Pergamon

Stud. Hist. Phil. Mod. Phys., Vol. 30, No. 1, pp. 1–40, 1999

Printed in Great Britain
1355-2198/99 $ – see front matter

EXORCIST XIV: The Wrath of Maxwell's Demon. Part II. From Szilard to Landauer and Beyond

*John Earman and John D. Norton**

In this second part of our two-part paper we review and analyse attempts since 1950 to use information theoretic notions to exorcise Maxwell's Demon. We argue through a simple dilemma that these attempted exorcisms are ineffective, whether they follow Szilard in seeking a compensating entropy cost in information acquisition or Landauer in seeking that cost in memory erasure. In so far as the Demon is a thermodynamic system already governed by the Second Law, no further supposition about information and entropy is needed to save the Second Law. In so far as the Demon fails to be such a system, no supposition about the entropy cost of information acquisition and processing can save the Second Law from the Demon.

1. Exorcising the Demon Without. Intelligent Intervention: A Dilemma for Information Theoretic Exorcisms

Whatever may be the threat posed by fluctuations to the Second Law of thermodynamics, the literature has clearly decided that there is a greater threat to be answered, if volume of writing is any gauge of the magnitude of a perceived threat. That threat is the possibility of external intervention in a thermodynamic system by an intelligent agent. The dominant response follows Szilard, at least in broad outline, in urging that there are entropy costs associated with information processing by the intelligence and that these entropy costs are sufficient to protect the Second Law in all cases. Leff and Rex (1990) supply an annotated bibliography of this literature that runs to 42 pages. In it, the rate of publication

(*Received 15 July 1997; revised 18 December 1997*)
* Department of History and Philosophy of Science, University of Pittsburgh, PA 15260, U. S. A. (*e-mail*: jearman+@pitt.edu and jdnorton+@pitt.edu).

PII: S1355-2198(98)00026-4

accelerates as we approach 1990. Seventeen of its pages—40% of the bibliography—are devoted to the 1980s alone. The vigour of this literature has shown no sign of tiring in the years since 1990.[1]

Our thesis in this paper is that information theoretic analyses provide largely illusory benefits: they are either essentially trivial restatements of earlier presuppositions or posits without proper foundation. To sharpen this thesis, we formulate it as a dilemma for any information theoretic exorcism of the Demon. The dilemma is based on the supposition that the Demon, the intelligence intervening, is itself a physical system. Thus the object system upon which the Demon acts and the Demon itself form a larger system and we may ask after the physical laws that govern the behaviour of this combined system. To this extent, the Demon is naturalised. The dilemma has a 'sound' and a 'profound' horn:

> *Dilemma for an information theoretic exorcism of the Demon*: Either the combination of object system and Demon forms a canonical thermal system or it does not. In the first case ('sound'), it follows that the Second Law of thermodynamics obtains (in suitable form) for the combined system so that there can be no total reduction in entropy no matter how the Demon may interfere with the object system, beyond that allowed by the applicable form of the Second Law. This result is automatic and no information theoretic notions are needed to generate it. In the second case ('profound'), we need a new physical postulate to ensure that the Second Law holds for the combined system. Any such postulate, either a general one or one specifically relating entropy and information, requires independent justification. We do not believe that the literature has succeeded in providing such justification. Moreover, there is reason to doubt that any such justification is possible for a postulate couched in terms of the entropy cost of information acquisition and/or processing. For having departed from canonical thermal systems, one must confront systems where anti-entropic behaviour can occur and where there is no natural way to identify a component of the system occupying the role of information gatherer/processor.

This dilemma is really a scheme of dilemmas according to how we identify a 'canonical thermal system' and this in turn fixes the appropriate sense of the Second Law of thermodynamics. We may define a canonical thermal system as one that obeys the standard laws of thermodynamics; or as one that obeys these laws with the Second Law weakened as in the previous part of this paper or we may identify it by specifying a microdynamics known to yield canonical thermal behaviour macroscopically.

For the case of the 'sound' horn of the dilemma, in Appendix 1: 'Deriving Principles of Entropy and Information' we illustrate how the derivation of the Demon's entropic limitations can proceed in the case of a particular choice for the combined systems' microdynamics that ensures canonical thermal

[1] We have not tried to prepare a comprehensive bibliography of the post-1990 literature. However the references given below will guide the reader to the main tributaries.

behaviour. We assume that the microdynamics obeys Hamilton's equations in its phase space and that it further has suitable properties (such as ergodicity and mixing) that ensure that the long term time averages of quantities approach phase averaged quantities. In particular we show how we can derive standard results of the information theory literature that equate an entropy cost with information acquisition and with information erasure.

Whatever interest we may find in these results connecting information and entropy, they serve no essential function in protecting the Second Law. The Second Law is already protected by the *supposition* of the 'sound' horn of the dilemma that our combined system is a canonical thermal system. At best these information theoretic results provide a picturesque way of understanding how it is that the operation of a plausible Demon is defeated. Thus their importance lies in their heuristic value in aiding the understanding of demonic interventions. And in this regard, their utility cannot be universal. For there are well-known cases of mechanical-demonic intervention in which the notion of information has no natural role. They are typified by Smoluchowski's one-way valve. It is a kind of mechanical demon since it is intended to pass faster molecules only in one direction. To be sure, we may conceive of the response of the valve's flapper to molecular impact as a kind of information gathering; and we may then somehow convince ourselves of an entropy cost associated with the information. But are our intuitions not better satisfied by Smoluchowski's beautifully simple observation that the thermal agitation of the flapper would nullify its discriminating power and defeat its demonic action?

Let us now consider the 'profound' horn of the dilemma. In so far as the literature accepts this horn, its burden is to provide a demonstration of the principles connecting information and entropy that are used to exorcise the Demon. We shall examine this literature more closely below to see the extent to which such demonstrations are provided. But we shall search in vain for an adequate, independent justification of these principles.

We do not believe that this lack of adequate justification can be readily remedied, for once we leave the realm of canonical thermal systems, it is easy to display demonic devices that induce anti-entropic behaviour. Skordos (1993) has produced a time-reversible and energy conserving microdynamics for a two-dimensional system of disks plus membrane. The membrane can be thought of as a Maxwell Demon because it produces a density differential which can be then used to run a perpetual motion machine of the second kind. In keeping with the Zhang and Zhang (1992) analysis, the dynamics of the Skordos disks does not leave phase volume invariant. Skordos himself makes a valiant but, we think, misguided attempt at exorcism by information-theoretic concepts. He imagines that the membrane is replaced by a tennis playing Demon who is able to place his racket at only a finite number of fixed locations so as to deflect the disks. This Demon gives rise to an irreversible dynamics since different trajectories are mapped on top of one another. Only in the limit when the spacing of the racket positions is allowed to go to zero does the dynamics go over to a reversible-but-non-invariant dynamics. Here is where Skordos (1993,

p. 783) thinks that information concepts provide an exorcism:

> [T]he reversibility of trajectories that is achieved in the limit comes at the expense of requiring the demon to operate with infinite information. Because Maxwell's demon can only operate with finite information (we can think of it as a microscopic computer) it follows that the tennis demon cannot imitate the membrane [...] reversibly.

We submit, however, that this anthropomorphising of the Demon is a mistake, and it is a mistake that runs throughout the Maxwell Demon literature. A membrane of finite thickness can be realised by an inanimate, non-anthropomorphic force field of the Zhang and Zhang type, as is shown in detail in Appendix 2. That the equations of motion for this model cannot be given in Hamiltonian form may lead us to doubt whether the field may be realised in nature. But that should not preclude it from our consideration here. Rather the issue is whether a valid principle concerning the entropy costs of information acquisition and processing can defeat demonic devices. It is difficult to see how such a principle could be applied in any non-contrived way to the system described in Appendix 2 since there is nothing in the system that can plausibly be taken to be a subsystem that acquires and manipulates information about the rest of the system.[2] And even if it did apply, such a principle, to succeed, would have to presuppose the non-existence of the non-Hamiltonian force fields used to produce the anti-entropic behaviour. Such a presupposition is quite substantive; a restriction to Hamiltonian systems is already a major part of what is needed to characterise the microdynamics of canonical thermal systems, for which no further restriction is needed to rule out anti-entropic behaviour. Again it is difficult to see how general truths about information acquisition and processing can delimit the dynamics in this way. Indeed, the thrust of Szilard's strategy suggests that the direction of explanation should go the opposite way: as a naturalised object, the Demon is subject to a naturalised version of information theory, where what information the Demon can acquire and process at what entropy cost is dictated by the details of the particular physics of the system of which the Demon is a part.

[2] Zhang and Zhang (1992) are clear about the irrelevance of information concepts to the one-way valve they base on a velocity-dependent force field:

> In our approach, there is no explicit formulation of how information about the movement of the particles is gathered, processed or used to control a door of some kind. Instead, we consider how the motion of the particles is affected by the velocity-dependent force fields such that overall effects of Maxwell's demon are achieved. (p. 4598)
>
> [I]n our models the demon is implemented simply as velocity-dependent force fields, which do not have internal states for information storage and processing. It seems rather natural to regard these models as purely mechanical devices. (p. 4601)

We agree with their assessment. Of course, information concepts are so broad and elastic that we do not doubt that a clever enough analysis can see information gathering and processing going on. But such an analysis would seem to us to add nothing essential while detracting attention from the correct explanation.

While there may be no universal principle relating an entropy cost to information processing, it remains open whether such a principle obtains in the restricted context of particular theories. A persistent but minority view holds that quantum mechanics provides such a context. We will register our dissatisfaction with this view in Section 3.

Since we take our dilemma to be evident, we owe an explanation for why it has been generally ignored. In part the answer lies in the fact that many researchers have seen it as an opportunity rather than a problem.[3] Following Szilard's lead, one tentatively assumes that the Second Law is secured from Demons and then one deduces what the hidden entropy cost of demonic operation must be. If one can find an independent justification for this cost, one then posits it independently and infers back to the protection of the Second Law. It is our perception that this research programme has been a disappointment if not an outright failure and this leads us to urge that the dilemma be taken seriously as a dilemma rather than an opportunity.

2. The Information Theoretic Exorcisms

We now turn to the task of deciding how the various information theoretic exorcisms fare in the face of this dilemma. While we cannot review all the exorcisms, none of the ones we review provide an escape from the dilemma and we have no confidence that others might.

Broadly speaking, there are two approaches. The first, embodied most clearly in von Neumann's and Brillouin's work, sees an entropy cost in information acquisition. Its basic postulate is that gaining information that allows us to discern between n equally likely states must be associated with a minimum entropy cost of $k\log n$, the entropy dissipated by the system that gains the information. Because this thesis is most commonly attributed to Szilard, we shall name it '*Szilard's Principle*'. The second approach, based on an idea of Landauer, sees an entropy cost in the erasure of the memory devices that stored that information. In erasing information that can discern n states, we dissipate at minimum entropy $k\log n$. We call this '*Landauer's Principle*'. This principle restricts the Demon's operation in so far as he is constrained to complete a cycle of operations. If the Demon records any information, such as the position of the molecule in Szilard's engine, in a physical storage device, then this information must be erased to complete the cycle. Landauer's principle imposes an entropy cost.

These two principles can be converted into assertions about minimum amounts of energy degradation in a special case. Imagine that a naturalised Demon can exchange heat with just one heat reservoir at temperature T and that the Demon's operation is optimal: it employs only reversible processes.

[3] We are grateful to an anonymous referee for this helpful elucidation.

Szilard's Principle asserts in this case that there must be a minimum thermal degradation of $kT\log n$ work energy when the Demon discerns among n equally likely cases. Landauer's principle asserts in this case that there must be a minimum thermal degradation of $kT\log n$ work energy when the Demon erases information that can discern among n cases.[4]

Our gloss suggests a greater degree of unanimity in the literature than there can be. There are many subtle and some not so subtle variants of these main themes. One line of analysis (see Section 3 below) holds that the information theoretic exorcisms of the Demon succeed only as long we allow for the quantum character of matter.

2.1. Szilard's Principle

Von Neumann (1932, pp. 398–400) gave one of the earliest statements of a form of Szilard's Principle. In considering Szilard's one-molecule engine, he noted that information on the position of the molecule in its chamber could be converted into an entropy reduction. Knowing the molecule's position, we insert a partition spontaneously compressing the gas to a lower entropy state. However in the footnote quoted earlier he attributed to Szilard the essential result that protects the Second Law: the entropy reduction is compensated by the entropy cost of acquiring this information. How does von Neumann's analysis fare against our dilemma? The difficulty is that von Neumann gives no hint as to the foundation of Szilard's Principle. If we assume he follows Szilard in its generation, then the principle is derived from the *assumption* that the total system obeys the Second Law in a suitable form. Thus the protection of the Second Law by Szilard's Principle amounts to an elaborate and indirect statement of 'If the Second Law holds, then the Second Law holds!' This corresponds to the 'sound' horn of the dilemma. If we presume that von Neumann intends to adopt the 'profound' horn with Szilard's Principle postulated independently, then we are left with no justification for it. A citation to Szilard is no assistance, since Szilard assumes the Second Law in his analyses.

Brillouin (1953, pp. 1152–1153) describes the simplest imaginable demonstration of Szilard's Principle and, because of its intuitive simplicity, it is, we believe, the one that lies behind the broad acceptance of Szilard's Principle. If we have a thermodynamic system whose state corresponds to W equiprobable microstates, then its entropy S is given up to an additive constant by the celebrated

[4] One sees this as follows. If the Demon dissipates entropy ΔS in a cyclic operation, then it follows from the classical thermodynamic definition of entropy that the Demon has communicated heat $T\Delta S$ to the reservoir in which the entropy increase occurs. Since the process is assumed cyclic, the internal energy of the Demon is unaltered. Therefore we must seek the source of this heat energy externally. Since we have presumed that the Demon exchanges heat with no other reservoir, the only source for this energy is work energy derived from elsewhere. That is, in this special case, dissipation of entropy ΔS must be associated with the degradation of an amount of work energy $T\Delta S$ to heat energy.

Boltzmann Principle

$$S = k \log W. \tag{1}$$

In a process that transforms the system from state 0 to state 1 and reduces the corresponding equiprobable microstates in number from W_0 to W_1, we represent the associated information as:

$$I = k \log W_0/W_1, \tag{2}$$

where the sign of (2) has been chosen to assure that a reduction in the number of equiprobable microstates corresponds to a positive quantity of information I. If we interpret the transition from state 0 to state 1 as exploitation of the information I, then we read directly from (1) and (2) that this use of information is associated with a reduction in the entropy of the system equal to I:

$$S_1 - S_0 = k \log W_1/W_0 = -\, k \log W_1/W_0 = -\, I. \tag{3}$$

In Brillouin's preferred terminology, this is understood as a conversion of the information I into 'negentropy' (negative entropy):

$$I = -\,(S_1 - S_0) = (\,-\,S_1) - (\,-\,S_0). \tag{4}$$

So far, we have essentially a matter of definition; Brillouin's analysis is unobjectionable in so far as we conceive the quantity of information defined in (2) merely as a way of redescribing the corresponding change of entropy. To be of use in exorcising Maxwell's Demon, conversion of information and entropy must be subject to constraints. They are given by Brillouin (1953, p. 1153) as

> Any experiment by which an (sic) information is obtained about a physical system corresponds *in average* to an increase of entropy in the system or in its surroundings. This average increase is always larger than (or equal to) the amount of information obtained. In other words, an information must always be paid for in entropy, the price paid being larger than (or equal to) the amount of information received. Vice versa, when the information is lost, the entropy of the system increase[s] [Brillouin's emphasis].

This crucial constraint is immediately recognisable as a statistical form of the Second Law of thermodynamics. It asserts that, on average, the total entropy of a closed system is non-decreasing. Therefore a decrease in entropy in one of its parts must be accompanied by an increase in entropy at least as great in magnitude. If we rewrite this last assertion in terms of information, we recover Brillouin's result: information I is accompanied by an entropy cost at least as great in magnitude. In association with (2) this last result is Szilard's Principle.

Brillouin now claims:

> These remarks lead to an explanation of the problem of the Maxwell's demon, which simply represents a device changing negentropy into information and back into negentropy.

That is, the Demon acquires information about the gas system; there is an associated increase in entropy, which is equivalent to a loss of negentropy. The information is then used in a process that reduces the entropy of the gas; the information is converted to a corresponding quantity of negentropy.

Brillouin's (1962, Ch. 12) later treatment extends these ideas. In particular, he distinguishes 'bound information' as a special case of 'free information', where bound information is encoded as the Boltzmann-type complexions of some physical system. As a result, Brillouin is able to split the total entropy of a system into the combination of a quantity of entropy S and a quantity of negentropy I_b, corresponding to bound information associated with the system. So this total entropy can be represented as $(S - I_b)$. Since this quantity is in effect the total physical entropy of the system, Carnot's Principle asserts that this quantity is non-decreasing for a closed system.

Since Brillouin employs the Second Law to impose the basic constraint on information and negentropy conversion, he has clearly chosen the 'sound' horn of our dilemma. As a result it is hard to agree with his suggestion that the notion of information explains the failure of Maxwell's Demon to effect net entropy reduction. That the Demon fails is surely explained in this account by the presumption that the Demon and gas system together are closed and obey the Second Law. Thus, if the Demon supplies negentropy to the gas, there must be a compensating entropy increase in another part of the system. *That* explains the Demon's failure.

Brillouin's labeling of the quantity in (2) as 'information' is intended, of course, to suggest our everyday notion of information as knowledge of a system. But those anthropomorphic connotations play no role in the explanation of the Demon's failure. All that matters for the explanation is that the quantity I is an oddly labelled quantity of entropy and such quantities of entropy are governed by the Second Law of thermodynamics. The anthropomorphic connotations of human knowledge play no further role. And that is good, for, in so far as they become essential, the analysis would become vague and intractable. Thus, in his later (1962, Ch. 12), Brillouin tries to couple a principle that non-physical, free information can only decrease with Carnot's Principle in order to yield a generalisation of Carnot's Principle. For the generalised principle to have as precise a meaning as the original principle, one would want a precise sense in which the negentropy of free information is interconvertible with that of bound information—much as the entropy of the surrounding air is transformed into that of a melting ice block if the melting is reversible. Brillouin's examples do not provide such precision. One example describes someone hearing a sound as the transformation of the bound information in the sound wave to the free information in the person's mind.

Here we concur with Denbigh's (1981, pp. 112–115)[5] critique of Brillouin's 'Negentropy Principle' that asserts the reversible interconvertibility of

[5] See also Denbigh and Denbigh (1985, §5.4).

information and negentropy:

> This has encouraged the idea [...] that almost any form of 'information' is freely interconvertable (after appropriate change of sign) into thermodynamic entropy, or vice versa. My own view is that the interconvertability thesis can be maintained only in certain special cases where it becomes trivially true, due to the reference to 'information' being unnecessary.

Denbigh proceeds to explain that the trivial truth lies in the fact that Brillouin's '[...] "bound information" is really nothing more than a *name* given by Brillouin to [an] entropy change [...]' [Brillouin's emphasis].

This claim is reinforced by a paraphrase of Brillouin's exorcism of Maxwell's Demon in which the notion of information is not invoked. It succeeds handily. Brillouin introduced the notion of information by renaming certain quantities of negative entropy as information. Information can be eliminated by reversing the renaming operation! This ease of elimination should come as no surprise. As we have emphasised, Brillouin's exorcism succeeds because he assumes that all the systems—including the Demon—are canonical thermodynamic systems. Thus they can yield no total entropy change that violates the Second Law, no matter how we rename the quantities of entropy involved.

Brillouin's treatment based on (1) and (2) is otherwise essentially sound. However it suffers two technical defects. First, Boltzmann's equation has been the subject of serious complaint; it is only meaningful if the context of its application is precisely specified (see Khinchin, 1949, p. 142). Second, statistical mechanical considerations can only support Szilard's Principle for the average of many trials, just as they can only support the Second Law statistically. However, attention to technical details can resolve both problems so that unobjectionable versions of Szilard's Principle can be derived, as is shown in Appendix 1: 'Deriving Principles of Information and Entropy'.

Biedenharn and Solem (1995, pp. 1227–1229) have suggested that there is a contradiction between the Third Law of thermodynamics and the equating of information and negentropy. The two cannot be identified, they argue, since they have different temperature sensitivities; information is presumably not temperature sensitive, but the Third Law requires the entropy of all systems to approach zero as the absolute zero of temperature is approached. To illustrate this objection, consider the one-molecule gas of Szilard's engine as the temperature approaches absolute zero. In reducing the volume of the cylinder by half, we gain one bit of information, $I = k \log 2$, on the position of the molecule. But the corresponding entropy change is $S_1 - S_{1/2} = k \log 2$ only for much greater temperatures. (S_1 is the entropy of the expanded one-molecule gas; $S_{1/2}$ is its entropy upon isothermal contraction to one half volume.) As we approach absolute zero, the entropy change $S_1 - S_{1/2}$ itself approaches zero.

The objection is weakened but not eradicated if we employ Brillouin's definition (2) of information, since a direct application of that definition does not return the value $k \log 2$ for the low temperature case. Since $S_1 - S_{1/2}$ approaches zero, it follows from (1) that the number of Boltzmann-style complexions,

W_1 and $W_{1/2}$, corresponding to the two states must approach one another. That is $W_1/W_{1/2} \to 1$. This behaviour cannot be secured for classical kinetic gases. We need to consider other types of matter to realise this consequence of the Third Law. For example, a quantum gas of a single particle exhibits this property in so far as the particle in both full and half-sized cylinder is increasingly likely to be in its lowest energy eigenstate as the temperature nears absolute zero. In each case, then, there will be just one complexion in which the particle is most likely to be. Thus W_1 and $W_{1/2}$ will each approach 1 and, if the Third Law is to be satisfied, in such a way that $W_1/W_{1/2}$ approaches 1 as well. Applying (2), it automatically follows that the associated information, $I = k \log (W_1/W_{1/2})$, also approaches zero, exhibiting the same temperature dependence as negentropy. Biedenharn and Solem's intriguing objection is not completely eradicated, however. It now survives in the following counterintuitive result for Brillouin's notion of bound information. Consider a quantum mechanical, one-particle gas in its lowest energy eigenstate, trapped in some spatial volume and the same gas trapped in half the volume. Even though there has been a localisation of the particle, there is no information associated with the localisation.

Brillouin's work is better remembered for his detailed analyses of the operation of Maxwell's Demon, all leading to the conclusion that the Demon cannot succeed. The simplest and best known example is his analysis in Brillouin (1951). He imagines a Maxwell Demon seeking to determine whether to open his trapdoor. In order to operate successfully, he must detect molecules within the chamber. Brillouin supplies a mechanism for detection: a torch emits light quanta from a heated filament. They are reflected off the approaching molecule and enter a detector. The decisive condition is that the quanta must be sufficiently energetic to be visible above the background thermal radiation. The resulting dissipation of the quanta's energy is easily shown to involve an entropy cost that outweighs any entropy reduction available from the operation of the trapdoor. This is the simplest example of a class of thought experiments of this type pursued by Brillouin (1951; 1953; 1962, Ch. 13) and Gabor (1951).

The principal difficulty with these demonstrations is that they do not provide a general proof. Rather they are suggestive examples. What they suggest is that if one models any demonic system sufficiently realistically, hidden entropy costs will be revealed that eventually defeat the Demon. The real force of the demonstrations lies in the induction to this latter general claim. (And we shall see below that counterexamples have been offered that challenge the induction.) In any case, Brillouin's induction does little to escape our dilemma. In so far as the demonic sensing apparatus is a canonical thermal system, then its failure is inevitable. Any appearance to the contrary must result from an erroneous analysis. The induction merely gives us weak inductive support for a result that can be had directly from the knowledge that the sensing apparatus is canonically thermal. If the demonic sensing apparatus is not canonically thermal, then the failure of one or other scheme cannot be parlayed into the general conclusion that no Demon is possible. Once we forgo the restriction to canonical thermal systems, we know already that demonic systems are possible. Our Appendix 2 has an example. At best, we might look for support for the idea that some wider class of Demons, but not all conceivable Demons, must fail. But that possibility would require a clear characterisation of the class in question. Since Brillouin's analysis is distinctive in calling on the *quantum* character of light, one possibility is that Brillouin's induction intends to show that no *quantum* Demon can succeed. But, as we shall see in Section 3 below, he explicitly disavows that conclusion.

Brillouin's analysis has been influential. Unfortunately some of the work it promoted proves to be incoherent. Rodd (1964), for example, seeks to demonstrate 'a definite physical relation' (p. 145) between entropy and information using Brillouin's devices. He proceeds from two definitions. The entropy S of a system is given as (1). Using standard information theoretic viewpoints, Rodd then concludes that the 'uncertainty' U associated with the same state is

$$U = k \log W\,. \tag{5}$$

In some process that takes a system in state 0 to state 1, the associated information I is just the reduction in uncertainty so that

$$I = U_0 - U_1\,. \tag{6}$$

It now follows immediately from (1) and (5) that this information I equals the change in negative entropy $(S_0 - S_1) = (-S_1) - (-S_0)$:

$$S_0 - S_1 = I. \tag{7}$$

Thus far, Rodd's analysis is unobjectionable. Rodd now argues that the equality of (6) must in general be replaced by an inequality: 'less than or equal to'. The argument recapitulates Brillouin's analysis of a Demon detecting molecules with a torch and thereby dissipating entropy in an irreversible process. This extra entropy dissipation of the detection system, apparently, is to be included in the left hand side of (7), supposedly generalising (7) to both reversible and irreversible processes. A simple rearrangement of the modified form of (7) gives 'a generalised second law of thermodynamics'

$$\Delta(S - U) \geqslant 0\,. \tag{8}$$

This last phase of analysis withstands all attempts at a charitable reading. Since, according to their definitions (1) and (5), S and U are state functions, we must have for all possible states that $S = U$. From this equality, one recovers trivially for all processes—reversible or irreversible—that

$$\Delta(S - U) = 0\,. \tag{9}$$

Rodd's replacement of the equality of (7) with an inequality is without basis. It commits at least two errors. First, it misreads the terms S_0 and S_1 in (7). They represent the entropy of the systems in states 0 and 1 alone. Rodd's argument somehow requires the term $S_0 - S_1$ to include entropy dissipated by other systems. Second, and more seriously, Rodd arrives fallaciously at an inequality in (8) because he ignores the change in uncertainty U that must accompany any entropy change in these other systems.

Given these deficiencies, there is little point is confronting Rodd's analysis with our dilemma. We will remark, however, that Rodd would appear to accept the 'profound' horn of the dilemma, in so far as he seeks 'a generalized second law of thermodynamics' (8). However the analysis is fully within the 'sound' horn. All the systems considered are canonical thermal systems; the only supplement is a definition of U and of I.

Raymond (1950, 1951) provides a treatment of the relationship of information and entropy that is superficially akin to Szilard's Principle, except that its interpretation of the entropy involved is rather different. In his view, information is used to cause a thermodynamic system to move away from equilibrium to a metastable but non-equilibrium state. His basic postulate involves an information theoretic extension of the customary notion of entropy to include non-equilibrium systems. He writes (1950, p. 273):

> [...] the entropy of a system may be defined quite generally as the sum of the positive thermodynamic entropy which the constituents of the system would have at thermodynamic equilibrium and a negative term proportional to the information necessary to build the actual system from its equilibrium state.

Raymond (1951) applies this notion to a simple demonically operated molecular engine. A molecular gas is enclosed in two communicating chambers whose connection can be closed by a door. When a pressure fluctuation generates a sufficient pressure differential, the door is shut, preserving the pressure differential. The resulting system clearly has a lower entropy than the initial system. Raymond portrays the reduction in entropy as arising from the storage of information in the system which is now in a non-equilibrium state. It is represented quantitatively as $-\ln P$, where P is the probability that the pressure fluctuation would arise on any random closing of the door. It turns out that the entropy reduction computed by ordinary thermodynamic means agrees with the information stored up to the multiplicative factor k.[6]

Thus far, we have no assurance that the demonic machine fails to achieve an entropy reduction. This possibility is blocked by a further assertion. Raymond now considers the further physical processes that bring about the storage of information in the system. Presumably this is the observer and apparatus used to close the door between the two chambers. 'No observer yet considered', he reports (p. 141), 'has proved capable of storing information in any system without degrading an amount of energy sufficient to make the total entropy change in a system, including the observer, positive. The Second Law is therefore not in danger through the treatment of information as a form of negative physical entropy'.

[6] This agreement will come as no surprise to those who know how to manipulate the Boltzmann equation $S = k \log W$.

In terms of our dilemma, Raymond has apparently chosen the 'profound' horn in addressing systems he regards as non-equilibrium. The exorcism of the Demon then depends on the above quoted principle, which is given without further justification. Raymond (1950, p. 275) gives some further justification in considering the example of the entropy cost of information transmission through electrical signalling. However the general result is not established. He concedes: 'If the theories of other communications devices, such as reading and speaking, were as well developed as that of electrical communication it might be possible to develop quantitative treatments of such information storage devices as the Maxwell demon [...]'.

While Szilard's Principle has been much celebrated in the literature, it has also been subject to persistent challenge. It has been suggested repeatedly that some Demons incarnated as mechanical contrivances might after all be able to acquire information without the requisite entropy cost. (These are the counter-examples to Brillouin's induction mentioned above.) These proposals concentrate on Szilard's one-molecule engine. The most intriguing challenge is due to Gabor (1951). He describes a scheme in which one end of a long chamber is illuminated by a circulating stream of nearly monochromatic light of sufficient intensity to rise above the background thermal radiation. The presence of the molecule is revealed when the molecule scatters some of this light into a photo-sensor. Gabor argues in detail that his device is able to detect the molecule at an entropy cost less than the entropy reduction recovered in the expansion of the gas—as long as the light behaves classically. If the light is treated quantum mechanically, then this result fails. Curiously, Brillouin's (1962, Ch. 13, pp. 179–181) re-analysis of Gabor's machine concludes that the machine must fail. The analysis appears to invoke no special assumption about light that would distinguish between the classical and quantum mechanical case, so that Brillouin's analysis contradicts Gabor's claim that his machine succeeds if the light has classical properties.

Other proposals abound, although they are less detailed in their analysis. Chambadal (1971) considers the moment in the cycle of the Szilard one-molecule engine at which a piston is at the midpoint of the cylinder with the molecule in one or other side.[7] If one now inserts two rods to touch the piston, then, as the one-molecule gas expands, the direction of expansion and location of the molecule is revealed by the motion of the rods. Jauch and Baron (1972) polish the proposal by replacing the rods with electrical contacts that are closed by the piston once it starts to move, revealing its direction of motion to the mechanism that is to extract work from the expanding gas. Bennett (1987, p. 114) has described another apparatus designed to locate the molecule in a Szilard one-molecule engine. Two pistons are lowered in each half of the cylinder in such a way that the pressure from the molecule will tip a delicately

[7] Not having access to Chambadal (1971), we proceed from the account and quote in Leff and Rex (1990, p. 25). See Leff and Rex (1990, p. 25) also for further related proposals from Popper and Feyerabend.

balanced keel, attached to the pistons. That tipping purportedly reveals the position of the molecule without entropy cost. (Bennett (1982, p. 240) describes an electronic analog of this device in which the location of a diamagnetic particle is revealed when it flips the state of a bistable ferromagnet.) The difficulty with Bennett's proposal is that the mechanical keel system described is an ordinary mechanical device that would be governed by a Hamiltonian mechanics. As a result we must presume that it would behave like a canonical thermal system. That would mean that it would be subject to the usual fluctuation phenomena. Intuitively, these fluctuations would arise as a wild rocking of the keel resulting from its recoils upon each of the many impacts with the molecule of the gas. If the keel is light enough to be raised by the pressure of the one-molecule gas, then it must have very little inertia and such rocking is to be expected. Presumably this wild rocking would obliterate the keel's measuring function. Similar fluctuation problems would trouble electronic analogues of this device. Whether these fluctuations problems would defeat Chambadal's proposal and Jauch and Baron's is unclear because their descriptions do not give sufficient detail of the complete apparatus proposed.

2.2. Landauer's principle

On the account of Bennett (1987, pp. 115–116; 1988, pp. 282–283) and Leff and Rex (1990, pp. 21–29), the 1980s saw a major shift in the strategy used to exorcise Maxwell's Demon, based on developments in the thermodynamics of computation. While agreeing that entropy reducing demonic interventions will be nullified by a compensating entropy increase elsewhere, Bennett (1982, Section 5) urged that the Brillouin tradition had mislocated the locus of entropy dissipation. It is not associated with the acquisition of information, as posited by Szilard's principle, but it arises in the erasure of information, as posited by Landauer's Principle defined above. To use the information it acquires, for example, about the location of a molecule in Szilard's one-molecule engine, the Demon must record that information in some physical memory storage device. If a thermodynamic cycle is to be completed, this memory must be erased. It is in this step, Bennett claims, that there is an inevitable entropy cost. Insofar as there is an official doctrine about the exorcism of Maxwell's Demon, this is it, as evidenced by the endorsements it has received in leading scientific journals and conference proceedings (see, for example, Caves (1993, 1994), Schumacher (1994), Zurek (1989a, 1989b, 1990)) and even by Feynman (1996, pp. 149–150). Bennett (1987, p. 116) summarised the argument for Landauer's Principle:

> Landauer's proof begins with the premise that distinct logical states of a computer must be represented by distinct physical states of the computer's hardware. For example, every possible state of the computer's memory must be represented by a distinct physical configuration (that is, a distinct set of currents, voltages, fields and so forth).
>
> Suppose a memory register of *n* bits is cleared; in other words, suppose the value in each location is set at zero, regardless of the previous value. Before the operation

> the register as a whole could have been in any of 2^n states. After the operation the register can be in only one state. The operation therefore compressed many logical states into one, much as a piston might compress a gas.
>
> By Landauer's premise, in order to compress a computer's logical state one must also compress its physical state: one must lower the entropy of its hardware. *According to the second law*, this decrease in the entropy of the computer's hardware cannot be accomplished without a compensating increase in the entropy of the computer's environment. Hence one cannot clear a memory register without generating heat and adding to the entropy of the environment. Clearing a memory is a thermodynamically irreversible operation [our emphasis].

Bennett's synopsis agrees with Landauer's own development such as in Landauer (1961, §4). It makes clear that Landauer's Principle depends on some very definite assumptions about the physical process of memory erasure. That is, the memory registers are thermalised for at least part of the erasure process. Without this thermalisation, the Second Law of thermodynamics could not be invoked and the entropy cost of erasure assessed. For example, we may represent n bits of information with the position of n particles in n chambers; a zero might be represented by the particle in the left side of its chamber and a one represented by the particle in the right side. To erase some memory state, we allow the particles to gain thermal energy, if they do not already have it, and then release them so they can move freely to both sides of their chambers with that thermal energy. The erasure is completed by compressing the thermalised particles to the zero state of all particles in the left of their chambers. An isothermal, reversible compression would release heat from the memory device and correspondingly reduce its entropy. There would be a compensating increase of entropy in the environment that absorbs the heat released during compression.[8]

In terms of our dilemma, this treatment again chooses the 'sound' horn. Its central principle, Landauer's Principle, depends on the assumption that the demonic apparatus is constituted of canonical thermal systems, at least in its crucial elements. Thus the explanation of the failure of the Demon to effect a net entropy reduction lies simply in the assumption that the Second Law governs the Demon as well as the system he acts upon. As before, the value of Landauer's principle to exorcising the Demon is heuristic. This value is limited in cases in which there is no natural way of seeing where information erasure occurs. For example, if we conceive Smoluchowski's one-way valve as a mechanical Demon, we could probably contrive to find a sense in which there is memory erasure in

[8] For a calculation of the quantities of entropy involved, see Part I, Appendix 3. Al Janis has pointed out to us that Bennett's synopsis may have mislocated the most serious entropy cost of the erasure process. In the compression phase, the entropy increase of the environment is compensated by an entropy reduction in the memory device so that the total entropy of the universe stays constant. However, in the first step of the erasure, the thermalised particles are released from confinement to one or other side of their chambers. This release corresponds to an irreversible expansion. The system's entropy increases without any compensating decrease in the entropy of the environment, so that the total entropy of the universe increases.

its motion. Heuristically, however, its failure to effect a net reduction in entropy is better explained by the presumption that the Second Law governs the operation of the valve itself and that its operation is compromised by fluctuations.

There are two further problems with Bennett's analysis. The first is an apparent discrepancy in the treatments Bennett accords to Szilard's and Landauer's Principles. While there seems to be no obvious incompatibility between Szilard's and Landauer's Principles, Bennett has gone to some pains to urge that Szilard's Principle fails and that Landauer's Principle replaces it in exorcisms of Maxwell's Demon. Bennett, Landauer and Leff and Rex (see Bennett (1982) and Leff and Rex (1990, pp. 27–29)) seek to explain the long acceptance of a principle they deem false through a widespread error: they argue that Brillouin's widely accepted induction fails. While the systems that he conceived could not effect measurement without corresponding entropy dissipation, they urge that there are systems that can effect measurement without an entropy cost. We have discussed Bennett's proposals for such systems above. These devices can only succeed in so far as we presume that they are not canonical thermal systems. Thus Bennett's logic is difficult to follow. Landauer's Principle is supported by arguments that require memory devices to be canonical thermal systems. But Szilard's Principle is defeated by the expedient of ignoring the canonical thermal properties of the sensing device.

The second problem leads us to doubt the success of Bennett's exorcism. There seems to be a way of programming a computerised Demon so that there are no erasures required. Whether this is so depends upon whether the following operation counts as an erasure. Assume we have a binary state memory device with the two states '*L*' and '*R*'. If the programme knows that the device is in state *L* and then switches this state to *R*, is this switching an erasure? Consistency with the claims of the Bennett–Landauer tradition would say it is not since the process is logically reversible and does not involve the mapping of several states onto one.[9] But if this process is not an erasure, then a computerised Demon can be devised that operates a Szilard one-molecule engine without the need for erasure. To see this, recall that the computerised Demon will, at some point in its operation, need to invoke one of two subprograms: programme-*L* if the molecule ends up on the left side of the partition or programme-*R* if the molecule ends up on the right side of the partition. The position of the molecule and, simultaneously, a record of which subprogramme is to run is held in the binary memory

[9] Landauer (1961, p. 193) addresses this very case:

> When the initial states [of a memory device] are all ZERO and we wish to go to ONE, this is analogous to a phase transformation between two phases in equilibrium, and can, presumably, be done reversibly and without an entropy increase in the universe, but only by a procedure specifically designed for that task.

Notice that in the computerised Demon described in the text, the two segments of the programme that effect the two switchings are specifically designed for just those tasks alone; the first always switches an *L* to an *R*, the second an *R* to an *L*.

device mentioned above. The programme is set up so that the memory device is assuredly in the L state at the starting point in the programme when the position of the molecule is determined. So the portion of the programme that records the molecule's position either does nothing to the memory device (if the molecule is on the left) or switches the state L to the state R (if the molecule is on the right). Neither operation involves erasure. Then, according to the contents of the memory device, programme-L or programme-R is executed. Programme-L leaves the memory register unaltered as it directs the expansion that yields a net reduction of entropy. Programme-R proceeds similarly. However, at its end, programme-R resets the memory register to L. This last resetting is again not an erasure. Programme-R knows that the memory register is in the R state; it just switches it to the L state. With the cycle complete, the memory register has been returned to the L state. Thus both subprogrammes leave the memory device in the L state. No erasure step is needed as preparation for the next cycle of operation. The net effect of each cycle is an entropy reduction of $k \log 2$ in violation of the Second Law.

According to recent work by Zurek (1989a, 1989b, 1990) and Caves (1993), the Bennett–Landauer analysis cannot provide a reliable exorcism of the Demon. It requires a repair provided by the theory of algorithmic complexity and supplied by means of a modification of the very notion of entropy itself. They now define the total physical entropy as $S = H + I$, where H is the Gibbs–Shannon statistical entropy[10] and I is the algorithmic information. The latter is defined as the length in bits of the shortest computer programme that a universal Turing machine needs to generate a description of the relevant state.[11] They claim that this move is essential to exorcising an information processing Demon. Although they employ a modified notion of entropy, we can affirm that their goal in exorcism is the same as Brillouin's and Bennett's. That is, they seek to demonstrate that a computerised Demon cannot violate the Second Law by allowing a thermodynamic cycle whose sole effect is the conversion of heat into work (see Part I, Appendix 1). Thus the Demon of a Szilard engine is conceived in Zurek (1989b, pp. 4743–4744) as operating reversibly at constant internal energy. The net work extracted from heat in the transition from the initial state s_{i} to the final state s_{f} is $\Delta W = T\Delta S = T(S_{\mathrm{f}} - S_{\mathrm{i}})$. The gain due to the change in the statistical entropy is $\Delta W^{+} = T(H_{\mathrm{f}} - H_{\mathrm{i}})$. The information bill for this gain comes from erasing the record r_{i} of s_{i} and introducing the record r_{f} of s_{f}, which at a minimum calls for an erasure of $|r_{\mathrm{i}}^{*}| - |r_{\mathrm{f}}^{*}|$ bits of information, where $|r_{\mathrm{i}}^{*}|$ and $|r_{\mathrm{f}}^{*}|$ are the lengths of the shortest programmes needed to describe s_{i} and s_{f} respectively (see Zurek (1989b)). If this is done in an environment at temperature T, the Landauer principle calls for an energy dissipation of $\Delta W^{-} = T(|r_{\mathrm{i}}^{*}| - |r_{\mathrm{f}}^{*}|)$. Thus, the net gain by the engine plus Demon is $\Delta W^{+} - \Delta W^{-} = T[(H_{\mathrm{f}} - H_{\mathrm{i}}) - (I_{\mathrm{i}} - I_{\mathrm{f}})] = T(S_{\mathrm{f}} - S_{\mathrm{i}}) = \Delta W$, which is taken to justify the definition of total entropy.

[10] The Gibbs–Shannon entropy $H = -k\Sigma_i p_i \ln p_i$ can be regarded as a generalisation of the Boltzmann entropy because in the case where the p_i are all equal (say) to $1/W$, H reduces to $k \ln W$.
[11] This number is the same, up to an additive constant, for all universal Turing machines.

Zurek and Caves show that for such an information processing Demon the net work extracted on average is guaranteed to be non-positive.

Our concern is the foundations of the Zurek-Caves programme.[12] In discussing one example of the programme's work, Caves, Unruh and Zurek (1990) have given us what is apparently an inventory of the programme's basic principles:

> Three ingredients go into Caves' [(1990)] analysis: (i) *Landauer's Principle* […] — to erase a bit of information at temperature T requires dissipation of energy $\geqslant k_B T \ln 2$; (ii) Bennett's […] observation that a complete engine cycle includes returning the demon's memory to its standard state, which requires the demon to pay a Landauer erasure cost, and that this erasure is the only necessary irreversible part of the cycle; and (iii) Zurek's […] realization that the demon can reduce its erasure cost by compressing reversibly its description of the observed submacrostate and, hence, that limits of principle require consideration of the most compact description […].

We see immediately that this programme will be of little assistance with our dilemma, since the programme simply imports Landauer's Principle as a result presumed to be widely accepted. The programme leaves others, apparently, to decide whether Landauer's Principle is to be justified within the confines of the sound or profound horn.[13] Further, the inventory makes clear that it is difficult for us to be optimistic about the programme. We have already expressed our reservations about two of the three 'ingredients'. Landauer's Principle faces our dilemma and, as we indicated above, there seem to be computerised Demons that function without erasure. This is not a firm foundation upon which to lay the third item, Zurek's modified notion of entropy.

For our purposes the most important consequence of the programme is the notion that the third item is needed at all. What this reveals is that the Landauer–Bennett exorcism of the Demon is as fragile as the Brillouin tradition that it overwhelmed. For it took only a modest concern for economy in programming to devise an alternative Demon whose efforts elude exorcism by the unaugmented Bennett–Landauer analysis. Such clever Demons were investigated by Caves (1990). These Demons seek to reduce the Landauer erasure cost by focusing on favourable states and compressing the descriptions used to record information. To illustrate, imagine a compound Szilard engine consisting of two Szilard one-molecule engines. Insert partitions in the middle of each and suppose, following Bennett, that the Demon can make cost free measurements

[12] While Zurek and Caves agree substantially in their outlook, our description of their programme may overstate the degree of agreement. Caves (1994, p. 93) outlines a disagreement with Zurek.

[13] There is some suggestion that Caves and Zurek might endorse the 'sound' horn. Caves (1993) explicitly acknowledges that cases of Maxwell's Demons can be analysed from an 'external' point of view (in which no information concepts are brought into play) as well as an 'internal' point of view (which sees the process from the perspective of the Demon who is gathering and processing information). What we have not been able to appreciate—perhaps a failing on our part—is how the internal perspective adds anything other than sometimes useful and sometimes misleading heuristics.

to determine on which sides of the partitions the gas molecules lie. Our clever Demon asks only: are both molecules on the left or not? A positive answer can be recorded using only one bit of information storage, whose Landauer energy dissipation erasure cost is $kT\log 2$. But the work obtained from the compound Szilard engine by isothermal expansion is $2kT\log 2$, leaving a net gain of $kT\log 2$. Caves *et al.* (1990) think that this reasoning has neglected a hidden erasure cost that emerges, apparently, with a full appreciation of the entropy costs of executing the Demon's algorithm. They say the Demon must record and then erase information that instructs it how to proceed. In our example, a '1', say, tells the Demon to record the presence of the favoured state, to extract work by isothermal expansion, and to erase the record of the favoured state. Otherwise a '0' tells the Demon to remove the partitions and start over. The additional Landauer erasure cost cancels out the apparent net gain in work.

Our reservations about the viability of this programme lie in two areas. First, the programme seems willing to proceed from presumptions that are just inadmissible on their face.[14] Second, even if we accept the entropy cost of erasure asserted in Landauer's Principle, there are computerised Demons whose design can be so economised as to yield long term violations of the Second Law in the sense of allowing the complete conversion of heat to work. Thus, no matter how Zurek and Caves adjust the notion of entropy, their efforts cannot be sufficient to ensure exorcism of the Demon. To see how more efficient programming can yield a Demon they cannot exorcise, consider the case of a compound Szilard engine that employs *three* single-molecule engines. The Demon waits for the favourable case in which all three molecules are trapped on the left side. Aside from erasure costs, in one of eight trials on average, the Demon extracts $3kT\log 2$ of work from heat; in the remaining seven of eight nothing is done. Now Caves *et al.* presume that the programme itself must record in addition to the gas state the decision to use programme-A (= try to extract work) versus programme-B (= do not try to extract work). That is, any computer executing instructions must record in some memory device which one it is executing—either programme-A or programme-B. But cannot the very same memory register that records 'all molecules on the left side' be used to record the decision to use programme-A rather than B? When the computer Demon needs to ascertain which subprogramme it is performing it reads the same memory register as is used to record whether the state is favourable. Therefore the erasure costs are reduced—but not enough. The Demon on average recovers (3/8) $kT\log 2$ work from heat for each cycle. But each cycle dissipates work $kT\log 2$ in erasure costs. These costs, however, can be reduced by relocating all erasure costs to the one in eight cycles that execute programme-A. To do this, we leave the memory device in state B as a default. If the programme finds an unfavourable state, it does nothing to the memory device. Similarly programme-B has no need to reset the device. Resetting and erasure

[14] Thus Caves' (1993) opening sentence proclaims: 'To say that a system occupies a certain state implies that one has the information necessary to generate a complete description of that state'.

arise only in the one in eight cycles in which a favourable state is found. First the memory device is erased and set to A when the favourable state is detected. Then, at the end of programme-A's execution, the memory device is set back to state B. Thus there is a total work dissipation of two erasures[15] in the cycle, that is, $2kT\log 2$. The net effect is that there is no dissipation if an unfavourable state is found; in the one in eight cycles in which a favorable state is found, there is a net recovery of $3kT\log 2 - 2kT\log 2 = kT\log 2$ work from heat. Spread over all cycles this amounts to an average conversion $(1/8)kT\log 2$ heat to work per cycle. The Second Law is violated. Caves (1990) is willing to allow violations of the Second Law from rare fluctuations. A violation on average in one of eight cycles is far from Caves' rare fluctuation, especially since their effects can be steadily accumulated.

These are fun ideas to play with, but we have a hard time believing that the fate of the Second Law turns on such details of computerese. The strategy being pursued by the clever Demon is to wait for an improbable state. If enough waiting time is allowed a spontaneous violation of the Second Law is (almost) sure to occur with no need for an information processing Demon; *a fortiori*, nothing about the thermodynamic cost of memory erasure can prevent the violation. Perhaps then the goal of the game is not to protect the Second Law against straight violations but rather to protect it against embellished violations in which work is continuously extracted from a macroscopic system. Our claim, to repeat it once again, is that such protection can be afforded if the Demon is naturalised as a canonical thermal system and that the proof of protection neither uses nor benefits from information theoretic concepts. And if the Demon is not so naturalised, then we have yet to see a demonstration that he must fail in his efforts to violate the Second Law.

3. Do We Need Quantum Theory to Exorcise the Demon?

The idea that quantum mechanics is needed to exorcise Maxwell's Demon has been a persistent though not widely popular idea. The profound horn of our dilemma asks for some physical principle upon which to base Szilard's and Landauer's Principles. Might we find this principle somewhere in the assertion that matter has a quantum character? The idea is certainly not beyond the pale since quantum mechanics allows for understanding of some thermodynamic phenomena, such as specific heats at low temperatures, and of related general principles, such as the Third Law, which asserts the impossibility of reaching absolute zero. To our knowledge, however, this idea has never been given satisfactory shape. Instead this corner of the literature is distinguished by a superficial agreement that quantum mechanics somehow enables us to exorcise the Demon, but no deeper agreement on precisely which aspect of quantum mechanics underwrites the exorcism. The arguments offered are often weak: they

[15] We set aside here the concern expressed above that both of these erasures are really switchings that have no associated entropy costs.

develop one instance of the Demon's defeat at the hands of quantum mechanics and we are left to infer by induction that such defeat is universal. In some cases the argument advanced is unconvincing or incoherent. Prominent voices in the information theoretic establishment continue to insist that quantum mechanics is irrelevant to the exorcism of the Demon. Worse, we saw in Section 2 that Biedenharn and Solem's (1995) remarks on the Third Law, which obtains in a quantum context, lead to counterintuitive results on Brillouin's equation of information and entropy. With the reader prepared for the worst, we review some of the proposals.

Slater (1939, Ch. 3, Sec. 4) held that Maxwell's Demon is stymied by the Heisenberg uncertainty relations. The demon he had in mind was not a creature who creates a perpetual motion machine of the second kind by sorting atoms but rather a creature who implements Loschmidt's reversibility objection by reversing the velocities of all the atoms. The impossibility of simultaneously determining the positions and velocities of the particles would, Slater thought, make the operation of the latter creature impossible. In fact, however, there are quantum mechanical experiments, such as the spin-echo effect, in which the initial state is recovered without having to ascertain the states of the individual atoms.

As noted in Section 2, Gabor (1951) held that if the principles of classical electromagnetism were correct, the energy dissipation bill for locating the molecule in the Szilard engine can be made smaller than the work gained by isothermal expansion. The quantum nature of radiation, he thought, comes to the rescue because of the need for the Demon to use a sufficiently energetic light quantum to distinguish the molecule against the blackbody radiation in the cavity (see also Demers (1944, 1945)). If correct, this line of reasoning would effectively blunt the profound horn of our profound versus sound dilemma. In some reader's minds this line of reasoning was given currency by Brillouin's (1951, 1953) examples of photon wielding Demons. But as we have already seen, Brillouin's (1962, Ch. 12) reanalysis of Gabor's version of the Szilard engine makes no specifically quantum mechanical assumptions but still claims to show that the Demon cannot violate the Second Law. And Brillouin, who was one of the chief exponents of Szilard's principle as the key to exorcism, specifically disavowed the notion that quantum mechanics was an essential ingredient in the exorcism:

> The limitation to the possibilities of measurement, as contained in our formulas, have nothing to do with the [quantum mechanical] uncertainty relations. They are based on entropy, statistical thermodynamics, Boltzmann's constant, all quantities that play no role in the uncertainty principle. (1951; Leff and Rex 1990, p. 137)
>
> In all our discussions we had to use quantum conditions, because we actually live in a quantized world. When, however, the experiment performed was on the classical level, we noted that Planck's constant h dropped out from the final results, which contained only Boltzmann's constant k. This proves that our 'Negentropy Principle of Information' really is a new principle and cannot be reduced to quantum conditions. (1953, p. 1162)

Bennett (1987), who started the fashion that Landauer's principle is the key to exorcism, is likewise explicit in rejecting the claim that a sound exorcism had to await the advent of quantum mechanics.

Zurek (1984) offers a quantum mechanical treatment of Szilard's one-molecule engine. The purpose, apparently, is to demonstrate how this quantum treatment escapes Jauch and Baron's objection that Szilard's engine requires a violation of the ideal gas law. (We have expressed our belief that such an escape is unnecessary in Part I, Section 10.) Zurek allows that this escape:

> [...] is not too surprising, for, after all, thermodynamic entropy which is central in this discussion is incompatible with classical mechanics, as it becomes infinite in the limit $\hbar \to 0$. (1984, p. 250)

However it is unclear whether quantum mechanics supplies any machinery for exorcism of the Demon that could not be supplied by the classical theory.

The particle in the cylinder is idealised as a solution of the non-relativistic Schrödinger equation in a square potential well. A thermalised particle at temperature T is represented by a canonically distributed ensemble of energy eigenstates and a high temperature approximation is used. The cycle proceeds exactly as the classical Szilard engine. The exception is in the step that corresponds to the reinsertion of the dividing barrier-piston at the cylinder's center. The reinsertion merely divides the system wave function into two components that persist on either side of the barrier. A quantum mechanical measurement operation then collapses the wave to one that fills one side of the cylinder only. This measurement completes the recompression of the gas.

Zurek shows that this recompression corresponds to a reduction in entropy of $\Delta S = -k \log 2$, exactly as in the classical case. His statement of the result is slightly different. The entropy change is represented as a change in free energy F, where $F = U - TS$, for U the internal energy, T the temperature and S the entropy of the gas. (In suitable circumstances, a change of free energy can be understood as a degradation of energy.) The free energy change is $\Delta F = kT \log 2$. This reduction in entropy is what enables the complete cycle to violate the Second Law, so Zurek needs to find a compensating increase in entropy. His proposal is that we find this compensating increase in the system that effects the measurement that collapses the particle to one or the other side of the barrier. After the measurement collapse, Zurek urges, this measurement system—in effect the Demon—is in a superposition of after-measurement states coupled with particles localised on either side of the barrier. To complete the cycle, we must reset the Demon-measuring system to its before-measurement state and, he urges, this resetting corresponds to the erasure of one bit of information. At this point Zurek *assumes* the truth of Landauer's Principle and announces that this resetting operation carries the hidden entropy cost of $k \log 2$ (pp. 257–258). This invocation of Landauer's Principle is the *deus ex machina* that saves the Second Law. We will not quibble over whether the principle is correctly applied here. We do insist, however, that, in spite of all the gestures to quantum mechanics, quantum mechanics plays no essential role in the exorcism. The Demon is exorcised by precisely the same method that Bennett uses for classical systems. In both Bennett's and Zurek's exorcisms, it is memory erasure governed by Landauer's principle that saves the Second Law. We also note that Landauer's principle is assumed and not derived.

Biedenharn and Solem (1995) give an analysis of essentially the same quantum mechanical Szilard engine as Zurek. They too find that understanding the effects of measurement is the decisive step in exorcising the demon but we are unable to follow the details of their analysis. Zurek showed that measurement results in an increase in *free* energy, $F = U - TS$, which in this case is really a disguised entropy change. Biedenharn and Solem, however, claim to show that measurement changes the energy of the gas; observation puts energy into the gas. We have been unable to follow the arguments for this claim and also for the conclusion they draw from it: this energy is the energy recovered during the expansion step, rather than heat energy from a reservoir converted by means of the expansion to work.

Beghian (1994) offers yet another means for quantum mechanics to exorcise Maxwell's Demon. He conceives the gas on which the Demon operates to be a Bose gas, that is, made up of indistinguishable quantum particles. In order to sort fast from slow molecules by manipulating his door, Beghian argues, the Demon must be able to distinguish the fast from slow molecules. But this requires a labelling of all particles and this labelling corresponds to a change in entropy as the statistics move from Bose–Einstein statistics to Maxwell–Boltzmann statistics. This entropy change is the hidden entropy cost that saves the Second Law from the Demon.

We are unconvinced by Beghian's analysis. It is based on a premise dependent on both classical and even anthropomorphic assumptions. Why must we assume that a successfully sorting Demon must be able to distinguish particles in order to succeed in his intended operation? As long as we think of the Demon as a tiny little man nervously scanning for molecules with his hand on the door handle, then the assumption is natural. But sorting Demons can have a quite different character. Zhang and Zhang's sorting Demon is simply a field with special properties. Presumably similar such fields could be described in the quantum context. On what basis would we demand that such fields can only function as intended if the particles are distinguishable? Why should we expect that our intuitions about the needs of tiny little men should carry over into the bizarre world of the quantum?

4. Conclusion

The idea that information theoretic notions hold the key to the exorcism of Maxwell's Demon is at once natural and astonishing. An intelligent Demon must, it seems, acquire information and then process and store it; and so it seems that an explanation of a failure of the Demon to violate the Second Law must appeal to principles which govern these steps. Once one has started down this road one should not be surprised to be led into Bennett's camp:

> The correct answer—the real reason Maxwell's demon cannot violate the second law—has been uncovered only recently. It is the unexpected result of a very different line of research: research on the energy requirements of computers. (1987, p. 108)

But a little reflection reveals why it is an astonishing notion that a proper understanding of Maxwell's Demon had to await the development of information theory in general or research on the energy requirements of computers in particular. First, one of the seeming strengths of the information approach is its generality. But in this strength lies its weakness. How can considerations stated at such a high level of generality distinguish between cases where exorcism is required and cases where the Second Law is simply false? And if such considerations cannot so distinguish, how can they yield a satisfying explanation of the core validity of the Second Law? Second, even if the operation of the Demon involves what deserves to be called an information processor or computer—a dubious assumption in some cases—the ultimate explanation for the possibility or impossibility of various operations must be traced to fundamental physical laws, laws which are stated without the mention or use of information concepts. One thus suspects that at best what information considerations can offer is a handy heuristic. We have sharpened these suspicions into the dilemma posed in Sections 1 and 2 and urge it be put to any information theoretic exorcism of Maxwell's Demon.

The literature that exorcises Maxwell's Demon has all the trappings of a mature and stable science. Like the sciences it emulates, its inferences proceed from a stable base of principle—Szilard's or Landauer's. The hesitation and doubts of skeptics are answered by a small repertoire of ingenious and vivid thought experiments in which many effects miraculously combine to defeat all challenges. But this literature is far from the stable sciences it emulates. It is an enterprise whose goals, methods and presumptions have been continuously subject to modification. It has so mutated that it scarcely resembles its origins. Maxwell thought his Demon illustrated the statistical character of thermodynamic laws. Smoluchowski and Szilard sought to limit the extent of the damage wreaked by fluctuations on phenomenological thermodynamics. By the time of Brillouin, the exorcism literature treated the threat of the Demon and the problem of fluctuations as distinct issues. Bennett then announced that Brillouin's analysis was just plainly wrong—both in his basic principle and his ingenious thought experiments—and replaced it with a new system that seems to us no more secure than Brillouin's. Zurek and Caves have already adjusted it with their notion of algorithmic information. All the while this mainstream of analysis is immersed in a proliferation of alternative proposals that contradict the notion of unanimity in the literature even at any one time. So Bennett (1988, p. 282) laments that 'Maxwell offered no definitive refutation of the demon, beyond saying that we lack its ability to see and handle individual molecules'.

Have we completely lost sight of the fact that Maxwell had no interest in exorcising the Demon? He was on the Demon's side!

We see the exorcism literature as driven by the latest fads, be it the alluring notion of information of the 1950s or the analogy to computers of the 1970s or the current fascination with complexity, and its major turnings governed by the momentary success of this or that thought experiment. To those who find deep insight in this literature, our portrait will no doubt seem cantankerous. But we will accept the mantle of the curmudgeon if we can help our readers to see that a literature, capable of such rapid and thorough mutation, has no stable core. In this century we have become used to the catastrophic overthrow of mature theories. But their overthrow is not simply the result of blunders of thought by the proponents of the old theory. Newton's mechanics did not fail because of errors of reasoning in his *Principia*. It failed because it could not be extended to the domain of the very fast, the very small, the very large and the very heavy. But Bennett assures us that Brillouin's brilliant star must fall because he was just wrong to think that information acquisition has an associated entropy cost and stumbled in reading the import of his clever thought experiments. If the literature can celebrate so defective a system, we had better be prepared for a similar and equally catastrophic fall of the latest star when the next fad takes hold. The Demon lives.

Appendix 1: Deriving Principles of Information and Entropy

If one naturalises Maxwell's Demon as an ordinary thermal system, then the results concerning the entropy cost of his operation are readily derivable from standard properties of thermal systems. These include the presumption of the Second Law. In this Appendix we make this claim a little more precise by sketching some of the results we have in mind and their derivations. They are, of course, neither new nor the most general form of the results; but they will suffice to illustrate our claim.

We assume that we have an object system O coupled with a demon system D where the demon system D is designed so that it brings about a process in which the entropy of O is decreased. For example, the object system O may be an ideal gas of n molecules at temperature T in a container of volume V. The demon system may consist of all the apparatuses needed to manipulate a door in a dividing wall within the container so that the final outcome is a concentration of the molecules to one side of the wall. In naturalising and thermalising the demon, we will adopt the statistical mechanical framework of Khinchin (1949).[16] We will assume that the combined system of O and D is

[16] We are grateful to an anonymous referee for emphasising the known deficiencies of Khinchin's system. Since the combined system's Hamiltonian is the sum of the component systems' Hamiltonians, these systems do not interact through an interaction Hamiltonian. Moreover, as we will note below, our use of the system precludes treatment of fluctuations. Our purpose in this section is not to give the most rigorous demonstrations possible, but to show that more rigour is readily achievable than, for example, in the treatments described in the literature in Section 2.

microcanonically distributed on a surface of constant energy E in the combined phase space. Moreover we will assume that O and D combined consist of a very large number of small component systems; that O is just one of them; and that D contains all the rest: that is, components D_1, D_2, ... It now follows that the state of O is distributed canonically in its own phase space, that is, according to the probability density

$$p_O = \exp\,(-e_O\theta)/\phi_O, \tag{10}$$

where e_O is the object system's energy at the relevant point in phase space, θ is more commonly recognisable as $1/kT$ (for k Boltzmann's constant and temperature T) and the generating function is

$$\phi_O = \int_{\gamma_O} \exp(-e_O\theta)\,dv_O, \tag{11}$$

with the integration extending over the volume of O's phase space γ_O. Similar distributions obtain for each of the demon subsystems D_1, D_2, ...

We will assume that the demon is so designed that in the course of its interaction with the object system O, the object system's entropy decreases. To avoid unnecessary entropy dissipation, we will assume that this interaction is reversible, so that thermal equilibrium is maintained throughout. Its course is governed by an external parameter λ that changes in value from λ_1 to λ_2 in the course of the process. The effect of varying λ is to alter the energy of points in the phase space. In general, work will be done by the agent manipulating λ and this will alter the combined system's energy—but we may choose to design the systems so that this work is arbitrarily small or even vanishes. Otherwise the combined system O and D is assumed thermally isolated, so that any change in its energy E is due to this work.

It now follows that any reduction in entropy in this process in the object system O must be matched by a dissipation of entropy in the demon system. To see this we follow Khinchin (1949, Ch. VII) in identifying the entropy S of the combined system and the entropies S_O and S_D of the object and demon systems as

$$S = k(E\theta + \log\Phi), \tag{12}$$

$$\bar{S}_O = k(\bar{e}_O\theta + \log\phi_O), \tag{13}$$

$$\bar{S}_D = k(\bar{e}_D\theta + \log\phi_D), \tag{14}$$

where Φ is the generating function for the combined system defined analogously to (11) and the overhead bar represents a phase average in the phase spaces of the object and demon systems respectively.

An important point: the identification of the total entropy S of the combined system by expression (12) precludes S representing any fluctuations in the entropy of the combined system. Take the special case in which we hold λ constant, so that the combined system's energy is constant and the combined system is confined to a single surface of constant energy in the combined phase space. All the quantities used to define S in (12)—E, θ and Φ—are constants of this energy surface. Thus, if the isolated, combined system can arise by spontaneous time development from a low entropy state in the energy surface, then we have from the recurrence lemma that it will eventually return arbitrarily close to that state. That fluctuation ought to be reflected in a reduction of the combined system's entropy, but the entropy S of (12) cannot reflect this reduction since it must remain constant. Thus the results derived below for processes from the constancy of S are only assured to the extent that the processes persist for sufficient time that the phase averaged, fluctuating entropy approaches the expression of (12). That is, sufficient time must elapse for the phase point to visit extensively throughout the phase space.[17]

As the interaction of demon and object system proceeds, the total entropy S remains constant. This follows from the Second Law of thermodynamics and application of the thermodynamic definition of entropy, since the process is by assumption adiabatic and reversible. This law and definition are also built into the identification of S in (12);[18] so constancy of entropy S also follows from (12):

$$dS/d\lambda = k(d/\mathrm{d}\lambda)[E\theta + \log \Phi] = 0. \tag{15}$$

To see this vanishing, recall that for this thermally isolated system, $dE/d\lambda - dW/d\lambda = 0$, where W is the phase average work supplied to the system by manipulation of λ. But we have from Khinchin (1949, p. 135) that this work satisfies $-\theta dW/d\lambda = d\log\Phi/d\lambda + E\mathrm{d}\theta/\mathrm{d}\lambda$. These equations immediately entail that $dS/d\lambda = 0$.

The combined entropy S is just the sum of the mean entropies $\bar{S}_O$ and $\bar{S}_D$ of the object and demon system,[19]

$$S = \bar{S}_O + \bar{S}_D, \tag{16}$$

so the constancy of S during the process—$\Delta S = 0$—entails that any change in mean entropy of the object system $\Delta\bar{S}_O$ is compensated by a corresponding change $\Delta\bar{S}_D$ in the demon system

$$\Delta\bar{S}_O + \Delta\bar{S}_D = 0. \tag{17}$$

Thus, if the demon reduces the object system's mean entropy, it does so through a corresponding dissipation of mean entropy in its system.

[17] Because of this restriction, we consider only the phase average entropies $\bar{S}_O$ and $\bar{S}_D$. Expressions are available for fluctuating entropies for the object system and demon system—simply replace the phase average energies $\bar{e}$ in (13) and (14) by the energies e at each phase point. But we forgo them since we have effectively decided to treat phase averaged entropies in choosing expression (12) for the combined system entropy.

[18] The identification of S as (12) in Khinchin (1949, Ch. 6) is based on showing that changes in this quantity equal the quotient of heat gained/temperature during a reversible process. The Second Law is enforced by presumption since the identification considers only the limit of very slow processes so that Second Law violating fluctuation phenomena are averaged away.

[19] This follows from (12), (13) and (14) directly since $\Phi = \phi_O\phi_D$.

Szilard's Principle

We can now convert (17) into a version of Szilard's Principle. The simplest form of Szilard's Principle would associate a quantity of information I with the object system O by means of the relation

$$I = k \log n, \tag{18}$$

where n is the number of equiprobable microstates associated with the state of O. If p_O is the probability that the system O is in one of these states, then we would have $p_O = 1/n$ and the information associated with that state as

$$I_O = -k \log p_O. \tag{19}$$

If we use the celebrated Boltzmann equation

$$S_O = k \log n = -k \log p_O \tag{20}$$

for the entropy of the state, we have that $I_O = S_O$ so that any information change ΔI_O associated with the object system O must be associated with an equal change of entropy ΔS_O. Result (17) tells us that, in phase averaged quantities, any such change in the object system is compensated by an entropy dissipation of opposite sign in the Demon system.

The difficulty of this analysis is that Boltzmann's equation does not have universal validity. Therefore we need to confirm compatibility of its definition of entropy with that of (13). Substituting for p_O from (10) we have

$$S_O = -k \log p_O = -k \log \left(\frac{\exp(-e_O \theta)}{\phi_O} \right) = k(e_O \theta + \log \phi_O). \tag{21}$$

The phase average of this entropy is the phase averaged entropy $\bar{S}_O$ of (13). Further we should restate the information I_O as a phase averaged quantity so that result (17) can be applied directly. We define its phase average

$$\bar{I}_O = -k \int_{\gamma_O} p_O \log p_O \, dv_O. \tag{22}$$

This expression has a natural interpretation. If a communications channel transmits symbols 1, 2, 3, ... such that each has probability p_i, $i = 1,2,3, \ldots$ of transmission, then information theory accords an information entropy of $-\sum_i p_i \log p_i$ to each transmitted symbol. If we now treat the many different states possible for the object system as symbols, then we arrive at (22) as the corresponding definition for the information entropy of the object system. Boltzmann's constant k is introduced as an arbitrary unit for simplicity.

Substituting for p_O with (10) one recovers after a brief manipulation[20] that $\bar{I}_O = \bar{S}_O$, so that the result of (17) becomes

$$\Delta\bar{S}_D = -\Delta\bar{S}_O = -\Delta\bar{I}_O\,. \tag{23}$$

This result is Szilard's Principle translated into our present context. The demonic interaction has reduced the object's entropy by $\Delta\bar{S}_O$ and, correspondingly, the associated information by $\Delta\bar{I}_O$. However there is no net reduction in entropy in the combined system, since the thermodynamic cost of the process is a dissipation of entropy $\Delta\bar{S}_D$ by the demonic system.

The practice in the literature is to associate the information term $\Delta\bar{I}_O$ with a more ordinary notion of information. Such interpretation is usually conducted in the context of an example. The canonical case is to take the object system O to be an ideal gas. We consider a monoatomic gas of n molecules each of mass m at temperature θ in a vessel of volume V. Its generating function is $\phi_O = V^n(2\pi\mathrm{m}/\theta)^{3n/2}$. Using $\bar{e}_O = 3n/2\theta$ and the expression (13) for the entropy $\bar{S}_O$, we readily recover a standard result. If the gas is compressed isothermally from volume V_1 to V_2, then its entropy changes by

$$\Delta\bar{S}_O = nk\log(V_2/V_1)\,. \tag{24}$$

In ordinary thermodynamics, one would effect this compression with a piston. Then the resulting decrease in entropy would be compensated by an increase in entropy elsewhere. A heat sink would need to absorb the heat given off during the compression and its resulting increase in entropy would effect the compensation.

In the Demonic context we might try a devious method. If the gas vessel is divided into two chambers with a connecting doorway, then, over time, we expect the numbers of molecules in each chamber to fluctuate. We may even expect a fluctuation in which all the molecules momentarily are in one chamber only. A demon closing the door at this moment has effected a reduction in entropy. Result (23) assures us that this reduction must be compensated by a corresponding dissipation of entropy in the demonic system. The result warrants a little less than it may first appear. Recall that (23) is only assured to the extent that identification of S by (12) is assured; and that is assured to the degree that sufficient time passes for the phase point to visit extensively through the phase space. Thus the result assures us that over the longer term, no naturalised and thermalised demon can exploit this fluctuation. But it can make no such assurance for the shorter term. Short term and correspondingly improbable violations of the Second Law remain. On this point, we have advanced little in our understanding beyond that of the 1910s.

[20] $\bar{I}_O = k\int_{\gamma_O} \frac{\exp(-e_O\theta)}{\phi_O}\log\frac{\exp(-e_O\theta)}{\phi_O}\,dv_O = k\int_{\gamma_O}\frac{\exp(-e_O\theta)}{\phi_O}[e_O\theta + \log\phi_O]\,dv_O$

$= k\theta\frac{1}{\phi_O}\int_{\gamma_O}\exp(-e_O\theta)e_O dv_O + k\log\phi_O\frac{1}{\phi_O}\int_{\gamma_O}\exp(-e_O\theta)\,dv_O = k[\bar{e}_O\theta + \log\phi_O] = \bar{S}_O.$

What of the information term in (23)? Following the style of analysis of Brillouin (1951), we might say that the proper operation of the machine requires that the demon knows *when* to close the door. The problem is one of information for we presume the demon able to close the door with minimum energy and entropy cost once the decision is taken. Thus the compensating dissipation of entropy $\Delta\bar{S}_D$ is really the entropy cost of deciding the right moment to close the door. We need to fix the location of n molecules; that is, we need n bits of information, which has an associated information entropy of $nk\log 2$. If the compression is to halve the volume, $\Delta\bar{I}_O$ obligingly takes the value $-nk\log 2$, which coincides with the value expected for $\Delta\bar{S}_O$ from (24). Appealing as this narrative may be, as we explain in the text, these anthropomorphic considerations of information as human knowledge play no role in the demonstration that the Demon must fail to effect an overall reduction in entropy. That burden is fully carried by (17) and the equation that expresses Szilard's Principle, (23), is merely a relabelling of a term in (17) as an information term.

Landauer's principle

This principle locates the relevant information theoretic result as pertaining not to the acquiring of information, but its erasure. 'Landauer's principle' can be illustrated through (17). We assume that the memory storage device can store n bits of information in the settings of n devices, each capable of two states. This device is capable of discriminating 2^n states so that its erasure should, by Landauer's Principle, require a dissipation of entropy

$$k\log 2^n = kn\log 2. \tag{25}$$

We assume that erasure of the device is through a process that first thermalises the memory and then uses some macroscopic process to return it to a null state—all zeros. For example, we might imagine that each bit in storage is represented by the location of a single molecule in its own chamber of volume V. Each chamber is divided into two halves by a partition. The molecule in the left half encodes a 0; the molecule in the right half a 1. We erase the memory by removing the partition so that the molecules are free to move to either side of each of their chambers. We then bring the chambers isothermally to half their volume—their left half—forcing the molecules to return to the null state. The generating function for each chamber will be $\phi = V(2\pi m/\theta)^{3/2}$ from above with $n = 1$. So the generating function for the system of n chambers will be a product of n such functions $\phi^n = V^n(2\pi m/\theta)^{3/2\,n}$ and the entropy reduction in the erasure the expression (24) as before. Since $V_2/V_1 = 1/2$, there is an entropy reduction of $kn\log 2$ in the memory system. This of course coincides with the change of information content corresponding to erasing n bits.

The obvious way to effect this compression from V_2 to V_1 is with a piston and the application of work. As before this will force an entropy dissipation of at least $kn\log 2$ in the heat sink that absorbs the heat discharged during the isothermal compression. Is there any other way of effecting the entropy reduction without a compensating increase elsewhere? If we take the memory device as the object system O, then our answer is that there is not—up to the constraints of the validity of (17). This relation assures us that any device that reduces the entropy of the memory cells by $\Delta\bar{S}_O$ must do so at the cost of a corresponding entropy dissipation elsewhere. Moreover we read in the equality $\Delta\bar{S}_O = \Delta\bar{I}_O$ that the entropy cost is equal to the information erased, scaled in appropriate units of Boltzmann's constant.

Appendix 2: How to Build a Mechanical Maxwell Demon

Zhang and Zhang (1992) and Skordos (1993) have described how one might construct systems that will most probably reverse the course of entropy growth. They are mechanical Maxwell Demons. Skordos' model is of a box containing hard core disks in motion. The demonic component is a dividing membrane that either reflects the disks or passes them, deflecting their trajectories, according to their angles of incidence. The effect of the membrane is to maintain a pressure difference across the membrane. Skordos supplies no account of the internal operation of the membrane. Zhang and Zhang describe force fields that act on moving particles in such a way that anti-entropic processes ensue if the fields act on a kinetic gas. In this Appendix, we will review the properties of one of Zhang and Zhang's fields and show how the field could be used to construct the sort of membrane described by Skordos. We will see that particle trajectories in the field will be time reversible but volumes in the phase space of the gas interacting with the field will not be invariant under time evolution.

The system we will consider is a pressure vessel maintained at constant temperature T by a heat sink as shown in Fig. 1. The vessel is filled with a kinetic gas. Dividing the vessel in half is a thin membrane that will turn out to pass the gas molecules more easily from left to right than right to left. The membrane itself is just a region in which one of Zhang and Zhang's fields prevails. The result will be a pressure differential between the two halves. After an equilibrium pressure differential has been achieved, that differential can be tapped to produce work by means of a device such as a turbine shown in Fig. 1, or a frictionless piston that expands reversibly under the higher pressure and recompresses reversibly to the lower pressure. The rate of flow through the device is kept sufficiently low so as not to disturb the equilibrium pressure differential materially. Through the device's action, energy is drawn from the gas which cools. As long as the temperature of the gas is maintained by heat supplied by the heat sink, the device will continuously convert heat energy from the heat sink into work energy, in violation of the Second Law of thermodynamics.

We shall idealise the gas molecules as point particles that do not interact with one another in order to avoid the complication of collision interactions between the molecules within the membrane field. The particles will be assumed to bounce off the vessel walls; this is the interaction that allows the exchange of

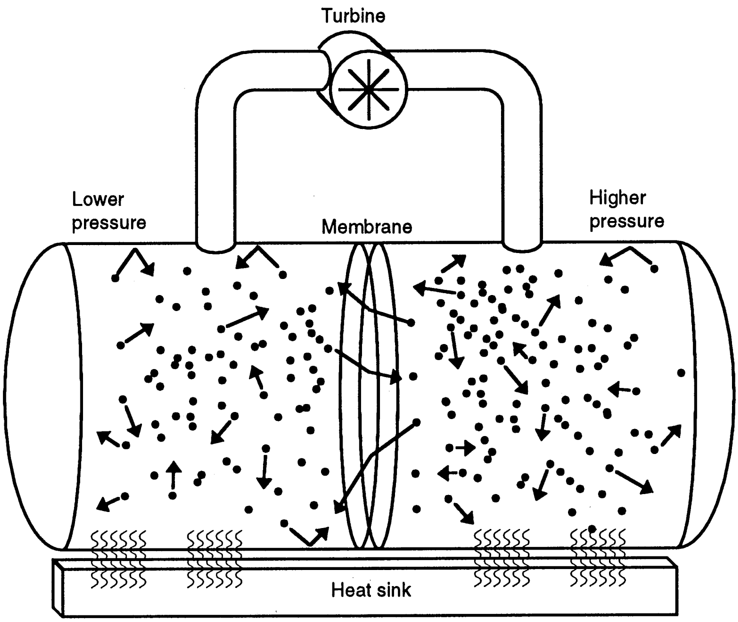

Fig. 1. *A mechanical Maxwell Demon.*

heat energy between the gas and heat sink. We will presume that the particles' velocities outside the membrane field are distributed isotropically according to the Maxwell–Boltzmann distribution. This distribution is enforced by the assumption of a very irregular inner surface of the vessel wall and other thermally agitated objects anchored to the inner wall surface.

We set Cartesian coordinates such that the x-axis aligns with the axis of the vessel in Fig. 1, with the $+x$ direction to the right, and the y and z coordinates are parallel to the membrane field. In this coordinate system, the force $\mathbf{F}$ on a particle in the membrane field is given in the usual vector notation as

$$\mathbf{F} = c\mathbf{v} \times [\hat{\mathbf{x}} \times \mathbf{v}] \tag{26}$$

where $\hat{\mathbf{x}}$ is a unit vector in the x direction, c is a coupling constant and the particles are all assumed to be of unit mass. It follows immediately from (26) that the force $\mathbf{F}$ is orthogonal to the velocity $\mathbf{v}$ of the particle so that no work is done by the field on the particles:

$$\mathbf{F} \cdot \mathbf{v} = c\mathbf{v} \times [\hat{\mathbf{x}} \times \mathbf{v}] \cdot \mathbf{v} = \mathbf{0}. \tag{27}$$

The scalar velocity $v = |\mathbf{v}|$ is constant since

$$0 = \mathbf{F} \cdot \mathbf{v} = d\mathbf{v}/dt \cdot \mathbf{v} = (1/2)\, d(v^2)/dt \tag{28}$$

for time t. Thus the sole effect of the membrane field is to deflect the direction of motion of the particles without altering their energy or speed. The field does not supply or draw energy from the gas. The deflections induced by (26) are time reversible in the usual sense: if $\mathbf{x}(t)$ describes a particle trajectory compatible with (26), then the time reversed trajectory $\mathbf{x}^*(t) = \mathbf{x}(-t)$ is also compatible with (26).[21]

To see the magnitude of the deflection we rewrite (26) in component form as

$$\dot{v}_x = c(v_y^2 + v_z^2), \quad \dot{v}_y = -cv_xv_y, \quad \dot{v}_z = -cv_xv_z, \tag{29}$$

where the overhead dot represents differentiation with respect to time t. Since the force field (26) is rotationally symmetric about the x-axis, the trajectory of each particle will be fully confined to a flat plane with one axis parallel to the x-axis. We rotate the y and z-axis about the x-axis until the y-axis is parallel to the plane as well. We label this rotated y-axis the new axis u. Because of the symmetry of (26) under this rotation, the form of (29) is preserved and it becomes[22]

$$\dot{v}_x = cv_u^2, \quad \dot{v}_u = -cv_xv_u. \tag{30}$$

Note that the xu-plane will in general be different for each particle according to its initial velocity. Within its xu-plane, assume that the particle's velocity vector $\mathbf{v}$ is at an angle θ with the x-axis, so that $v_x = v\cos\theta$ and $v_u = v\sin\theta$. Substituting these expressions into (30) both equations reduce to $d\theta/dt = -cv\sin\theta$. If we now introduce the arc length $ds^2 = dx^2 + du^2$ along the curve, then we have $v = ds/dt$ so that the deflection is given as

$$d\theta/ds = -c\sin\theta. \tag{31}$$

This equation is independent of the particle's velocity. Therefore the field-induced deflection of the particle trajectory is independent of the particle's speed and fixed solely by the direction of the particle's motion.

We will consider the special case of a membrane field of very small thickness Δl as shown in Fig. 2. As long as the angular direction of the particle's trajectory has an angle θ not close to $\pi/2$, the particle's path length in the membrane field will be given to good approximation as $\Delta l/\cos\theta$, where θ remains approximately constant in the membrane field. Therefore each trajectory is deflected by a small angle

$$\Delta\theta = -c\Delta l\tan\theta. \tag{32}$$

[21] This follows from the invariance of (26) under the substitution $t \to -t$. For the time reversed trajectory we have $\mathbf{F}^*(t) = d^2\mathbf{x}^*(t)/dt^2 = d^2\mathbf{x}(-t)/dt^2 = d^2\mathbf{x}(-t)/d(-t)^2 = cd\mathbf{x}(-t)/d(-t) \times [\hat{\mathbf{x}} \times d\mathbf{x}(-t)/d(-t)] = cd\mathbf{x}(-t)/dt \times [\hat{\mathbf{x}} \times d\mathbf{x}(-t)/dt] = c\mathbf{v}^*(t) \times [\hat{\mathbf{x}} \times \mathbf{v}^*(t)]$.

[22] To see this same result analytically, note that we have from (29) that $dv_y/dv_z = v_y/v_z$ so that v_y is proportional to v_z. Thus the particle's motion is constrained to a flat plane parallel to the x-axis. The new u coordinate is given as $u = Ay + Bz$, where $A = v_y(0)/\sqrt{v_y(0)^2 + v_z(0)^2}$ and $B = v_z(0)/\sqrt{v_y(0)^2 + v_z(0)^2}$. Therefore, directly, $v_u = Av_y + Bv_z$ and $\dot{v}_u = A\dot{v}_y + A\dot{v}_z$. Straightforward manipulation also gives us that $v_u^2 = v_y^2 + v_z^2$. These results suffice to allow (30) to be derived from (29).

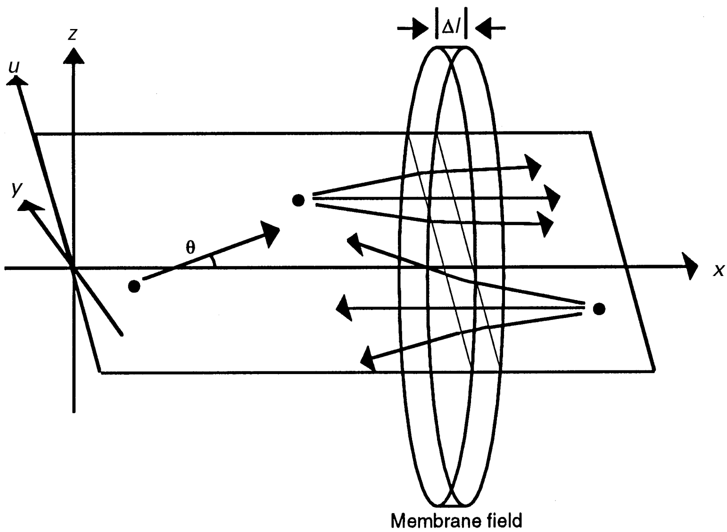

Fig. 2. *Deflection of all particles by a membrane field towards the* $+x$ *direction.*

Thus each trajectory is deflected back towards the $+x$ direction as shown in Fig. 2.

This deflection towards the $+x$ direction is the effect that allows the membrane to pass particles more easily in the $+x$ direction than in the $-x$ direction. All particles moving in the $+x$ direction and approaching the membrane field from the left will pass. This is not so for particles attempting to pass in the other direction, the $-x$ direction. If they do not approach too steeply, their motion will be deflected towards the $+x$ direction but they will still pass. However, if the approach is steep—that is the θ angle is just past $\pi/2$—then the particle will be fully deflected back to the right hand side of the vessel. Thus all the particles impinging on the membrane field from the left side of the vessel will pass to the right. But not all impinging on the membrane from the right will pass to the left.

We now need to estimate how closely a particle moving in the $-x$ direction may come to a θ angle of $\pi/2$ before it is deflected back into the right hand side of the vessel. Let $\theta = \pi/2 + \Delta\phi$ for $\Delta\phi > 0$ be the smallest angle of approach for which the particles may pass from right to left. To estimate $\Delta\phi$ we can no longer use (31) since this relation was derived using an approximation of $\Delta l/\cos\theta$ for the path length. This approximation fails when the particle's path in the field is no longer approximately straight such as will be the case when the particle is returned to the right side. Instead we assume that a particle approaches the membrane from the right at the critical angle $\theta = \pi/2 + \Delta\phi$ as in Fig. 3 so that the trajectory just penetrates the width Δl of the membrane field before being turned back. We ask after the form of the trajectory in the membrane field; that is, we seek the functional dependence of x on u for the trajectory. For

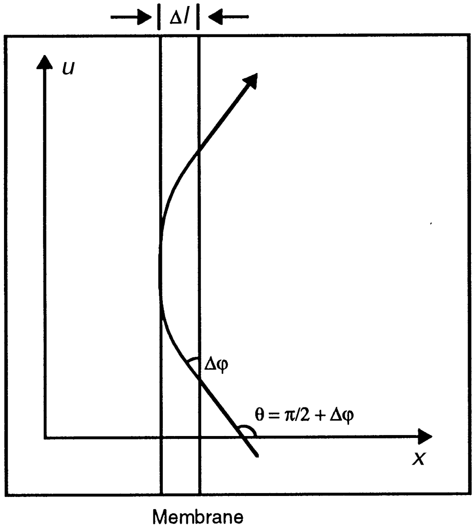

Fig. 3. *Trajectory with greatest θ that is still fully deflected back into right hand side of vessel.*

convenience, we locate the origin of the x and u coordinates at the turning point of the trajectory. Since the trajectory is time reversible, we would expect the function $x(u)$ to be even and, since the membrane is assumed very thin, only its lowest order terms to be appreciable. That is, $x(u)$ is quadratic in u and the trajectory is a parabola.

We derive a result that affirms our expectations from the force equation (30). Since the velocity ds/dt is a constant for the trajectory, the first equation of (30) becomes

$$d^2x/ds^2 = c(du/ds)^2 . \tag{33}$$

Multiplying both sides by $(ds/du)^2$ and retaining only lowest order terms we recover

$$d^2x/du^2 = c . \tag{34}$$

Solving with $u = x = 0$ at the turning point we have

$$x = cu^2/2 . \tag{35}$$

The value of $\Delta\phi$ is recovered from this quadratic by identifying $\tan\Delta\varphi$ with the slope of the parabola at the point at which it enters the membrane field, that is the point at which $x = \Delta l$. Since $\Delta\phi$ is small, we approximate $\tan\Delta\phi$ as $\Delta\phi$ and we have

$$\Delta\varphi = dx/du|_{\mathrm{x=\Delta l}} = cu|_{x=\Delta l} = \sqrt{2cx}|_{x=\Delta\mathrm{l}} = \sqrt{2\mathrm{c}\Delta l} . \tag{36}$$

Thus all particles approaching the membrane from the left with angles $\theta = 0$ to $\pi/2$ will pass to the right. But only particles approaching from the right with angles $\theta = \pi/2 + \sqrt{2c\Delta l}$ to π will pass.

A necessary condition for equilibrium is that the particle flux from left to right equals the particle flux from right to left, that is,

$$j_{left \to right} = j_{right \to left}\,. \tag{37}$$

Since particles pass more easily from left to right than right to left, some mechanism must compensate to allow (37) to obtain. The temperature on each side is maintained at a constant T by the heat sink. The only remaining variables are the volume density of particles of the gas ρ and the gas pressure P, which stand in direct proportion according to the ideal gas law. If the density and pressure of gas in the vessel is initially equal on both sides of the membrane field, the net passage of particles from left to right will cause an increase in pressure and particle density on the right hand side with respect to the left until the resulting imbalance in the flow of particles allows the equilibrium condition (37) to obtain. If the resulting density difference is $\Delta\rho$ and the pressure difference ΔP, then they are related through the ideal gas law by

$$\Delta P = \Delta\rho k T\,. \tag{38}$$

The velocity distribution of the particles in both sides of the vessel will be the Maxwell–Boltzmann distribution.[23] Recalling that ρ is the volume density of particles in the gas, then, according to the Maxwell–Boltzmann distribution, the volume density of particles per unit scalar velocity v and per unit solid angle Ω for the direction of particle velocity will be

$$\frac{\partial^2 \rho}{\partial v \partial \Omega} = \rho \left(\frac{m}{2\pi k T}\right)^{3/2} \exp\left(-\frac{mv^2}{2kT}\right) v^2\,, \tag{39}$$

where k is Boltzmann's constant, T temperature and the particles' mass $m = 1$. The flux of particles at the membrane, that is, the flow per unit membrane area, due to particles with angular direction θ will be given as the product of the density of particles with angular direction θ and the component of their velocity in the direction θ, that is, $v\cos\theta$. The solid angle associated with a θ differential

[23] While the membrane will introduce deviations from this distribution by not passing all particles, we can bring the distribution as close as we like to the Maxwell–Boltzmann distribution by making the size of each side of the vessel suitably large with respect to the membrane. Thus particles passing into the left hand side of the vessel that are not distributed according to the Maxwell–Boltzmann distribution will soon revert to this distribution under collisions with the vessel's irregular walls and other thermally agitated objects within the vessel. For the calculation that follows, our concern is that the velocity distribution of the particles approaching the membrane accord with the Maxwell–Boltzmann distribution. That will be so since these approaching particles come from deep within the vessel. Since the Maxwell–Boltzmann distribution of particle velocities entails the ideal gas law, these considerations also justify its use in (38).

$d\theta$ is $d\Omega = 2\pi \sin\theta \, d\theta$. Therefore, if ρ is the particle density on the left hand side of the vessel, the flux of particles from left to right is given by

$$j_{left\to right} = \int_{\theta=0}^{\pi/2} \int_{v=0}^{\infty} \rho \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left(-\frac{mv^2}{2kT}\right) v^2 \cdot 2\pi \sin\theta \cdot v\cos\theta \cdot dv d\theta$$

$$= \frac{\rho}{2}\sqrt{\frac{2kT}{m\pi}}. \tag{40}$$

If $\rho + \Delta\rho$ is the particle density on the right hand side of the vessel, the flux of particles from right to left is given by

$$j_{right\to left} = \int_{\theta=\pi/2+\sqrt{2c\Delta l}}^{\pi} \int_{v=0}^{\infty} (\rho + \Delta\rho) \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left(-\frac{mv^2}{2kT}\right)$$

$$\times v^2 \cdot 2\pi \sin\theta \cdot v\cos\theta \cdot dv d\theta = \frac{\rho + \Delta\rho}{2}\sqrt{\frac{2kT}{m\pi}}(1 - 2c\Delta l), \tag{41}$$

for Δl small. At equilibrium, these two expressions must be equal and we have

$$\frac{1}{2}\sqrt{\frac{2kT}{m\pi}}\left(\frac{\Delta\rho}{\rho} - 2c\Delta l\right) = 0. \tag{42}$$

Therefore, recalling the ideal gas law (38), we have that the membrane supports a pressure and density differential of

$$\frac{\Delta\rho}{\rho} = \frac{\Delta P}{P} = 2c\Delta l. \tag{43}$$

This pressure differential supports the violation of the Second Law of thermodynamics since it can operate the turbine shown in Fig. 1. As long as the turbine is operated at a low flow rate, the membrane will be able to preserve the pressure differential of (43). The work extracted by the turbine is at the cost of the thermal energy of the gas, slightly lowering its temperature. This thermal energy will be replenished with heat supplied from the heat sink. Since the process can continue indefinitely, the device allows the unrestricted conversion of heat energy fully into work energy, in violation of the Second Law.

A characteristic of Skordos' membrane and of Zhang and Zhang's fields is that their dynamics does not leave phase space volume invariant under time development and, consequently, cannot be expressed in Hamiltonian form. We review this failure for the force field (26)/(29). In the membrane force field, we have assumed that the individual particles do not interact. Therefore it is sufficient to demonstrate lack of phase space volume invariance for the phase spaces of the particles individually. The corresponding result for their combined phase space follows immediately.

We take the phase space of a single particle to have coordinates (x, y, z, v_x, v_y, v_z).[24] The equations of motion (29) define a velocity vector field on the phase space, $(\dot{x}, \dot{y}, \dot{z}, \dot{v}_x, \dot{v}_y, \dot{v}_z)$; its integral curves are the trajectories of phase points. The expansion

$$\frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{y}}{\partial y} + \frac{\partial \dot{z}}{\partial z} + \frac{\partial \dot{v}_x}{\partial v_x} + \frac{\partial \dot{v}_y}{\partial v_y} + \frac{\partial \dot{v}_z}{\partial v_z} \tag{44}$$

measures the scalar change in a volume of phase space moving with the flow. If the flow were Hamiltonian, then this expansion (44) would vanish immediately as a result of Hamilton's equations.[25] This result fails if the flow is governed by (29). We read directly from (29) that the only non-zero terms in the expansion (44) are $\partial \dot{v}_y / \partial v_y = -cv_x$ and $\partial \dot{v}_z / \partial v_z = -cv_x$ so that the expansion is

$$\left(\frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{y}}{\partial y} + \frac{\partial \dot{z}}{\partial z} + \frac{\partial \dot{v}_x}{\partial v_x} + \frac{\partial \dot{v}_y}{\partial v_y} + \frac{\partial \dot{v}_z}{\partial v_z} \right) = -2cv_x. \tag{45}$$

We now read directly from (45) that the phase volume either increases or decreases according to whether the phase point is associated with a particle motion in the $-x$ or $+x$ direction; that is, according to whether the particle moves from right to left or left to right. This result applies to the phase spaces of the individual particles. Since the particles do not interact in the membrane field, the phase volume of the combined phase space will be the product of the volumes from the individual phase spaces. Therefore volume in that combined phase space is unlikely to remain invariant under time development. Whether it increases or decreases will depend momentarily on the particular distribution of particle motion over the $+x$ and $-x$ directions. This combined phase space volume will remain constant at best on average when the gas is in an equilibrium state.

Finally, the vanishing of the expansion (44) follows if the equations of motion have Hamiltonian form. For the field (29), the expansion does not vanish. Therefore we conclude that the equations of the field (29) cannot be rewritten in a Hamiltonian form.

[24] Recall that the mass of particle $m = 1$. Thus the momentum $\mathbf{p} = m\mathbf{v} = \mathbf{v}$.

[25] Hamilton's equation are

$$\dot{x} = \frac{\partial H}{\partial p_x}, \ \dots, \ \dot{p}_x = -\frac{\partial H}{\partial x}, \ \dots,$$

for some Hamiltonian H. Therefore we have

$$\frac{\partial \dot{x}}{\partial x} = \frac{\partial}{\partial x}\frac{\partial H}{\partial p_x} = -\frac{\partial \dot{p}_x}{\partial p_x} = -\frac{\partial \dot{v}_x}{\partial v_x}, \ \dots,$$

and all the terms on the right hand side of (44) cancel.

References

Beghian, L. E. (1994) 'Maxwell's Demon Revisited', *Nuovo Cimento*, **109,** 611–616.
Bennett, C. H. (1982) 'The Thermodynamics of Computation—a Review', *International Journal of Theoretical Physics* **21,** 905–940. Reprinted in Leff and Rex (1990), pp. 213–248.
Bennett, C. H. (1987) 'Demons, Engines and the Second Law', *Scientific American* **257,** 108–116.
Bennett C. H. (1988) 'Notes on the History of Reversible Computation', *IBM Journal of Research and Development* **32,** 16–23. Reprinted in Leff and Rex (1990), pp. 281–288.
Biedenharn, L. C. and Solem, J. C. (1995) 'A Quantum-Mechanical Treatment of Szilard's Engine: Implications for Entropy of Information', *Foundations of Physics* **25,** 1221–1229.
Brillouin, L. (1951) 'Maxwell's Demon Cannot Operate: Information and Entropy. I', *Journal of Applied Physics* **22,** 334–337. Reprinted in Leff and Rex (1990), pp. 134–137. 'Maxwell's Demon Cannot Operate: Information and Entropy. II', *Journal of Applied Physics* **22,** 338–343.
Brillouin, L. (1953) 'The Negentropy Principle of Information', *Journal of Applied Physics* **24,** 1152–1163.
Brillouin L. (1962) *Science and Information Theory* (New York: Academic Press).
Caves, C. M. (1990) 'Quantitative Limits on the Ability of Maxwell Demon to Extract Work from Heat', *Physical Review Letters* **64,** 2111–2114.
Caves, C. M. (1993) 'Information and entropy', *Physical Review* **E47,** 4010–4017.
Caves, C. M. (1994) 'Information, Entropy and Chaos', in J. J. Halliwell, J. Pérez-Mercader and W. H. Zurek (eds), *Physical Origins of Time Asymmetry* (Cambridge: Cambridge University Press), pp. 47–87.
Caves, C. M., Unruh, W. G., and Zurek, W. H. (1990) 'Comment on "Quantitative Limits on the Ability of a Maxwell Demon to Extract Work from Heat"', *Physical Review Letters* **65,** 1387.
Chambadal, P. (1971) *Paradoxes of Physics* (London: Transworld Publishers).
Demers, P. (1944) 'Les Démons de Maxwell et le Second Principe de la Thermodynamique', *Canadian Journal of Research* **A22,** 27–51.
Demers, P. (1945) 'Le Second Principe et la Théorie des Quanta', *Canadian Journal of Research*, **A23,** 47–55.
Denbigh, K. G. (1981) 'How Subjective is Entropy?', *Chemistry in Britain* **17,** 168–185. Reprinted in Leff and Rex (1990), pp. 109–115.
Denbigh, K. G. and Denbigh, J. S. (1985) *Entropy in Relation to Incomplete Knowledge* (Cambridge: Cambridge University Press).
Feynman, R. (1996) *Feynman Lectures on Computation* eds A. J. G. Hey and R. W. Allen (Reading, Mass.: Addison–Wesley).
Gabor, D. (1951) '§5 A Further Paradox: "A Perpetuum Mobile of the Second Kind"', Reprinted in Leff and Rex (1990), pp. 148–159.
Jauch, J. M. and Baron, J. G. (1972) 'Entropy, Information and Szilard's Paradox', *Helvetica Physica Acta* **45,** 220–232. Reprinted in Leff and Rex (1990), pp. 160–172.
Khinchin, A. I. (1949) *Mathematical Foundations of Statistical Mechanics* (New York: Dover).
Landauer, Rolf (1961) 'Irreversibility and Heat Generation in the Computing Process', *IBM Journal of Research and Development* **5,** 183–91. Reprinted in Leff and Rex (1990), pp. 188–196.
Leff, H. S. and Rex, A. F. (1990) *Maxwell's Demon*: *Entropy*, *Information*, *Computing* (Princeton: Princeton University Press).
Leff, H. S. and Rex, A. F. (1994) 'Entropy of Measurement and Erasure: Szilard's Membrane Model Revisited', *Americal Journal of Physics* **62,** 994–1000.

Neumann, J. von (1932) *Mathematical Foundations of Quantum Mechanics* (Princeton: Princeton University Press, 1955).

Raymond, R. (1950) 'Communication, Entropy, and Life', *American Scientist* **38,** 273–278.

Raymond, R. (1951) 'The Well-Informed Heat Engine', *American Journal of Physics*, **19,** 109–112. Reprinted in Leff and Rex (1990), pp. 138–141.

Rodd, P. (1964) 'Some Comments on Entropy and Information', *Americal Journal of Physics* **32,** 333–335. Reprinted in Leff and Rex (1990), pp. 145–147.

Schumacher, B. W. (1994). 'Demonic Heat Engines', in J. J. Halliwell, J. Pérez-Mercader and W. H. Zurek (eds), *Physical Origins of Time Asymmetry* (Cambridge: Cambridge University Press), pp. 90–97.

Skordos, P. A. (1993) 'Compressible Dynamics, Time Reversibility, Maxwell's Demon, and the Second Law', *Physical Review* **E 48,** 777–784.

Slater, J. C. (1939) *Introduction to Chemical Physics* (New York: McGraw–Hill).

Szilard, L. (1929) 'On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings', in *The Collected Works of Leo Szilard*: *Scientific Papers* (MIT Press, 1972), pp. 120–129. Reprinted in Leff and Rex (1990), pp. 124–133.

Zhang, K and Zhang, K. (1992) 'Mechanical Models of Maxwell's Demon with Non-invariant Phase Volume', *Physical Review* **A 46,** 4598–4605.

Zurek, W. H. (1984) 'Maxwell's Demon, Szilard's Engine and Quantum Measurements', in G. T. Moore and M. O. Seully (eds), *Frontiers of Nonequilibrium Statistical Mechanics* (New York: Plenum). Reprinted in Leff and Rex (1990), pp. 249–259.

Zurek, W. H. (1989a) 'Thermodynamic Cost of Computation, Algorithmic Complexity and the Information Metric', *Nature* **341,** 119–124.

Zurek, W. H. (1989b) 'Algorithmic Randomness and Physical Entropy', *Physical Review* **A40,** 4731–4751.

Zurek, W. H. (1990) 'Algorithmic Information Content, Church–Turing Thesis, Physical Entropy, and Maxwell's Demon', in W. H. Zurek (ed.), *Complexity*, *Entropy and the Physics of Information* (New York: Addison–Wesley), pp. 73–89.