Chapter 1

Overview

Maxwell's demon . . . *[after J. C. Maxwell, its hypothecator]: 'A hypothetical being of intelligence but molecular order of size imagined to illustrate limitations of the second law of thermodynamics.' Webster's Third New International Dictionary*

1.1 Introduction

Maxwell's demon lives on. After more than 130 years of uncertain life and at least two pronouncements of death, this fanciful character seems more vibrant than ever. As the dictionary entry above shows, Maxwell's demon is no more than a simple idea. Yet it has challenged some of the best scientific minds, and its extensive literature spans thermodynamics, statistical physics, quantum mechanics, information theory, cybernetics, the limits of computing, biological sciences, and the history and philosophy of science.

Despite this remarkable scope and the demon's longevity, coverage in standard physics, chemistry, and biology textbooks typically ranges from cursory to nil. Because its primary literature is scattered throughout research journals, semipopular books, monographs on information theory, and a variety of specialty books, Maxwell's demon is somewhat familiar to many but well known only to relatively few. Two Scientific American articles on the demon (Ehrenberg, 1967; Bennett, 1987) have been helpful, but they only scratch the surface of the existing literature. A recent popularization of the subject (von Baeyer, 1998) helps but, being directed at a lay audience, is limited in both depth and breadth.

The main purpose of this reprint collection is to place in one volume: (1) important original papers covering Maxwell's demon, (2) an overview of the demon's life and current status, and (3) an annotated bibliography that provides perspective on the demon plus a rich trail of citations for further study.

The life of Maxwell's demon can be viewed usefully in terms of three major phases. The first phase covers its 'birth' in approximately 1867 through the first 62 years of its relatively quiet existence. The flavor of the early history is reflected in Thomson's classic paper on the dissipation of energy (1874, Article 2.1). A modern perspective of automatic (non-intelligent) mechanical demons along the line of those addressed by Smoluchowski (1912, 1914) is given by Skordos and Zurek (1992, Article 3.1) and Rex and Larsen (1992, Article 3.2).

The second phase began with an important paper by Leo Szilard (1929, Article 3.3). The entry on Szilard in Scribner's Dictionary of Scientific Biography cites his '...famous paper of 1929, which established the connection between entropy and information, and foreshadowed modern cybernetic theory.' Notably, Szilard discovered the idea of a 'bit' of information, now central in computer science. His discovery seems to have been independent of earlier identifications of logarithmic forms for information by Nyquist (1924) and Hartley (1928). The term 'bit' (= binary digit) was suggested approximately 15 years after Szilard's work by John Tukey. Subsequently Rothstein formulated fundamental information-theoretic interpretations of thermodynamics, measurement, and quantum theory (1951, Article 2.5). The history of the demon during the first two phases is described by Daub (1970, Article 2.2), Klein (1970, Article 2.3), and Heimann (1970, reprinted in MD1).

After a hiatus of about 20 years, Leon Brillouin (1949, Article 2.4) became involved in the Maxwell's demon puzzle through his interest in finding a scientific framework to explain intelligent life. At roughly the same time Dennis Gabor (1964, reprinted in MD1), based on his 1951 lecture, and Brillouin (1951, Article 3.4) focused on the demon's acquisition of information. Both Brillouin and Gabor assumed the use of light signals in the demon's attempt to defeat the second law of thermodynamics. The result was a proclaimed 'exorcism' of Maxwell's demon, based upon the edict that information acquisition is dissipative, making it impossible for a demon to violate the second law.

Independently of Brillouin, Raymond (1951, reprinted in MD1) published an account of a clever variant of Maxwell's demon that did not explicitly entail light signals—a 'well-informed heat engine' using density fluctuations in a gas. Raymond found that 'an outside observer creates in the system a negative information entropy equal to the negative entropy change involved in the operation of the engine'. His work, though less influential than Brillouin's, also made the important connection between information and entropy. Finfgeld and Machlup (1960, reprinted in MD1) analyzed Raymond's model further, assuming that the necessary demon uses light signals, and also obtained an estimate of its power output.

The impact of Brillouin's and Szilard's work was far-reaching, inspiring numerous investigations of Maxwell's demon, including those by Rodd (1964) and Rex (1987), both reprinted in MD1. Weinberg (1982, reprinted in MD1) broadened the demon's realm to include 'macroscopic' and 'social' demons. Despite some critical assessments of the connections between information and entropy by Denbigh, (1981, reprinted in MD1), Jauch and Báron (1972, Article 3.5), and Costa de Beauregard and Tribus (1974, Article 3.6), those linkages and applications to the Maxwell's demon puzzle became firmly entrenched in the scientific literature and culture.

The third phase of the demon's life began at age 94 when Rolf Landauer made the important discovery (1961, Article 4.1) that memory erasure in computers feeds entropy to the environment. This is now called 'Landauer's principle'. Landauer referred to Brillouin's argument that measurement requires a dissipation of order kT, and observed: 'The computing process ... is closely akin to a measurement'. He also noted that: '...the arguments dealing with the measurement process do not define measurement very well, and avoid the very essential question: When is a system A coupled to a system B performing a measurement? The mere fact that two physical systems are coupled does not in itself require dissipation'.

Landauer's work inspired Charles Bennett to investigate logically reversible computation, which led to Bennett's important demonstration (Bennett, 1973, reprinted in MD1) that reversible computation, which avoids erasure of information, is possible in principle. The direct link between Landauer's and Bennett's work on computation and Maxwell's demon came with Bennett's observation (1982, Article 7.1) that a demon 'remembers' the information it obtains, much as a computer records data in its memory. Bennett argued that erasure of a demon's memory is the fundamental act that saves the second law because of Landauer's principle. This was a surprising, remarkable event in the history of Maxwell's demon. Subsequent analyses of memory erasure for a quantum mechanical Szilard's model by Zurek (1984, Article 5.1) and Lubkin (1987, Article 5.2) support Bennett's finding.

A key point in Bennett's work is that, in general, the use of light signals for information acquisition can be avoided. That is, although such dissipative information gathering is sufficient to save the second law of thermodynamics, it is not necessary. Bennett's argument nullifies Brillouin's 'exorcism', which was so ardently believed by a generation of scientists. The association of the Maxwell's demon puzzle with computation greatly expanded the audience for the demon, and articles by Bennett (1988, Article 7.3) and Landauer (1961, 1987a, 1996a, 1996b, Articles 4.1, 7.2, 7.4, 7.5, respectively) illustrating that association are reprinted here.

These three phases of Maxwell's demon life are described in further detail in Sections 1.2–1.5. Section 1.6 provides an examination of the foundations of Landauer's principle and Section 1.7 contains a discussion of the role of quantum mechanics in the Maxwell's demon puzzle. Section 1.8 deals with aspects of the demon not treated in the earlier sections. This includes the concept of *process* demons that do not collect information, but nevertheless threaten the second law of thermodynamics. Chapters 2–7 contain reprinted articles covering, respectively: historical and philosophical considerations; information acquisition; and information erasure and computing. This is followed by a chronological bibliography, with selected annotations and quotations that provide a colorful perspective on the substantial impacts of Maxwell's demon. An alphabetical bibliography and an extensive index are also included.

1.2 The Demon and Its Properties

1.2.1 Birth of the Demon

The demon was introduced to a public audience by James Clerk Maxwell in his 1871 book, *Theory of Heat*. It came near the book's end in a section called 'Limitation of the Second Law of Thermodynamics'. In one of the most heavily quoted passages in physics, Maxwell wrote:

> Before I conclude, I wish to direct attention to an aspect of the molecular theory which deserves consideration.
>
> One of the best established facts in thermodynamics is that it is impossible in a system enclosed in an envelope which permits neither change of volume nor passage of heat, and in which both the temperature and the pressure are everywhere the same, to produce any inequality of temperature or of pressure without the expenditure of work. This is the second law of thermodynamics, and it is undoubtedly true as long as we can deal with bodies only in mass, and have no power of perceiving or handling the separate molecules of which they are made up. But if we conceive a being whose faculties are so sharpened that he can follow every molecule in its course, such a being, whose attributes are still as essentially finite as our own, would be able to do what is at present impossible to us. For we have seen that the molecules in a vessel full of air at uniform temperature are moving with velocities by no means uniform, though the mean velocity of any great number of them, arbitrarily selected, is almost exactly uniform. Now let us suppose that such a vessel is divided into two portions, A and B, by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole, so as to allow only the swifter molecules to pass from A to B, and only the slower ones to pass from B to A. He will thus, without expenditure of work, raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics.
>
> This is only one of the instances in which conclusions which we have drawn from our experience of bodies consisting of an immense number of molecules may be found not to be applicable to the more delicate observations and experiments which we may suppose made by one who can perceive and handle the individual molecules which we deal with only in large masses.
>
> In dealing with masses of matter, while we do not perceive the individual molecules, we are compelled to adopt what I have described as the statistical method of calculation, and to abandon the strict dynamical method, in which we follow every motion by the calculus.

Maxwell's thought experiment dramatized the fact that the second law is a statistical principle that holds almost all the time for a system composed of many molecules. That is, there is a nonzero probability that anisotropic molecular transfers, similar to those accomplished by the demon, will occur if the hole is simply left open for a while. Maxwell had introduced this idea in a 1867 letter to Peter Guthrie Tait (Knott, 1911) '...to pick a hole' in the second law. There he specified more detail about the sorting strategy intended for the demon:

> Let him first observe the molecules in A and when he sees one coming the square of whose velocity is less than the mean sq. vel. of the molecules in B let him open the hole and let go into B. Next let him watch for a molecule of B, the square of whose velocity is greater than the mean sq. vel. in A, and when it comes to the hole let him draw the slide and let it go into A, keeping the slide shut for all other molecules.

This allows a molecule to pass from A to B if its kinetic energy is less than the average molecular kinetic energy in B. Passage from B to A is allowed only for molecules whose kinetic energies exceed the average kinetic energy per molecule in A. In the same letter Maxwell emphasized the quality of 'intelligence' possessed by the demon:

> Then the number of molecules in A and B are the same as at first, but the energy in A is increased and that in B diminished, that is, the hot system has got hotter and the cold colder and yet no work has been done, only the intelligence of a very observant and neat-fingered being has been employed.

William Thomson (1874, Article 2.1) subsequently nicknamed Maxwell's imaginary being 'Maxwell's intelligent demon.' He apparently did not envisage the creature as malicious: 'The definition of a demon, according to the use of this word by Maxwell, is an intelligent being endowed with free-will and fine enough tactile and perceptive organization to give him the faculty of observing and influencing individual molecules of matter.' He expounded further on his view of 'the sorting demon of Maxwell' (Thomson, 1879):

> The word 'demon', which originally in Greek meant a supernatural being, has never been properly used to signify a real or ideal personification of malignity.
>
> Clerk Maxwell's 'demon' is a creature of imagination having certain perfectly well defined powers of action, purely mechanical in their character, invented to help us to understand the 'Dissipation of Energy' in nature.
>
> He is a being with no preternatural qualities and differs from real living animals only in extreme smallness and agility. ... He cannot create or annul energy; but just as a living animal does, he can store up limited quantities of energy, and reproduce them at will. By operating selectively on individual atoms he can reverse the natural dissipation of energy, can cause one-half of a closed jar of air, or of a bar of iron, to become glowingly hot and the other ice cold; can direct the energy of the moving molecules of a basin of water to throw the water up to a height and leave it there proportionately cooled ...; can 'sort' the molecules in a solution of salt or in a mixture of two gases, so as to reverse the natural process of diffusion, and produce concentration of the solution in one portion of the water, leaving pure water in the remainder of the space occupied; or, in the other case separate the gases into different parts of the containing vessel.
>
> 'Dissipation of Energy' follows in nature from the fortuitous concourse of atoms. The lost motivity is essentially not restorable otherwise than by an agency dealing with individual atoms; and the mode of dealing with the atoms to restore motivity is essentially a process of assortment, sending this way all of one kind or class, that way all of another kind or class.

Following Thomson's introduction of the term 'demon,' Maxwell clarified his view of the demon (quoted in Knott, 1911) in an undated letter to Tait:

> Concerning Demons.
>
> 1. Who gave them this name? Thomson.
> 2. What were they by nature? Very small BUT lively beings incapable of doing work but able to open and shut valves which move without friction or inertia.
> 3. What was their chief end? To show that the 2nd Law of Thermodynamics has only a statistical certainty.
> 4. Is the production of an inequality of temperature their only occupation? No, for less intelligent demons can produce a difference in pressure as well as temperature by merely allowing all particles going in one direction while stopping all those going the other way. This reduces the demon to a valve. As such value him. Call him no more a demon but a valve like that of the hydraulic ram, suppose.

In light of Maxwell's intentions, it is interesting to examine the accuracy of dictionary definitions. The *Webster's Third New International Dictionary* definition quoted at the beginning of this chapter, though brief, properly cites Maxwell's intention to 'illustrate limitations of the second law of thermodynamics.' In contrast, *The Random House Dictionary of the English Language* (Second Edition, 1988) contains the definition:

> A hypothetical agent or device of arbitrarily small mass that is considered to admit or block selectively the passage of individual molecules from one compartment to another according to their speed, constituting a violation of the second law of thermodynamics.

And the Second Edition (1989) of The Oxford English Dictionary describes it in the entry for James Clerk Maxwell:

> ...a being imagined by Maxwell as allowing only fast-moving molecules to pass through a hole in one direction and only slow-moving ones in the other direction, so that if the hole is in a partition dividing a gas-filled vessel into two parts one side becomes warmer and the other cooler, in contradiction to the second law of thermodynamics.

Despite the emphasis on violating rather than illustrating limitations on the second law in these two definitions, there is no indication that Maxwell intended his hypothetical character to be a serious challenge to that law. Nevertheless, the latter two definitions reflect the interpretation by many subsequent researchers that Maxwell's demon was a puzzle that must be solved: If such a demon cannot defeat the second law, then why not? And if it can defeat the second law, then how does that affect that law's status?

Maxwell did not relate his mental construction to entropy. In fact, he evidently misunderstood the Clausius definition of entropy and went out of his way to adopt a different definition in early editions of his *Theory of Heat*. He wrote: 'Clausius has called the remainder of the energy, which cannot be converted into work, the Entropy of the system. We shall find it more convenient to adopt the suggestion of professor Tait, and give the name of Entropy to the part which can be converted into mechanical work.' He then argued that entropy *decreases* during spontaneous processes. Later Maxwell recanted: 'In former editions of this book the meaning of the term Entropy, as introduced by Clausius, was erroneously stated to be that part of the energy which cannot be converted into work. The book then proceeded to use the term as equivalent to the available energy; thus introducing great confusion into the language of thermodynamics.'

Maxwell's discomfort and confusion with entropy is ironic, for his demon has had a profound effect on the way entropy is viewed. In particular, Maxwell's demon led to an important linkage between entropy and information. Unfortunately, Maxwell did not live long enough to see this outgrowth of his thought experiment. It is also noteworthy that his originally adopted definition of entropy gave rise to a function that decreases during spontaneous processes. Many years later, Brillouin found it useful for interpretive purposes to define a function, negentropy ($= -$entropy), with this property (see Section 1.4 for more on negentropy).

1.2.2 Temperature and Pressure Demons

Maxwell's specification of the demon was brief enough to leave considerable room for interpretation. As envisioned, his creature was a temperature-demon that acts within a thermally isolated system of which it is an integral part. Its task was to generate a temperature difference without performing work on the gas. In effect this is the equivalent of producing heat flow from a lower to a higher temperature with no other effect, in conflict with the Clausius form of the second law.

In his later clarification (recall 'Concerning demons' in Section 1.2.1), Maxwell recognized that 'less intelligent' demons could generate differences in pressure. The first detailed investigation of a pressure-demon was by Leo Szilard (1929, Article 3.1). Szilard's work is discussed further in Section 1.3. A pressure demon operates in a system linked to a constant-temperature reservoir, with the sole net effect of converting energy transferred as heat from that reservoir to work on an external object, in conflict with the Kelvin-Planck form of the Second Law. The 'Maxwell's demon puzzle' is to show why neither a temperature nor pressure demon can operate outside the limits imposed by the second law of thermodynamics.

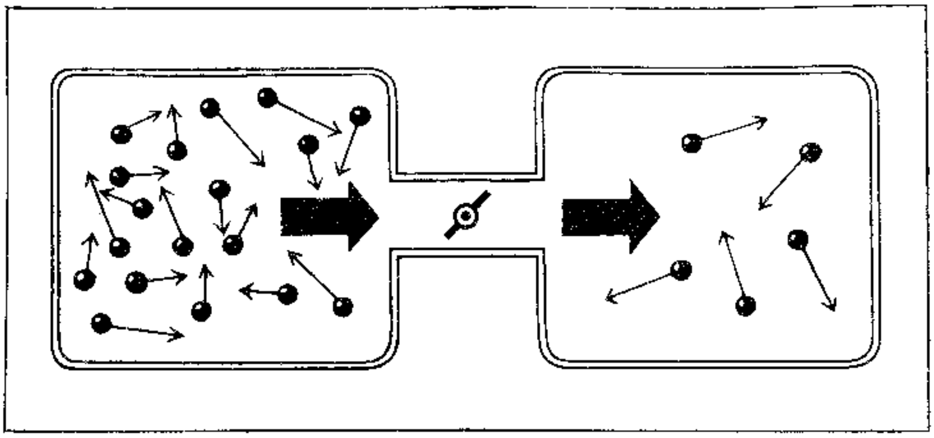


Figure 1.

1.2.3 Depictions of the Demon

Maxwell described his proposed being as 'small'. The dictionary definitions above suggest 'molecular' or 'arbitrarily small' size. Various authors have included cartoon depictions of Maxwell's demon with their writings. Figures 1–9 illustrate some of the ways the demon has been portrayed. Figure 1 (Pekelis, 1974) is in accord with Maxwell's view that the demon is nothing more than a valve, but does not show any control mechanism. Figures 2 (Pierce, 1961) and 3 (Angrist and Hepler, 1967) show characters operating trap doors manually from within one of the chambers, but without any obvious means of detecting molecules. Figure 4 (Angrist and Hepler, 1967) shows the demon wearing a helmet with a built in light source. Figure 5 (Lerner, 1975) shows a satanic character with a flashlight, operating a shutter from inside one of the chambers.

Figure 6 (Darling and Hulburt, 1955) shows the demon *outside* the two chambers. Figure 7 (Morowitz, 1970) depicts a pressure demon controlling a shutter between two chambers that are in contact with a constant-temperature heat bath. Figure 8 (Gasser and Richards, 1974) shows yet another view of an external demon, here operating a valve, allowing one species of a two component gas (hot and cold) through a partition separating a gas from an initially evacuated chamber. Only fast molecules are allowed through, resulting in a cold gas in one chamber and a hot gas in the other. More cartoons can be found in 'Information physics in cartoons,' honoring Rolf Landauer (Bennett, 1998, Article 7.6).

Typically, cartoon depictions show the demon as being relatively large compared to the shutter, sometimes with a light source to detect molecules, and sometimes located outside the system. Placing a temperature-demon outside the gas is questionable because of the need for thermal isolation. Depictions with a light source are not surprising in view of Maxwell's specification of a 'being who can see the individual molecules ....' Because his intent was to dramatize the statistical nature of the second law rather than to exorcise the demon, Maxwell had no reason to address the question of whether a demon could detect molecules by any means other than vision.

1.2.4 Means of Detection

Leon Brillouin (1951, Article 3.4), closely following the work of Pierre Demers (1944, 1945), took Maxwell's specification of 'seeing' molecules seriously and assumed the use of light signals. Dennis Gabor did the same, apparently independently. Others have considered detecting molecules via their

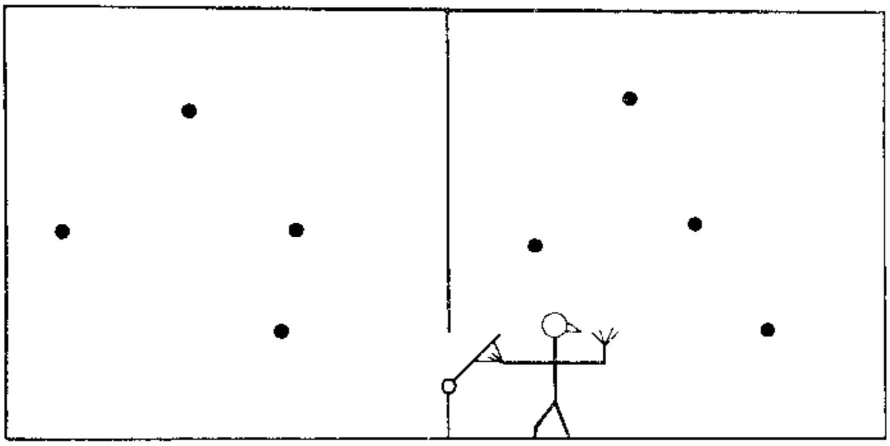


Figure 2.

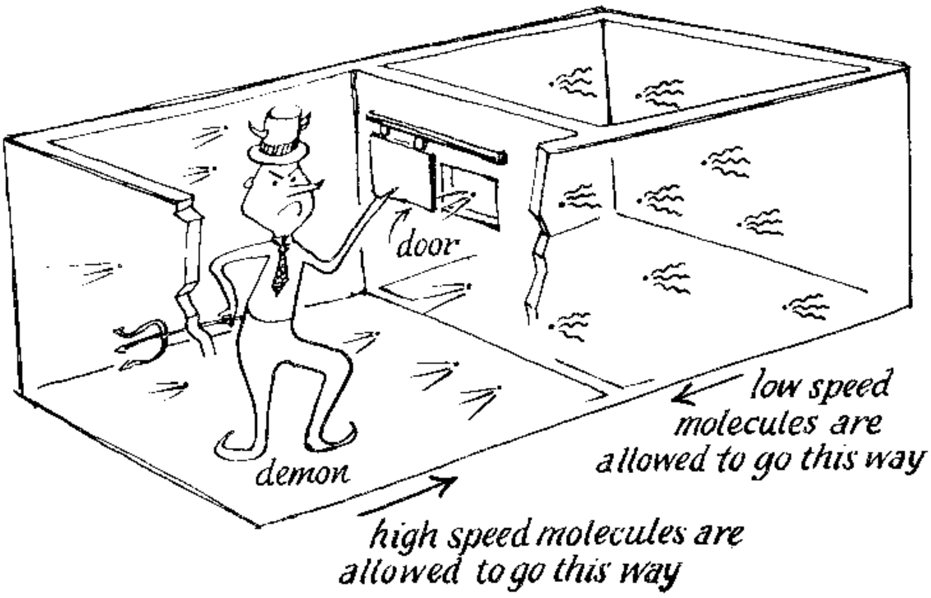


Figure 3.

magnetic moments (Bennett, 1982, Article 7.1), Doppler-shifted radiation (Denur, 1981; Chardin, 1984; Motz, 1983), and even via purely mechanical means (Bennett, 1987). The prevailing modern view is that one must not prejudice the demon's operation by assuming the use of light signals, for that is too restrictive. The fundamental question is whether measurement in general is necessarily irreversible.

The clever mechanical detector (Bennett, 1987) proposed in the context of Szilard's 1929 model suggests that, in principle, the presence of a molecule can be detected with arbitrarily little work and dissipation. Bennett's scheme is compelling, but is limited to a one-molecule gas. The general question of whether measurement in a many-particle gas must be irreversible lacks a correspondingly compelling

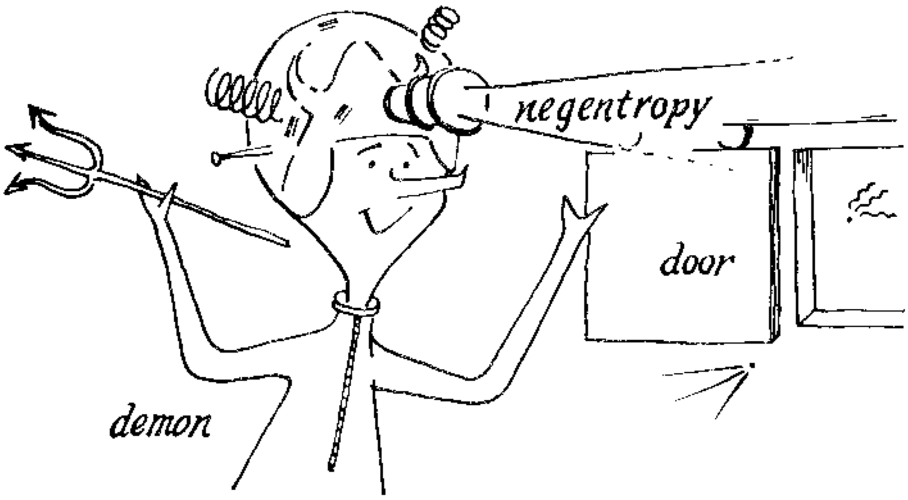


Figure 4.

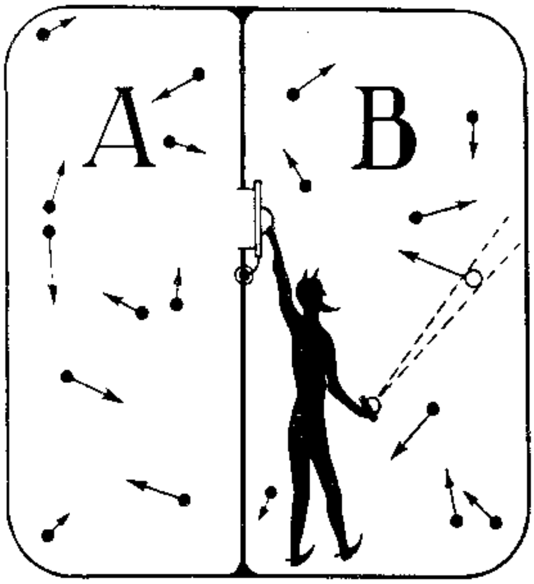
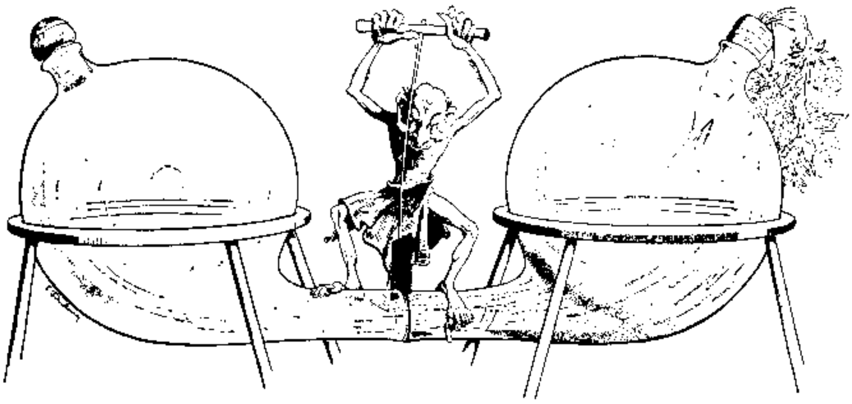


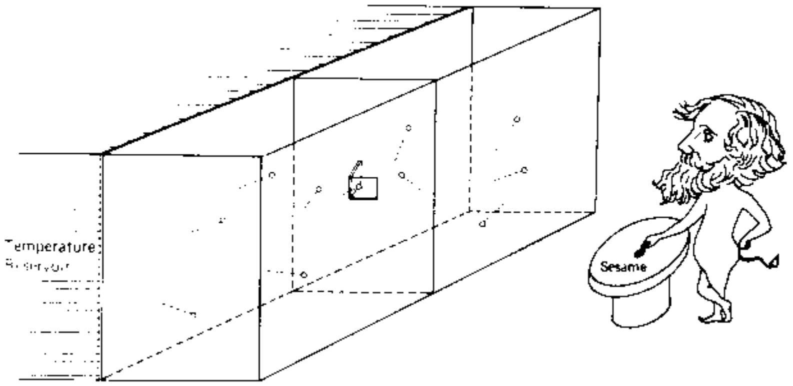
Figure 5.

answer. Maxwell's original temperature-demon must distinguish between molecular velocities among numerous molecules, a more complex task than detecting the presence of a single molecule. To our knowledge no specific device that can operate with arbitrarily little work and dissipation has been proposed for such velocity measurements. Given this void, the possibility of measurement without entropy generation in a macroscopic system is not universally accepted. See for example Rothstein (1988), Porod, *et al* (1984a) and responses thereto (1984b). Yet it is also true that there is no reason to believe that work and entropy thresholds exist for such measurements.



Maxwell's demon at work

Figure 6.



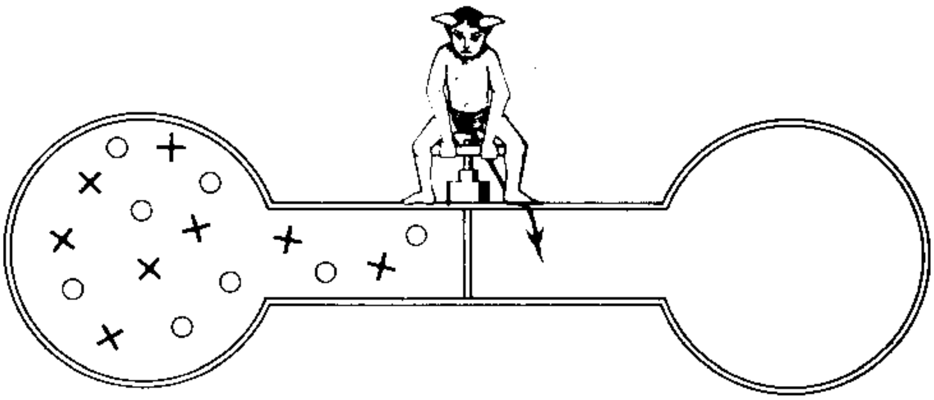
A Maxwell demon controlling a door between two chambers each initially at temperature T_1 and pressure P_1

Figure 7.

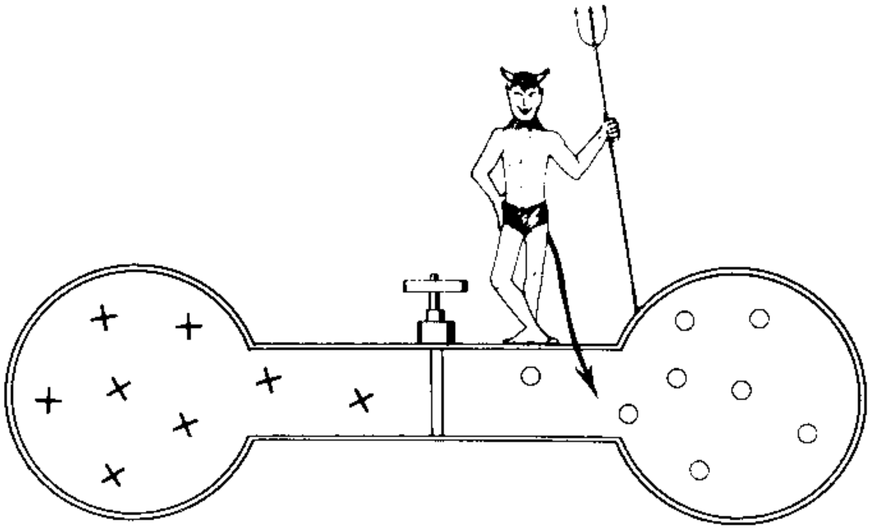
1.2.5 Thermal Equilibrium and Fluctuations

The demon must be in thermal equilibrium with the gas in which it resides, or an irreversible energy transfer between gas and demon would occur, clouding the basic puzzle. As a temperature-demon generates a temperature gradient within a gas, its own temperature presumably changes with its host gas. The heat capacity of a 'small' demon is presumably much less than that of the gas, and its temperature can vary with that of the gas via negligibly small energy exchanges. Except for the receipt of light signals, energy exchanges with the demon are usually neglected.

A demon that continually receives energy input via light signals (or other means) must eventually experience a temperature rise unless it transfers energy (via heat) to its surroundings. Additionally, if a lamp is used to generate light signals, photons that miss the demon will add energy to the gas and or container walls. Such phenomena threaten the assumption of constant-temperature operation, and most treatments of Maxwell's temperature-demon ignore these details. Of course such photons could heat one chamber directly, with no need for a demon—a phenomenon for which there is obviously no challenge to the second law.



Maxwell's Demon



Maxwell's Demon: later

Figure 8.

Located within a gas, a Maxwell's demon is continually bombarded by gas molecules and by photons from the blackbody radiation field within the container. It can be jostled around by this bombardment, impeding the accuracy of its measuring activities. Long ago it was pointed out (Smoluchowski, 1912, 1914) that thermal fluctuations would prevent an automatic device from operating successfully as a Maxwell's demon.

A modern discussion of Smoluchowski's ideas was given by Richard Feynman (1963), who compared Maxwell's demon with a ratchet and pawl and an electrical rectifier, neither of which can systematically transform internal energy from a single reservoir to work. He wrote: 'If we assume that the specific heat of the demon is not infinite, it must heat up. It has but a finite number of internal gears and wheels, so it cannot get rid of the extra heat that it gets from observing the molecules. Soon it is shaking from Brownian motion so much that it cannot tell whether it is coming or going, much less whether the molecules are coming or going, so it does not work.'

If a demon absorbs energy, periodic dumping of energy to an external reservoir is needed to keep its temperature approximately equal to the temperature of the gas in which it resides. For a temperature-demon this violates assumed thermal isolation, and the 'system' must be expanded to include the demon, gas, and external reservoir. Of course feeding entropy to the reservoir helps to keep the second law intact. Modern day computer simulations dramatically reveal the fluctuation phenomena envisaged by Smoluchowski and Feynman (Skordos and Zurek, 1992, Article 3.1; Rex and Larsen, 1992, Article 3.2).

Smoluchowski's observation regarding thermal fluctuations suggested that Maxwell's demon ought to be buried and forgotten. But that did not happen, apparently because Smoluchowski left open the possibility that somehow, a perpetual motion machine operated by an intelligent being might be achievable. It was the fascinating idea of using intelligence that captured Leo Szilard's interest.

1.2.6 Intelligence

The demon must have sufficient 'intelligence' to discern fast from slow molecules, right-moving from left-moving molecules, or (in Szilard's model) simply the presence or non-presence of a molecule. In normal parlance intelligence is considered to include, among other things, ability to learn, reason, and understand relationships. But none of these seems to be required by Maxwell's demon. One feature associated with intelligence that is needed by a demon is memory: it must 'remember' what it measures, even if only briefly. Indeed without somehow recording a result, one can argue that a measurement has not been completed.

Despite the title 'The decrease of entropy by intelligent beings' of his classic 1929 paper, Leo Szilard wrote that physics is not capable of properly accounting for the biological phenomena associated with human intervention. Szilard asserted, 'As long as we allow intelligent beings to perform the intervention, a direct test (of the second law) is not possible. But we can try to describe simple nonliving devices that effect such coupling, and see if indeed entropy is generated and in what quantity.' In 1929, prior to the development of solid state electronics, that was a fanciful thought.

If the demon were an automaton, it would perform preprogrammed functions upon receipt of certain well-defined signals. Evidently a Maxwell's demon need not be any more intelligent than an electronic computing machine connected to some type of transducer that detects molecular phenomena and puts out electrical signals signifying detection. Certainly it need not possess human intelligence. The concept of Maxwell's demon as a computer automaton was explored by Laing (1974; reprinted in MD1) who, unfortunately, was unaware of Landauer's important finding (1961, Article 4.1) that memory erasure in computers feeds entropy to the environment.

In recent years some researchers have investigated the feasibility of quantum mechanical computers that operate via changes in the states of individual atoms. Feynman (1986) wrote '... we are going to be even more ridiculous later and consider bits written on one atom instead of the present 10^{11} atoms. Such nonsense is very entertaining to professors like me. I hope you will find it interesting and entertaining also ... it seems that the laws of physics present no barrier to reducing the size of computers until bits are the size of atoms, and quantum behavior holds dominant sway.' This suggests the possibility of Maxwell's demon being a quantum automaton of microscopic size, if such a microscopic demon could avoid devastation from fluctuations.

1.2.7 Interplay Between the First and Second Laws

Before continuing with the demon's history, it is helpful to examine implications of the first and second laws of thermodynamics on its actions. Consider first a temperature-demon that sorts molecules, lowering the entropy of a gas without altering its energy. The term 'demon' here includes any peripheral equipment used to effect sorting. What do the first and second laws of thermodynamics imply? Because the demon-gas system is energetically isolated, the second law requires the demon's entropy to increase at least as much as the gas entropy decreases (which is assumed here) during sorting. The first law implies that a temperature-demon's energy is unchanged by sorting because the gas and gas-demon system energies are both fixed. Thus, the demon's entropy must increase at fixed energy.

Can the demon be returned to its initial state without disturbing the gas? Such 'resetting' of the demon is desirable for two reasons. First, if the demon is to operate repeatedly, its entropy cannot be allowed to increase indefinitely or it will ultimately become too 'disordered' and unable to operate (see Section 1.2.5). Second, resetting simplifies the thermodynamic analysis, which can focus on the gas and its environment, without regard for the demon's details. Resetting the demon requires an exchange of energy with other objects. For example, the demon's excess entropy might be dumped as heat to a constant-temperature reservoir, with an external work source subsequently increasing the demon's energy at constant entropy, returning it to its initial state.

Evidently the first and second laws of thermodynamics assure that: (1) a temperature-demon cannot sort molecules without increasing its entropy, (2) the demon cannot return to its initial state without external energy exchanges, and (3) the combination of sorting and resetting generates an energy transfer from an energy source to a reservoir.

Next consider a pressure-demon, operating a cyclic process in a constant temperature ideal gas. Contact with a thermal reservoir assures that the temperature will be constant. Initially the gas pressures and densities are equal on each side of a central partition. The cyclic process is defined as follows:

(a) The demon reduces the gas entropy at fixed temperature and energy by letting molecules through the partition in one direction only. This sorting process generates pressure and density differences across the partition.
(b) The gas returns to its initial state by doing isothermal, reversible work on an external load. Specifically, the partition becomes a frictionless piston coupled to a load, moving slowly to a position of mechanical equilibrium (away from the container's center) with zero pressure and density gradients across the piston. The piston is then withdrawn and reinserted at the container's center.
(c) The demon is returned to its initial state.

What do the laws of thermodynamics imply? The process sequence (a)–(c) results in a load with increased energy. The first law of thermodynamics requires that this energy come from some well-defined source. It cannot be supplied by the reservoir or the entropy of the universe would decrease in the cyclic process, in violation of the second law. Apparently, resetting the demon in (c) requires use of a work source which, in effect, supplies the energy to the load. It is helpful to look at the thermodynamic details of steps (a)–(c).

The second law implies that the demon's entropy increases in (a) to 'pay' for the entropy decrease of the gas. That is, sorting must increase the pressure-demon's entropy. In (b) work W is done by the gas on the load, inducing heat $Q = W$ from reservoir to gas. The load's energy increases with its entropy unchanged, and the gas is returned to its initial state. Withdrawing and replacing the piston has zero thermodynamic effect. In step (b) the work on the load is compensated by the diminished energy (and entropy) of the reservoir. The demon's entropy increase offsets the reservoir's entropy decrease to maintain the second law's integrity. Now suppose that in (c) the demon is reset, returning to its initial state by energy exchanges with the reservoir and a reversible work source, with work E done on the demon.

The demon's entropy decrease here must be compensated by an entropy increase in the reservoir. We conclude that resetting the demon results in heat to the reservoir.

Overall, in (a)–(c) the entropy change of the universe equals that of the reservoir. The second law guarantees this is nonnegative; i.e., the reservoir cannot lose energy. The cyclic process results in an increased load energy and a reservoir internal energy that is no lower than its initial value. The first law implies that the work source loses sufficient internal energy to generate the above gains; in particular, the source does positive work in (c). The relevant energy transfers during the cycle are: work $W > 0$ by gas on load, work $E > 0$ by work source on demon, and energy $E - W \geq 0$ added to the reservoir. The entropy change of the universe is $(E - W)/T \geq 0$, where T is the reservoir temperature. Maxwell apparently envisioned a being who could run on arbitrarily little energy, an assumption that is implicit in most treatments of Maxwell's demon. Thomson assumed demons could store limited quantities of energy for later use, implying a need for refueling. Our analysis here illustrates that if the first and second laws of thermodynamics are satisfied, the refueling (i.e. resetting) energy to a Maxwell's pressure-demon is transferred to the load as the gas and demon traverse their cycles. This suggests that resetting a demon is of fundamental importance, a view that is strengthened considerably in Section 1.5.

1.3 Szilard's Model: Entropy and Information Acquisition

Sixty-two years after Maxwell's demon was conceived, Leo Szilard introduced his famous model in which an 'intelligent' being operates a heat engine with a one-molecule working fluid (1929, Article 3.3). We briefly outline that model. Initially the entire volume V of a cylinder is available to the fluid, as shown in Figure 9a. Step 1 consists of placing a partition into the cylinder, dividing it into two equal chambers. In Step 2 a Maxwell's demon determines which side of a partition the one-molecule fluid is on (for the sake of illustration, Figure 9b shows the molecule captured on the right side), and records this result. In Step 3 the partition is replaced by a piston, and the recorded result is used to couple the piston to a load upon which work W is then done (Figures 9d and 9c). Strictly speaking, the load should be varied continuously to match the average force on the piston by the fluid, enabling a quasistatic, reversible work process. The gas pressure moves the piston to one end of the container, returning the gas volume to its initial value, V (Figure 9a). In the process the one-molecule gas has energy $Q = W$ delivered to it via heat from a constant-temperature heat bath.

After Step 3 the gas has the same volume and temperature it had initially. The heat bath, which has transferred energy to the gas, has a lower entropy than it had initially. It appears that without some other effect, the second law of thermodynamics has been violated during the cyclic process. Szilard observed: 'One may reasonably assume that a measurement procedure is fundamentally associated with a certain definite average entropy production, and that this restores concordance with the second law. The amount of entropy generated by the measurement may, of course, always be greater than this fundamental amount, but not smaller.' He further identified the 'fundamental amount' to be $k \ln 2$. His observation was the beginning of information theory.

The ingenuity of Szilard's engine is striking. His tractable model allows thermodynamic analysis and interpretation, but at the same time entails a binary decision process. Thus, long before the existence of modern information theory and the computer age, Szilard had the foresight to focus attention on the 'information' associated with a binary process. In doing so he discovered what is now called the binary digit—or 'bit'—of information. Szilard's observation that an inanimate device could effect the required tasks—obviating the need to analyze the thermodynamics of complex biological systems—was a precursor to cybernetics.

Szilard examined two other models involving memory in his 1929 paper. Unfortunately, his arguments are sometimes difficult to follow, and it is unclear whether the thermodynamic cost is from

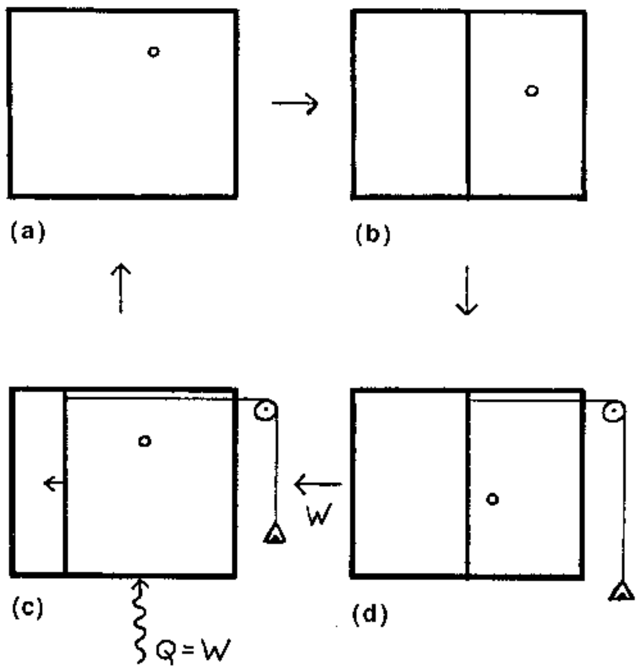


Figure 9.

measurement, remembering, or forgetting. In the course of his analyses, Szilard observed: 'Having already recognized that the only important factor (of intervention) is a certain characteristic type of coupling, a 'measurement', we need not construct any complicated models which imitate the intervention of living beings in detail. We can be satisfied with the construction of this particular type of coupling which is accompanied by memory.' His concluding sentence is, 'We have examined the 'biological phenomena' of a nonliving device and have seen that it generates exactly that quantity of entropy which is required by thermodynamics.'

Thus, Szilard regarded memory as an important feature in a demon's operation, but he did not identify its specific role in saving the second law. His writing implies the production of entropy during measurement, along with an undefined, but important, effect of the memory process. While he did not fully solve the puzzle, the tremendous import of Szilard's 1929 paper is clear: *He identified the three central issues related to information-gathering Maxwell's demons as we understand them today—measurement, information, and memory—and he established the underpinnings of information theory and its connections with physics.*

Szilard's work met with mixed response. Some researchers felt that it put the final nail in the coffin of Maxwell's demon. For example, Jordan (1949) wrote, 'This ... stands rather isolated apart from the flow of modern physical ideas; but I am inclined to regard it as one of the greatest achievements of modern theoretical physics, and believe that we are still very far from being able to evaluate all its consequences.' Much later Peter Landsberg (1982) wrote, 'Maxwell's demon died at the age of 62 (when a paper by Leo Szilard appeared), but it continues to haunt the castles of physics as a restless and lovable poltergeist.' Although it is unclear how much he was influenced by the work of Szilard, Brillouin broke new ground by developing an extensive mathematical theory connecting measurement and information. Demers (1944, 1945), Brillouin (1951a,b, 1956), and Gabor (1964) were led to the conclusion that the second law is linked to the quantum nature of light.

On the negative side, there were criticisms of Szilard's efforts to link entropy and information. Popper (1974) described Szilard's suggestion that knowledge and entropy are related as 'spurious'. Similar criticisms may be found elsewhere (see Popper, 1957; Feyerabend, 1966; Chambadal, 1971; Jauch and Báron, 1972, Article 3.5). Some objections emanate from the view that thermodynamical entropy is a measurable quantity (within an additive constant) that is independent of an observer's knowledge, and any other definition of entropy that is observer-dependent is unacceptable.

Rothstein (1957) clarified this point as follows: 'Demons do not lower entropy; the information they act on defines a lower entropy state of the system than one not subject to the restrictions in which the information consists.' Later, Rothstein (1971) elaborated further on this point: 'Physical information and its associated entropy reduction, localized in the system to which the information refers, can be expressed via specifications or constraints taken as part of the description of a system, or can be obtained from measurement. Measuring a system and thus finding it to be in some state is formally equivalent ... to preparing it to be in that state, specifying it to be in that state, or constraining it in a manner so that it can be in no other state (the state in question can, of course, be mixed).' The intimate connections between entropy, a system property, and information, a property of the observer, are discussed also by Morowitz (1970).

Along similar lines, E T Jaynes (1979) wrote, 'The entropy of a thermodynamic system is a measure of the degree of ignorance of a person whose sole knowledge about its microstate consists of the values of the macroscopic quantities X_i which define its thermodynamic state. This is a completely "objective" quantity, in the sense that it is a function only of the X_i, and does not depend on anybody's personality. There is then no reason why it cannot be measured in the laboratory.'

Jaynes (1965) also observed that a given physical system corresponds to many different thermodynamic systems. Entropy is not a property simply of the system, but of the experiments chosen for it. One normally controls a set of variables, and measures entropy for that set. A solid with N atoms has approximately $6N$ degrees of freedom, of which only a few (e.g., temperature, pressure, magnetic field) are usually specified to get the entropy. By expanding that set (say, to include components of the strain tensor), we could get a sequence of entropy values, each of which corresponds to a different set of constraints. Extension of this process ultimately gets one outside the normal domain of thermodynamics, for which the number of degrees of freedom greatly exceeds the number of thermodynamic variables.

A one-molecule system never satisfies the latter description because experiments do not even exist. Thus the use of the entropy concept—or any other thermodynamic concept—must be clarified. One possible clarification envisages an ensemble of similar systems, the average behavior of which is related to a 'typical' single system. In ordinary statistical mechanics of macroscopic systems, the system of interest is typically in contact with a constant-temperature reservoir. Energy exchanges between system and reservoir go on continually, and observations over a long time period can in principle detect fluctuations about a well-defined, time-averaged energy. The ensemble description replaces the single system, viewed over an extended time by a collection of many similar systems, all viewed at a chosen time. Use of the ensemble approach assumes the equality of time and ensemble averages.

In the present context one may consider taking a large number of one-molecule gases through Szilard's cycle. Using statistical mechanics, entropy and average pressure may be defined as meaningful thermodynamic properties of the ensemble. One must choose 'appropriate' variables of the system over which to take statistical averages. In the Szilard cycle, left (L) and right (R) side indexes are appropriate. In a sense these are measurable 'macroscopic' variables. Being outside the normal domain of thermodynamics, the Szilard model can be criticized as having no thermodynamic significance. An alternative viewpoint, which we take here, is that it gives an opportunity to extend thermodynamic concepts into interesting new territory, with some help from information theory. Jordan (1949) described this well: '...the tendency of Szilard's views is to acknowledge also a microphysical applicability of thermodynamics.' We discuss this further in Section 1.5.

In summary, Szilard's contributions have influenced the way we think about entropy. Through Szilard's ideas, Maxwell's demon led to the concept of a 'bit' of information and to key concepts in information theory, cybernetics, and computing. In a remarkable and fitting reciprocation, modern day theories of computing have led to a new understanding of the Maxwell's demon puzzle. A new, fundamentally different resolution of that conundrum involves erasure of the demon's memory, a point that Szilard just narrowly missed in 1929. We return to this in Section 1.5.

1.4 Information Acquisition via Light Signals: A Temporary Resolution

As mentioned earlier, Leon Brillouin (1951, Article 3.4) and (independently) Dennis Gabor (1951; published in 1964, reprinted in MD1) followed up on the measurement aspect of Maxwell's demon about 20 years later using the quantum nature of light. Because quantum theory was still not invented during his lifetime, Maxwell could not have foreseen that his demon would provide a path to the quantum domain. But a small demon exists in a sea of gas molecules and photons. The photons, quanta of the blackbody electromagnetic radiation within the vessel, have a well-defined energy distribution dictated by quantum theory. In the mid 1940s, Pierre Demers recognized that because of this, a high temperature lamp is needed to provide signals that are distinguishable from the existing blackbody radiation. Brillouin, who was influenced by Demers' studies, adopted this assumption.

Consider the nonuniform wavelength distribution of blackbody radiation. For a gas temperature T, Wien's law gives the wavelength of maximum spectral density, $\lambda_m(T) \approx 2900/T$ [μm]. Assuming an ambient temperature $T = 290$ K, the wavelength region in the vicinity of $\lambda_m(290) \approx 10$ μm can be avoided by having the lamp emit substantial radiation power with $\lambda \ll \lambda_m$. Using a constant-emissivity approximation, a lamp with radiating temperature 1500 K has $\lambda_m(1500) \approx 2$ μm, and an incandescent light bulb with filament temperature $3,000$ K has $\lambda_m(3,000) \approx 0.1\mu$m. Whether a lamp's radiation is distinguishable from ambient blackbody radiation depends on the power incident to the demon's eyes in the low wavelength region with and without the lamp. The radiating area of the lamp and geometrical considerations can be important, but the details of this complicated problem do not appear to have been pursued in the literature. It is clear that a lamp giving distinguishable signals can be chosen: humans regularly use high-temperature incandescent lamps.

Could a low temperature radiator, say, with $T = 100$ K and $\lambda_m = 29$ μm be used? This is less satisfactory for two reasons. First, the total power radiated is proportional to $A_s T^4$, where A_s is the radiating surface area—and a low-temperature lamp must have a much larger radiating surface area to emit the same total power as a high temperature source. The radiating surface of a 100 K radiator must be $810,000$ times larger than that for a $3,000$ K lamp with the same total power output. Second, higher wavelength radiation is accompanied by more pronounced diffraction effects than low wavelength light, decreasing the demon's ability to resolve signals.

Brillouin assumed a high temperature lamp and melded the developing field of information theory with the Maxwell's demon puzzle. The first assumption, together with judicious use of the quantum nature of radiation, enabled an explicit demonstration that information gathering via light signals is accompanied by an entropy increase. This increase is sufficient to save the second law.

The mathematical theory of information had been solidified by Claude Shannon (Shannon and Weaver, 1949) in connection with communication processes. Shannon introduced a mathematical function, which he called information entropy, to analyze the information carrying capacity of communication channels. Although Shannon's function bears a striking mathematical resemblance to the canonical ensemble entropy of statistical mechanics, Shannon's stimulus, method of attack, and interpretation were very different. Brillouin boldly postulated a direct connection between information entropy and thermodynamic entropy.

Suppose a physical system can be in any of P_0 states with equal likelihood, and we do not know which state is actually occupied. Brillouin assigned information $I_0 = 0$ to signify total ignorance. If by measurement we eliminate some of the states as possibilities, reducing the number to $P_1 < P_0$, the information so gathered is defined as $I_1 \equiv K' \ln(P_0/P_1) > 0$. K' is an undesignated positive constant. Had the number of states increased, I_1 would be negative; i.e., we would have lost information. These ideas are described in more detail in Article 3.4.

Five years after his path-breaking article, Brillouin published *Science and Information Theory*, which solidified his ideas on the subject. There he distinguished between two kinds of information, 'free' and 'bound', in order to handle information that did not have thermodynamic significance. Free information (I_f) was regarded as abstract and without physical significance. Bound information (I_b) was defined in terms of the possible states of a physical system. Brillouin gave as an example of free information the knowledge possessed by an individual. That knowledge is transformed into bound information when it is transmitted from one individual to another via physical signals.

According to Brillouin it is the physical character of signals that makes the information they carry 'bound.' In the communication process, the information might get distorted or partially lost; i.e., I_b can decrease. When the resulting bound information is received by another individual, it is again considered to be free information. Brillouin linked changes in bound information to changes in entropy of a physical system via the hypothesis: $I_{b1} - I_{b0} \equiv k(\ln P_0 - \ln P_1) = S_0 - S_1 > 0$, where the initially arbitrary constant K' has been chosen to be Boltzmann's constant, k; and S_0 and S_1 are the initial and final entropy values for the physical system. Choosing $K' = k$ makes information entropy and physical entropy comparable in the sense that they have the same units.

Brillouin's hypothesis implies that gaining bound information about a physical system decreases its physical entropy. He then made two further important steps. First, he defined 'negentropy' $\equiv N \equiv -(entropy)$ and thus negentropy change $\Delta N = -$ (entropy change) $= -\Delta S$. Second, he applied his negentropy principle of information to an isolated physical system. Suppose this system's entropy is $S_1 = S_0 - I_{b1}$, as above. The second law of thermodynamics is then written:

$$\Delta S_1 = \Delta(S_0 - I_{b1}) = \Delta S_0 - \Delta I_{b1} = -\Delta N_0 - \Delta I_{b1} \geq 0,$$

or simply,

$$\Delta(N_0 + I_{b1}) \leq 0.$$

With this last result Brillouin gave a new interpretation of the second law of thermodynamics: The quantity (negentropy + information) can never increase, and in a reversible transformation, the sum remains fixed. He applied these ideas to 'exorcise' Maxwell's demon.

As might have been anticipated, Brillouin's proposal to generalize and reinterpret the second law got considerable attention, splitting the scientific community into groups of believers and nonbelievers. If the subsequent literature accurately reflects level of belief, the believers are more numerous, for Brillouin's method is widely quoted (see for example: Barrow, 1986; Bell, 1968; Ehrenberg, 1967; Dugdale, 1966; Rex, 1987; Waldram, 1985; Zemansky, 1981; Yu, 1976; Rodd, 1964). Unqualified acceptance is evident in a paragraph labeled, 'Obituary: Maxwell's Demon (1871-c.1949)' (Bent, 1965), reprinted in the chronological bibliography. Acceptance is evident also in all 20 articles that comprise Chapters 4–7 in this book.

Though smaller in numbers, nonbelievers leveled thoughtful criticisms of the subjectivity implied by Brillouin's theory. (In contrast, recall arguments illustrating and supporting objectivity of entropy within the informational approach in Section 1.3.) Among the most vociferous critics of Brillouin's theory is Kenneth Denbigh, (1981, reprinted in MD1) who rejects the view that entropy is subjective. He emphasizes that Brillouin's exorcism of Maxwell's demon can be accomplished solely using thermodynamic principles, without need for information theory or negentropy. Denbigh's dismay with subjectivism led to a book on the subject (Denbigh and Denbigh, 1985). Karl Popper has leveled harsh criticisms at attempts to link information and thermodynamics (Popper, 1957, 1974, 1982). Much of this is focused on Szilard's 1929 paper, which began the process of associating information and entropy (see Section 1.3).

Rudolph Carnap (1977) wrote, 'Although the general identification of entropy (as a physical concept) with the negative amount of information cannot be maintained, there are certainly important relations between these two concepts.' He praised Szilard's work analyzing the paradox of Maxwell's demon as showing an important connection between entropy and information. He summarized Brillouin's ideas, which '...are certainly interesting and clarify the situation with respect to Maxwell's paradox in the direction first suggested by Szilard.' Despite this commendation Carnap also took issue with Brillouin's identification of negentropy with information: 'However, when Brillouin proceeds to identify negentropy with an amount of information, I cannot follow him any longer ... He does not seem to be aware that the definition of S which he uses (and which he ascribes to Boltzmann and Planck) makes S a logical rather than a physical concept.' We return to connections between logical and physical concepts in the next section. More recent critiques of the information-entropy connection are discussed in Section 1.8.1.

In work complementary to Brillouin's, Gabor (1964, reprinted in MD1) analyzed the use of light signals to operate the Szilard engine. Although that work was not published until 1964, it was actually reported in lectures Gabor presented the same month that Brillouin's paper was published. The point of Gabor's treatment was to illustrate that a Maxwell's demon could in principle violate the second law if the light used satisfies *classical* laws. Employing a cleverly designed system with an incandescent lamp, mirrors, and photodetector, Gabor found that if the light intensity can be made arbitrarily large relative to the background blackbody radiation, then the second law is vulnerable. He argued, however, that this is prohibited by quantum theory because 'Very weak beams of light cannot be concentrated.'

Attempted resolution of the Maxwell's demon puzzle by focusing on information acquisition was an important phase of the demon's life. It is interesting that the focus on information acquisition seemed to eliminate all interest in the memory aspects that Szilard emphasized. This is nowhere more clear than in Brillouin's decision to define two types of information, one of which ('free' information, I_f) was designed explicitly to deal with 'knowledge,' and the other ('bound' information, I_b) was linked to entropy changes. In effect this inhibited considerations of the *physical* aspects of memory. Ironically, it is these physical effects of memory that subsequently led to an overthrow of the resolutions proposed by Brillouin and Gabor!

1.5 Computers and Erasure of Information: A New Resolution

1.5.1 Memory Erasure and Logical Irreversibility

Recall that after Step 3 in the Szilard model discussed in Section 1.3, the demon retains the memory of its finding, plus any other effects of the measurement process. We assume the demon has experienced zero temperature change and negligible, if any, 'other' effects of the measurement. In order to make the process within the gas-demon system cyclic, thereby restricting all thermodynamic changes to the environment, the memory evidently must be erased. The thermodynamic consequences of this process become of fundamental interest.

Landauer (1961, Article 4.1) introduced the concept of 'logical irreversibility' in connection with information-discarding processes in computers. Memory erasure, which takes a computer memory from an (arbitrary) existing state A, to a unique, standard reference state L discards information in a logically irreversible way. Logical irreversibility means that the prescription 'Map the existing state A to the state L' has no unique inverse because state A can be any of many possible states in the computer's memory.

Put differently, starting from state L, one cannot get to the state A without using further information - e.g., the computer program and the initial data that led to state A in the first place.

Landauer argued that to each logical state there must correspond a physical state. Logical irreversibility carries the implication of a reduction of physical degrees of freedom, resulting in 'dissipation'. This is a subtle concept. We show shortly that logical irreversibility does not necessarily imply physical irreversibility in the thermodynamic sense. Rather, it can manifest itself in terms of a thermodynamically reversible conversion of work to heat. That is, the work of erasure W can result in heat $Q = W$ going to the environment. Landauer also argued that computation steps that do not discard information, e.g., writing and reading, can be thermodynamically reversible in principle.

Charles Bennett (1973, reprinted in MD1) extended Landauer's work, arguing that a computing automaton can be made logically reversible at every step. This allows an in-principle thermodynamically reversible computer that saves all intermediate results, avoiding irreversible erasure, prints out the desired output, and reversibly disposes of all undesired intermediate results by retracing the program's steps in reverse order, restoring the machine to its original condition.

Subsequently Bennett (1982, Article 7.1) argued that a demon's memory may be viewed as a two-state system that is set in a standard state prior to measurement. The measurement process increases the available phase space of the memory from one state to two (e.g., in an ensemble of systems, measurement can lead to either state). Memory erasure returns it to the standard state L, thereby compressing a two-state phase space to a single state. This is a logically irreversible act that is accompanied by an entropy transfer to the reservoir.

Bennett showed that if all steps in the Szilard model are carried out slowly, the resulting entropy increase of the reservoir compensates exactly for the entropy decrease of the demon's memory and saves the second law. Strictly speaking this cyclic process is thermodynamically reversible: the gas, demon, and reservoir are all returned to their initial states. Bennett uses a phase space that contains both the states of the particle (left or right) and demon memory (also left and right). Fahn (1994, 1996) expanded this idea, and argued that the expansion step, which decorrelates the gas from the memory is the source of the dissipation that saves the second law.

In his 1970 book *Foundations of Statistical Mechanics*, Oliver Penrose independently recognized the importance of 'setting' operations that bring all members of an ensemble to the same observational state. Applied to Szilard's heat engine, this is nothing more than memory erasure. Penrose was unaware of Bennett's work, and it is interesting that he arrived at similar conclusions but by rather different arguments. Penrose wrote:

> The large number of distinct observational states that the Maxwell demon must have in order to make significant entropy reductions possible may be thought of as a large memory capacity in which the demon stores the information about the system which he acquires as he works reducing its entropy. As soon as the demon's memory is completely filled, however, ... he can achieve no further reduction of the Boltzmann entropy. He gains nothing for example, by deliberately forgetting or erasing some of his stored information in order to make more memory capacity available; for the erasure being a setting process, itself increases the entropy by an amount at least as great as the entropy decrease made possible by the newly available memory capacity.

Penrose did not go as far as Bennett, who argued that measurement can be done with arbitrarily little dissipation and that erasure is the fundamental act that saves Maxwell's demon. Published within a rather abstract, advanced treatment of statistical mechanics, Penrose's modest but important treatment of memory erasure went largely unnoticed among Maxwell's demon enthusiasts. We discuss his approach further in Section 1.6.1.

1.5.2 Logical versus Thermodynamic Irreversibility

Because the concept of memory erasure has generated considerable debate, further clarification is appropriate. Motivated by the Szilard model, suppose we choose our memory device to be a box of volume V, partitioned down its middle, and containing a single molecule. The molecule is either in the left (L) side or the right (R) side, and the container walls are maintained at temperature T. In effect the molecule is in a double potential well whose middle barrier potential is infinite. As before, let the standard reference state in this example be L, and consider an ensemble (see Section 1.3) of demon memories in which some of the ensemble members can occupy state L and others occupy state R.

Erasure and resetting takes each memory from its existing state and bring it to the standard state L. A crucial observation is this: It is not possible to use a specific erasure process for an L state and a different one for the R state. Why? Because that would necessitate first determining the state of each memory. After erasure, the knowledge from that determination would remain; i.e., erasure would not really have been accomplished. (In Section 1.8.2 an attempt to circumvent this restriction is addressed.)

An acceptable erasure and resetting process must work equally well for either initial memory state (L or R). For example, this can be accomplished by the following two-step algorithm applied to each ensemble member:

(i) To effect erasure, remove the central partition from each ensemble member.
(ii) To effect resetting, slowly compress each gas isothermally to the left half of the box.

The diffusion process in the erasure step (i), eradicates the initial memory state. Despite the fact that this process is logically irreversible, it is thermodynamically reversible for the special case where the ensemble has half its members in state L and half in state R. This is evident from the fact that partition replacement leads to the initial thermodynamic state (assuming fluctuations are negligibly small). Isothermal compression in (ii) means that the walls of the box are maintained at temperature T. Each gas molecule's energy, on average, is determined by the wall temperature, and the work of compression on each memory results in a transfer of energy to the constant-temperature reservoir. For the ensemble, the average work W must equal the average heat Q to the reservoir. Thermodynamically, work has been 'converted' to heat, and entropy $\Delta S_{res} = Q/T = W/T = k \ln 2$ has been delivered to the reservoir. This example illustrates how the act of blurring the distinction between L and R can be linked to the delivery of entropy to the reservoir.

How has the ensemble entropy of the memory changed during the erasure process? Under our assumptions, the initial ensemble entropy per memory associated with the equally likely left and right states is $S_{LR}(\text{initial}) = k \ln 2$. After erasure and resetting, each ensemble member is in state L, and $S_{LR}(\text{final}) = 0$. Therefore, $\Delta S_{LR} = -k \ln 2 = -\Delta S_{res}$. In this sense the process is *thermodynamically reversible*; i.e., the entropy change of the universe is zero. This counterintuitive result is a direct consequence of the assumed uniform initial distribution of ensemble members among L and R states. During erasure, work from an external source has been used to effect energy transfer to the reservoir, but without altering the entropy of the universe.

Further understanding of the erasure-resetting procedure's thermodynamically reversible character for a uniform initial distribution of L and R states can be gained by reversing the steps of that procedure. Starting with all ensemble memories in state L ($S_{LR} = 0$), let each gas in the ensemble slowly expand isothermally to the full volume V. The performance of average work $W = kT \ln 2$ by each gas on its work source (now a work *recipient*) induces energy transfer $Q = W$ from the reservoir to the gas. The average gas entropy increases by $\Delta S_{LR} = k \ln 2 = -\Delta S_{res}$. Subsequent placement of the partition has zero entropic effect, because (approximately) half the ensemble members are likely to end up in each of the two states.

The fact that some specific systems that were initially L become R, and vice versa, illustrates that the process is *logically* irreversible. However, it is *thermodynamically* reversible in the sense that carrying out the steps in reversed order: (a) re-establishes the initial distribution of L and R states among ensemble members; (b) returns energy Q, transferred from gas to reservoir during resetting, back to the gas; (c) returns energy $W = Q$ used to effect erasure-resetting back to the external work source; and (d) leaves the entropy of the universe unaltered.

We emphasize that memory erasure and resetting is always *logically* irreversible. It is *thermodynamically* reversible only when the initial memory ensemble is distributed uniformly among L and R states. To see how erasure can be thermodynamically irreversible, consider the case where all ensemble memories are initially in state L. In the above two-step erasure-resetting procedure, partition removal in step (i) is thermodynamically irreversible, with the entropy change of the universe equaling $\Delta S_{LR} = k \ln 2$. During the subsequent compression of each ensemble member in step (ii), external work W results in heat $Q = W$ to the reservoir. The initial and final ensemble entropy values of the gas are both zero, and the average entropy change of the universe equals that of the reservoir, namely, $k \ln 2$, which is attributable to irreversible partition removal. Similar reasoning shows that the erasure-setting combination is both thermodynamically *and* logically irreversible whenever the initial ensemble of memories is not distributed equally among L and R states.

One might argue that prior to an erasure procedure, the memory of a single memory device (rather than an ensemble of memory devices) is in a fixed state and its entropy S_{LR} must be zero. (Note, however, that it is shown in Section 1.8.1 that *algorithmic* entropy is *nonzero* for a given memory state.) With this view, erasure brings the memory to another single state with zero entropy, and the entropy change of the memory is zero. The only entropy change is the positive one in the reservoir, and the process must be viewed as *thermodynamically* irreversible. Whether this or the previous interpretation is used, the crucial point is that memory erasure saves the second law, and discarding information results in energy dissipation as heat to the environment.

The foregoing analysis suggests that the entropy of a collection of Szilard gases does not change when partitions are installed or removed. Without partitions installed, and without the use of special measurements, we expect half the boxes in our ensemble to have their molecules on the left and half on the right at any chosen time, giving an ensemble entropy $S_{LR} = k \ln 2$. This is unchanged by placement of a partition in each box and is unchanged again upon partition removal. Thus, for both replacement and removal of partitions, the change in the ensemble entropy of the gas is zero. John von Neumann (1955; originally published in German, 1932) recognized this in his *Mathematical Foundations of Quantum Mechanics*, writing:

> ... if the molecule is in the volume V, but it is known whether it is in the right side or left side ... then it suffices to insert a partition in the middle and allow this to be pushed ... to the left or right end of the container ... In this case, the mechanical work $kT \ln 2$ is performed, i.e., this energy is taken from the heat reservoir. Consequently, at the end of the process, the molecule is again in the volume V, but we no longer know whether it is on the left or right ... Hence there is a compensating entropy decrease of $k \ln 2$ (in the reservoir). That is, we have exchanged our knowledge for the entropy decrease $k \ln 2$. Or, the entropy is the same in the volume V as in the volume $V/2$, provided that we know in the first mentioned case, in which half of the container the molecule is to be found. Therefore, if we knew all the properties of the molecule before diffusion (position and momentum), we could calculate for each moment after the diffusion whether it is on the right or left side, i.e., the entropy has not decreased. If, however, the only information at our disposal was the macroscopic one that the volume was initially $V/2$, then the entropy does increase upon diffusion.

It is notable that von Neumann associated entropy decrease with the demon's knowledge. Had he addressed the process of *discarding* information, which is needed to bring the demon back to its initial state, he might have discovered the Landauer–Penrose–Bennett (LPB) resolution of the puzzle decades earlier.

The idea that neither partition placement nor removal changes the one-molecule gas entropy is supported and clarified by a quantum mechanical analysis of entropy changes for the gas (and memory) given by Zurek (1984, Article 5.1). His work was evidently inspired by a criticism of the Szilard model by Jauch and Báron (1972; Article 3.5), who argued that the Szilard model is outside the realm of statistical physics, and should be dismissed altogether! That opinion was subsequently rebuked by Costa de Beauregard and Tribus (1974; Article 3.6). Zurek viewed partition insertion in terms of the introduction of a thin potential barrier of increasing strength V_0. When $V_0 = 0$ there is no barrier, and when V_0 is made sufficiently large, the barrier is effectively impenetrable. As V_0 is increased, the wave function of the molecule distorts, and Zurek shows the entropy to be unchanged by partition insertion. Zurek's work is discussed further in Section 1.7.2.

1.5.3 Role of Measurement in Szilard's Model

Some authors have argued that in Szilard's engine no measurement is needed prior to coupling the piston to the external load. They argue that clever design of the engine would enable the proper coupling to be made automatically. For example, Chambadal (1971) wrote as follows:

> As far as the location of the molecule is concerned, that is determined after its first collision with the piston, since the latter experiences a very small displacement in one direction or the other. We may suppose that the work supplied by the molecule can be absorbed by two gears situated in the two parts of the cylinder. After the piston has experienced the first impact we connect it, according to the direction of its motion, to one or another of these gears which will thereafter absorb the work supplied by the movement of the molecule. This coupling of the piston to the parts which it drives can also be achieved automatically.
>
> But, in fact, it is not even necessary to solve the problem of the location of the molecule. Indeed we can, without altering the principle of the apparatus at all, visualize it in the following way. When the piston is placed in the cylinder, we fix two shafts on its axis, one on either side. These shafts make contact with the piston, but are not connected to it. Consequently, whatever the position of the molecule, the piston, moving in either direction, pushes one of the two shafts and so engages the gears which make use of the work produced.

Chambadal concluded that neither entropy nor information is involved in this model.

Popper (1974) and Feyerabend (1966) proposed similarly modified Szilard engines that couple the piston via pulleys to equal weights on either side of it. The weights can be lifted by the pulley system but are constrained such that they cannot be lowered (see Figure 10). If the engine's molecule is in the left chamber, the piston moves to the right, raising the left weight, leaving the right weight unmoved. If the molecule is in the right chamber, the reverse happens; i.e., the right weight gets raised and the left weight stays fixed. Feyerabend wrote 'The process can be repeated indefinitely ... We have here a "perpetual source of income" of the kind von Smoluchowski did not think to be possible.'

Jauch and Báron (1972, Article 3.5) imagined a similar situation (see Figure 11), writing: 'Near the mid-plane of the cylinder and on both its sides are electrical contacts in its walls. When activated by the piston's motion along them, they operate mechanisms which attach a weight to the piston in whichever direction it moves. Thus a weight is lifted and the engine performs work, without interference by a conscious observer.'

An ingenious coupling was illustrated by Rothstein (1979). His intent was not to argue against Szilard's work but rather to rebut Popper who had attempted to do so. Rothstein couples the piston to two racks that alternately engage a pinion gear as it moves left or right (see Figure 12). When it moves left, one

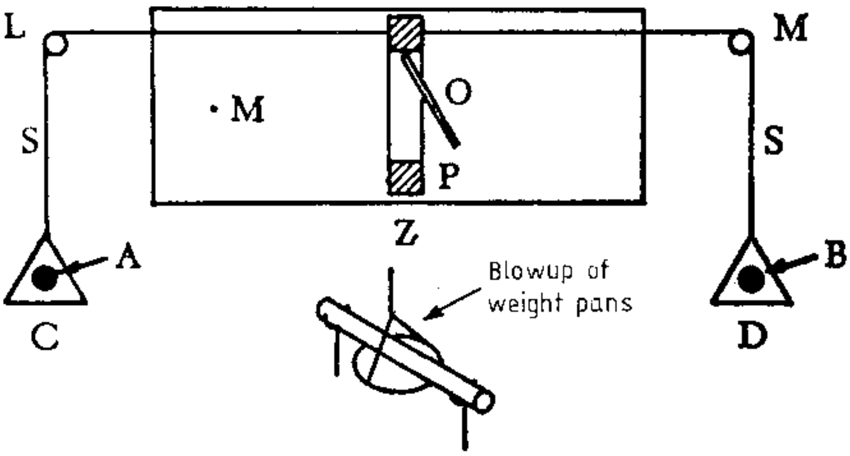


Figure 10.

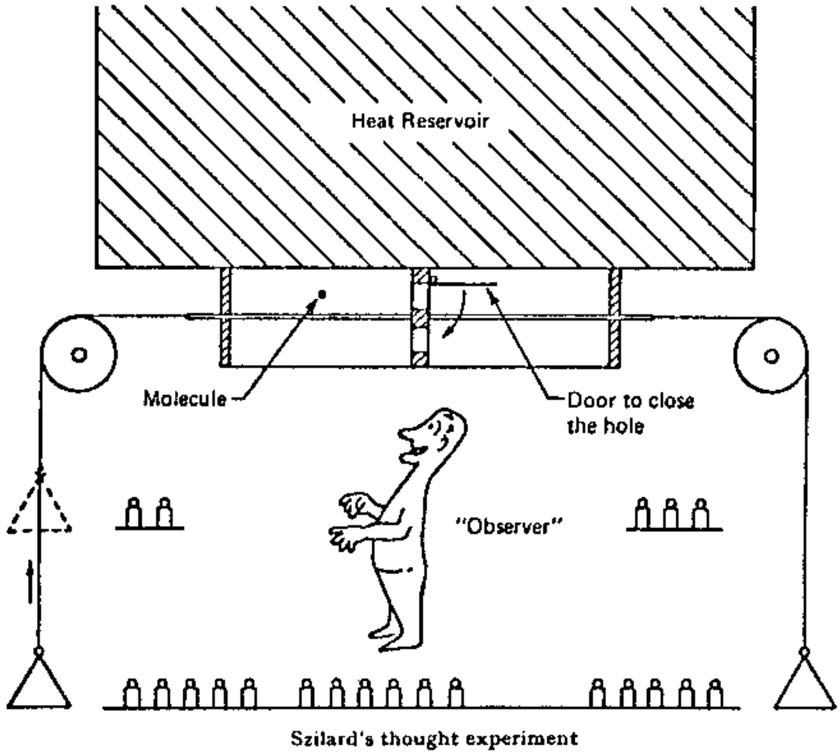
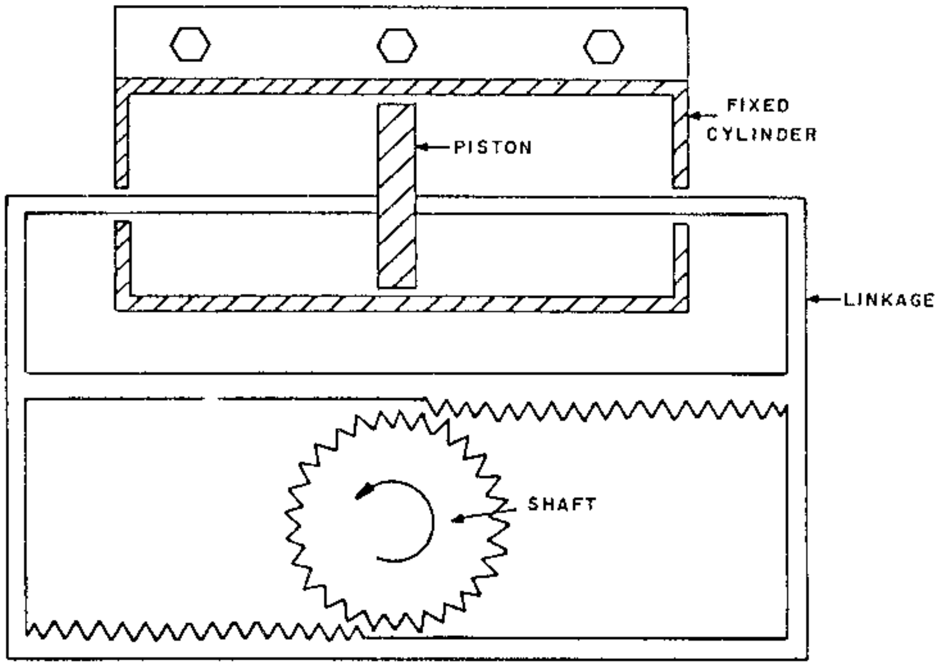


Figure 11.

rack rotates the pinion gear counterclockwise while the other rack is disengaged. When the piston moves right, the second rack (diametrically opposed to the first) rotates the pinion gear, again counterclockwise. Thus, regardless of whether the molecule is in the left or right chamber, the design of the racks assures



Thought experiment illustrating Popper's refutation of Szilard's assignment of an entropy equivalent to physical information.

Figure 12.

counterclockwise motion, suggesting an automatic machine for converting heat from a single reservoir to work.

The above examples are clever and do indeed *seem* to challenge the second law. However, there is more to the story. In Figure 10, after the work has been done, one of the weight hangers is raised. The pulley string at the other side is relaxed and limp. In essence, this configuration stores information about the molecule's previous location: it serves as a memory. Put differently, the process is not truly cyclic. The memory must be reset enabling commencement of the next cycle. Because resetting of the pulley system was not accounted for, the arguments by Popper and Feyerabend—while rather convincing at first look—must be considered incomplete.

Along the same lines, the Jauch-Báron idea leads to an asymmetric situation (Figure 11), with a weight attached on one side only. This is a physical embodiment of a memory that must be reset in order to make the process cyclic. Chambadal's analysis similarly overlooks the need for resetting the apparatus. In Rothstein's example, Figure 12, each cycle moves the rack assembly either left or right, where it stays until it is reset. Again this constitutes a memory register that must be zeroed periodically.

In summary, Szilard's engine requires a binary decision process in order to couple the piston and load. This requires information acquisition, memory, and subsequent information erasure. Although examples based upon macroscopic analogues involving gears and pulleys suggest that resetting can be done at arbitrarily little cost, that is misleading. Maxwell's demon actually entails a memory in which the relevant energy modes are *microscopic*. Erasure must act upon hidden degrees of freedom, without knowledge of their existing states. One cannot simply examine a register and zero it in the least energy-consuming or entropy-producing way. That examination would transfer information to another memory that would still have to be erased subsequently.

Our algorithm must be such that erasure occurs independently of the existing state of the memory. As the example above suggests, this can entail first randomizing the memory's hidden energy modes and then using a work process to bring the memory to the standard state. In a perceptive discussion of information and thermodynamics, Rothstein (1952a) observed: 'From an information viewpoint quantity of heat is thus energy transferred in a manner which has eluded mechanical description, about which information is lacking in terms of mechanical categories.' Given this observation, it is interesting that memory erasure-resetting via rearrangement of hidden degrees of freedom entails energy transfer (to the environment) via heat. In Section 1.8.2 we present a somewhat different view of memory erasure and Landauer's principle in connection with other criticisms of the LPB approach to the Maxwell's demon puzzle.

1.5.4 Entropy of Measurement Revisited

What about the entropy of measurement? As discussed earlier, Landauer showed that, in contrast with memory erasure, most computer operations could in principle be performed with arbitrarily little energy dissipation per bit. Bennett argued that a demon can do its measurements with arbitrarily little dissipation, in analogy with reading instructions in a computer. The act of 'reading' can in fact be viewed as a measurement process. Bennett proposed idealized magnetic and mechanical detection devices to buttress his argument that a Maxwell's demon can accomplish dissipationless measurement.

Were dissipationless means of detection not possible, we could simply append Bennett's erasure-based resolution to Brillouin's measurement-based resolution. But if detection can in fact be done with arbitrarily little dissipation, the Landauer–Penrose–Bennett (LPB) viewpoint implies that the exorcism accepted by a generation of researchers and teachers must now be rejected. In retrospect it is clear that assuming the use of light signals is not sufficient to rule out *all* demonic operations.

Remarkably, this lack of generality was not recognized by most researchers prior to Bennett's work (for an exception, see Penrose, 1970, p 236). Light signals became widely accepted as *the* way a Maxwell's demon collects information. In *Science and Information Theory*, after showing that detection via light signals saves the second law, Brillouin extrapolated his result: 'We have ... discovered a very important physical law ... every physical measurement requires a corresponding entropy increase, and there is a lower limit below which the measurement becomes impossible.'

Why generalization based upon a special case achieved such wide acceptance is puzzling. Landauer (1989b) wrote in this regard, 'Brillouin ... and others found dissipative ways of transferring information, and without further justification, assumed that they had discovered a minimally dissipative process. It is one of the great puzzles in the sociology of science why this obviously inadequate argument met with wide and uncritical acceptance. Only in recent years have clearer discussions emerged, and these are not yet widely appreciated.'

Had the use of light signals been questioned earlier, Brillouin's method of attack might have achieved far less credibility. Yet, despite its overly narrow view, Brillouin's work brought Maxwell's demon considerable attention in 1951 and subsequent years. The demon's popularity seemed to grow as its perceived challenge to the second law diminished. Bennett's 1982 overthrow of Brillouin's exorcism enhanced the popularity of Maxwell's demon, and more reason to retain it as a tool for understanding.

As mentioned already there is not universal agreement on the thesis that measurement can in principle be accomplished with arbitrarily little dissipation. Rothstein (1952) argued that 'the accuracy of any measurement is limited by how much entropy can be usefully expended in order to perform the measurement.' More recently (Rothstein,1988) he wrote:

> Despite several ingenious attempts to achieve reversible computation, including conceptual designs for quantum mechanical computers, we remain convinced that an entropy price for unavoidable selection, measurement, or preparation acts must be paid for every such act in physical communication or computation. ... We are willing to grant that for limited kinds of computation physical systems can be set up in principle whose dynamical equations will generate a succession of states isomorphic to the computation, and, as an idealization, such systems can be reversible. We deny that possibility for a true general purpose computer. Information must be generated and stored until it needs to be consulted. The storage is writing, i.e., a preparation of some subsystem. The consultation is reading, i.e., a measurement on some subsystem. Both kinds of operation are selective and thus demand their entropy costs.

To close this section we point out that the distinction between entropy of data acquisition and entropy of data erasure is not sharp. In Szilard's model, when a demon determines the side (L or R) in which the molecule resides, its own memory state changes from a unique, known reference state to either of two possible states. This generates an entropy increase, in an ensemble sense, that in essence 'pays' entropically for the diminished entropy of the gas. This 'entropy of measurement' is stored by the demon, and ultimately becomes entropy of erasure—which is passed on to the environment when the demon's memory is reset. In this sense, the entropy of erasure feeds to the environment entropy gained by the demon via information acquisition.

1.6 Foundations of Landauer's Principle

1.6.1 Landauer's Pioneering Work and Penrose's Independent Development

In his seminal article, Landauer (1961, Article 4.1) introduced what is now commonly known as Landauer's principle, namely, that erasure of one bit of information increases the entropy of the environment by at least $k \ln 2$. Landauer views storage of a bit of information in terms of a particle in a bistable potential well whose potential barrier is high relative to kT, to assure that a ONE (ZERO) will remain a ONE (ZERO) over time with high probability. Erasure entails modification of the bistable potential so as to bring all ensemble members to the state ONE. This compresses the available phase space and lowers the entropy of the memory. From the Clausius inequality (assuming the second law of thermodynamics holds), for any isothermal entropy change ΔS of the memory, $\Delta S \geq Q/T$, where Q is the heat *to* the memory. If $\Delta S < 0$, then $Q_{res} \equiv -Q \geq T|\Delta S|$. That is the heat ($Q_{res}$) *to* the reservoir is at least $T|\Delta S|$. In the case where the entropy of the demon's memory has been reduced by one bit, namely, $|\Delta S| = k \ln 2$, this gives Landauer's principle: $Q_{res} \geq k \ln 2$. Landauer's arguments in Article 4.1 are based upon these ideas, aided by simple models, under the assumption that the second law of thermodynamics is valid.

As mentioned already, independently of Landauer's work, and prior to Bennett's publications on Maxwell's demon, Oliver Penrose (1970) argued in his book, *Foundations of Statistical Mechanics*, that a demon must store information in a memory, and erasure of that memory sends entropy of at least $k \ln 2$ to the environment. This solution to the Maxwell's demon puzzle emanated from Penrose's examination of the possibility that a *perpetual motion machine* could be operated using statistical fluctuations such as Brownian motion. Penrose considers a composite system $R' = R + M$, where R is a composite system, consisting of the system of interest, plus the constant-temperature reservoir, plus the load to be lifted. M is a 'machine' that monitors statistical fluctuations and uses them to do work on the load. R' is an isolated system—i.e., a mini-universe.

Penrose defines the entropy of R' at time t to be $\tilde{S}_t = \langle S_B \rangle_t + S_d(t)$, where $\langle S_B \rangle_t$ is the Boltzmann entropy, averaged over observational states at time t, and $S_d(t)$ is the statistical entropy of the distribution for observational states at time t. Using the underlying postulates of his unique development, he shows that $\tilde{S}_{t'} \geq \tilde{S}_t$ for $t' \geq t$, consistent with the second law. In essence the second law is implied by Penrose's postulates.

The main point is that an entropy reduction in $\langle S_B \rangle_t$ requires a compensating increase $\Delta S_d(t) \geq -\Delta \langle S_B \rangle_t$ in the entropy of the distribution. This increase in $S_d(t)$ is interpreted as the number of *observational* states the machine M needs to reduce $\langle S_B \rangle_t$. Penrose calls $\Delta S_d(t)$ a 'latent contribution' to the entropy, and argues that completion of the cyclic process requires 'resetting' the machine M to its original state. The *latent contribution* to the entropy then becomes manifest as entropy of the environment. It is clear that Penrose discovered Landauer's principle, albeit in a somewhat different context than that used by Landauer.

Typically, the number of such observational states required to reduce the entropy of a macroscopic system *measurably* is enormous. An example is the cooling of 0.001 kg of water by 1 K, which requires at least $exp(10^{21})$ observational states! Penrose applies the above logic to a traditional Maxwell's demon who operates a shutter, letting faster molecules go one way and slower ones the other. He relates the increased number of observational states to the demon's memory, which stores the observed information, and calls for its erasure. This erasure saves the second law of thermodynamics.

1.6.2 Recent Proofs of Landauer's Principle

In recent years, a very different proof of Landauer's principle was given by Shizume (1995, Article 4.3), with no explicit appeal to any form of the second law of thermodynamics. He considers a *classical* particle in a bistable potential well. The particle is subjected to three forces: a random thermal force, a damping force, and the force from the bistable potential well. Shizume uses the Langevin equation of motion and the concomitant Fokker–Planck equation to prove that erasure of one bit of information leads to energy $Q_{res} \geq kT \ln 2$ sent to the constant-temperature reservoir. This is Landauer's principle. The details are in Article 4.3.

Piechocinska (2000, Article 4.4) provides proofs of Landauer's principle in the domains of both classical and quantum mechanics. In the quantum mechanical proof, she makes the following assumptions: (1) the memory consists of a particle in a bistable potential well that is in contact with a constant-temperature reservoir. (2) The canonical ensemble formalism of equilibrium statistical mechanics holds. (3) The erasure process is governed by a unitary time evolution operator. (4) The reservoir decoheres in the sense that its off diagonal density matrix elements vanish. With these assumptions, Piechocinska exploits the convexity of the exponential function $exp(-x)$ to obtain Landauer's principle. The details are in Article 4.4.

These proofs of Landauer's principle suggest that its validity is *universal*. However, in Section 1.7.4 we find that this is not necessarily so under extreme quantum conditions.

1.6.3 Zero-Work Erasure Using a Reservoir

Lubkin (1987, Article 5.2) examines erasure by putting the system in contact with a reservoir at temperature T, and others (Vedral, 2000, Article 5.5; Plenio and Vitelli, 2001) have done similarly. We assume here that the reservoir temperature is *low* relative to room temperature. The result of this zero-work type of erasure is a bit different from that for Landauer's principle. Before erasure, the memory state is mixed, with density operator ρ. Each pure component has zero entropy. During erasure each component undergoes entropy change $\Delta S_{sys} = -k_B tr(\omega \ln \omega)$ where $\omega = exp(-H/k_B T)/Z$. The concomitant entropy change of the reservoir is $\Delta S_{res} = k_B[tr(\omega \ln \omega) - tr(\rho \ln \omega)] \geq k_B[tr(\omega \ln \omega) - tr(\rho \ln \rho)]$. The total entropy change for erasure is $\Delta S = -k_B tr(\rho \ln \omega) \geq -k_B tr(\rho \ln \rho)$. This result extends

Landauer's principle to zero-work erasure with a low-temperature reservoir, showing that erasure yields a *total* entropy increase at least as large as the memory's entropy before erasure.

1.7 Quantum Mechanics and Maxwell's Demon

1.7.1 Quantum Conundrums: Szilard and Einstein

Albert Einstein devised a Gedankenexperiment that bears resemblance to Szilard's 1929 model heat engine. Fine (1986) discusses the idea as outlined in a letter from Einstein to Schrödinger in 1935. A ball is located in one of two closed boxes, but it is not known which. We might expect that the probability is $1/2$ for either possibility. Einstein was concerned with two concepts, incompleteness and separation. By definition, a complete theory would have unit probability for the ball being in one of the boxes. Otherwise the theory is incomplete. Separation means that if two objects are separated spatially by a sufficient amount, they must become independent of one another. If knowledge about one of the boxes provides information regarding the other, then separation does not hold.

Suppose the ball cannot be destroyed or duplicated, and the two boxes are separated. If observation of box 1 gives a complete answer (YES or NO) as to whether it contains the ball, it is known instantly whether the distant box 2 contains the ball. Thus completeness implies that separation is impossible. On the other hand, if separation holds, then a measurement of one box cannot give a certain answer, for that would lead to an inconsistency. That is, separation implies that completeness is impossible. The conclusion is that separation and completeness are incompatible, a perplexing finding if one believes that both must be valid. The act of observation seems inconsistent with expected physical properties.

A familiar problem of the same ilk arises when a measurement of which box contains the ball is interpreted in terms of 'collapse of the wave function.' That is, if the particle is found in box 1, the wave function collapses to zero in box 2, and it appears that the system's state is intimately connected with the observer's knowledge. Yet we expect the state of the system to be independent of the observer's knowledge.

In the Szilard problem, a possible interpretation is that the entropy decreases by $k \ln 2$ upon determination of which chamber the molecule is in. Once again the state of the observer seems to determine the state of the system. With this approach, the measurement induces an increase in the demon's accessible phase space, and thus the entropy of its memory. This increase equals the entropy decrease of the gas (assuming a nondissipative measurement). Is this a subjectivist view that 'works' but should be avoided? Not necessarily, because *any* demon making a measurement of the molecule's location would reach the same conclusion (see Section 1.3). The gas entropy is 'known' to be less to any observer who measures the left-right state.

Another interpretation is suggested for an observer *outside* the gas–demon system. To this observer, the coupling of the gas to the demon produces a correlation between the two objects but no entropy change overall, consistent with a nondissipative measurement. This outsider cannot discuss the demon and gas independently because they are linked. The coupling between gas and demon is broken when the gas does work against the piston, expanding to fill the volume. Then to the outside observer, the gas has attained the same entropy it had initially. The demon's entropy has increased, compensating for the entropy decrease of the reservoir during the isothermal expansion.

Such conundrums are not easy to resolve, or even to accept. They seem to be an integral part of microscopic, probabilistic physics, and often lead to mental discomfort. They extend from quantum mechanics to thermal physics through Maxwell's playful, imaginary demon.

1.7.2 Quantum Mechanical Treatments of Szilard's Engine

As mentioned already, Zurek (1984, Article 5.1) examined a quantum mechanical version of the Szilard one-particle gas. He replaced the partition with a slowly generated, narrow potential barrier of height V_0, centered about the container's center. The probability density $|\psi(x)|^2$ is symmetric about the container's center, as expected. Zurek assumes high temperature ($kT \gg \epsilon$), where ϵ is the ground state energy; i.e., the particle exists in the *classical* domain.

The quantum mechanical features of Zurek's treatment lie in his representation of the particle's position ket vectors $|L_n\rangle$ and $|R_n\rangle$ when the particle is in quantum state n and on the right or left sides, respectively. The corresponding demon ket vectors are $|D_L\rangle$ and $|D_R\rangle$. Zurek argues that the demon's density operator goes from $\rho_D = |D_0\rangle\langle D_0|$ initially (i.e., a pure standard state) to $\rho' = (|D_L\rangle\langle D_L| + |D_R\rangle\langle D_R|)/2$ after it measures, with a concomitant entropy increase of $k \ln 2$. If the measurement is reversible, then the total entropy change of the gas plus demon cannot change (the reservoir is not involved here), and Zurek argues that there is a compensating increase in *mutual* information. Interested readers are referred to article 5.1 for the details, and to the work of Lloyd (1989, Article 5.3) for a detailed treatment of the mutual information concept. It is fair to say that the quantum elements of Zurek's treatment in Article 5.1 are primarily useful as *interpretive* rather than *calculational* tools.

A very different quantum model has been published by Lloyd (Lloyd, 1997, Article 5.4). This is a magnetic resonance model of a quantum Maxwell's demon, which focuses on particles with spin-1/2 and magnetic moment μ in a constant magnetic field B pointing in the negative-z direction. The magnetic interaction energy is $(-\mu B, +\mu B)$ when the magnetic moment and field are (parallel, antiparallel). Experimental magnetic resonance techniques can induce spin flips from $|\uparrow\rangle$ to $|\downarrow\rangle$ or *vice versa* using a so-called π pulse with Larmor precession angular frequency $\omega = 2\mu B/\hbar$. If a spin is flipped from an antiparallel to parallel configuration, its spin state goes from $|\uparrow\rangle$ to $|\downarrow\rangle$, and the interaction energy with the field decreases by $2\mu B$. A photon of energy $\hbar\omega$ is emitted, and the spin does positive work, adding energy to the electromagnetic radiation field.

Using established conditional spin-flipping techniques, one can effect measurement of a particle's spin using a *reference* spin, and can subsequently use a π pulse to induce spin flips that result in work done on the electromagnetic radiation field. Measurement generates an 'entangled' state (see Section 1.7.3), and decoherence occurs when the reference spin is erased. Lloyd examines the complex details, using two electromagnetic-field reservoirs, in which case the demon's information gathering facilitates operation of a heat engine, and he proposes ways to achieve such a Maxwell's demon in a laboratory.

Evidently all quantum mechanical treatments of the Szilard one-particle engine to date have treated the constant-temperature reservoir in classical terms—as a nearly independent entity that interacts weakly with the system. In contrast, one might imagine a situation where the system and reservoir are inextricably linked in an 'entangled' quantum state. This is the topic of Section 1.7.4. But first we provide a brief introduction to quantum information theory.

1.7.3 Quantum Information, Maxwell's Demon, and Landauer's Principle

Landauer strongly and repeatedly made the point that 'information is physical'. That is, information is stored in material systems in accordance with the laws of physics. (Landauer, 1996a,b, Articles 7.4, 7.5) Information is also transferred from place to place by physical means—e.g., electromagnetic radiation. Information in *not* a mystical, non-physical entity.

At a fundamental level, the world we live in is quantum mechanical. Like quantum systems, quantum systems can store information—in this case, *quantum* information (Nielsen and Chuang, 2000). Quantum information is described using vectors in Hilbert space and must reduce to classical information in the classical limit. Thus the Hilbert space *contains* the classical information and, generally, also contains *much more* information than that.

The measure of classical information is the bit, an amount of information associated with binary choices between, say, 0 and 1. In contrast, the measure of quantum information is the *qubit* (quantum bit), the amount of information associated with the direction of a vector, whose tip is confined to a sphere of radius 1. Sometimes the qubit state is written $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, where the complex numbers α and β satisfy $|\alpha|^2 + |\beta|^2 = 1$. An example is an atomic electron that is in a superposition of its ground state $|0\rangle$ and its first excited state $|1\rangle$.

Unlike classical information, which can be copied perfectly, quantum information *cannot* generally be copied with perfect fidelity. Noise from the environment can cause continual bit flips within qubits. Furthermore the noise problem is exacerbated by the fact that *observing* a quantum state to see if it needs correction generally destroys that state! Nevertheless, error correcting codes that operate on *probabilistic* information can diminish such unwanted effects.

Interestingly, error correction can be viewed as a *friendly* Maxwell's demon that observes and then corrects errors. Suppose the system starts in state i and is driven, via noise, to state j, with entropy $S_j > S_i$. The demon (i.e., error-correcting code) measures the state and finds state m with probability p_m. This demon then operates on the system, which means it applies a unitary operator that brings the system to the final state f. The demon is successful only if $f = i$, in which case the system has undergone a cyclic process. The demon's memory contains information from the measurement process and must be erased to make the process cyclic for the system + demon. It can be shown (Nielsen, 2000) that, on average, the entropy sent to the environment by erasure is at least as great as the entropy reduction achieved by the demon's error-correcting operations. Thus the second law of thermodynamics is satisfied amidst the benevolent actions of the friendly demon. This is yet another example of the utility of Landauer's principle.

The domain of application of Landauer's principle is larger yet. Vedral (2000, Article 5.5) has argued that measurement can involve quantum entanglement (defined later in this section) between a system and the measuring apparatus. In such cases one might expect the degree of entanglement to be related to the amount of information that gets stored in the memory of the apparatus. Once again Landauer's principle links the latter information to the entropy sent to the environment upon erasure. Vedral argues that such erasure entropy limits the possible increase in entanglement. He goes on to link the second law of thermodynamics to the principle that 'entanglement cannot increase locally.'

In an article entitled, 'The physics of forgetting: Landauer's erasure principle and information theory,' Plenio and Vitelli (2001) argue that Landauer's principle is important throughout the new field called *the physics of information*. They argue that the amount of classical information that can be encoded in a *mixed* quantum state, described by the density operator, $\hat{\rho} = \sum_j q_j |e_j\rangle\langle e_j|$, is given by the von Neumann entropy $-tr[\hat{\rho}\log_2 \hat{\rho}]$. The argument is similar to that in Section 1.6.3. They then use Landauer's principle to obtain a form of the so-called Holevo bound of information theory (see Nielsen and Chuang, 2000). They also use Landauer's principle to argue that 'the efficiency of quantum data compression is limited by the von Neumann entropy ... just as classical data compression is limited by the Shannon entropy.'

In 1935, Erwin Schrödinger wrote, 'I would not call [entanglement] *one* but rather *the* characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought.' Two particles are entangled quantum mechanically if their state vector cannot be written as a tensor product of the state vector for one times the state vector for the other. An example is the pure state $|\psi_{ent}\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B)$, where subscripts A and B connote particles A and B.

Entangled states exhibit nonlocal correlations that cannot be understood classically. The density operator for $|\psi_{ent}\rangle$ is $\rho_{AB} = |\psi_{ent}\rangle\langle\psi_{ent}|$. If we 'trace out' particle B, to get the density operator ρ_A, we find the remarkable result, $\rho_A = \frac{1}{2}(|0\rangle_A\langle 0|_A + |1\rangle_A\langle 1|_A)$. The quantity in parentheses is a sum of two projection operators that span the A-space; i.e., it is the identity operator. Thus measurements of particle A yield each of the two possible A-states half the time and provide the observer with *no* information about the state $|\psi_{ent}\rangle$. Indeed an entangled state is a weird thing.

If one could arrange conditions such that a system of interest becomes entangled with its environment, we might then not be surprised to find extraordinary, nonclassical behavior. This is precisely the case in theoretical work by Allahverdyan and Nieuwenhuizen, described in the next section.

1.7.4 Quantum Entanglement and the Second Law: Can the Demon Win?

Is the second law of thermodynamics sacrosanct? Recent theoretical work by Allahverdyan and Nieuwenhuizen (2000a, 2000b, 2001, 2002, 2002) suggests that the answer is *no*. They find that the second law of thermodynamics is violated for some models in the extreme quantum domain. The source of the violation is *quantum entanglement* between the system and the constant-temperature reservoir with which it interacts. As described in Section 1.7.3 when such entanglement exists, the system and reservoir cannot be usefully examined separately; the entangled systems act as a single entity.

For example, energy of (system + reservoir) $\neq$ energy of (system) + energy of (reservoir), as is usually assumed in thermodynamics. The last sentence holds also if we replace energy by entropy. Allahverdyan and Nieuwenhuizen show in particular that the Clausius inequality is violated and, of most interest here, Landauer's principle does not hold. Thus in contrast to the results of Section 1.6, which suggest the *universal* validity of Landauer's principle, the results of Allahverdyan and Nieuwenhuizen indicate that there are exceptions.

Allahverdyan and Nieuwenhuizen (2001, 2002) discuss a Brownian particle in a harmonic potential, in thermal interaction with a large reservoir. Their method of attack is via *microscopic* physics, namely quantum mechanics, with relevant thermodynamics *emerging* from the analysis rather than being assumed at the outset. Traditional statistical mechanics, for example in the form of the canonical ensemble, cannot be used because it is based on three assumptions that do not apply here: (i) the entropy for the system and bath is additive, (ii) the system plus reservoir achieve thermal equilibrium quickly, with $dQ_{res} = TdS_{res}$, and (iii) the interaction energy is sufficiently 'small.'

Allahverdyan and Nieuwenhuizen use the Wigner phase space distribution to show that at sufficiently low temperatures, the system can have $dQ > 0$ while $dS < 0$. This implies that, as already stated, Clausius's inequality and Landauer's principle are both violated. This violation is referred to by Allahverdyan and Nieuwenhuizen as being due to an 'entanglement demon.' A Maxwell's demon of this ilk does not gather information, but rather functions because of the idiosyncrasies of quantum mechanics. Other such 'process' or 'disembodied' demons are discussed further in Section 1.8.3.

1.8 Other Aspects of Maxwell's Demon and Computation

1.8.1 Algorithmic Entropy

Charles Bennett (1982, Article 7.1) observed that under certain circumstances algorithmic entropy, defined within the framework of algorithmic information theory, is a microscopic analogue of ordinary statistical entropy. Zurek (1989a, 1989b) extended the definition of algorithmic entropy to physical systems. He argued that physical entropy must consist of two distinct contributions: (i) a term $I(s)$ that represents the randomness of known aspects of the system in microstate s; and (ii) a term H, representing the remaining ignorance of the observer about the system's actual state. In his words, 'This recognition of the dual nature of physical entropy allows one to consider 'engines' operated by a modern day 'Maxwell demon'—a universal Turing machine capable of measuring and processing information—without endangering the validity of the second law.'

To a Maxwell's demon, after a measurement, a reliable memory is in a *specific* memory state and no ensemble is necessary to describe it. The quantity $I(s)$ in the previous paragraph is the so-called algorithmic entropy (also called algorithmic information), defined as the length, in bits, of the shortest computer program that runs on a *universal* computer and fully specifies the state s. For a memory, state s is a string of ONES and ZEROES. Algorithmic entropy is maximal for a *random* string s, in which case $I(s) = \log$*(length of the string in bits)*.

Despite the fact that algorithmic entropy is defined for a system in a *specific* microstate, it is useful nevertheless to consider an ensemble of M systems. If state s_i occurs with probability p_i in the ensemble, the *average* algorithmic entropy is defined as $\langle I \rangle \equiv \sum p_i I(s_i)$, where $i = 1, 2, \ldots M$. The reason this is useful is the following important result from algorithmic information theory: For sufficiently large M, $\langle I \rangle \approx -\sum p_i \log_2 p_i = H$, the Shannon entropy in bits. These ideas, and those that follow, are elaborated in work by Caves (1993, Article 6.1) and Zurek (1999, Article 6.3). Mathematical details concerning the goodness of the approximation $\langle I \rangle \approx H$ are given by Schack (1997, Article 6.2).

The Shannon entropy H is proportional to the Gibbs entropy S_G. Specifically, $S_G = k_B(\ln 2)H$ and thus, $S_G = -k_B \sum p_i \ln p_i = k_B(\ln 2)H \approx k_B(\ln 2)\langle I \rangle$; i.e., $\langle I \rangle \approx H = S_G/(k_B \ln 2)$. This result shows that the *average* algorithmic entropy $\langle I \rangle$ for an ensemble of similar systems can be well approximated using the Gibbs entropy obtained via statistical mechanics. This is fortuitous because there is no known way to systematically *calculate* I. If I for a system is well approximated by $\langle I \rangle$ for the ensemble, then I can be determined from a knowledge of S_G. In such cases, algorithmic entropy provides a definition of entropy for a microstate, just as an energy eigenvalue is defined for each microstate.

The above definition of physical entropy, in bits, written as $S_{phys} = H + I$, is useful to discuss Maxwell's demon. If the demon obtains information about a gas reversibly, then we might expect the 'physical entropy' S_{phys} to be unchanged by the measurement. Specifically we expect that measurement increases I (the information known about the system) and therefore must diminish H (the missing information) *in the view of the demon*. This expectation is borne out by the following mathematical result: For an ensemble of similar gases and demons, *reversible* information gathering induces $\Delta S_{phys} = \Delta H + \langle \Delta I \rangle \approx 0$.

Thermodynamically, the isothermal work W made possible by a Shannon entropy reduction $-\Delta H$ is limited to $W \leq -k_B T(\ln 2)\Delta H$. Given the average erasure cost of $W_{erase} \geq k_B T(\ln 2)\langle \Delta I \rangle$ from Landauer's principle, the *net* work possible is $W_{net} = W - W_{erase} \leq -k_B T(\ln 2)\Delta S_{phys} \approx 0$, or $W \leq W_{erase}$. Thus the erasure cost is at least as great as the work made possible by the demon's observations, and the demon's actions *cannot* turn heat from the reservoir into an equal quantity of work with no other effect. This is consistent with the second law of thermodynamics. Zurek (1999, Article 6.3) applies these ideas to a *demon of choice* that places the partition in the Szilard gas well off center, trying to obtain work only when unlikely states (with the particle in the smaller volume) are observed. Zurek shows that this strategy does not succeed, and that the demon cannot defeat the second law with such shenanigans.

1.8.2 Challenges to the LPB View

Although the Landauer–Penrose–Bennett (LPB) solution of the Maxwell's demon puzzle has been accepted widely, some researchers have argued against its validity. One such challenge has come from an in-depth critical analysis by Earman and Norton (1998, 1999). They allege a number of deficiencies in the LPB solution, writing:

> ...exorcisms of the demon provide largely illusory benefits ... they either return a presupposition that can be had without information theoretic consideration or they postulate a broader connection between information and entropy than can be sustained.

> 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.

A major point of criticism by Earman and Norton is that Landauer's principle emerges from the demand that the second law of thermodynamics must be satisfied; i.e., it is not *independent* of the second law. Thus, it appears that the second law is being used to save itself! They were evidently unaware of the proof of Landauer's principle by Shizume (1995, Article 4.3), described in Section 1.6.2. That proof does *not* assume the second law. Earman and Norton's work also preceded the proofs (classical and quantum mechanical) by Piechocinska (2000, Article 4.4), described in Section 1.6.2. Those proofs also did not entail an explicit assumption of the second law. Earman and Norton also seem to have been unaware of the work of Penrose (1970), whose postulational development of statistical mechanics leads to a proof of the second law in the form of the principle of entropy increase, and then to Landauer's principle. The existence of these proofs weakens Earman and Norton's criticisms of the LPB approach.

Another thought-provoking criticism by Earman and Norton concerns a modified Szilard engine, which requires information gathering by a Maxwell's demon, but which (they argue) does *not* require logically irreversible erasure. It is claimed that this model violates the second law of thermodynamics and is a counterexample to the LPB framework. Specifically, the model is defined as follows.

The standard memory state is taken to be left (L). The demon places the partition in the container's center and then determines the side where the particle resides. If the result is L, the memory remains in the standard state, and the demon fits the engine with pulleys that enable it to do isothermal work $W = kT \ln 2$ on an external load. Memory erasure is unnecessary because the memory state is already L, the standard memory state. If the particle is on the right (R), the demon fits the engine with pulleys appropriately, enabling it to do work W. In this case, Earman and Norton argue, the memory is *known* to be in state R because the R-algorithm (subroutine) is being run. It is possible to design the R-algorithm to transform R to L with arbitrarily little entropy production and outside work.

Earman and Norton propose including such memory-specific algorithms within the L and R subroutines so that the memory ends up in state L regardless of the particle's observed position—and with arbitrarily little entropy sent to the reservoir. Landauer's principle does not apply here, and they argue that work W is done by the gas *without* sending entropy $k \ln 2$ to the environment. Their main point is that the net result is transfer of energy Q from the reservoir and performance of work $W = Q$, with no other effect. If this argument is correct, then the second law is violated.

The Earman and Norton argument has been rebutted by Bub (2001) and Bennett (2002). In the operation of a computer, various types of information-bearing degrees of freedom exist. One type, discussed already, relates to data bits stored in memory. Another relates to *control variables*, for example the position of the magnetic head in a Turing machine. In this kind of computer a program is written in a sequence of squares on a linear magnetic tape (Bennett, 1988, Article 7.3), and the magnetic head advances sequentially through the program until it is directed to a subroutine. Then the head must find, and execute, that subroutine, which is located in some region of the tape. In this case, the position of the magnetic head deviates from its sequential path, because it must move to the location of the subroutine, and when the subroutine is completed, it must return to the main program.

If the phase volume of *any* information-bearing degree of freedom is compressed, there must be a corresponding increase in the phase volume of another part of the universe. Such compression occurs whenever two possible predecessor paths connect to one successor path, effecting a compression of phase space volume (Bennett, 2002). This is precisely what happens as the magnetic head of a Turing machine moves from the last step of a subroutine, back to the main program. The head could have come from *either* of the two subroutines and the step that returns it to the main program is logically irreversible. Effectively this step constitutes an erasure of information.

Applying Landauer's principle, this compression of the head position's phase space must be accompanied by an entropy increase 'elsewhere.' In the Szilard model, the only possible 'elsewhere' is the reservoir. When the entropy increase, $k \ln 2$, of the reservoir is taken into account, the *apparent* violation of the second law disappears. What seemed initially to be a second law violation is transformed into an instructive example that illustrates the importance of accounting for *all* information-bearing degrees of freedom. Bennett's main point is that if one does a full accounting along the lines described here, the second law of thermodynamics remains intact.

Earman and Norton leveled other thoughtful criticisms at the LPB approach to the Maxwell's demon puzzle. Interested readers are encouraged to read their articles (Earman and Norton, 1998, 1999).

Gyftopoulos leveled two different criticisms against the LPB solution of the Maxwell's demon puzzle. One criticism (Gyftopoulos, 2002a, 1993) is purely thermodynamic while the other (Gyftopoulos, 2002b) is attributed to quantum mechanics. In the first, Gyftopoulos appeals to his own development of thermodynamics (Gyftopoulos and Beretta, 1991). In particular, he examines an energy *versus* entropy diagram for a container of air, with stable equilibrium states lying on a convex curve with positive first and second derivatives. The slope of this curve is the absolute temperature.

The action of a Maxwell's demon would move the air's thermodynamic state from a point on the latter curve, leftward to a non-equilibrium state with the same energy, but lower entropy. Gyftopoulos assumes the demon does its sorting 'without any contribution on his behalf,' interpreting Maxwell's specification of the demon to mean that it must operate with neither energy expenditure nor anything else (including entropy). Disallowing any entropic effects for the demon itself, he concludes correctly that the process is impossible. Alternatively, he argues that the demon can be considered to undergo a cyclic process, and comes to the same conclusion because at fixed energy, the entropy cannot decrease. Anticipating claims that his argument is circular because he uses the laws of thermodynamics to show that they cannot be violated, Gyftopoulos enunciates six points of rebuttal and an example to ward off such claims.

The second criticism (Gyftopoulos, 2002b) is of a very different nature. He adopts a *unified* theory, in which quantum theoretic and thermodynamic laws do *not* apply to density operators for mixed states. With this choice, he is restricted to use *homogeneous* ensembles for which the ensemble average of an observable whose hermitian operator is A is $\langle A \rangle = Tr[\rho A] = \sum a_j/N$ where a_j is the measured result for ensemble member j, with $j = 1, 2 \ldots M$, and $M \to \infty$.

With these assumptions, it is shown that the average value of any momentum component of a molecule is identically zero. Gyftopoulos argues that there are in fact no swift and slow molecules for the demon to sort and that, in fact, each molecule is 'at a standstill.' Gyftopoulos again anticipates criticism by readers that the vanishing of the expectation value does not imply that *each* ensemble member will have zero momentum, and he provides a rebuttal in advance. To accept this argument, one must be willing to reject the existence of mixed states.

The critiques by Earman and Norton and by Gyftopoulos are well researched, articulately written, and well worth reading. Many thoughtful and thought-provoking issues are raised. A similar statement holds for pointed criticisms addressed by Shenker (1999b, 2000). Interested readers are encouraged to study the articles by these authors to develop an appreciation of the points of contention. Indeed the Maxwell's demon literature shows many foibles, and there are grounds for criticism and skepticism on various issues. Critical assessments can give rise to healthy debate and sometimes, deeper understanding of fundamental issues. To date the criticisms alluded to have not convinced the bulk of the scientific community that the LPB framework should be rejected.

In their book, *The Refrigerator and the Universe*, Goldstein and Goldstein (1993) argue that the distinction made between non-intelligent and intelligent demons is artificial. They cite the spring-loaded trapdoor as an example for which (they claim) measurement occurs when a particle pushes the door open; information storage is related to the energy given to the door; and erasure is effected by the subsequent sharing of this energy with many other particles. However, this so-called information storage is uncontrollable and ephemeral, and therefore fundamentally different from *reliable* information storage using a potential well (Leff, 1995b). Thus Goldstein and Goldstein's conclusion, 'It is ironic that von Smoluchowski fully resolved the paradox of the demon in 1912 and then undermined the force of his reasoning by suggesting that it might not apply to intelligent beings,' is difficult to accept by advocates of the LPB approach.

Biedenharn and Solem (1995) examine a quantum mechanical Szilard engine, namely, a one-dimensional variant of the model examined by Zurek (1984, Article 5.1). They assume observation is adiabatic but, in contrast with Jauch and Báron (1972, Article 3.5) and Popper (1957), they argue that measurement *is* necessary to extract work. The wave function collapses when the particle is observed in one side of the container, and all energy eigenvalues become 4 times their original value. Biedenharn and Solem write that the quantum state does not change (quantum mechanical adiabaticity) and thus the energy and temperature both increase by a factor of 4. They assume the subsequent gas expansion is also adiabatic, so that the complete cycle proceeds with zero entropy change. Distinguishing between *physical* and *information* entropy, they argue that the k in the information entropy $k \ln 2$ cannot be Boltzmann's constant. They write that 'there is simply no way to compare the [gas and information] entropies.' Biedenharn and Solem do not mention, and evidently do not accept, Landauer's principle; therefore they do not account for the changed state of the demon's memory.

1.8.3 Process (Disembodied) Maxwell's Demons

Maxwell's original description of his *demon* as 'a being who can see the individual molecules' suggests to some that it is an entity that gathers information. Certainly the demons considered in connection with the Szilard one-particle engine have been of this type and, indeed, this book focuses on information-gathering demons. In contrast Allahverdyan and Nieuwenhuizen have argued that Maxwell intended his demon to be something other than a miniature physical system. They posit that Maxwell was motivated by theological considerations, and intended his *demon* to be a psyche lacking a body. Such an ethereal entity could operate without input of energy and would not require erasure of information because it has no physical memory. With this interpretation, Maxwell's *original* demon can be a *physical process* rather than a small physical system. The notion of this type of demon goes back at least to Whiting (1885).

A substantial class of such demons has surfaced in the literature. These *process* or *disembodied* demons threaten the second law of thermodynamics, not by gathering information and using it cleverly, but rather by taking advantage of specific conditions, including, but not limited to, special geometries, force fields, or the oddities of quantum physics—such as quantum entanglement.

In recent years, Sheehan and co-workers have proposed a variety of process demon challenges to the second law. These proposals are notable in that they are aimed at ultimate experimental testing. In each case, *naturally-occurring* steady state potential gradients (electrostatic, chemical, gravitational) induce matter flows from which macroscopic work can be generated. For each it appears possible that a cyclic process can convert heat to an equivalent amount of work, with no other effect, in violation of the second law of thermodynamics. In some cases, preliminary experiments have been performed. The experiments are difficult and (with the exception of the model in the next paragraph) require uncommon, sometimes extreme, physical conditions. Primarily because the conditions are so different from ambient conditions, entropy generation by the apparatus has exceeded the entropy decrease that is theoretically possible.

The most recent proposal (Sheehan, *et al* 2002) deals with a 'solid state Maxwell's demon.' The main apparatus is a solid-state electrostatic motor that utilizes the electric field energy in an open-gap p-n junction. The model has undergone detailed numerical studies using a commercial semiconductor device-simulator code, yielding results that agree with those for a soluble one-dimensional analytic model. An experimental test is thought to be feasible within the next 5 years. Other second law violation proposals by Sheehan's group deal with plasmas (Sheehan, 1995, 1996a,b, 2002a; Sheehan and Means, 1998; Čápek and Sheehan, 2002); chemical physics (Sheehan, 1998a, 2001, 2002a); and systems under the influence of gravity (Sheehan, Glick, and Means, 2000; Sheehan, *et al* 2002, Sheehan, 2002a).

Various second law violations, based upon theoretical models, have been reported. An example is a violation of Thomson's formulation of the second law, discovered by Allahverdyan, Balian, and Nieuwenhuizen (2002). They consider a thermodynamic system interacting with a macroscopic reservoir at temperature T, and able to exchange energy with a work source. They find that when the work source is *mesoscopic*, a violation of Thomson's principle is possible. They call this a 'mesoscopicity demon'.

Čápek and co-workers have investigated a number of abstract models of open quantum systems, defined by specific second quantized Hamiltonians. These systems show second law violations, in that potential gradients are self-generated or the principle of detailed balance is violated under isothermal conditions (Čápek, 1997a,b, 1998, 2001, 2002; Čápek and Bok, 1998, 1999, 2001, 2002; Čápek and Tributsch, 1999; Čápek and Mancal, 1999, 2002; Čápek and Frege, 2000, 2002).

Zhang and Zhang (1992) proposed a second law-defying mechanical model with velocity-dependent forces, whose classical phase space volume is *not* invariant under time evolution. Liboff (1997) found behavior inconsistent with the second law for a two dimensional classical mechanical model consisting of three disks that can move in any of three narrow (one-dimensional) channels. This weird geometry causes the three disks to collect in one of the channels, and the motion is *not* reversible.

Gordon (1981) considers a molecular machine, patterned after the Szilard engine, that generates a chemical potential gradient and seems to function as a perpetual motion machine. More recently, Gordon (2002) examines a molecular rotor model that generates a temperature gradient from Brownian motion in an equilibrated, isolated system, in violation of the second law. Goychuk and Hanggi (2000) investigate fluctuation-induced currents in a non-dissipative system with initially localized particles.

The examples above are illustrative and surely not exhaustive. Other examples exist in the bibliography and the general literature. Research on possible violations of the second law of thermodynamics has become a burgeoning field.

We close this section with mention of a related line of theoretical research. Scully and coworkers (2001, 2002a, 2002) have studied a quantum heat engine whose working fluid is a single-mode radiation field, and work is extracted from a single reservoir. However, in this case there is no violation of the second law because entropy is generated as the heat engine operates.

1.8.4 Maxwell's Demon, Efficiency, Power, and Time

Although the bulk of the work on Maxwell's demon has centered on its exorcism, one can ask how effective a demon can be whether or not it can defeat the second law. The following questions have been posed (Leff, 1987a, 1990) in this regard: What rate of energy transfer is attainable by a Maxwell's demon who sorts gas molecules serially, and how much time does it take it to achieve a designated temperature difference, T, across a partition? The assumption of serial processing enables an estimate of minimal effectiveness. By the use of two or more demons operating in parallel, improved performance is possible.

Numerical estimates have been made using the energy-time form of Heisenberg's uncertainty principle and also using classical kinetic theory. For a dilute gas at 300 K, the uncertainty principle implies that power $< 1.5 \times 10^{-6}$ W. If the gas volume is the size of a large room, and $\Delta T = 2$ K, then the demon's processing time $> 10^3$ years. With similar assumptions classical kinetic theory implies much tighter bounds, namely, power $< 10^{-9}$ watt and processing time $> 4 \times 10^6$ years. The latter power level, which is comparable to the average dissipation per neuron in a human brain, illustrates the impotence of a lone Maxwell's demon using serial processing.

Once a temperature difference exists between two portions of a gas, it is possible in principle to run a heat engine using this difference. The available energy and efficiency for delivery of this energy as work has been determined (Leff, 1987a). The maximum efficiency for operating a reversible heat engine between two identical chambers at initial temperatures T_+ and $T_- < T_+$, and equal final temperatures, has the simple form $\eta = 1 - (T_-/T_+)^{1/2}$. As expected, this maximum efficiency is lower than that from a Carnot cycle operating between infinite reservoirs at fixed temperatures T_- and T_+. It is noteworthy that the same efficiency expression arises in other contexts, including the reversible Otto and Joule cycles (Leff, 1987b), other reversible cycles (Landsberg and Leff, 1989), and the irreversible Curzon–Ahlborn cycle (Curzon and Ahlborn, 1975) at maximum power.

1.8.5 Limits of Computation

Landauer's principle places a lower bound on the energy dissipated to the environment for each information bit that is erased. Real computers dissipate much more energy per erased bit. Gershenfeld (1996, Article 7.7) points out that laptop computers consume about 10 W, desktop computers consume $\sim$100 W, and supercomputers consume as much as $\sim$100 kW of electric power intermittently, pushing the limits of energy transfer via heat to avoid excessive temperatures.

As computer speeds increase and chip sizes diminish, it is natural to ask: What are the fundamental *thermodynamic* limits to computation? Physicists are typically interested in the limits of single logic gates, while engineers tend to focus on incremental design improvements. Gershenfeld begins a study of fundamental *thermodynamic* limits, sketching a theory of computation that can handle both *information-bearing* and *thermal* degrees of freedom. In order to approach the physical limits of computation, Gershenfeld observes that a computer should not erase its internal states, its processing speed should be no faster than required, and reliability should be no greater than necessary. He provides some interesting back of the envelope calculations to help make his case.

Lloyd (2000, Article 7.8) investigates *ultimate* physical limits on computation in terms of the speed of light c, Planck's (reduced) constant $\hbar \equiv h/(2\pi)$, the universal gravitational constant G, and Boltzmann's constant k. Lloyd observes that if Moore's law is extrapolated into the future, 'then it will take only 250 years to make up the 40 orders of magnitude in performance between current computers that perform 10^{10} operations per second on 10^{10} bits and our 1-kg ultimate laptop that performs 10^{51} operations per second on 10^{31} bits.'

1.8.6 Physics Outlaw or Physics Teacher?

As we have seen, Maxwell's demon was invented to illustrate the statistical nature of the second law of thermodynamics. It ultimately became viewed as a potential physics outlaw that had to be defeated. Now, more than 130 years later, the widely accepted Landauer–Penrose–Bennett solution has shown that information gathering demons do not threaten the second law, except in the extreme quantum domain, where there are indications that standard thermodynamics—including the second law—seem to break down. On the other hand, process (i.e., disembodied) Maxwell's demons are alive and well, and continue to pose potential threats to the second law. It is impossible to know how this will play out, but it will surely be intriguing to follow.

In the closing paragraph of his Scientific American review article, Ehrenberg (1967) captured the spirit of what has kept Maxwell's demon alive: 'Let us stop here and be grateful to the good old Maxwellian demon, even if he does not assist in providing power for a submarine. Perhaps he did something much more useful in helping us to understand the ways of nature and our ways of looking at it.'

Regardless of whether Maxwell's demon is a physics outlaw, it has surely been a potent physics teacher! Though merely a simple idea, it has been a vehicle for relating measurement and information to thermodynamics, quantum mechanics, and biology. Modern electronic computing seems at first thought to be totally unrelated to Maxwell's demon. Yet important connections exist, with the demon illuminating the binary decision process and the computer amplifying the importance of information erasure. Remarkably, Maxwell's microscopic demon has even played a role in the development of black hole thermodynamics (Bekenstein, 1972, 1980).

Will Maxwell's demon play a role in future progress? Considering its rich history and present research trends, a continuing active life is likely. Whether or not it is proven 'guilty' of being an outlaw, we expect Maxwell's demon to remain a potent teacher for many years to come!