\ M \in \mathcal{M}(a \to b)\,\}. \tag{2.4}$$

Via the obvious extension of this definition we also define quantities such as $S_{\mathcal{M}}(a \to b \to c)$ for processes with any number of intermediate steps. It is assumed that manipulations can be composed, that is, that any manipulation that takes a to b can be followed by one that takes b

[1] Here, "l.u.b" means *least upper bound*, and, in eq. (2.10), "g.l.b." means *greatest lower bound.*

to c to form a manipulation that takes a to b and then to c. It follows from this composition assumption and the definition of the entropies that

$$S_{\mathcal{M}}(a \to b \to c) = S_{\mathcal{M}}(a \to b) + S_{\mathcal{M}}(b \to c), \tag{2.5}$$

and similarly for processes consisting of longer chains of intermediate states.

One version of the second law of thermodynamics says that, for any cyclic process, the sum of Q_i/T_i over all heat exchanges cannot be positive. Since we're working in the context of statistical mechanics and we do not want to ignore fluctuations, the appropriate revision of the second law involves expectation values of heat exchanges. A cyclic process will be one that restores the marginal probability distribution of the system A to the one it started out with. The revised second law that we will prove in the next section states that, for any cyclic process, the sum of $\langle Q_i \rangle/T_i$ over all heat exchanges cannot be positive. In the notation we have introduced, this is:

The statistical second law. For any state a, $S_{\mathcal{M}}(a \to a) \le 0$.

It follows from this that, for any states a, b,

$$S_{\mathcal{M}}(a \to b \to a) = S_{\mathcal{M}}(a \to b) + S_{\mathcal{M}}(b \to a) \le 0, \tag{2.6}$$

and similarly for processes involving longer chains of intermediate states.

In any process M that takes a state a to a state b, some of the work done, or heat discarded into a reservoir, might be recoverable by some process that takes b back to a. If the process can be reversed with the signs of all $\langle Q_i \rangle$ reversed, then full recovery (on average) is possible. If full recovery is not possible, and cannot even be approached arbitrarily closely, we will say that the process is *dissipatory*.

From the statistical second law it follows that, for any manipulations M, M',

$$\sigma_M(a \to b) + \sigma_{M'}(b \to a) \le 0. \tag{2.7}$$

Given a manipulation M that takes a to b, a manipulation M' that recovers, on average, work done and heat discarded in the course of manipulation M would be one that saturates this upper bound; that is, it would be one such that

$$\sigma_M(a \to b) + \sigma_{M'}(b \to a) = 0. \tag{2.8}$$

For a given manipulation M, there might be a bound on how closely we can approximate complete recovery. This would mean that there is an upper bound less than zero to $\sigma_M(a \to b) + \sigma_{M'}(b \to a)$, taken over all M' in $\mathcal{M}(b \to a)$. Define the *dissipation* $\delta_M(a \to b)$ associated with M as the absolute value of this bound. That is, for all $M' \in \mathcal{M}(b \to a)$,

$$\sigma_M(a \to b) + \sigma_{M'}(b \to a) \le -\delta_M(a \to b), \tag{2.9}$$

and $\delta_M(a \to b)$ is the largest quantity for which this is true. More compactly,

$$\begin{aligned} \delta_M(a \to b) &= \text{g.l.b.}\{\,|\sigma_M(a \to b) + \sigma_{M'}(b \to a)|\,,\ M' \in \mathcal{M}(b \to a)\,\} \\ &= -\left(\sigma_M(a \to b) + S_{\mathcal{M}}(b \to a)\right). \end{aligned} \tag{2.10}$$

The dissipation $\delta_M(a \to b)$ is an indicator of the extent of departure from reversibility of the process of going from a to b via manipulation M. Define

$$D_{\mathcal{M}}(a \to b) = \text{g.l.b.}\{\,\delta_M(a \to b),\ M \in \mathcal{M}(a \to b)\,\} = -S_{\mathcal{M}}(a \to b \to a). \tag{2.11}$$

This is the minimal dissipation incurred in any manipulation that takes a to b.

Recall that, from the statistical second law, $S_{\mathcal{M}}(a \rightarrow b \rightarrow a) \leq 0$. The quantity $D_{\mathcal{M}}(a \rightarrow b)$ is, therefore, always nonnegative. If we have

$$D_{\mathcal{M}}(a \rightarrow b) = 0, \tag{2.12}$$

this means that there is no limit to how much the dissipation associated with processes that connect a to b can be diminished. When this holds, it is traditional to say that a and b can be connected reversibly, and to imagine a fictitious process that can proceed in either direction, reversing the signs of all heat exchanges. There is no harm in doing so, as long as this is not taken too literally.[2] Following convention, we will say, for any a, b for which (2.12) is satisfied, that a and b can be connected reversibly. When this locution is used, bear in mind that it is shorthand for (2.12), and does not presume the existence of an actual reversible process.

From the statistical second law it follows that, if all states can be connected reversibly — that is, if, for all a, b, $D_{\mathcal{M}}(a \rightarrow b) = 0$ — then there exists a state function $S_{\mathcal{M}}$, defined up to an additive constant, such that

$$S_{\mathcal{M}}(a \rightarrow b) = S_{\mathcal{M}}(b) - S_{\mathcal{M}}(a). \tag{2.13}$$

This is the familiar entropy function. The reason we have been expressing things in an unfamiliar way is that we *don't* want to assume reversibility as a general rule.

Any manipulation that takes a to b must have dissipation of at least $D_{\mathcal{M}}(a \rightarrow b)$. Define the *inefficiency* associated with a manipulation that takes a to b as the amount by which its dissipation exceeds this minimal value.

$$\begin{aligned} \eta_M(a \rightarrow b) &= \delta_M(a \rightarrow b) - D_{\mathcal{M}}(a \rightarrow b) \\ &= S_{\mathcal{M}}(a \rightarrow b) - \sigma_M(a \rightarrow b). \end{aligned} \tag{2.14}$$

If a and b can be connected reversibly, the distinction between dissipation and inefficiency vanishes, and the inefficiency associated with a process that takes a to b is equal to the dissipation associated with it.

It might happen that a probability distribution encodes details about the state that are irrelevant to the results of subsequent manipulations. As an example, consider a gas that is initially confined to one side of a container by a partition. The partition is removed, and the gas allowed to diffuse throughout the container while remaining isolated from its environment. Two probability distributions over initial conditions with disjoint supports, corresponding to the gas being in different sides of the container initially, evolve into distributions with disjoint supports. The usual sorts of manipulations, however, will be insufficient to distinguish them, and hence the sort of detailed knowledge of the state that stems from knowledge about the initial state will be irrelevant to possibilities of extracting work from the gas.

With these considerations in mind, we say that two statistical mechanical states $b = (\rho, H)$, $b' = (\rho', H)$, with the same values of manipulable variables but differing probability distributions, are *thermodynamically equivalent* with respect to a class $\mathcal{M}$ of manipulations if and only if, for any manipulation M in $\mathcal{M}$, the expectation values for work, $\langle W \rangle_M$, and heat exchanges $\langle Q_i \rangle_M$ are the same for both states. It follows from this definition that, if two states b, b' are thermodynamically equivalent with respect to $\mathcal{M}$, then, for any state a, $S_{\mathcal{M}}(b \rightarrow a) = S_{\mathcal{M}}(b' \rightarrow a)$.

For a device to implement a logical operation L, which maps inputs $\{\alpha_i\}$ to outputs $\{\beta_i = L(\alpha_i)\}$, there must be a conventional association of logical states with sets of physical states. Distinct logical states are to be represented by distinguishable states, which in the classical

[2] As Norton ([2016]) has argued, taking talk of irreversible processes too literally can lead to contradictions.

context means probability distributions with non-overlapping support, and in the quantum-mechanical context, orthogonal density operators. For any logical state γ, let $[\gamma]$ be the corresponding set of physical states. An implementation of a logical operation L that maps inputs $\{\alpha_i\}$ to outputs $\{\beta_i = L(\alpha_i)\}$ is a manipulation M_L that maps each physical state a_i in the class $[\alpha_i]$ to a physical state in $[\beta_i]$. A logically irreversible operation maps two or more inputs $\{\alpha_i\}$ to the same output β. For such inputs, an implementation of L must map each state a_i in $[\alpha_i]$ to a state b_i in $[\beta]$.

We ask: can the manipulation M_L do this without dissipation? That is, can we have $\delta_{M_L}(a_i \rightarrow b_i)$ equal to zero, for every a_i? Failing that, can we, by appropriate choice of manipulation, make every element of the set $\{\delta_{M_L}(a_i \rightarrow b_i)\}$ arbitrarily small?

The question is a bit subtle. The concept of dissipation is defined with respect to a class $\mathcal{M}$ of manipulations. The concept of implementation of a logical operation is defined with respect to a conventional association of logical states with sets of physical states. Landauer's principle says that there is dissipation associated with loss of distinguishability. The sense of distinguishability relevant to this context is distinguishability by operations within the set $\mathcal{M}$ of manipulations used to defined thermodynamic concepts. The reason that the question is a bit subtle is that the operation of a computing device might lump together into a single logical state physical states that are counted as thermodynamically distinct with respect to the set of manipulations one is using to define thermodynamic concepts.

We will say that a manipulation M is *logically irreversible* with respect to a class $\mathcal{M}$ of manipulations if there are two or more thermodynamically distinct states $\{a_i\}$ that get mapped into states $\{b_i\}$ that are thermodynamically equivalent with respect to $\mathcal{M}$. On this definition, it can be proven that there is dissipation associated with logical irreversibility. In the next section we will prove the following.

> **Landauer bound on dissipations** Suppose a manipulation M takes a distinguishable set of states $\{a_i,\ i = 1, \ldots, n\}$ to a set of states $\{b_i,\ i = 1, \ldots, n\}$, which are thermodynamically equivalent with respect to a set $\mathcal{M}$ of manipulations. Then
>
> $$\sum_{i=1}^{n} e^{-\delta_M(a_i \rightarrow b_i)/k} \leq 1.$$

The Landauer bound on dissipations entails that every member of the set $\{\delta_M(a_i \rightarrow b_i)\}$ is greater than zero. As proven in Appendix B, it also entails a formulation that is often presented as a gloss of Landauer's principle, that the mean of the set is not smaller than $k \log n$.[3]

$$\frac{1}{n}\sum_{i=1}^{n} \delta_M(a_i \rightarrow b_i) \geq k \log n. \tag{2.15}$$

That is, there is an *average* dissipation, taken over members of the set $\{a_i\}$, of at least $k \log n$. The Landauer bound means that, though we might be able to reduce the dissipation associated with any particular member of the set as much as we like, we cannot simultaneously make all of the dissipations arbitrarily small. For the case of $n = 2$, the most commonly discussed case, the constraint is graphed in Figure 1. The shaded region is the set of permitted pairs $(\delta_1, \delta_2) = (\delta_M(a_1 \rightarrow b_1)/k, \delta_M(a_2 \rightarrow b_2)/k)$.

The dissipations and the relation of thermodynamic equivalence that are invoked in the Landauer bound are defined with respect to a class of manipulations. If a computing device employs

[3] In this article, all logarithms are natural logarithms, that is, logarithms to the base e.

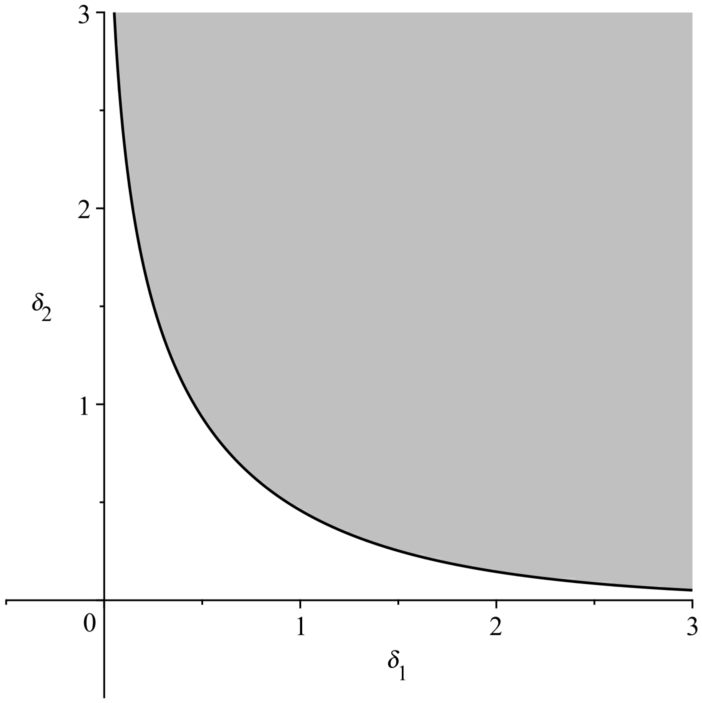

Figure 1: Values of (δ_1, δ_2) permitted by Landauer's principle.

a class $\mathcal{M}$ of manipulations in its operation, which do not distinguish between physical states $\{b_i\}$ lumped together into a single logical state of the device, there might be a wider class $\mathcal{M}^+$ of manipulations that do distinguish the output states $\{b_i\}$. We might then have logically irreversibility with respect to $\mathcal{M}$ without dissipation with respect to $\mathcal{M}^+$. What we cannot have is logical irreversibility with respect to $\mathcal{M}$ without dissipation with respect to the same class $\mathcal{M}$ of manipulations. This is an illustration of the dependence of the concept of dissipation on the class of manipulations considered; see Myrvold ([2020], §7) for further discussion.

If, as is usually assumed in these discussions, the states $\{a_i\}$ can be connected reversibly to the output states $\{b_i\}$, then any dissipation is inefficiency, and bounds on dissipations are bounds on inefficiencies. If reversibility is not assumed, there may be unavoidable levels of dissipation associated with some state transitions; if this is the case, not every dissipation represents an inefficiency. We can re-state the Landauer principle in terms of inefficiencies.

> **Landauer bound on inefficiencies.** If manipulation M takes a distinguishable set of states $\{a_i,\ i = 1, \ldots, n\}$ to thermodynamically equivalent states $\{b_i\}$, then
>
> $$\sum_{i=1}^{n} e^{-(\eta_i + D_M(a_i \to b_i))/k} \leq 1,$$
>
> where η_i is the inefficiency $\eta_M(a_i \to b_i)$.

If we have reversibility, the Landauer bound entails that all of the inefficiencies $\eta_M(a_i \to b_i)$ must be positive, and that they cannot all be made arbitrarily small in the same process. Far enough from reversibility, it places no constraint on inefficiencies at all. The condition for the Landauer bound to place a constraint on inefficiencies is,

$$\sum_{i=1}^{n} e^{-D_{\mathcal{M}}(a_i \to b_i)/k} > 1. \tag{2.16}$$

A necessary condition for (2.16) to be satisfied, and thus for the Landauer principle to have teeth, is the condition that, for some a_i,

$$D_{\mathcal{M}}(a_i \to b_i) < k \log n. \tag{2.17}$$

If there is a sufficiently large in-principle bound on the minimum dissipation required for carrying out processes at the molecular level, then (2.17) is *not* satisfiable. Any process would then depart from reversibility by an amount that exceeds the Landauer bound. In section 4 it will be argued that this is not correct, and the Landauer principle does have teeth.

The Landauer bound we have stated involves a distinguishable set of states. Distinguishability, like reversibility, is something that we should not expect to hold perfectly; in actual implementations it will be approximate at best. For this reason, the theorem that we will prove in the next section will not require perfect distinguishability, and will entail the version of the Landauer bound we have stated in this section as a special case.

3 Proving the statistical second law, and Landauer's principle

The theorems we will be concerned with come in two versions, classical and quantum, each proven in pretty much the same way. To avoid saying everything twice, we adopt a systematically ambiguous notation, and state each theorem in such a way that it can be read either as a theorem of classical statistical mechanics, or as a theorem of quantum statistical mechanics.

In what follows, ρ will be used either for a density function, with respect to Liouville measure, on a classical phase space, or, in the quantum context, a density operator on a Hilbert space. $S\,[\rho]$ is the Gibbs entropy (classical), or the von Neumann entropy (quantum).

$$S\,[\rho] = -k\,\langle \log \rho \rangle_\rho. \tag{3.1}$$

We also define the relative entropy of two distributions.

$$S\,[\rho\,||\,\sigma] = -k\Big(\langle \log \sigma \rangle_\rho - \langle \log \rho \rangle_\rho\Big). \tag{3.2}$$

$S\,[\rho\,||\,\sigma]$ is one way to measure how much the distribution represented by σ departs from that represented by ρ. It is equal to zero if σ and ρ represent the same distribution, and is positive otherwise.

Suppose $\bar{a}$ is a probabilistic mixture of states $\{a_i\}$.

$$\rho_{\bar{a}} = \sum_{i=1}^{n} p_i\, \rho_{a_i}, \tag{3.3}$$

where $\{p_i\}$ are positive numbers that add up to one. Then the Gibbs/von Neumann entropy of $\bar{a}$ is related to that of the a_i's via,

$$S\,[\rho_{\bar{a}}] = \sum_{i=1}^{n} p_i\, S\,[\rho_{a_i}] + \sum_{i=1}^{n} p_i\, S\,[\rho_{a_i}\,||\,\rho_{\bar{a}}]. \tag{3.4}$$

If the states $\{a_i\}$ are distinguishable, then $S[\rho_{a_i} \| \rho_{\bar{a}}] = -k \log p_i$, and so

$$S[\rho_{\bar{a}}] = \sum_{i=1}^{n} p_i \, S[\rho_{a_i}] - k \sum_{i=1}^{n} p_i \log p_i. \tag{3.5}$$

As outlined in the previous section, we are concerned with a system A evolving between times t_0 and t_1 according to a time-varying Hamiltonian, and interacting successively with one or more heat reservoirs $\{B_i\}$, which initially have canonical distributions at temperatures T_i. The Hamiltonians of the heat reservoirs remain fixed throughout the evolution. We define

$$\langle Q_i \rangle = -\Delta \langle H_{B_i} \rangle = -\left(\langle H_{B_i} \rangle_{\rho_{B_i}(t_1)} - \langle H_{B_i} \rangle_{\rho_{B_i}(t_0)} \right). \tag{3.6}$$

This is the expectation value of the heat energy obtained by A from B_i.

Our first theorem relates the entropies as defined in the previous section to the Gibbs/von Neumann entropies. Though a simple one, it is of fundamental importance in the foundations of statistical mechanics, and deserves to be called the *fundamental theorem of statistical thermodynamics*.[4]

Proposition 1 (Fundamental theorem of statistical thermodynamics)
If $\mathcal{M}$ is a class of manipulations of the sort outlined in section 2, then, for any states a, b,

$$S_{\mathcal{M}}(a \to b) \leq S[\rho_b] - S[\rho_a].$$

Proof is given in Appendix A. The following are immediate corollaries of this.

Corollary 1.1 (Statistical second law of thermodynamics)
For any state a,

$$S_{\mathcal{M}}(a \to a) \leq 0.$$

Corollary 1.2
If a and b can be connected reversibly—that is, if

$$S_{\mathcal{M}}(a \to b \to a) = 0,$$

then

$$S_{\mathcal{M}}(a \to b) = S[\rho_b] - S[\rho_a].$$

Thus, the Gibbs/von Neumann entropy is the state function whose existence is guaranteed by the second law plus reversibility.

Now, to the Landauer principle. Suppose a manipulation M takes a sets of states $\{a_i, i = 1, \ldots, n\}$ to states $\{b_i\}$. Let $\bar{a}$ be a probabilistic mixture of the states $\{a_i\}$ with weights $\{p_i\}$, and let $\bar{b}$ be a mixture of the states $\{b_i\}$, with the same weights. Since the manipulation M takes each a_i to b_i, it takes $\bar{a}$ to $\bar{b}$. The expectation value of heat exchanges when M is applied to this mixture is a weighted average of exchanges associated with the states $\{a_i\}$, and so

$$\sigma_M(\bar{a} \to \bar{b}) = \sum_{i=1}^{n} p_i \, \sigma_M(a_i \to b_i). \tag{3.7}$$

[4] 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*.

We must have, of course,

$$\sigma_M(\bar{a} \to \bar{b}) \leq S_M(\bar{a} \to \bar{b}). \tag{3.8}$$

This gives us,

$$\sum_{i=1}^{n} p_i \, \sigma_M(a_i \to b_i) \leq S_M(\bar{a} \to \bar{b}). \tag{3.9}$$

Adding $\sum_i p_i \, S_M(b_i \to a_i)$ to both sides yields,

$$\sum_{i=1}^{n} p_i \, (\sigma_M(a_i \to b_i) + S_M(b_i \to a_i)) \;\leq\; S_M(\bar{a} \to \bar{b}) + \sum_{i=1}^{n} p_i \, S_M(b_i \to a_i). \tag{3.10}$$

Recalling the definition (2.10) of dissipations, this is,

$$-\sum_{i=1}^{n} p_i \, \delta_M(a_i \to b_i) \;\leq\; S_M(\bar{a} \to \bar{b}) + \sum_{i=1}^{n} p_i \, S_M(b_i \to a_i), \tag{3.11}$$

or,

$$\sum_{i=1}^{n} p_i \, \delta_M(a_i \to b_i) \;\geq\; -S_M(\bar{a} \to \bar{b}) - \sum_{i=1}^{n} p_i \, S_M(b_i \to a_i). \tag{3.12}$$

We can re-write this as,

$$\sum_{i=1}^{n} p_i \delta_M(a_i \to b_i) \;\geq\; -\left(S_M(\bar{a} \to \bar{b}) + \sum_{i=1}^{n} p_i \, S_M(\bar{b} \to a_i)\right) - \sum_{i=1}^{n} p_i \left(S_M(b_i \to a_i) - S_M(\bar{b} \to a_i)\right). \tag{3.13}$$

Applying the Fundamental Theorem twice and invoking (3.4) gives us,

$$S_M(\bar{a} \to \bar{b}) + \sum_{i=1}^{n} p_i \, S_M(\bar{b} \to a_i) \;\leq\; \sum_{i=1}^{n} p_i \, S\,[\rho_{a_i}] - S\,[\rho_{\bar{a}}] = -\sum_{i=1}^{n} p_i \, S\,[\rho_{a_i} \,||\, \rho_{\bar{a}}]. \tag{3.14}$$

Plugging this into (3.13) yields,

$$\sum_{i=1}^{n} p_i \, \delta_M(a_i \to b_i) \;\geq\; \sum_{i=1}^{n} p_i \, S\,[\rho_{a_i} || \rho_{\bar{a}}] - \sum_{i=1}^{n} p_i \, \left(S_M(b_i \to a_i) - S_M(\bar{b} \to a_i)\right). \tag{3.15}$$

Equation (3.15) is completely general, and holds for any states whatsoever. If, now, the states $\{b_i\}$ are thermodynamically equivalent, then, for all i,

$$S_M(b_i \to a_i) = S_M(\bar{b} \to a_i), \tag{3.16}$$

and so (3.15) becomes,

$$\sum_{i=1}^{n} p_i \, \delta_M(a_i \to b_i) \;\geq\; \sum_{i=1}^{n} p_i \, S\,[\rho_{a_i} || \rho_{\bar{a}}]. \tag{3.17}$$

Thus, we have,

Proposition 2 (Landauer bound on dissipations)
For any manipulation M that takes states $\{a_i, \; i = 1, \ldots, n\}$ to states $\{b_i\}$ that are thermodynamically equivalent to each other, and any positive numbers $\{p_i\}$ such that

$$\sum_{i=1}^{n} p_i = 1,$$

we have

$$\sum_{i=1}^{n} p_i \, \delta_M(a_i \to b_i) \geq \sum_{i=1}^{n} p_i \, S\,[\rho_{a_i} \,||\, \rho_{\bar{a}}],$$

where $\bar{a}$ is a mixtures of the states $\{a_i\}$, with weights $\{p_i\}$.

As an immediate corollary we get the special case in which the manipulation M takes all of the input states to the same state.

Corollary 2.1
For any manipulation M that takes states $\{a_i, \; i = 1, \ldots, n\}$ to the same state b, and any positive numbers $\{p_i\}$ such that

$$\sum_{i=1}^{n} p_i = 1,$$

we have

$$\sum_{i=1}^{n} p_i \, \delta_M(a_i \to b) \geq \sum_{i=1}^{n} p_i \, S\,[\rho_{a_i} \,||\, \rho_{\bar{a}}],$$

where $\bar{a}$ is a mixtures of the states $\{a_i\}$, with weights $\{p_i\}$.

If we apply our result to the case in which the states $\{a_i\}$ are distinguishable, we get the following corollary.

Corollary 2.2
For any manipulation M that takes each of a distinguishable set of states $\{a_i, \; i = 1, \ldots, n\}$ to states $\{b_i\}$ that are thermodynamically equivalent to each other, and any positive numbers $\{p_i\}$ such that

$$\sum_{i=1}^{n} p_i = 1,$$

we have

$$\sum_{i=1}^{n} p_i \, \delta_M(a_i \to b_i) \geq -k \sum_{i=1}^{n} p_i \log p_i.$$

As shown in Appendix B, this is equivalent to the following.

Corollary 2.3
For any manipulation M that takes each of a distinguishable set of states $\{a_i, \; i = 1, \ldots, n\}$ to states $\{b_i\}$ that are thermodynamically equivalent to each other,

$$\sum_{i=1}^{n} e^{-\delta_M(a_i \to b_i)/k} \leq 1.$$

This is the version stated in the previous section.

The Landauer principle has a generalization to the case in which thermodynamic equivalence of the output states is not exact, but approximate.

Proposition 3
For any manipulation M that takes states $\{a_i, \; i = 1, \ldots, n\}$ to states $\{b_i\}$, and any positive numbers $\{p_i\}$ such that

$$\sum_{i=1}^{n} p_i = 1,$$

we have

$$\sum_{i=1}^{n} p_i\, \delta_M(a_i \to b_i) \geq \sum_{i=1}^{n} p_i\, S\left[\rho_{a_i} \,\|\, \rho_{\bar{a}}\right] - \sum_{i=1}^{n} p_i \Big(S_{\mathcal{M}}(b_i \to a_i) - S_{\mathcal{M}}(\bar{b} \to a_i)\Big),$$

where $\bar{a}$ and $\bar{b}$ are mixtures of the states $\{a_i\}$ and $\{b_i\}$, respectively, with weights $\{p_i\}$.

We can also get a generalization in terms of the relative entropies $S\left[\rho_{b_i} \,\|\, \rho_{\bar{b}}\right]$. The quantity $\sum_i p_i\, S\left[\rho_{b_i} \,\|\, \rho_{\bar{b}}\right]$ is equal to zero when all of the distributions $\{b_i\}$ are the same, and is positive otherwise, and thus is an indicator of the degree of distinctness of the distributions $\{b_i\}$.

$$\begin{aligned} -S_{\mathcal{M}}(\bar{a} \to \bar{b}) \;&\geq\; S\left[\rho_{\bar{a}}\right] - S\left[\rho_{\bar{b}}\right]; && (3.18) \\ -S_{\mathcal{M}}(b_i \to a_i) \;&\geq\; S\left[\rho_{b_i}\right] - S\left[\rho_{a_i}\right]. && (3.19) \end{aligned}$$

Inserting these into (3.12) yields,

$$\begin{aligned} \sum_{i=1}^{n} p_i\, \delta_M(a_i \to b_i) \;&\geq\; S\left[\rho_{\bar{a}}\right] - S\left[\rho_{\bar{b}}\right] + \sum_{i=1}^{n} p_i \left(S\left[\rho_{b_i}\right] - S\left[\rho_{a_i}\right]\right) \\ &=\; S\left[\rho_{\bar{a}}\right] - \sum_{i=1}^{n} p_i\, S\left[\rho_{a_i}\right] - \left(S\left[\rho_{\bar{b}}\right] - \sum_{i=1}^{n} p_i\, S\left[\rho_{b_i}\right]\right). \end{aligned} \tag{3.20}$$

Using (3.4), this gives us,

$$\sum_{i=1}^{n} p_i\, \delta_M(a_i \to b_i) \geq \sum_{i=1}^{n} p_i \left(S\left[\rho_{a_i} \,\|\, \rho_{\bar{a}}\right] - S\left[\rho_{b_i} \,\|\, \rho_{\bar{b}}\right]\right), \tag{3.21}$$

This gives us a version of the Landauer principle in terms of reduction of distinguishability of the sets of states $\{a_i\}$ and $\{b_i\}$.

Proposition 4
For any manipulation M that effects state transitions $a_i \to b_i$, and any positive numbers $\{p_i\}$ such that

$$\sum_{i=1}^{n} p_i = 1,$$

we have

$$\sum_{i=1}^{n} p_i\, \delta_M(a_i \to b_i) \geq \sum_{i=1}^{n} p_i \left(S\left[\rho_{a_i} \,\|\, \rho_{\bar{a}}\right] - S\left[\rho_{b_i} \,\|\, \rho_{\bar{b}}\right]\right),$$

where $\bar{a}$ and $\bar{b}$ are mixtures of the states $\{a_i\}$ and $\{b_i\}$, respectively, with weights $\{p_i\}$.

4 Approximating reversibility

The second law of statistical thermodynamics entails that, for any a, b,

$$S_{\mathcal{M}}(a \to b \to a) \leq 0. \tag{4.1}$$

We do not expect there to be any process that takes a to b and then back to a without any dissipation. However, if the array of permitted manipulations is sufficiently rich, there might be no bound on dissipation short of zero, and we may have $S_{\mathcal{M}}(a \to b \to a) = 0$.

One way to have a process that proceeds with negligibly small dissipation is to keep the system A in contact with a heat reservoir large and noisy enough that the reservoir may be regarded as canonically distributed throughout the process, and to vary the parameters $\boldsymbol{\lambda}$ slowly enough that the time it takes for any appreciable change in these parameters is long compared to the equilibration time-scale of the system A. Then the system A may be treated as if it is in equilibrium with the reservoir at each stage of the process.[5] We can also consider slowly varying the temperature of the reservoir. For a process like that, at any time t during the process, A may be treated as having a canonical distribution for the instantaneous parameter values $(\boldsymbol{\lambda}(t), \beta(t))$.

If ρ_1 is a canonical distribution for parameters $(\boldsymbol{\lambda}, \beta)$, and ρ_2 a canonical distribution for slightly differing parameters $(\boldsymbol{\lambda} + d\boldsymbol{\lambda}, \beta + d\beta)$, then, to first order in the parameter differences,[6]

$$d\langle H\rangle = \langle H_2\rangle_{\rho_2} - \langle H_1\rangle_{\rho_1} = \sum_i \left\langle \frac{\partial H}{\partial \lambda_i}\right\rangle_{\rho_1} d\lambda_i - \beta^{-1}\, d\langle \log \rho\rangle. \tag{4.2}$$

The first term on the right-hand side of this equation is the expectation value of the work done by changing the external parameters; the remainder is the expectation value of the heat obtained from the reservoir.

$$\langle đQ\rangle = -kT\, d\langle \log \rho\rangle, \tag{4.3}$$

where $kT = \beta^{-1}$. This means that, for a process in the course of which the system A is in continual contact with a heat reservoir at temperature T and the parameters $\boldsymbol{\lambda}$ are varied slowly from values $\boldsymbol{\lambda}_a$ to $\boldsymbol{\lambda}_b$, the expectation value of total heat absorbed will have the approximate value

$$\langle Q(a \to b)\rangle \approx -kT(\langle \log \rho_b\rangle_{\rho_b} - \langle \log \rho_a\rangle_{\rho_a}) = T\left(S\left[\rho_b\right] - S\left[\rho_a\right]\right). \tag{4.4}$$

As long as there is no in-principle limit to how much time a state-transformation may take, there is no in-principle limit to how closely this approximation can hold, and equality will be approached as the time-scale of the changes in the parameters $\boldsymbol{\lambda}$ is increased, relative to the time-scale of equilibration of the system A.

The result (4.4) is a result about expectation values. It is *not* assumed that the actual value of heat exchanged will be close to its expectation value, or even that it will *probably* be close to its expectation value. The probability distribution for the heat exchange may have a large variance, and probabilities of large deviations from the expectation value may be far from negligible. That is, the result does *not* depend on disregard of fluctuations. When we say that the system has time to equilibrate, this does not mean that it is ever in a quiescent state, only that its distribution may be treated as canonical at each stage of the process.

Let a, b be canonical states with parameters $(\boldsymbol{\lambda}_a, \beta_a)$, $(\boldsymbol{\lambda}_b, \beta_b)$. We will say that a class of manipulations $\mathcal{M}$ connects a and b quasi-statically if

1. $\mathcal{M}$ contains manipulations of the following form
   (a) During time interval $[t_0, t_0 + T]$, the parameters undergo smooth evolution $\boldsymbol{\lambda}(t)$, with $\boldsymbol{\lambda}(t_0) = \boldsymbol{\lambda}_a$ and $\boldsymbol{\lambda}(t_0 + T) = \boldsymbol{\lambda}_b$.
   (b) At time t the system A is in thermal contact with a heat reservoir at inverse temperature $\beta(t)$, where $\beta(t)$ is a smooth function with $\beta(t_0) = \beta_a$ and $\beta(t_0 + T) = \beta_b$.

[5] This does not, of course, mean that it *is* in equilibrium, only that, for the purposes at hand, differences between quantities calculated on the basis of the equilibrium distribution and quantities calculated on the basis of the actual distribution are small enough that they may be neglected.

[6] The classical version of this is eq. (112) on p. 44 of Gibbs ([1902]), and the quantum, eq. (121.8) on p. 534 of Tolman ([1938]).

2. For any such manipulation, there is one that proceeds twice as slowly. That is, there is a manipulation that takes place in time interval $[t_0, t_0 + 2T]$, with parameter values λ', β', where
$$\lambda'(t_0 + t) = \lambda(t_0 + t/2); \quad \beta'(t_0 + t) = \beta(t_0 + t/2)$$
for $t \in [0, 2T]$.

Then we have the following result.

Proposition 5
If a, b are canonical states, and $\mathcal{M}$ is a class of manipulations that connects a to b quasi-statically, then
$$S_{\mathcal{M}}(a \to b) = S[\rho_b] - S[\rho_a].$$

We have, as a trivial corollary,

Corollary 5.1
If a, b are canonical states, and $\mathcal{M}$ is a class of manipulations that connects a to b quasi-statically, and also connects b to a quasi-statically, then
$$D_{\mathcal{M}}(a \to b) = 0.$$

Suppose that we have a system to which can be applied a manipulable external potential V_λ, and which can also be confined, by suitable barriers, to various regions $\{\Gamma_i\}$ of its state space. Let $\{a_i\}$ be a finite set of canonical states, confined to the regions $\{\Gamma_i\}$, with values $\boldsymbol{\lambda}_a$ of the manipulable parameters $\boldsymbol{\lambda}$ on which the external potential depends, and let $\{b_i\}$ be a set of canonical distributions confined to the same regions, with parameter values $\boldsymbol{\lambda}_b$. Then, for any desired degree of approximation to the quasi-static limit, we can find a sufficiently slow variation of the parameters $\boldsymbol{\lambda}$ that yields the desired degree of approximation for *all* of the transitions $a_i \to b_i$. We will say, of such a situation, that $\mathcal{M}$ uniformly quasi-statically connects $\{a_i\}$ to $\{b_i\}$. We have, as another corollary to Proposition (5):

Corollary 5.2
Let $\{a_i\}$, $\{b_i\}$ be sets of canonical states, such that $\mathcal{M}$ uniformly quasi-statically connects $\{a_i\}$ to $\{b_i\}$ and $\{b_i\}$ to $\{a_i\}$. Let $\{p_i\}$ be a set of non-negative numbers that sum to 1, and let $\bar{a}$ and $\bar{b}$ be probabilistic mixtures of $\{a_i\}$ and $\{b_i\}$ with weights $\{p_i\}$. Then
$$D_{\mathcal{M}}(\bar{a} \to \bar{b}) = 0.$$

5 Example: the one-particle gas

The simplest example for illustrating erasure is that of a single particle in a box, with a partition that can be inserted and removed. If this is the only available manipulation, then we have a rather boring and uninteresting thermodynamical theory. To get a thermodynamically interesting theory, we need to introduce the possibility of doing work on and obtaining work from the system.

Suppose that the particle can be subjected to an external potential V_λ, that varies in the x-direction only. We take the system to be in thermal equilibrium with a heat reservoir at temperature T. On a canonical distribution, the distributions of the momentum $\mathbf{p}$ and the coordinates other than x are unchanged when the potential V_λ is varied. We therefore integrate these out, and consider the marginal distribution of the coordinate x.

$$\rho_{\lambda,\beta}(x) = \begin{cases} Z^{-1}_{\lambda,\beta}\, e^{-\beta\, V_\lambda(x)}, & \text{inside the container;} \\ 0, & \text{outside.} \end{cases} \tag{5.1}$$

Take the x-coordinate within the container to range from $-l$ to l. The partition function is

$$Z_{\lambda,\beta} = \int_{-l}^{l} e^{-\beta\, V_\lambda(x)}\, dx. \tag{5.2}$$

Suppose the force on the particle is constant within the box, and may be varied in both strength and direction. The particle could, for example, be a charged particle, and the applied field an electric field. Then the external potential varies linearly with x. Take it to be

$$V_\lambda(x) = \lambda\, kT\, x/l, \tag{5.3}$$

where λ is a dimensionless parameter.

The analogue of compressing or expanding the one-particle gas is varying the external potential. As λ is increased from zero, the distribution of the particle becomes more and more concentrated towards the left end of the container. We can make the probability that it is to the left of any chosen location as high as we want by taking λ sufficiently large. Similarly, for negative values of λ, the distribution is concentrated towards the right end of the container.

Relative to a canonical distribution with $\lambda = 0$, a distribution for a large value of λ has a large value of free energy, and so we have to do work on the gas while increasing the potential. The work done may be recovered by decreasing the potential back to zero. If the process is done slowly enough that the particle can be treated as canonically distributed at each stage, the expectation value of the work recovered while decreasing the potential is equal to the expectation value of the work done in increasing it: the process is thermodynamically reversible.

Let b be a state in which no partition is present and the applied potential is zero. The probability distribution of the particle is evenly distributed throughout the container. Now insert a partition that divides the container into subvolumes with ratio $p : (1 - p)$. Let $a_1(p)$ be a state in which the particle is to the left of the partition, and let $a_2(p)$ be a state in which the particle is to the right of the partition.

The states $a_1(p)$ and $a_2(p)$ are perfectly distinguishable states. There's a complication, however: given our class of manipulations, we have no way to prepare them, starting from state b. If we start from b and increase the potential, we can make the probability that the particle is to the left of where we intend to drop the partition as high as we like, but it can never be equal to 1.

In place of these states $a_1(p)$ and $a_2(p)$, which are perfectly distinguishable but not preparable using the manipulations considered, we consider a pair of states that are *almost* distinguishable, and are preparable. Let ϵ be a small positive number, and let $a^\epsilon_1(p)$ be a state in which V_λ is zero, and a partition is present, dividing the container into subvolumes with ratio $p : (1 - p)$, and in which there is a probability of $1 - \epsilon$ that the particle is to the left of the partition, and probability ϵ that it is to the right. Define $a^\epsilon_2(p)$ similarly, with the probabilities reversed.

One manipulation that takes $a^\epsilon_1(p)$ to b is removal of the partition, after which the particle equilibrates. This is an inefficient operation, as we could have performed an expansion of the gas, in the course of which work is obtained and heat enters the gas from the reservoir.

To see how much inefficiency, we consider the following process, which is analogous to a controlled expansion of a gas. We start in state $a^\epsilon_1(p)$.

1. We first slowly increase λ to the point at which, on the canonical distribution for V_λ, the particle has probability $1-\epsilon$ of being to the left of the partition, and probability ϵ of being to the right.

2. We remove the partition, allowing the particle to move freely throughout the container. The probability distribution does not change, as the probability, on the equilibrium distribution, of the particle being on the left of the former location of the partition is the same as it was before the partition was removed.[7]

3. The potential is slowly decreased to zero.

The process can be performed in reverse order to create $a_1^\epsilon(p)$ from b. If we have available to us arbitrarily slow processes,

$$S_{\mathcal{M}}(a_1^\epsilon(p) \to b \to a_1^\epsilon(p)) = S_{\mathcal{M}}(a_2^\epsilon(p) \to b \to a_2^\epsilon(p)) = 0. \tag{5.4}$$

The expectation value of heat gained in the process of expansion is, in the quasi-static approximation,

$$\langle Q(a_1^\epsilon(p) \to b)\rangle = T(S\,[\rho_b] - S\,[\rho_{a_1^\epsilon(p)}]) = -kT[(1-\epsilon)\log p + \epsilon \log(1-p) - v(\epsilon)], \tag{5.5}$$

where

$$v(\epsilon) = \epsilon \log \epsilon + (1-\epsilon)\log(1-\epsilon). \tag{5.6}$$

We can make $\langle Q(a_1^\epsilon(p) \to b)\rangle$ as close to $-kT\log p$ as we like by taking ϵ sufficiently small.

Therefore, erasure by removing the partitions has associated with it inefficencies,

$$\begin{aligned} \eta_1 &= -k[(1-\epsilon)\log p + \epsilon\log(1-p) - v(\epsilon)] \approx -k\log p, \\ \eta_2 &= -k[\epsilon\log p + (1-\epsilon)\log(1-p) - v(\epsilon)] \approx -k\log(1-p). \end{aligned} \tag{5.7}$$

Suppose that we want an erasure process that takes both $a_1^\epsilon(p)$ and $a_2^\epsilon(p)$ to the state b. One such process goes by removal of the partition. This has the inefficiencies exhibited in (5.7). But we have only availed ourselves of a fairly limited set of operations. Would it be possible to concoct a different set of operations, which might include the employment of auxiliary systems subject to any sort of Hamiltonian we might dream up, whether or not realization of such Hamiltonians is even remotely feasible, and thereby construct an operation that takes both $a_1^\epsilon(p)$ and $a_2^\epsilon(p)$ to b, with lower inefficiency for both input states than the lossy removal-of-partition operation, which has the inefficiencies (5.7)?

Alas, the answer is negative. As the reader can verify, as long as $\epsilon < p < 1-\epsilon$, the pair of inefficiencies (5.7) saturate the Landauer bound exhibited in Proposition 2.[8] This means that no process, no matter how elaborate, will achieve a lower inefficiency for both input states, so long as all exchanges of heat are with canonically distributed reservoirs, there are at the beginning of the process no dynamically relevant correlations between the state of A and either the auxiliary systems or the reservoirs, the evolution of the total system is Hamiltonian, and at the end of the evolution the auxiliary systems are restored to their initial states.

[7] General rule: if we take state space Γ and partition the space into disjoint regions Γ_i, a canonical distribution ρ defined on Γ is a mixture of canonical distributions ρ_i confined to the regions Γ_i, with weights being the probabilities, on ρ, that the system is in Γ_i.

[8] Because we have reversibility, inefficiencies and dissipations are equal.

6 The LPSG proof vindicated

The LPSG proof proceeds as follows.[9] Suppose we have a manipulation M_L that takes each of a distinguishable set of states $\{a_i, i = 1, \ldots, n\}$ of a device D to a common destination state b. The proof employs as an auxiliary system a one-molecule gas in a box into which partitions may be inserted and removed, and which can be expanded reversibly. LPSG reason that, on pain of violating the statistical second law of thermodynamics, the manipulation M_L must satisfy the Landauer principle. This involves considering the following cycle of operations (performed with both the device D and the gas G in contact with a heat reservoir at temperature T at all times). The starting state is one in which device D is in state b, and there are no partitions in the box.

1. $n - 1$ partitions are inserted into the box, dividing its volume into n subvolumes, with volumes that are fractions p_i of the total volume. With probability p_i, the gas molecule is in the ith subvolume.

2. A controlled operation is performed on D, using the state of the gas G as control. If the gas molecule is in the ith subvolume, b is taken into the state a_i. The heat exchange with the reservoir can be made arbitrarily close to $TS_M(b \to a_i)$.

3. A controlled operation is performed on the gas G, using the state of D as control. The ith subvolume is expanded reversibly, obtaining heat $-kT \log p_i$ from the reservoir. The gas has now been restored to its initial state.

4. The operation M_L is performed, restoring the device D to the state b, with heat transfer $\sigma_{M_L}(a_i \to b)$.

If one works through the expectation values of heat exchanges in the course of this cycle, assuming the statistical second law but not assuming reversibility of the processes $b \to a_i$, then what is obtained is precisely our Corollary 2.2 of section 3. Obviously, if one replaces the assumption that heat $-kT \log p_i$ can be obtained in step 3 with the assumption that there are operations such that the expectation value of heat obtained can come arbitrarily close to $-kT \log p_i$, the result still obtains.

The point of contention is whether expansion of a one-molecule gas can be performed in such a way that the expectation value of heat obtained is arbitrarily close to $-kT \log p_i$. Norton, in the works cited, contends that this is false. In my opinion Ladyman and Robertson ([2014]) are right when they say that he has not established this. However, if one has doubts about this being true for a one-molecule gas expanded by a piston, because of lack of control over a sufficiently sensitive piston, our example from the previous section of a one-molecule gas subjected to an external potential may be substituted.

We replace step 3 with the following process. For simplicity we illustrate it for the case of a single partition; extension to multiple partitions is straightforward. Suppose the particle is found to be to the left of the partition. The initial state is $a_1(p)$.

1. Slowly increase λ to a high positive value λ^*.

2. Remove the partition, and allow the system to equilibrate. Some heat is absorbed from the reservoir, but, for large λ^*, this is a small amount.

[9] LPSG present the argument for the case $n = 2$, but the generalization to an arbitrary number of input states is obvious and straightforward.

3. Slowly decrease λ to zero.

If the particle is found to the right of the partition, one takes λ to a large negative value instead. It is not difficult to calculate the expectation value of heat obtained in such a process in the quasi-static limit. The details of this calculation need not concern us; what matters if that it can be made arbitrarily close to $-kT \log p$ by taking λ^* sufficiently large.[10]

7 Demonology

As Landauer's principle is often discussed in connection with the literature on Maxwell's demon, the reader might be wondering what, if any, connection what is done here has with that literature.

A Maxwell demon is meant to produce violations of the second law of thermodynamics. It is useful to distinguish between two sorts of feats that a demon might be imagined to accomplish.

Earman and Norton ([1998]) distinguish between *straight* and *embellished* violations of the second law of thermodynamics. A straight violation decreases the entropy of an adiabatically isolated system, without compensatory increase of entropy elsewhere. An embellished violation exploits such decreases in entropy reliably to provide work. In a similar vein, David Wallace ([2018]) distinguishes between two types of demon. A demon of the first kind decreases a coarse-grained entropy, either a Boltzmann entropy or a coarse-grained Gibbs entropy, of an isolated system. A demon of the second kind violates the Carnot bound on efficiency of a heat engine over a repeatable cycle that restores the state of the demon plus any auxiliary system utilized to its original thermodynamic state.

A demon of the first kind illustrates the dependence of entropy on the class of manipulations considered. A manipulation M outside of a class $\mathcal{M}$ might adiabatically decrease a system's thermodynamic entropy, as defined with respect to $\mathcal{M}$, but it will not decrease the thermodynamic entropy, as defined with respect to a wider class $\mathcal{M}^+$ that includes M, because it follows from the definition of thermodynamic entropy, either the standard textbook definition, which presumes that the thermodynamic states involved can be connected by a reversible process, or the definition adopted here, that the entropy of an isolated system cannot decrease.[11] See Myrvold ([2020], §8) for further discussion of this point.

Boltzmann entropy is another matter. As has been pointed out by Oliver Penrose ([1970], Ch. V), by David Albert ([2000], Ch. 5), and by Meir Hemmo and Orly Shenker ([2012], Ch. 13), if the macroevolution of a system is not predictable—that is, if there is a plurality of macrostates that it may end up in, with nonzero probability, from a given initial macrostate—then it is consistent with Hamiltonian evolution that its Boltzmann entropy will with certainty decrease in the course of isolated evolution. Considerations such as these illustrate the somewhat tenuous nature of the connection between Boltzmann entropy and thermodynamic entropy. While it is true that, for many macroscopic systems subjected to feasible manipulations

[10] For those who are interested, the result is

$$\langle Q \rangle = -kT \log p - kT \log \left(\frac{1 - e^{-2\lambda^*}}{1 - e^{-2p\lambda^*}} \right).$$

For any p, $0 < p < 1$, for large λ^* we have

$$\langle Q \rangle \approx -kT \log p - kT e^{-2p\lambda^*}.$$

Therefore, $\langle Q \rangle$ approaches $-kT \log p$ exponentially with increase of λ^*.

[11] That is, if there is a process that takes a state a to state b with no heat exchange with any reservoir, $S_{\mathcal{M}}(a \to b) \geq 0$. This follows from the definition of $S_{\mathcal{M}}(a \to b)$, and does not depend on the statistical second law.

and measurements, if Boltzmann entropy is defined with respect to a partition corresponding to observationally distinguishable states, differences in Boltzmann entropy will approximate differences in thermodynamic entropy, for other situations, the connection between the two can come apart, as the authors cited demonstrate.

In a similar vein, Hemmo and Shenker ([2012], Ch.12) address the question of whether there must be an increase of Boltzmann entropy associated with a logically irreversible operation. They show that it is consistent with conservation of Liouville measure that one can designate certain degrees of freedom as information-bearing and effect erasure with respect to those degrees of freedom without an increase of total Boltzmann entropy of all systems involved. They count this as a counterexample to Landauer's principle. It is, however, consistent with the theorem proven in this paper, which is concerned, not with Boltzmann entropy, but with dissipations, as defined by (2.10), in terms of expectation values of heat exchanges.

Is a demon of the second kind possible? It follows from the statistical second law that there can be no system that operates in a cycle, exchanging heat with any number of reservoirs, that reliably violates the Carnot bound on efficiency of heat engines. Landauer's principle is not needed to see this. The principle may, however, serve a heuristic role, in analyzing some proposed device that may appear at first sight to violate the statistical second law, as a reminder that the device should operate in a cycle and that dissipations associated with resetting its state to the initial state should not be neglected. This is accepted in at least some of the literature on the thermodynamics of computation. Bennett ([2003]), for example, suggests that, though Landauer's principle is in a sense a "a straightforward consequence or restatement of the Second Law," it nevertheless has considerable pedagogic value.

As is usual in thermodynamics, we have employed a division of the world into the system of interest and the remainder. We have considered systems subjected to external time-varying potentials, without including the sources of those potentials in our analysis. Of course, these sources might have dissipations associated with their operation. Landauer's principle tells us that, in addition to whatever dissipation is occurring outside the system of interest, there is an additional dissipation associated with implementation of operations that are not logically reversible. One of the concerns of the literature on the thermodynamics of computation is whether there is any in-principle minimal heat generation internal to the computing machinery. In such a context, it is entirely appropriate to analyze a system supplied with an external power supply that is itself left out of the analysis.[12]

Norton ([2013c]) has recommended restricting consideration to self-contained processes, internalizing all driving potentials. The thought motivating this seems to be that, if the process considered is not self-contained, one might be able to construct a system that appears to violate the second law if external dissipations are neglected, but can be seen not to do so if all dissipations are taken into account. If this is the motivation, it rests on a false premise. The statistical second law holds for systems subjected to an external potential. It is necessary to internalize any control mechanism responsible for controlled operations, but it is not necessary to internalize the driving potential.

Norton also offers an argument, which he takes to be a 'no-go' theorem for the thermodynamics of computation, to the effect that, if self-contained processes are considered, there is always dissipation associated with any process that far exceeds the Landauer bound. I do not believe that he has established this conclusion, but an analysis of that argument would take us beyond the scope of this paper, and will be left for a sequel.

[12] This point has been made by Ladyman ([2018], p. 235).

8 Conclusion

Landauer's principle, as a statement about dissipations defined, as above, in terms of expectation values of heat exchanges, is a theorem of statistical mechanics. If a manipulation takes each of a set of n distinguishable input states to the same output state, or to output states that are thermodynamically equivalent, it is not possible for the manipulation to be dissipationless for all of the inputs; there must be dissipation, averaged over the set of input states, of at least $k \log n$. Our proof of Landauer's principle does not rely on an assumption of the availability of thermodynamically reversible processes, or even an approximation to them, though, unless the processes involved can be effected with sufficiently small dissipation, the principle places no bounds on *extra* dissipation associated with logical irreversibility. Worries about whether a sufficiently close approximation to thermodynamic reversibility can be achieved in the face of molecular-scale fluctuations can be alleviated. If we define reversibility in terms of reversal of *expectation values* of heat exchanges, reversibility can be approximated as closely as one likes, if the processes proceed slowly enough.

Appendix A Proof of the Fundamental Theorem

To be proven: If $\mathcal{M}$ is a class of manipulations of the sort outlined in section 2, then, for any states a, b,

$$S_{\mathcal{M}}(a \to b) \leq S[\rho_b] - S[\rho_a].$$

We use the following lemmas.

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

$$\langle H \rangle_\rho - TS[\rho].$$

Lemma 2 (Subadditivity)
For a composite system AB,

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

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

Lemma 3
$S[\rho]$ is conserved under Hamiltonian evolution.

We consider some manipulation $M \in \mathcal{M}$ that takes a state a of A at t_0 to a state b at t_1. At time t_0 the composite system consisting of A and $\{B_i\}$ has distribution represented by density $\rho_{tot}(t_0)$. At time t_1 the density is $\rho_{tot}(t_1)$. We will write $S_{tot}(t)$ as an abbreviation for $S[\rho_{tot}(t)]$, and similarly for $S_A(t)$ and $S_{B_i}(t)$.

By Lemma 1 we have, for each reservoir B_i,

$$\langle H_{B_i}(t_0) \rangle - T_i S_{B_i}(t_0) \leq \langle H_{B_i}(t_1) \rangle - T_i S_{B_i}(t_1), \tag{A.1}$$

or,

$$\Delta \langle H_{B_i} \rangle - T_i \Delta S_{B_i} \geq 0. \tag{A.2}$$

Since $\langle Q_i \rangle = -\Delta \langle H_{B_i} \rangle$, this gives

$$\frac{\langle Q_i \rangle}{T_i} \leq -\Delta S_{B_i}. \tag{A.3}$$

Because A is uncorrelated with each B_i at t_0,

$$S_{tot}(t_0) = S_A(t_0) + \sum_{i=1}^{n} S_{B_i}(t_0). \tag{A.4}$$

Because of subadditivity,

$$S_{tot}(t_1) \leq S_A(t_1) + \sum_{i=1}^{n} S_{B_i}(t_1). \tag{A.5}$$

Because Hamiltonian evolution conserves S,

$$S_{tot}(t_1) = S_{tot}(t_0). \tag{A.6}$$

Taken together, (A.4), (A.5), and (A.6) yield,

$$\Delta S_A + \sum_{i=1}^{n} \Delta S_{B_i} \geq 0. \tag{A.7}$$

This, together with (A.3), gives us the result,

$$\sigma_M(a \to b) = \sum_{i=1}^{n} \frac{\langle Q_i \rangle}{T_i} \leq \Delta S_A. \tag{A.8}$$

Since this must hold for every manipulation in the set $\mathcal{M}$, it must hold also for $S_{\mathcal{M}}(a \to b)$, which we defined as the least upper bound of the set of all $\sigma_M(a \to b)$ for $M \in \mathcal{M}$. This gives us the desired result,

$$S_{\mathcal{M}}(a \to b) \leq \Delta S_A. \tag{A.9}$$

Appendix B Proof of equivalence of two formulations.

We wish to prove the following.

Lemma 4
Let $\{x_i, i = 1, \ldots, n\}$ be any sequence of n real numbers. The following are equivalent.

(A) For all positive $\{p_i, i = 1, \ldots, n\}$ such that $\sum_i p_i = 1$,

$$\sum_{i=1}^{n} p_i\, x_i \geq -\sum_{i=1}^{n} p_i \log p_i.$$

(B)

$$\sum_{i=1}^{n} e^{-x_i} \leq 1.$$

To prove this, we prove,

Lemma 5
Let $\{q_i, i = 1, \ldots, n\}$ be any sequence of n positive real numbers. The following are equivalent.

(A) For all positive $\{p_i, i = 1, \ldots, n\}$ such that $\sum_i p_i = 1$,

$$\sum_{i=1}^{n} p_i \,\log\,(p_i/q_i) \geq 0.$$

(B)

$$\sum_{i=1}^{n} q_i \leq 1.$$

From Lemma 5, Lemma 4 follows immediately by taking $q_i = e^{-x_i}$. To prove Lemma 5 we will invoke the log sum inequality.

Lemma 6 (**Log Sum Inequality**)
For any two sequences of positive real numbers $\{p_i, i = 1, \ldots, n\}$, $\{q_i, i = 1, \ldots, n\}$,

$$\sum_{i=1}^{n} p_i \log(p_i/q_i) \geq \left(\sum_{i=1}^{n} p_i\right) \log\left(\sum_{i=1}^{n} p_i \bigg/ \sum_{i=1}^{n} q_i\right).$$

The proof of this can be found in many textbooks of information theory; see, for example, Cover and Thomas ([1991]), Theorem 2.7.1.

We now prove Lemma 5.

Proof that $(A) \Rightarrow (B)$. Let $\{q_i, i = 1, \ldots, n\}$ be a sequence of positive real numbers such that (A) holds. Take

$$p_i = q_i \bigg/ \sum_{j=1}^{n} q_j. \tag{B.1}$$

Then $\sum_i p_i = 1$, and

$$\sum_{i=1}^{n} p_i \log(p_i/q_i) = -\log\left(\sum_{j=1}^{n} q_j\right) \geq 0, \tag{B.2}$$

or,

$$\sum_{i=1}^{n} q_i \leq 0. \tag{B.3}$$

Proof that $(B) \Rightarrow (A)$. Suppose that $\{q_i\}$ is a sequence of positive numbers such that (B) holds. Then, by the log sum inequality, for any sequence $\{p_i\}$ of positive numbers such that $\sum_i p_i = 1$,

$$\sum_{i=1}^{n} p_i \log(p_i/q_i) \geq -\log\left(\sum_{i=1}^{n} q_i\right) \geq 0. \tag{B.4}$$

Acknowledgements

I am grateful to a number of people with whom I have discussed these matters over the years. In particular, I thank Owen Maroney for drawing my attention to what I have called the Fundamental Theorem, David Wallace for discussions of Maxwell's demon, John Norton and Giovanni Valente for discussions of reversible processes, and Katie Robertson for comments on an earlier draft of this article. I thank Sona Ghosh for editorial assistance.

Wayne C. Myrvold
Department of Philosophy
The University of Western Ontario
London, ON, Canada
N6A 5B8
wmyrvold@uwo.ca

References

Albert, D. [2000]: *Time and Chance*, Cambridge: Harvard University Press.

Bennett, C. H. [2003]: 'Notes on Landauer's principle, reversible computation, and Maxwell's Demon', *Studies in History and Philosophy of Modern Physics*, **34**, pp. 501–510.

Cover, T. M. and Thomas, J. A. [1991]: *Elements of Information Theory*, New York: John Wiley & Sons.

Earman, J. and Norton, J. D. [1998]: 'Exorcist XIV: The Wrath of Maxwell's Demon. Part I. From Maxwell to Szilard', *Studies in History and Philosophy of Modern Physics*, **29**, pp. 435–471.

Earman, J. and Norton, J. D. [1999]: 'Exorcist XIV: The Wrath of Maxwell's Demon. Part II. From Szilard to Landauer and Beyond', *Studies in History and Philosophy of Modern Physics*, **30**, pp. 1–40.

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 Shenker, O. R. [2012]: *The Road to Maxwell's Demon: Conceptual Foundations of Statistical Mechanics*, Cambridge: Cambridge University Press.

Hemmo, M. and Shenker, O. [2013]: 'Entropy and Computation: The Landauer-Bennett Thesis Reexamined', *Entropy*, **15**, pp. 3297–331.

Hemmo, M. and Shenker, O. [2019]: 'The physics of implementing logic: Landauer's principle and the multiple-computations theorem', *Studies in History and Philosophy of Modern Physics*, **68**, pp. 90–105.

Ladyman, J. [2018]: 'Intension in the Physics of Computation: Lessons from the Debate about Landauer's Principle', in M. E. Cuffaro and S. C. Fletcher (*eds*), *Physical Perspectives on Computation, Computational Perspectives in Physics*, Cambridge: Cambridge University Press, pp. 219–239.

Ladyman, J., Presnell, S. and Short, A. J. [2008]: 'The use of the information-theoretic entropy in thermodynamics', *Studies in History and Philosophy of Modern Physics*, **39**, pp. 315–324.

Ladyman, J., Presnell, S., Short, A. J. and Groisman, B. [2007]: 'The connection between logical and thermodynamic irreversibility', *Studies in History and Philosophy of Modern Physics*, **38**, pp. 58–79.

Ladyman, J. and Robertson, K. [2013]: 'Landauer defended: Reply to Norton', *Studies in History and Philosophy of Modern Physics*, **44**, pp. 263–271.

Ladyman, J. and Robertson, K. [2014]: 'Going Round in Circles: Landauer *vs.* Norton on the Thermodynamics of Computation', *Entropy*, **16**, pp. 2278–2290.

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

Maroney, O. [2007]: 'The Physical Basis of the Gibbs-von Neumann entropy', arXiv:quant-ph/0701127v2.

Maroney, O. J. E. [2009]: 'Generalizing Landauer's principle', *Physical Review E*, **79**, 031105.

Maxwell, J. C. [1871]: *Theory of Heat*, London: Longmans, Green, and Co.

Maxwell, J. C. [1878]: 'Tait's "Thermodynamics", II', *Nature*, **17**, pp. 278–280.

Myrvold, W. C. [2020]: 'The science of $\Theta\Delta^{cs}$', *Foundations of Physics*, **50**, pp. 1219–1251.

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**, pp. 375–411.

Norton, J. D. [2011]: 'Waiting for Landauer', *Studies in History and Philosophy of Modern Physics*, **42**, pp. 184–198.

Norton, J. D. [2013a]: 'Author's Reply to Landauer Defended', *Studies in History and Philosophy of Modern Physics*, **44**, p. 272.

Norton, J. D. [2013b]: 'The End of the Thermodynamics of Computation: A No-Go Result', *Philosophy of Science*, **80**, pp. 1182–1192.

Norton, J. D. [2013c]: 'All Shook Up: Fluctuations, Maxwell's Demon and the Thermodynamics of Computation', *Entropy*, **15**, pp. 4432–4483.

Norton, J. D. [2016]: 'The Impossible Process: Thermodynamic Reversibility', *Studies in History and Philosophy of Modern Physics*, **55**, pp. 43–61.

Norton, J. D. [2018]: 'Maxwell's Demon Does Not Compute', in M. E. Cuffaro and S. C. Fletcher (*eds*), *Physical Perspectives on Computation, Computational Perspectives in Physics*, Cambridge: Cambridge University Press, pp. 240–256.

Penrose, O. [1970]: *Foundations of Statistical Mechanics: A Deductive Approach*, Oxford: Pergamon Press.

Szilard, L. [1925]: 'Über die Ausdehnung der phänomenologischen Thermodynamik auf die Schwankungserscheinungen', *Zeitschrift für Physik*, **32**, pp. 753–788 English translation in Szilard (1972).

Szilard, L. [1972]: 'On the Extension of Phenomenological Thermodynamics to Fluctuation Phenomena', in B. T. Feld, G. W. Szilard and K. R. Winsor (*eds*), *The Collected Works of Leo Szilard: Scientific Papers*, Cambridge, MA: The MIT Press, pp. 70–102.

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

Wallace, D. [2018]: 'Thermodynamics as control theory', Lecture delivered June 21, 2018, at conference, *Thermodynamics as a Resource Theory: Philosophical and Foundational Implications*, The University of Western Ontario. Available at https://www.youtube.com/watch?v=TnZTIZN2LiQ.