Logic and Entropy

Orly R. Shenker[†]

Abstract

A remarkable thesis prevails in the physics of information, saying that the *logical* properties of operations that are carried out by computers determine their *physical* properties. More specifically, it says that logically irreversible operations are dissipative by $k\log 2$ per bit of lost information. (A function is logically irreversible if its input cannot be recovered from its output. An operation is dissipative if it turns useful forms of energy into useless ones, such as heat energy.) This is Landauer's dissipation thesis, hereafter LDT. LDT underlies and motivates numerous researches in physics and computer science. Nevertheless, this paper shows that is it plainly wrong. This conclusion is based on a detailed study of LDT in terms of the various notions of entropy used in main stream statistical mechanics. It is supported by a counter example for LDT. Further support is found in an analysis of the phase space representation on which LDT relies. This analysis emphasises the constraints placed on the choice of probability distribution by the fact that it has to be the basis for calculating phase averages corresponding to thermodynamic properties of individual systems. An alternative representation is offered, in which logical irreversibility has nothing

[†] Department of Philosophy, Logic and Scientific Method, The London School of Economics and Political Science, Houghton Street, London WC2A 2AE, Britain. O.shenker@lse.ac.uk.

.

to do with dissipation. The strong connection between logic and physics, that LDT implies, is thereby broken off.

1. Introduction

A remarkable thesis in the physics of computation states that the *logical* properties of operations carried out by computers determine some *physical* properties of these processes. The logical property in question is *logical irreversibility*. A logical function is irreversible if it is not 1:1, so that its input cannot be uniquely recovered from its output. The physical property in question is *dissipativity*. A physical process is dissipative if it is entropy increasing, that is, if it takes in energy in a form usable as work and turns it into useless heat energy. That logically irreversible computation is dissipative has been the prevalent opinion ever since Landauer proposed it in (1961). Landauer's Dissipation Thesis (hereinafter: *LDT*) is best summarized in Landauer's (1992, p.2) own words:

> Consider a typical logical process, which discards information, e.g., a logical variable that is reset to *0*, regardless of its initial state. Fig. [1.1, based on Landauer 1992, p.2] shows, symbolically, the phase space of the complete computer considered as a closed system, with its own power source. The erasure process we are considering must map the *1* space down into the *0* space. Now, in a closed conservative system phase space cannot be compressed, hence the reduction in the [horizontal] spread must be compensated by a [vertical] phase space expansion, i.e., a heating of the [vertical] irrelevant degrees of freedom, typically thermal lattice vibrations. Indeed, we are involved here in a process which is similar to adiabatic magnetization (i.e., the inverse of adiabatic demagnetization), and we can expect the same entropy increase to be passed to the thermal background as in adiabatic magnetization, i.e., $k\ln 2$ per erasure process. At this point, it becomes worthwhile to be a little more detailed. Fig. [1.3] shows the end-result of the erasure process in which the original *1* and *0* spaces have both been mapped into the [horizontal] range originally occupied by the *0*. This is, however, rather like the isothermal compression of a gas in a cylinder into half its original volume. The entropy of the gas has been reduced and the surroundings have been heated, but the process is not irreversible, the gas can subsequently be expanded again. Similarly, as long as *1* and *0* occupy distinct phase space regions, as shown in Fig. [1.3], the mapping is reversible. The real irreversibility comes from the fact that the *1* and *0* spaces will subsequently be treated alike and will eventually diffuse into each other.

INSERT FIGURE 1 ABOUT HERE.

Landauer correctly adds that this description does not repeat *all* that is known and understood in the field. Still, the *essence* of LDT is summarized in these words.[1] LDT is truly remarkable for its generality. The dissipation depends only on the logical properties of the operations that are carried out by the physical computer. It is the
same for all computations, regardless of the technology used to carry the computation out.

Landauer's Dissipation Thesis is clearly empirical: it predicts phenomena. Yet, at the present stage of technology, it does *not* express *experimental* facts of any kind, for it has *not yet been empirically tested*. It is *not* experimentally known that logically irreversible computaion is associated with any minimum amount of dissipation.[2] Since LDT has not been empirically tested, it rests on theoretical arguments *only*. Many people find these arguments very intuitive and compelling. Nevertheless, various assumptions underlying them are questionable. In particular, the notion of entropy that they use is, at best, puzzling (sections 5-6). Moreover, the phase space representation of logical operations that they employ is problematic. It does not take proper account of some constraints that ought to be placed on the choice of probability distribution. The probability distribution over phase space is used to calculate phase averages that ought to correspond to thermodynamic magnitudes of individual systems. It seems that the probability distribution used in LDT is one appropriate for spaces of abstract bits, but not for the physical phase space (sections 9-10). Because of these difficulties LDT ought to be rejected. The strong connection between logic and entropy is consequently broken off.

The paper begins with a detailed and critical exposition of LDT (sections 2-7), followed by a counter example (section 8) and a reexamination of the phase space representation of memory cells on which LDT relies (section 9). It then proposes an alternative (10). The paper ends with a brief discussion of the information theoretic approach to Maxwell's Demon (section 11).

2. What Exactly Is Landauer's Dissipation Thesis?

To avoid a common confusion, I repeat the outline of some currently accepted ideas. They are composed of two premises and a conclusion.

(A) *Landauer's Dissipation Thesis:*

Logically irreversible computation is dissipative. It increases entropy by $k\ln 2$ per bit of lost information, regardless of the technology by which the computation is carried out. This dissipation is a consequence of the logical irreversibility of the computation in question.

---

[1] LDT, as well as all the existing arguments for it, are classical. In some cases, LDT is automatically carried over to the quantum domain. The quantum mechanical LDT involves a different kind of reasoning, and is not addressed here.

[2] Devices operating on logically *reversible* principles have been developed, but - even if we had the technology to measure a dissipation of $k\ln 2$ - they cannot support the LDT nor refute it, for obvious logical reasons, known as Hempel's ravens paradox (Hempel 1965). This kind of devices is mentioned as a support for LDT in Landauer (1996).

Therefore, logically *reversible* computation is *not* subject to dissipation of this origin. This is LDT, and its details are discussed below.

To this dissipation one must obviously add dissipation of other origins, for LDT certainly does not imply that absolutely dissipationless computation is possible. Yet, for the sake of simplicity only, in this paper we shall *ignore* other sources of dissipation, and call processes in which the Landauer dissipation does not occur (*ideally*) dissipationless.

(B) A Theorem Concerning the Logical reversibility of computations:

All computations can be carried out in a logically reversible way. To every logically irreversible algorithm there corresponds a logically reversible one having the same output plus an output that enables to reverse the computation. By replacing a logically irreversible algorithm by its corresponding reversible one we carry out the original computation reversibly. (Bennett 1973, Fredkin and Toffoli 1982).

This theorem is a logical one. The only way to endow it with a physical implication is through associating it with LDT. Therefore, rejecting LDT would leave *II* as a theorem of *logic* that has nothing to do with dissipation of *energy*.

(C) *Conclusion from I and II: The reversible computation thesis:*

I and *II* together imply that all computations can (ideally) be carried out without dissipation, that is, without the sort of dissipation associated with logical irreversibility.

There are two different ways to conclude that computation is, in the ideal case, dissipationless. One is rejecting *I.* Another is accepting *I* and *II* and therefore *III*. The currently prevalent opinion is the latter: it accepts *III*, the reversible computation thesis, as grounded in both *I* and *II*. The present paper, on the other hand, questions *I,* and does not address *II* and *III*.

3. Physical Implementation of Logical Operations

3.1. Bits as Macrostates

LDT is a thesis about the *physical* properties of *physical* operations, that implement *logical* operations. Logical operations map abstract inputs (typically bits) to abstract outputs. Their physical implementations are processes transforming input physical states (typically having binary information bearing degrees of freedom) to output physical states.[3] The physical states in question are of memory cells, and the transformation is carried out by logical gates acting on the cells.

The evolution of the computer as a physical system is a series of microstates. However, since it is intended to manipulate bits, the fine distinction between microstates is unnecessary, and the computer's mechanism need not be sensitive to the difference between microscopic trajectories. It is enough to tune the computer to be sensitive to macroscopic states, called *1* and *0*. Each of them is a set of microstates, and the two sets

[3] By using the term *physical implementation*, I make no claim regarding the controversial idea that *information is physical*. This idea deserves a separate philosophical investigation.

differ in the value of the binary degree of freedom in which the information is stored. The computer's phase trajectories can then be described in terms of series of macrostates, or of arrays of bits.

On the one hand, a distinction between the macrostates *1* and *0* is sufficient for the computer to manipulate bits; there is no need for it to be sensitive to the difference between the microstates belonging to these macrostates. On the other hand, the distinction capability between *1* and *0* is certainly a minimum. The computer cannot be less sensitive than that, being a device that manipulates bits. If macrostates are determined by distinguishability, then *1* and *0* are *natural macrostates* of the computer. (Their characterization as natural macrostates does not refer to the subjective capabilities of any particular observer, although they can be so interpreted if one prefers this point of view.)[4]

3.2. Memory as Inaccessibility

For a system to be a memory cell, its *0* or *1* macrostates (the information bearing degree of freedom) must be (*a*) *stable and reliable* when intended to store information, and yet (*b*) *amenable to manipulation* during computation. These requirements have to be satisfied by the model of the memory cell that is used to explain its physics. A model not satisfying them is not a model of a memory cell that acts in a computer, but of something else. To discuss the entropy of computation in the framework of statistical mechanics we need to represent the memory cell in its phase space. This representation must take into account *both* requirements, (*a*) and (*b*). Let us see how this is done.

A fundamental postulate of statistical mechanics states that a trajectory can uncontrollably evolve throughout the phase space region accessible for it. External agents (constraints) can fix the boundaries of the accessible region, but once these are determined, the external agents can no longer control the trajectory's evolution. Let us call this the *uncontrollability postulate*.[5] An obvious consequence of the uncontrollability postulate is that, to ensure stability and reliability of a memory state, phase space regions belonging to

---

[4] These ideas apply also when bits are stored in microscopic objects, as in the (currently preliminary attempts to construct) bits for quantum computers. The reason is that macrostates are defined by distinguishability, not by the number of degrees of freedom (Shenker 1997). The notion of macrostate is further discussed in sections 5.29, and 10 below.

[5] Three remarks are in place regarding this postulate. First, uncontrollability is not lawlessness. The trajectory is subject to the laws of nature. Second, uncontrollability places no constraints on the probability distribution of states, and should not be confused with ergodicity and the like. Third, a version of the uncontrollability postulate is consistent with the interventionist (or open systems) approach to statistical mechanics. In this approach, entropy increase reflects the interaction of the system of interest with unknown external systems whose effects, while often dramatic, cannot be screened out. Here, the focus is on the unknowability of the agents acting on (or, rather, in) the system, and the crucial point is being unable (even in principle) to control and direct the system's state.

different memory states *must* be *disjoint and not-inter-accessible*, for as long as the cell is intended to store information. Whenever inter-accessibility is allowed, the device no longer has (or no longer is) a memory.[6] On the other hand, inter-accessibility must be allowed when the cell is required to change its state during a computation. The inter-accessibility in that case must be *controlled*.[7]

As an illustration, consider Figure 1.1.[8] It depicts two disjoint regions in phase space, one containing all the microscopic states corresponding to the macroscopic state called *0*, and the other containing all the microstates corresponding to the macrostate called *1*. Due to the continuity of trajectories in phase space on the one hand, and the disconnectedness of the two regions on the other hand, a trajectory starting in one such region cannot evolve into the other. Figure 2.1 depicts the same idea for a combination of two bits. The *0* and *1* states are non-inter-accessible.

INSERT FIGURE 2 ABOUT HERE.

At this point, we must distinguish *inter-accessibility* from *accessibility*. The notion of inter-accessibility is a conditional one: *if* the memory cell is in one macrostate, *then* it can (or cannot) access another. The notion of accessibility, on the other hand, focuses on the *if*-clause only: it determines the states the cell can initially assume. The two notions appear as two aspects of the phase space representation (Figures 1.1 and 2.1). *Non-inter-accessibility* is represented by the disconnectedness of the phase space regions corresponding to the different memory macrostates. *Accessibility* is represented by the selection, among the several disconnected regions, of those belonging to the cell's accessible region. There are two options to be considered here. One: the accessible region consists of a single disconnected region at a time, the region containing the cell's *actual* state. This idea is illustrated in Figure 3. The other option to be considered: the accessible region is a union of all the disconnected regions, here: *both 1* and *0*, with a probability distribution over them. This is illustrated in Figures 1.1 and 2.1. (The notion of accessibility is further discussed in section 9 below.)

---

[6] In some realistic cells inter-accessibility is not absolutely impossible, but very improbable, so that one may reasonably expect no changes of memory state to occur during the period of interest. (E.g., Landauer (1961) p. 184 Fig. 2.) In this case the two macrostates are not inter-accessible for as long as the device has (or is) a memory; once inter-accessibility is allowed, the device is no longer a memory. With this reservation in mind, we can go on focusing on absolute non-inter-accessibility during information storage.

[7] *Control* is the action of an external agent. The arguments for LDT treat the computer as a close system, and so the computer as a whole evolves uncontrollably (spontaneously). We can speak of control here only with respect to a sub-system of the computer, such as a memory cell, in which case the controlling external agent is another sub-system of the computer.

[8] Figure 1.1 is essentially similar to Landauer (1971) p. 49 and Landauer (1992) p. 2.

INSERT FIGURE 3 ABOUT HERE.

Landauer's Dissipation Thesis is based on preferring the latter option, that is, on the idea that the accessible phase space region is a disjoint union of the phase regions corresponding to macrostates *1* and *0*, with some probability distribution (typically uniform) over them. The justification normally given for this choice is the following. "In most instances a computer pushes information around in a manner that is independent of the exact data which are being handled, and is only a function of the physical circuit connections."[9] In other words, logical gates do not contain measuring elements, and carry out their computation regardless of the macrostate of the input cells. (This idea is further discussed in section 10.1 below.) A modelling of computation must allow for this mode of operation: it must not involve measuring the actual memory macrostates. Figure 3 seems to go against this idea, for it represents a situation where the actual macrostate is given and can be used as the basis of computation.

It is normally taken for granted that the phase space representation of a memory is as in Figure 1.1. In section 9 I show that this representation is fraught with difficulties, and argue that the Figure 3 type of representation has some significant advantages. However preferring the Figure 3 type of representation over Figure 1.1 involves a rejection of LDT, since the Figure 1.1 type of phase space representation is crucial for LDT, as I presently show.

4. The Main Argument for LDT

Logical operations map M n-tuples of bits to N n-tuples of bits. In logically reversible operations (1:1), $M=N$. Example: in the function called *Not*, NOT(1)=0, NOT(0)=1, $n=1$, $M=2$, $N=2$. In logically *irreversible* operations (not 1:1), $M>N$. Examples: in the function called *Erasure*, ER(1)= ER(0)= 0, we have $n=1$, $M=2$, $N=1$; in the function called *Or*, where the first bit stores the output and the second remains unchanged, which is OR(<0,0>)= <0,0>, OR(<1,0>)= <1,0>, and OR(<0,1>)= OR(<1,1>)= <1,1>, we have $n=2$, $M=4$, $N=3$. Discussions of LDT often focus on erasure for its simplicity.

Note, in passing, that there are two ways to *erase* information. One is by way of *destruction*, e.g., pulling out the electrical plug, or blowing up the whole computer. The other is by way of a *logical operation*: ER(1)= ER(0)= 0. LDT is about the logical operation type of erasure, not about destruction. Actually, LDT is not specifically about erasure at all, but about logical irreversibility in general. Note also that there is no counterpart for the destruction type of erasure in other logically irreversible operations.

---

[9] Landauer (1961), p. 184.

Consider, for simplicity, the case where the memory states *1* and *0* both have phase space volumes V=1.[10], [11] Given that the accessible phase space region is the disjoint union of macrostates *1* and *0* (as in Figures 1.1 and 2.1), the phase space volume accessible for a memory cell seems to *decrease* when logically irreversible operations are performed. The volume MV is mapped to NV, for $M>N$ (see Figures 1.2 and 2.2).

At this point in the argument for LDT one turns to Liouville's theorem,[12] which entails that in a closed conservative system the accessible phase space volume cannot be compressed. The seeming reduction in phase space volume in the transformation 1.1□1.2 or 2.1□2.2 must, therefore, be compensated by an expansion in some other degrees of freedom, which do not carry the information. Thus, instead of 1.1□1.2 we must have 1.1□1.3 *directly*, and instead of 2.1□2.2 we must have 2.1□2.3, again: *directly*. The intermediate stage in both cases, Figures 1.2 and 2.2, never actually occurs, for it violates Liouville's theorem. Stages 1.2 and 2.2 are presented for explanatory purposes only. In actuality, says the LDT, stages 1.1 and 2.1 are transformed to stages 1.3 and 2.3 *directly*.

At this point the main argument for LDT almost ends. Its last point (expressed in the last sentence from Landauer in section 1 above) is subtle (and rarely mentioned), and will be discussed below (section 6).

So far, the word *dissipation* has not yet been mentioned.

5. Entropy

Nobody knows what entropy really is.
J. von Neumann[13]

Dissipation, which is the degradation of energy from useful to useless forms, is quantified by the increase of entropy. Thermodynamics focuses on entropy differences, which are functions of state functions. For instance, the entropy difference between two equilibrium states A and B of an ideal gas is given by $\Delta S=C_V\log(T_A/T_B)+R\log(V_A/V_B)$, where C_V is thermal capacity for constant volume, T is temperature, R is the gas constant and V is volume. One central aim of statistical mechanics is to recover thermodynamic laws like this

---

[10] The term *volume* here is not very precise. A more precise treatment involves the choice of measure. But an intuitive understanding of multidemensional Euclidean volume will suffice for the present discussion.

[11] The binary case, volume(0)=volume(1)=V, can be generalized to n-ary states, e.g., the ternary case discussed in some versions of Maxwell's Demon, like Bennett (1982, 1987). The simplicity assumption of equal volumes is not trivial, for assigning different volumes to different memory states has significant implications with respect to LDT, as shown by Fahn (1996). But the simple case is enough to reveal the main difficulties.

[12] For a proof and discussion of this theorem see, for example, Tolman (1938) Ch. 3 and Lanczos (1970) Ch. 6.

[13] Quoted by Tribus and McIrvine (1971) p. 180.

from the laws of dynamics acting on molecules, possibly with some additional postulates regarding the possible initial states of the system and the probability distribution over them.[14] There are two main schools in main stream statistical mechanics which differ on how to do this. They are named after their founders, Gibbs and Boltzmann.[15] Crudely, Boltzmann accounts for the properties of an individual system by reference to its *macrostate*, while Gibbs accounts for them using the notion of *ensemble*. One consequence of these fundamental conceptual differences is that the magnitude called *entropy* has different meanings and different values in the two appraoches.

5.1. Gibbsian dissipation and LDT

In Gibbs's approach, canonical fine-grained entropy (for systems with constant temperature) is $-k\int\rho(X)\ln\rho(X)dX$, where ρ is the probability distribution function and X stands for all the degrees of freedom and the integration is taken over all phase space. For a microcanonical ensemble (representing isolated systems), fine grained entropy is $k\ln V_{(a.r.)}$ where $V_{(a.r.)}$ is the volume of the *accessible region*. By Liouville's theorem both expressions are constant in time. Hence no dissipation, in the sense of an increase in Gibbsian fine grained entropy, can *ever* take place. (The method of coarse graining is discussed below.)

Consider Figure 1, which depicts the phase space of the whole computer. This system is subject to Liouville's theorem. That is a central assumption in the argument for LDT, as we have seen above (section 4), for it is Liouville's theorem that dictates the constancy of phase volume, forbidding Figure 1.1 to be transformed to 1.2 and dictating its direct transformation to 1.3. Since the phase volumes of 1.1 and 1.3 are equal throughout the transformation (with uniform probability distribution inside and zero outside), their Gibbs microcanonical (as well as canonical) entropies in them are equal . *No Gibbsian dissipation takes place at the transformation from 1.1 to 1.3*. Hence, the dissipation that Landauer's Thesis proclaims cannot be of *this* sort. Let us proceed to search for other notions of dissipation.

The fact that Gibbsian entropy is conserved by Liouville's themorem, and that this precludes a recovery of the Second Law of thermodynamics, was known to Gibbs. To account for changes of entropy, as described by the Second Law, he proposed the idea of coarse graining. This idea is problematic, and collapses entirely in the case of the spin echo experiments.[16] But even if one tends to accept it, the idea is not applicable for the LDT case, for the following reason. A change of entropy due to coarse graining occurs when we start out in a non-equilibrium state, in which the probability distribution is non-zero in one part of the accessible region and zero in another. (Such a state can be brought about

---

[14] Of course, one then has to *justify* these postulates, a very non-trivial business. I do not address these difficulties here.

[15] For the two approaches and the differences between them see, for example, Gibbs (1902), Ehrenfest (1912), Tolman (1938), Jaynes (1965), Lebowitz (1993), Sklar (1993), and Callender (1999).

[16] see Ridderbos and Redhead (1998).

by, for example, a sudden increase of volume.) In such a case the non-zero part of the probability distribution is claimed to evolves such that it fills up the whole accessible region in the fibrillated way that Gibbs compares to mixing two colours. Our case is not like that. Probability is (generally) equally distributed between the *1* and *0* parts of the accessible region, and a uniform distribution persists when Figure 1.1 evolves to 1.3 (directly). Therefore, dissipation in the sense of coarse graind Gibbsian entropy never takes place.

LDT turns out to be patently false in the Gibbsian framework. Whence the prevalent error? I return to this interesting question later.

5.2. Boltzmannian dissipation and LDT

In Boltzmann's approach, entropy is $k\ln V_{(mac)}$ where $V_{(mac)}$ is the phase space volume of a *macrostate*. Here, entropy can change spontaneously, when the system evolves from one macrostate to another. (The problem, not addressed here, is to prove that it normally evolves in accordance with the Second Law.) Whereas Gibbs's entropy is determined by the volume of the accessible region, and hence this volume must be carefully identified, Boltzmann's entropy is determined by the volume of the macrostate, and so the macrostates have to be carefully identified. Above (section 3.1) I presented considerations to the effect that the memory states *1* and *0* are the natural macrostates of a memory cell: distinction between them is sufficient for manipulating bits, and is also a minimum for this.

Since this is an important point, a possible objection ought to be addressed. In the computers we currently use individual memory cells are certainly inaccessible for the normal user at any practical level. In what sense, then, are the *1*s and *0*s, or arrays of them, distinguishable macrostates? The input and output of a computer that performs an algorithm are very well defined arrays of bits. A flip of a bit may change the input and output in a way that is very significant at the pragmatic level. Therefore the user must be able to distinguish *1* from *0* at both ends. It makes no difference whether the distinction can be directly perceived with naked human senses or using complex and theory-dependent devices. The use of automated error correction processes should make no difference either. The point here is only the very distinguishability: if and only if *1* and *0* can *somehow* be distinguished, they are different macrostates. And if they are distinguishable at the input stage there is no reason why they should cease to be so during computation (but see section 10.1 below). And so, prior to the erasure (Figure 1) the Boltzmann entropy is either $k\ln V(0)$ or $k\ln V(1)$, and after the erasure it is $k\ln V(0)$. In the simplest or extreme case $V(0)=V(1)$, and the entropy difference is null. (Since we are in search of minima, focusing on extreme cases is desirable.) In other words, entropy is the same before and after erasure. *No Boltzmannian dissipation takes place.*

Once again, the LDT turns out to be plainly false. And once again, the interesting question is about the source of the prevalent error.

6. Diffusion and Dissipation in LDT

Since - by the above results - no Gibbsian fine or coarse grained dissipation, and no Boltzmannian dissipation, necessarily occurs in logically irreversible operations (like erasure), it is important to discover what is *dissipation* in LDT.

In thermodynamics, *dissipation* means *degradation in the exploitability of energy*, as in Joule's famous experiment, where gravitational potential energy is transformed into heat energy. None of this necessarily happens in the case of logical irreversibility. The energy stored in the non-information bearing degrees of freedom may have a form that is valuable and exploitable from a thermodynamic point of view. Indeed, Landauer (1992, p. 2) notes that the dissipation is *not* identical with the expansion into the non-information bearing degrees of freedom. It does *not* occur at the transformation from Figure 1.1 to 1.3, but at a later stage. At that later stage, *after* 1.3, the top and bottom regions of 1.3 "*diffuse into each other*" (see quote at the introduction). What does this diffution mean, and how is it associated with disspation?

Since the top and bottom regions in Figure 1.3 are inter-accessible (unlike the two regions in 1.1), the uncontrollability postulate entails that the system may (and usually does) lose memory of its initial state. The top and bottom regions at 1.3 may be distinguishable macrostates, that is, we may know where the system is at each given time along the non-information bearing degrees of freedom. But since they are uncontrollably inter-accessible, we cannot deduce, from where the system *is* sometime after the erasure, anything about where it *has been* right after the erasure, and hence we can also not deduce its pre-erasure macrostate. In this sense, the information about the input, pre-erasure state, is irrecoverably lost. It is lost *not* during the transformation from 1.1 to 1.3, but *later*, as an outcome of the interaccessibility in the vertical direction. (We could make the top and bottom regions of 1.3 non-inter accessible, as in 1.1, but then we would not have had real and complete memory erasure.)

So we indeed have a well defined and very useful notion of *diffusion*, as Landauer suggests.

But does this diffustion mean, or entail, *dissipation*? As before, let us examine the different notions of dissipation that appear in main stream statistical mechanics. The Gibbsian fine grained microcanonical entropy $k\ln V_{(a.r.)}$ does not change during the diffusion, because the diffusion occurs within unchanging boundaries of the accessible region. The fine grained canonical entropy $-k\int\rho(X)\ln\rho(X)dX$ doesn't change either, since the probability distribution remains uniform throughout the process. Gibbs's coarse grained entropy doesn't change, too, since the distribution is spread out all over the accessible region, already at the beginning of the diffusion stage. And so the diffusion does not mean nor entail a Gibbsian dissipation. LDT fails here.

The Boltzmann entropy depends on the new macrostates. Suppose, for simplicity, that the top and bottom regions of 1.3 are macroscopically distinguishable and have equal phase volumes. In this case the Boltzmann entropy is constant: $k\ln V(0) = k\ln V(1) = k\ln V(\text{top}) = k\ln V(\text{bottom})$. (Again, the volumes *could* be different, but we are looking for minima, and so must focus on extreme cases.) The result: a*lthough macroscopic memory is lost (for the macroscopic distinction between top and bottom tells us nothing about the system's history), entropy is (or can be) conserved*. And so the diffusion of top and bottom does not entail a Boltzmann dissipation. LDT fails here, too.

7. Alternative Arguments for LDT

Landauer's Dissipation Thesis has other arguments for it, but they all suffer from the similar sort of fundamental problems, that make them unacceptable. In one of them, the focus on a single memory cell is replaced by focusing on an array of N memories. In the case of erasure, instead of the two regions, of 0 and 1, being mapped into region 0, one speaks of 2^N N-bits arrays mapped into an array of all 0. The above problems remain.

A second argument is that Figure 1 resembles an isothermal compression of a gas, such as the transformation 4.1*a*□4.2*a* in Figure 4.[17] This argument is very confused, and the notion of resemblance that it uses is unclear. Let us first examine resemblance in appearance and then resemblance in function.

The *prima facie* similarity between Figures 1 (parts 1.1, 1.2 and 1.3) and 4 (parts 4.1*a*, 4.2*a* and 4.3, respectively) is very misleading. They are significantly *dissimilar*, and once we try to make them similar, dissipation disappears. Let us see how. Consider Figure 4.1*a* and compare it to 1.1. The analogy of appearance seems to be between individual gas molecules and their trajectories in bounded physical space, and individual phase points and the trajectories starting from them in the accessible phase space region. The two cases differ, however, with respect to inter-accessiblity of macrostates: regions *1* and *0* in Figure 1.1 are non-inter accessible, whereas in 4.1*a* all states are uncontrollably interaccessible. So the structure of the phase space is significantly different from the structure of the physical space; there is no similarity of appearance, and so no ground for an analogy based on such an appearance.

INSERT FIGURE 4 ABOUT HERE.

To make the two cases *appear* similar we need to insert a partition in the container. Let us do this (replacing 4.1*a* by 4.1*b*). Now, however, with the partition, we can no longer perform the compression to 4.2*a*. Stage 4.2*a* becomes inaccessible from 4.1*b*, in (a very imperfect and perhaps misleading) analogy to the inaccessibility of 1.2 from 1.1.

Let us try another resemblance of appearance. Since the transformation 1.1□1.2 is forbidden (by Liouville's theorem) and what actually happens (by LDT) is 1.1□1.3, we need to focus on the seemingly-analogous transformation 4.1*a*□4.3 (or even 4.1*b*□4.3). This transformation is, however, a paradigmatic case of a dissipationless process.

Another aspect of the dissimilarity between Figures 1 and 4 takes us from the analogy by *similarity of form* to *similarity of function*. The macrostate of the gas in the container at stage 4.1 stores no information, with or without a partition. If we want it to store at least one bit we must first compress the gas into one chamber, as in 4.2*a*, and then keep it there with a partition, 4.2*b*. And so it turns out that, from the perspective of functioning as a memory, the analogy ought to be between 1.1 and 4.2*b* (not 1.1 and 4.1). Now, if 4.2*b* is the pre-erasure state, which is the erased state? The transformation 4.2*b*□4.1 (or, equally, 4.2*b*□4.3) is a process of erasure by *destruction*, not a case of the logical

[17] This analogy is mentioned in the quote from Landauer, at the opening of this paper

operation ER(0)=E(1)=0 (for the notion of destruction see section 4 above). To perform the logical operation we need to follow this destruction by compression into the chamber called *0*. This is an erausre by thermalization. Nothing, however, in the general scheme of Figure 1 dictates or entails that this is the only method of erasure.[18] An alternative process on erasure, not involving thermalization, is discussed in section 8 below.

Let's sum up the argument by analogy from Figure 4 type of processes. Figures 4 and 1 are *dissimilar in form*, due to the inter-accessibility of the macrostates regions in Figure 1.1d. Restoring similarity by inserting a partition (4.1*b*) makes the process (4.1*b*□4.2*a*) impossible; restoring similarity by transforming 4.1*a* directly to 4.3 makes the process dissipationless. Figures 4 and 1 are also *dissimilar in function*, for the state in 4.1 stores no information. Restoring similarity here destroys the similarity of form: in form, 1.1 resembles 4.1; in function, it resembles 4.2. And even then, the gas container turns out not to be the most general case of erasure, for it requires a not-generally-necessary stage of thermalization (see Section 8).

A third alternative argument for LDT is by analogy to (the reverse of) cooling by adiabatic demagnetization.[19] This argument suffers from problems similar to those of the previous example. In particular, (to take a semi-classical view of this quantum mechanical phenomenon) in the stage analogous to pre-erasure, the spins system has no memory (just like the gas in 4.1). This uncontrollable inter-accessibility is *essential* for the cooling process.[20] If we destroy the interaccessibility, the system ceases to cool. Again, once we make the systems analogous in either appearance or function, dissipation dissappears.[21]

8. A Counter Example for LDT

Before proceeding to propose an alternative, let us support the conclusion regarding the failure of LDT by a counter example. Consider the memory cell in Figure 5. Its contents is determined by whether the notch marked *R* or the one marked *L* in the key shaped board rests on the peg. This memory stores the outcome of measuring the position of a particle in a container: right or left. The symbol *?* indicates the *ready* state. The symbols *R, L, ?* are reminiscent of the original context in which this device originally appeared (see below) and can be thought of as *1*, *0* and *ready*. To erase, lift the key using gear *A*, then turn gear *B* in the indicated direction. The structure of the grooves will ensure that the key will stop with the *?* notch facing the peg, regardless of the initial state. Then Lower the key back using

---

[18] Incindentally, this process requires a measurement in order to pull the partition quasi statically in the right direction. But see section 10.1 below.

[19] This analogy is mentioned in the quote from Landauer, at the opening of this paper.

[20] To see that it is enough to consider a simplified semi-classical description of the adiabatic demagnetization method of cooling, such as in Reif (1981) pp. 445-451.

[21] Other arguments, not discussed here, are formally more complex, but on close examination subject to similar objections. E.g., Shizume (1995) and Lubkin (1987). They are discussed in Shenker (1997).

gear *A*. The information (*R* or *L*) is thereby erased, and the only source of dissipation is friction, which has no minimum (see Krim 1996).[22]

INSERT FIGURE 5 ABOUT HERE.

This device is based on one by Bennett (1987, p. 94). Bennett proposed it as a counter example for the claim that *measurement* is dissipative. The original context is a discussion of Maxwell's Demon (see section 11 below), but Bennett's counter example has implications for the physics of information which are quite general, and therefore it can and should be considered in its own right, regardless of its original context. Bennett's construct is a measuring device that determines which of two chambers contains a particle, and stores the information in a memory (the key with notch and peg). The device is a clever combination of gears, in which the only possible source of dissipation is friction. Bennett rightly concludes that this device shows that measurement cannot be associated with any minimum amount of dissipation.[23] This conclusion is based – as it should - on the details of the counter example, regardless of any general or abstract argument that might be offered regarding the entropy of measurement.[24]

If we add the Figure 5 device to Bennett's apparatus, the stored information can be erased, in a process where, again, friction is the only possible source of dissipation. From a logical point of view, the conclusions ought to be the same, namely, that the device shows

---

[22] The key acquires momentum, which indicates the original information stored in the memory. (Chris Moore and Michael Lachman raised this objection.) The key can be stopped using friction, which can be made as small as we want, in particular, smaller than $k\ln 2$, by turning gear *B* very slowly.

[23] Earman and Norton (1999, pp. 13-14 and 16) find the following flaw in Bennett's device, when this device is used as part of Maxwell's Demon. The keel shaped object with the two pistons that is lowered onto the chambers is subject to thermal fluctuations, they say, for it must have very little mass, to allow for a quasi-static compression against the particle's pressure. If, however, we disassociate Bennett's apparatus from Maxwell's Demon, we are not obliged to insert a light and fluctuating gas particle into the chamber. We can measure something else instead, for instance: determine which chamber contains a heavy rock. The dissipation associated with the piston's collision can be made as small as we like, provided the rock is very rigid relative to the piston's weight so that the collision is elastic, and that the piston's weight is not small enough to make it subject to thermal fluctuations, and that the piston's velocity is very close to zero at the time of collision.

[24] The usual argument supporting entropy of measurement is this. "The Second Law of Thermodynamics forbids a net gain of information. Yet a measurement 'provides information'. Measurement itself thus becomes paradoxical, until one reflects that the gain in information about the system of interest might be offset by a gain in entropy of some 'garbage can'. Indeed, it *must* be so offset to save the bookkeeping of the Second Law." Lubkin (1987), p. 523. This argument is confused; I discuss it in Shenker (1997).

that *erasure* cannot be associated with any minimum amount of dissipation. To a comment in the same spirit, by Parke (1988), Bennett (1988) replied that the erasure is nevertheless dissipative, *just because* it maps two memory states into one (by the above main argument for the LDT). This, however, is a reply to a counter example by reference to the universal claim being challenged. One ought, instead, to study the details of the counter example.

It is sometimes maintained that the erasure carried out by the Figure 5 device is not a complete one. For since the particle is still in its original chamber, the information regarding its position can be recovered. To obtain a complete and irreversible operation, it is claimed, we ought to erase also the information stored in the measured position itself. This amounts to the problematic claim that the state of a system is also storage of information regarding that system and that state. This claim involves a very general question about the meaning of measurement and information storage. I shall not address it here. Instead, I shall add an element to the Figure 5 device that will answer the said objection.

The idea here is simple, while its graphic illustration is cumbersome; I therefore describe it without an accompanying figure. Our attention is focused on the following elements: the key, which stores the original position; the particle, which is in the full chamber of the container, the other half being empty; and the two pistons at the right and left hand sides of the container. At the initial state the key still stores the which-chamber information. In the first stept of the erasure, a set of gears couples the side pistons to the key's position, so that the piston at the empty side is pushed in against vacuum. This involves no dissipation, apart from friction. Second, the (very thin) partition is pulled out. This involves no dissipation since the inserted piston replaces the partition. Third, the two side-pistons are shifted together, keeping their mutual distance constant, until the center of the full chamber is positioned where the partition has originally been. Then, the walls of the empty chamber are folded out or taken apart, with the result that the appearance of the container is symmetric and gives no indication whether the particle has originally been at the right or the left hand side chamber. This completes the erasure of the information stored in the particle's state. Next, we turn to erase the information stored in the key, as explained above and in Figure 5. By the end of this stage the information is erased and irretrievable. This does not close an operation cycle, for the particle's container doesn't have its original form. However, LDT is not about operation cycles but about logical irreversibility.

If Bennett's device shows that measurement is not necessarily dissipative, *then* it *also* shows that erasure is not necessarily dissipative.

9. Reexamination of the Phase Space Representation of Memory Cells

So far we have seen that LDT fails because, *given* the phase space representation of a memory cell which is uses (Figures 1 and 2), erasure does not involve any necessary change of entropy (as this term is understood in main stream statistical mechanics). I now turn to

show that problems start already in the preparatory stage of the argument for LDT. The phase space representation of a memory cell on which it relies is problematic.

In classical statistical mechanics probability distributions appear in two sorts of expressions. One is a family of *phase averages* which correspond to directly measureable thermodynamic properties of individual systems, like pressure and temperature. (Why do *averages* correspond to the properties of *individual* systems? This is an open question.[25] I don't address all its aspects here, but it will play an important role in our considerations below.) Another family of expressions, in which probability distributions appear, characterizes these distributions *themselves*. Some members of this family are called *entropy* (like the ones we have seen in section 5 above). *Entropy is a property of a probabilty distribution, regardless of the subject matter over which the probability is distributed*. (See below a discussion of *thermodynamic* entropy.)

The importance of the distinction between the two types of expressions in which probability appears cannot be overemphasised. In phase averages which are supposed to stand for the directly measureable thermodynamic quantities, the nature of that space is of major importance. For it is (for instance) the fact that we are dealing with (generalized) positions and mementa which makes our discussion part of statistical *mechanics*, rather than an abstract exercise in probability. In expressions for entropy, on the other hand, all that matters is the probability distribution itself, and there is no role whatsoever to the nature of the space over which the probability is distributed. For this reason the notion of entropy can be generalized and abstracted, as is done (for instance) in Shannon's communication theory and elsewhere. Such an abstraction is possible and legitimate only as long as the nature of the space, over which the probability is distributed, is immaterial.

Of course, entropy originates in classical thermodynamics. Notice, however, that thermodynamics discusses entropy changes, and that these changes are functions of the changes in the directly measureable quantities (entropy change is not a directly measureable quantity). The reason of this nature of thermodynamic entropy is this. Thermodynamic entropy describes the way that the directly measureable quantities change, and this way of changing is common to them all. It is common to all of them *because* the directly measureable quantities are averages, calculated using the probability distribution, that entropy characterizes.[26] If the distribution changes, all the quantities based on it change in a similar way.

Let us apply these considerations for LDT. As long as we focus our attention on entropy only, it is hard to see that anything is wrong with phase space representations like Figure 1.1. It can easily be interpreted as expressing ignorance with respect to the initial macrostate of the memory cell. (See, however, the problem with input data, section 5.2 above.) But since LDT is an argument about dissipation of *energy*, it is an argument in

---

[25] For an overview of the attempts to answer it see Sklar (1993).

[26] The word *because* here is problematic. As things stand now, thermodynamics is certainly not reducible to the underlying mechanics, in any sense of reduction. See surveys of this problem in Sklar (1993) and Guttmann (1999). Albert (1994) is an attempt to make such a reduction; this attempt is discussed in Hemmo and Shenker (2000).

statistical *mechanics.* It needs to account for the fact that the probability is distributed over *phase space* of position and momentum, and not over just any abstact space of logical possibilities. We ought to demand that the probability distribution used in LDT will work properly when we want to calcualte phase averages meant to stand for directly measureable thermodynamic properties of individual systesm. And here we face one of the central difficulties at the foundations of statistical mechanics, for which we need to make a brief digression.

One problem at the foundations of (Gibbsian) statistical mechanics is to justify *why* phase averages correspond to thermodynamic magnitudes pertaining to individual systems. A closely related question (in the Boltzmannian framework) is, *why* does the phase volume of a macrostate correspond to the probability of finding the system in that macrostate. (The questions are *why* and not *whether*, for the recipes of statistical mechanics are clearly successful in the appropriate circumstances.) Many have come to recognize that ergodicity cannot be the key for soveing this problem. Ergodicity is not necessary for explaining the success of statistical mechanics since many interesting and relevant systems are not ergodic. And it is not sufficient either since, for example, ergodicity is about infinite time averages while we are interested in finite relaxation times.[27] No sufficiently fault free alternative is currently available, and the justification for using probability distributions in predicting the behavior of individual systems is still an open question.[28]

Still, one basic intuition behind the ergodic approach seems to underlie every objectivist attempt to justify the use of probability distributions to predict the behavior of individual systems. This intuition appears to be indispensable, and it is the following. Consider some phase function f that corresponds to the thermodynamic magnitude F of system S. And consider a region R in the phase space of S, which has a non-zero weight in calculating f. By assigning R a non-zero weight we seem to claim that the states in R are *not altogether irrelevant* for the dynamics of S. In other words, we seem to claim that *it is not completely out of the question* that S will, at some point of time, assume a state belonging to that region.[29]

---

[27] For problems in the ergodic approach see, e.g., Earman and Redei (1996), Sklar (1973), and Guttmann (1999).

[28] A possibly path breaking proposal here, based on quantum mechanical considerations, is Albert (1994). See discussion and an alternative in Hemmo and Shenker (2000). An alternative direction is interventionism or the open systems approach; see Bergmann and Lebowitz (1955), Blatt (1959), Sklar (1993), Ridderbos and Redhead (1998), Shenker (2000), Hemmo and Shenker (2000).

[29] As long as we deal with essentially isolated systems, this may lead to something like the requirement of indecomposability, which can mean back to ergodicity with its problems. An attempt to avoid this outcome may lead to the open systems approach or to postulating inderterministic underlying dynamics (see footnote Error: Reference source not found). I shall not address this problem here, for the present purpose is not to solve the difficulties at the foundations of statistical mechanics, but to point at problems in LDT.

Consider, now, Figure 1.1. A memory cell that - possibly unknown to us - started out in macrostate *0*, will *never* be in macrostate 1. For such a memory cell, macrostate 1 is completely out of the question. And *vice versa*, of course, for a memory that started out in macrostate *1*. Suppose, now, that regions *1* and *0* in Figure 1.1 differ in their microstates in such a way that the phase average corresponding to, say, pressure in region *0* is P_0 and the phase average corresponding to pressure in region *1* is P_1. Because of the non-inter-accessibility of regions *1* and *0*, the *real* pressure of the system at any given point of time *t* (that is, pressure as a *directly measureable property of an individual system*) will be *either* P_0 *or* P_1. It will *not* be $(P_0+P_1)/2$. The magnitude $(P_0+P_1)/2$ does not correspond to any directly or indirectly measureable physical property of the system described in Figure 1.1, at any point of time. To generalize: *a phase average in which regions 1 and 0 both have a non-zero probability does not correspond to any physical-thermodynamic magnitude.*

There is one notable exception, of course: a phase average in which regions *1* and *0* both have a non-zero probability can correspond to *entropy*. But this is a pathological case, in which we can speak of entropy only as long as we do not associate it with the *thermodynamic* entropy. To emphasise this point think of expressions like $\checkmark S=C_V\log(T_A/T_B)+R\log(V_A/V_B)$, the entropy difference in an ideal gas. Since T and V, being phase averages, are not well defined in cases like Figure 1.1, $\checkmark S$ is not well defined either. Thus, magnitudes like $-k\int\rho(X)\ln\rho(X)dX$, $k\ln V_{(a.r.)}$, and $k\ln V_{(mac)}$, while being very interesting characterizations of probability distributions in general, cannot possibly represent the *thermodynamic* entropy in this case.

What are the consequences for phase space representations like Figure 1.1? The aim of this representation, in the framework of LDT, is to learn something about *dissipation of energy*, about *thermodynamic dissipation*. But, since phase averages over cases like Figure 1.1 do not stand for thermodynamic-physical magnitudes, arguments based on Figure 1.1 can teach us absolutely nothing about the *thermodynamics* of information processing. Figure 1.1 can teach us a lot about ignorance and probability, but not about thermodynamics, energy, and heat.

What ought we to do if we want to study the thermodynamics of information processing? Obviously, we must reexamine and change the phase space representation of memory cells.

10. An Alternative Phase Space Representation

The above difficulties indicate that the model or phase space representation of Figure 1, which is the central pillar on which the LDT rests, is problematic. Figure 1 is based on two ideas, which together lead to the above difficulties. One: to ensure reliability of memory storage, the memory states ought to occupy *disjoint* phase space regions. I think this idea is correct. Two: The accessible region is a *union* of all logical possibilities at every stage of the computation. Here, I suggest, is the problem.

A possible alternative phase space representation is Figure 3, in which the accessible region at every point of time consists only of the region containing the *actual* macrostate of the memory cell at that point of time. It consists of *either* region *0 or* region *1*, never both.

Such a representation solves the above difficulties.[30] It leads, however, to the conclusion that (ideally) computation – even logically *ir*reversible computation – is not necessarily associated with any minimum amount of dissipation; hence, to a rejection of LDT (above and beyond the considerations of sections 58 above). Let us now go over considerations that support Figure 3, and then turn to solve a difficulty that this representation *prima facie* raises.

As said in section 5.2 above, for a computer to operate as such, on the intended input and with the intended algorithm, the user must be able to *distinguish 1* from *0* at the input stage (whether with naked human senses or using measuring devices). This determines that *1* and *0* are distinct macrostates.

For the same reason, in order to *feed in* the right input, the user must be able to *control* whether the input state of a memory cell is *1* or *0*. This - I now claim - determines that *1* and *0* belong to different accessible regions. That is, *at each point of time the accessible region is either region 1 or region 0*. The boundaries of the accessible region change with time during computation, such that it is never the case that both macrostaes belong to the accessible region. This idea is expressed by Figure 3. I now argue for this claim.

Recall what is the meaning of an accessible region: the boundaries of the accessible region express the extent of the control that external agents (like the user) can have on a system. By definition, the user *can* determine the boundaries of the accessible region, but *cannot* determined which of the macrostates, that are inside this accessible region, the system will assume. If a user's control appears to overstep the boundaries of the accessible region, that is, if the user seems to be able to determine which of the macrostates the system will assume, then this indicates that we are mistaken about the actual boundaries of the accessible region; they are narrower than we thought. We must redraw the boudaries of the accessible region, so that they will express the extent to which an external agent can control the system.

If we want the user to dictate the input state of a memory cell, this state must be determinable by manipulating the boudaries of the accessible region, without overstepping them. If both macrostates *1* and *0* belong the the accessible region at the same time, the user cannot dictate which of them the memory will assume. The user, acting as an external agent, can only determine the boundaries of the accessible region. And so, if we want the user to control the contents of the memory cell at the input stage, we must construct the computer such that the macrostates *1* and *0* will inhabit distinct accessible regions i.e., that they will be accessible *one at a time*, depending on the external constraints acting on them, that are controlled by the user. Only if *1* and *0* belong to distinct accessible regions can a user determine and control the contents of the memory at the input stage.

These consideratiosn support a phase space representation of the Figure 3 type, in which regions *1* and *0* are accessible one at a time.

---

[30] It does not fully *guarantee* the *not completely out of the question* condition, however, for we still have to look into the dynamics of the system inside the disjoint regions. This is a major difficulty at the foundations of statistical mechanical and I do not address here.

10.1.A Difficulty and Its Solution

The Figure 3 type of phase space representation raises a difficulty when we move on from the input (and output) stage to the intermediate computation that takes place inside the computer. The computation is a series of physical interactions, in which the logical gates manipulate the states of memory cells. These interactions do not seem to include what we would nomally call *a measurement*, and in this respect it appears that "in most instances a computer pushes information around in a manner that is independent of the exact data which are being handled, and is only a function of the physical circuit connections."[31] Figure 3 seems to disagree with this idea, for it may be interpreted as suggesting that the memory's state is measured before the logical operation takes place. There is no ignorance regarding the cell's macrostate prior to the operation, but perfect knowledge.[32] A possible solution for this difficulty requires an understanding of what *measurement* means here and of its thermodynamic significance. Two points are involved.

(i) In Classical physics, *measurement* is coupling. Interestingly, classical physics does not distinguish couplings that are measurements from other couplings, and the only criteria for such a distinction are pragmatic, pertaining to human interests. In this sense, *measurement is not a natural kind from the point of view of classical physics*. (Whether the quantum case is different or not is an open question.) The phase space representation of Figure 3 does not reflect knowledge by any user. It only reflects the idea that the logical gate acting on the memory cell is coupled to the cell in its input state and operates accordingly. The operation of Bennett's measuring device discussed in section 8, with the way the position of the particle is coupled to the memory cell, is an example.[33] Therefore, the decision whether or not to *call* a certain coupling *a measurement* (or *use* it as such) cannot possibly have distinctive physical properties, such as a characteristic entropic behavior, that will distinguish it from other couplings. And therefore, whether or not we decide to call the cell-gate coupling *a measurement* cannot possibly affect the entropy of the logical operation.

(ii) The conclusion of (i) is also one justification for concurring with the opinion that classical measurement is *not* necessarily associated with any minimum amount of dissipation.[34] This opinion is supported by Bennett's (1987) thought experiment described in section 8. If, however, measurement is (ideally) entropy conserving, it ought to be entropically insignificant whether it takes place or not. In particular, whether or not the

---

[31] Landauer (1961), p. 184.

[32] It is sometimes claimed that this knowledge has to be erased, otherwise the process is not logically irreversible. See, for instance, the discussion of Bennett's though experiment in section 8 above, and the distinction made there between logical reversibility and cyclic operations.

[33] The details of the coupling mechanism were not described in Section 8, see Bennett (1987).

[34] See footnote Error: Reference source not found for the claim that measurement is dissipative by $k\ln 2$ per *gained* bit of information.

state of a memory cell is measured before we carry out a logical operation needn't affect the entropic properties of that operation.

11. A Remark Concerning Maxwell's Demon

Maxwell's Demon is often taken to illustrate the import of LDT, for it is held that dissipation in memory erasure is the key to solving the Demon conundrum (e.g., Bennett 1982, 1987). It may therefore be instructive to show that the LDT, even if it were correct,[35] couldn't have solved the Demon. This, for several reasons.

One is, that the meaning of *dissipation* in LDT is not connected to the usual notion used in thermodynamics and statistical mechanics, in any clear or straightforward way (see sections 5 and 6 above).

However, suppose that we somehow make a connection between the two notions of entropy. Then, the entropy balance of the Demon *with* the LDT is not different from the balance *without* the LDT. The dissipation, claimed in the LDT, compensates for the (alleged) reduction in phase space volume associated with discarding information; stage 3 compensates for stage 2 in both Figures 1 and 2. The Landauer dissipation is intended to make sure that erasures (and other logically irreversible operations) do not turn out to be perpetual motion machines *in their own right,* above and beyond Maxwell's Demon. Therefore, by the LDT, the *net* entropic effect of erasure is null, not positive. Erasure has no spare $k\log 2$ of dissipation that could be used to compensate for the Demon's operations. This is clearly seen by comparing the phase volumes of stages 1 and 3 in Figures 1 and 2. Therefore, whether or not erasure is dissipative is irrelevant for the Demon.

Finally, Earman and Norton (1999) claim that a solution for the Demon based on the LDT is circular in an unacceptable way. The Demon is intended to be a counter example for the Second Law, and therefore this law must not be assumed in its solutions. They emphasize that the dissipation-in-erasure argument relies, crucially and indispensably, on the Second Law. The dissipation accompanying erasure is intended to compensate for the (alleged) reduction in phase volume, and such a compensation is only required if we assume that the Second Law is valid in the system. Normally, it is very reasonable to assume the applicability of the Second Law; this assumption is, however, unacceptable when discussing Maxwell's Demon. [36]

Acknowledgements. I have greatly benefitted from the comments of the following people on various (some very old and different) drafts of this paper and am grateful to them all: Jacob Bekenstein, Harvey Brown, Jeffrey Bub, Carlton Caves, Chris Fuchs, Michael Lachman, Cris Moore, John Norton, Itamar Pitowsky, Sandu Popescu, Simon Saunders.

---

[35] And even if supplemented by the idea of algorithmic complexity; see discussion in Earman and Norton (1999) pp. 17-20.

[36] An alternative proposal for solving the Demon is given in Shenker (1999).

References

Albert, D. (1994), "The Foundations of Quantum Mechanics and the Approach to Thermodynamic Equilibrium", *British Journal for the Philosophy of Science* **45**, 669-677.

Arnold, V.I. and Avez, A. (1968), *Ergodic Problems of Classical Mechanics* (N.Y.: W.A.Benjamin).

Bennett, C.H. (1973), "Logical Reversibility of Computation", *IBM Journal of Research and Development* **17**, 525-532.

-- (1982), "The Thermodynamics of Computation: A Review", *International Journal of Theoretical Physics* **21**, 905-940.

-- (1987), "Demons, Engines and the Second Law", *Scientific American* **257**, 88-96.

-- (1988), "Letter to the Editor", *Scientific American* **258**, 6.

Bergmann P.G. and Lebowitz, J.L. (1955) "New approach to nonequilibrium processes"
Physical Review 99(2), 578-587.

Blatt, J.M. (1959) "An alternative approach to the ergodic problem", Progress of Theoretical Physics 22(6) pp. 745-756.

Callender, C. (1999), "Reducing Thermodynamics to Statistical Mechanics: The Case of Entropy", *Journal of Philosophy* **96**, 348-373.

Earman, J. and Norton, J. (1998), "Exorcist XIV: The Wrath of Maxwell's Demon, Part I", *Studies in the History and Philosophy of Modern Physics* **29**, 435-471.

-- (1999), "Exorcist XIV: The Wrath of Maxwell's Demon, Part II", *Studies in the History and Philosophy of Modern Physics* **30**, 1-40.

Earman J. and Redei M. (1996), Why Ergodic Theory Does Not Explain the Success of Equilibrium Statistical Mechanics", *British Journal for the Philosophy of Science* **47**, 63-78.

Ehrenfest P. and Ehrenfest T. (1912), *The Conceptual Foundations of the Statistical Approach in Mechanics*, trans. M.J. Moravcsik (Leipzig: 1912, N.Y: Dover, 1990)

Fahn, P. (1996), "Maxwell's Demon and the Entropy Cost of Information", *Foundations of Physics* **26**, 71-93.

Fermi, E. (1936) *Thermodynamics* (N.Y.: Dover)

Fredkin, E. and Toffoli, T. (1982), "Conservative Logic", *International Journal of Theoretical Physics* **21**, 219-253.

Gibbs, J.W. (1902), Elementary Principles in Statistical Mechanics (New Haven: Yale University Press; N.Y.: Dover, 1960).

Guttmann, Y. (1999) The Concept of Probability in Statistical Physics (Cambridge: Cambridge University Press).

Hemmo, M. and Shenker, O. (2000), "Quantum Decoherence and the Approach the Thermodynamic Equilibrium", *forthcoming*.

Hempel, K. (1965), "Studies in the Logic of Confirmation", in *Aspects of Scientific Explanation* (New York: The Free Press).

Jaynes, E.T. (1965), "Gibbs vs Boltzmann Entropies", *American Journal of Physics* **33**, 392. Reprinted in E.T.Jaynes (1983), *Papers on Probability, Statistics and Statistical Physics* (Dordrecht: Reidel), pp. 79-88.

Krim, J.(1996), "Friction at the Atomic Scale", *Scientific American* (Oct. 1996), 48-56.

Lanczos, C. (1970), *The Variational Principles of Mechanics*, 4th ed. (N.Y.: Dover)

Landauer, R. (1961), "Irreversibility and Heat Generation in the Computing Process", *IBM Journal of Research and Development* **3**, 183-191.

-- (1992), "Information is Physical", *Proceedings of PhysComp 1992* (Los Alamitos: IEEE Computer Society Press), pp. 1-4.

-- (1996) "Minimal Energy Requirements in Communication", *Science* **272**, 1914-1918.

Lebowitz, J. (1993), "Macroscopic Laws, Microscopic Dynamics, Time's Arrow and Boltzmann's Entropy", *Physica A* **194**, 1-27.

Lubkin, E. (1987), "Keeping the Entropy of Measurement: Szilard Revisited", *International Journal for Theoretical Physics* **26**, 523-535.

Parke, H.G. (1988), "Letter to the Editor", *Scientific American* **258**, 5.

Reif, F. (1981) *Fundamentals of Statistical Mechanics* (Auckland: McGraw Hill).

Ridderbos, T.M. and Redhead, M.L.G. (1998), "The Spin Echo Experiments and the Second Law of Thermodynamics", *Foundations of Physics* **28**, 1237-1270.

Shenker, O.R. (1999), "Maxwell's Demon and Baron Munchausen: Free Will as a *perpetuum mobile*", *Studies in the History and Philosophy of Modern Physics* **30**, 347-372.

-- (1997) *Maxwell's Demon* (PhD Dissertation, The Hebrew University of Jerusalem).

-- (2000) "Some Philosophical Remarks on Interventionism", *forthcoming*.

Shizume, K. (1995), "Heat Generation Required by Information Erasure", Review of Physics E 52, 3495-3499.

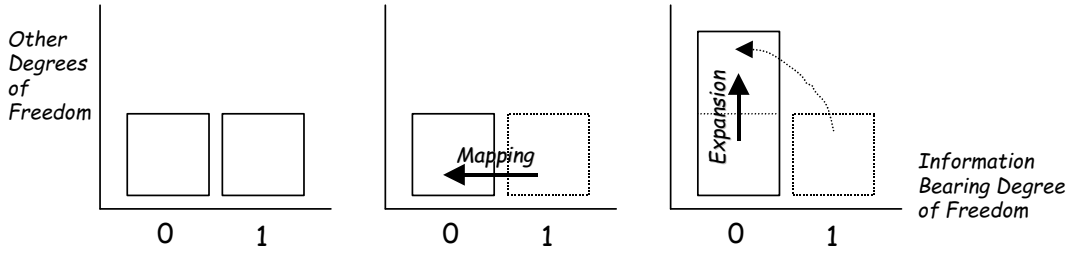
Simon F.E., Kurti N., Allen J.F. and Mendelssohn K. (1952), *Low Temperature Physics* (London: Pergamon Press)

Sklar, L. (1973), "Statistical explanation and ergodic theory" *Philsophy of Science* **40**, 194-212.

Sklar, L. (1993), *Physics and Chance* (Cambridge: Cambridge University Press).

Tolman, R. C. (1938), *The Principles of Statistical Mechanics* (N.Y.: Dover).

Tribus, M. and McIrvine E.C. (1971), "Energy and Information", *Scientific American* **225**, 179-188.

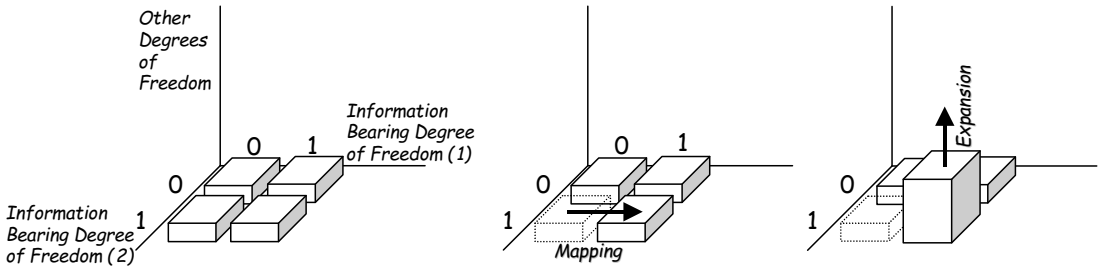


Stage 1:
Before erasure

Stage 2:
Mapping 0&1 to 0. This stage **never actually occurs** (in LDT); stage 1 is directly transformed into stage 3.

Stage 3:
After erasure, Expansion.

Figure 1



Stage 1:
Before "OR"

Stage 2:
Mappin: <0,0> to <0,0>; <1,0> to <1,0>; <1,1> ***and*** <0,1> to <1,1>. This stage **never actually occurs** (in LDT); stage 1 is directly transformed into stage 3.

Stage 3:
After "OR": distribution depending on details of mapping.

Figure 2

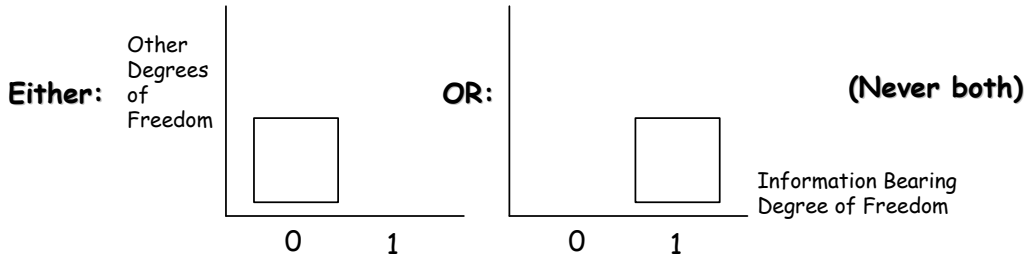


Figure 3

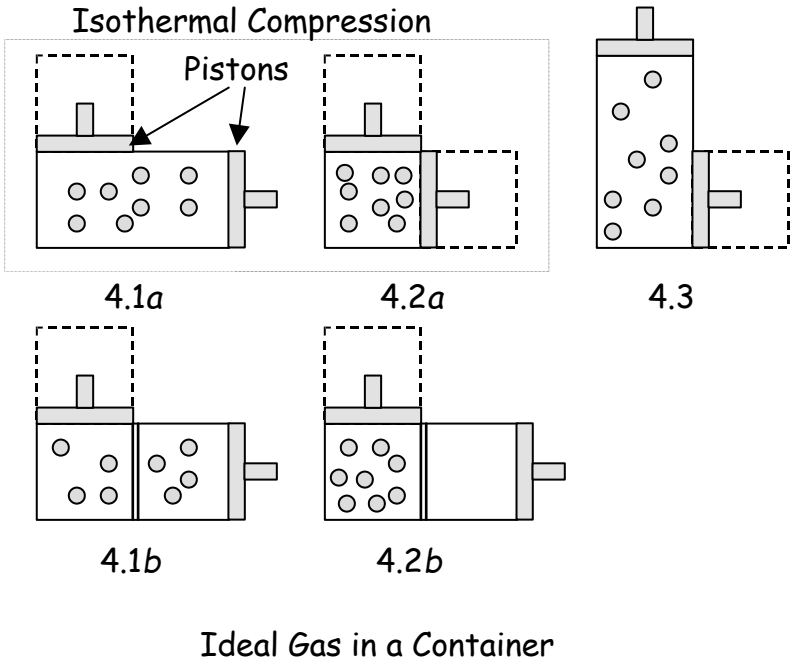

Isothermal Compression
Pistons
4.1a
4.2a
4.3
4.1b
4.2b
Ideal Gas in a Container

Figure 4

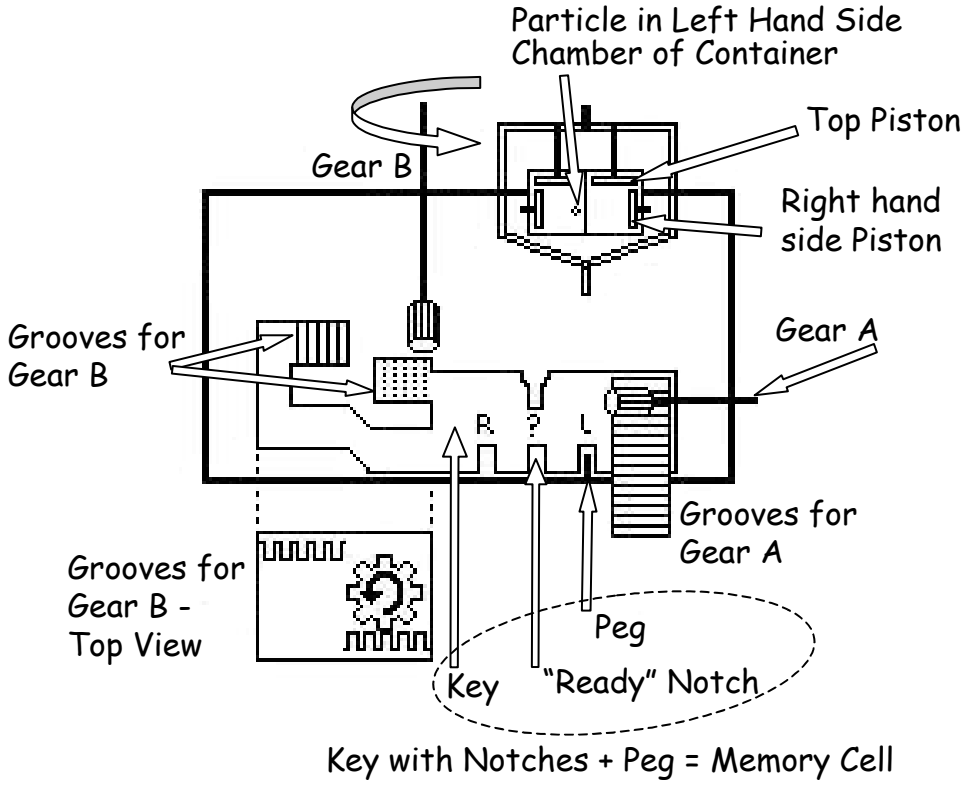

Particle in Left Hand Side Chamber of Container
Top Piston
Gear B
Right hand side Piston
Grooves for Gear B
Gear A
R
?
L
Grooves for Gear A
Grooves for Gear B - Top View
Peg
Key
"Ready" Notch
Key with Notches + Peg = Memory Cell

Figure 5