Chapter 2
Review of Maxwell's Demon

Abstract The paradox of Maxwell's demon was proposed in a letter from James C. Maxwell to Peter G. Tait for the first time. In the letter, Maxwell mentioned his gedankenexperiment of "a being whose faculties are so sharpened that he can follow every molecule" [1]. The being may be like a tiny fairy, and may violate the second law of thermodynamics. In 1874, William Thomson, who is also well-known as Lord Kelvin, gave it an impressive but opprobrious name—"demon." Later, Leo Szilard proposed an important model of the demon, which quantitatively connects the thermodynamic work to information [2]. Since then, numerous researchers have been discussed the foundation of the second law of thermodynamics in terms of Maxwell's demon [3–16]. In this chapter, we review the historical arguments and the basic ideas related to the problem of the demon. The modern aspects of the demon [5, 6, 17–25] will be discussed in the following chapters.

2.1 Original Maxwell's Demon

First of all, we consider the original version of the demon proposed by Maxwell (see also Fig. 2.1) [1]. A classical ideal gas is in a box that is adiabatically separated from the environment. In the initial state, the gas is in thermal equilibrium at temperature T. Suppose that a barrier is inserted at the center of the box, and a small door is attached to the barrier. A small being, which is named as a "demon" by Kelvin, is in the front of the door. It has the capability of measuring the velocity of each molecule in the gas, and it opens or closes the door depending on the measurement outcomes. If a molecule whose velocity is higher than the averaged one comes from the left box, then the demon opens the door. If a molecule whose velocity is slower than the average one comes from the right box, then the demon also opens the door. Otherwise the door is closed. By repeating this operation again and again, the gas in the left box gradually becomes cooler than the initial temperature, and the gas in the right box becomes hotter. After all, the demon is able to adiabatically create the temperature

T. Sagawa, *Thermodynamics of Information Processing in Small Systems*,
Springer Theses, DOI: 10.1007/978-4-431-54168-4_2,

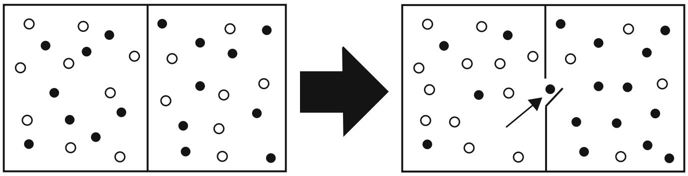

Fig. 2.1 The original gedankenexperiment of Maxwell's demon (reproduced from Ref. [25] with permission). A *white* (*black*) *particle* indicates a molecule whose velocity is slower (faster) than the average. The demon adiabatically realizes a temperature difference by measuring the velocities of molecules and controlling the door based on the measurement outcomes

difference starting from the initial uniform temperature. In other words, the entropy of the gas is more and more decreased by the action of the demon, though the box is adiabatically separated from the outside. This apparent contradiction to the second law has been known as the paradox of Maxwell's demon.

The important point of this gedankenexperiment is that the demon can perform the measurement at the single-molecule level, and can control the door based on the measurement outcomes (i.e., the molecule's velocity is faster or slower than the average), which implies the demon can perform feedback control of the thermal fluctuation.

2.2 Szilard Engine

The first crucial model of Maxwell's demon that quantitatively clarified the role of the information was proposed by Szilard in 1929 [2]. The setup by Szilard seems to be a little different from the Maxwell's one, but the essence—the role of the measurement and feedback—is the same.

Let us consider a classical single molecule gas in an isothermal box that contacts with a single heat bath at temperature T. The Szilard's engine consists of the following five steps (see also Fig. 2.2).

Step 1: *Initial state*. In the initial state, a single molecule is in thermal equilibrium at temperature T.

Step 2: *Insertion of the barrier*. We next insert a barrier at the center of the box, so that we divide the box into two boxes. In this stage, we do not know which box the molecule is in. In the ideal case, we do not need any work for this insertion process.

Step 3: *Measurement*. The demon then measures the position of the molecule, and finds whether the molecule is in "left" or "right." This measurement is assumed to be error-free. The information obtained by the demon is 1 bit, which equals $\ln 2$ nat in the natural logarithm, corresponding to the binary outcome of "left" or "right." The rigorous formulation of the concept of information will be discussed in the next chapter.

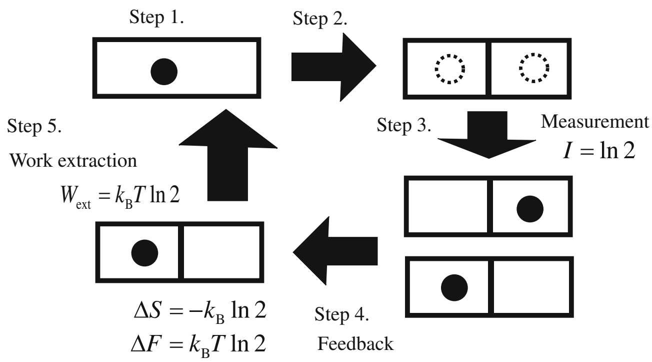

Fig. 2.2 Schematic of the Szilard engine (reproduced from Ref. [25] with permission). *Step 1*: Initial equilibrium state of a single molecule at temperature T. *Step 2*: Insertion of the barrier. *Step 3*: Measurement of the position of the molecule. The demon gets $I = \ln 2$ nat of information. *Step 4*: Feedback control. The demon moves the box to the left only if the measurement outcome is "right." *Step 5*: Work extraction by the isothermal and quasi-static expansion. The state of the engine then returns to the initial one. During this isothermal cycle, we can extract $k_B T \ln 2$ of work

Step 4: *Feedback*. The demon next performs the control depending on the measurement outcome, which is regarded as a feedback control. If the outcome is "left," then the demon does nothing. On the other hand, if the outcome is "right," then the demon quasi-statically moves the right box to the left position. No work is needed for this process, because the motion of the box is quasi-static. After this feedback process, the state of the system is independent of the measurement outcome; the post-feedback state is always "left."

Step 5: *Extraction of the work*. We then expand the left box quasi-statically and isothermally, so that the system returns to the initial state. Since the expansion is quasi-static and isothermal, the equation of states of the single-molecular ideal gas always holds:

$$pV = k_B T, \tag{2.1}$$

where p is the pressure, V is the volume, and k_B is the Boltzmann constant. Therefore, we extract $W_{ext} = k_B T \ln 2$ of work during this expansion, which is followed from

$$W_{ext} = \int_{V_0/2}^{V_0} dV \frac{k_B T}{V} = k_B T \ln 2, \tag{2.2}$$

where V_0 is the initial volume of the box.

During the total process described above, we can extract the positive work of $k_B T \ln 2$ from the isothermal cycle with the assistance of the demon. This apparently contradicts the second law of thermodynamics for isothermal processes known as Kelvin's principle, which states that we cannot extract any positive work from any isothermal cycle in the presence of a single heat bath. In fact, if one could violate Kelvin's principle, one was able to create a perpetual motion of the second kind. Therefore, the fundamental problem is the following:

- Is the Szilard engine a perpetual motion of the second kind?
- If not, what compensates for the excess work of $k_B T \ln 2$?

This is the problem of Maxwell's demon.

The crucial feature of the Szilard engine lies in the fact that the extracted work of $k_B T \ln 2$ is proportional to the obtained information $\ln 2$ with the coefficient of $k_B T$. Therefore, it would be expected that the information plays a key role to resolve the paradox of Maxwell's demon. In fact, from Step 2 to Step 4, the demon decreases $k_B \ln 2$ of physical entropy corresponding to the thermal fluctuation between "left" or "right," by using $\ln 2$ of information. Immediately after the measurement in Step 3, the state of the molecule and the measurement outcome are perfectly correlated, which implies that the demon has the perfect information about the measured state (i.e., "left" or "right"). However, immediately after the feedback in Step 5, the state of the molecule and the measurement outcome is no longer correlated. Therefore, we can conclude that the demon uses the obtained information as a resource to decrease the physical entropy of the system. This is the bare essential of the Szilard engine. On the other hand, the decrease of $k_B \ln 2$ of the entropy means the increase of $k_B T \ln 2$ of the Helmholtz free energy, because $F = E - TS$ holds with F being the free energy, E being the internal energy, and S being the entropy. Therefore, the free energy is increased by $k_B T \ln 2$ during the feedback control by the demon, and the increase in the free energy has been extracted as the work in Step 5. This is how the information has been used in the Szilard engine to extract the positive work.

Szilard pointed out that the increase of the entropy in the memory of the demon compensates for the decrease of the entropy of $k_B \ln 2$ by feedback control. In fact, the memory of the demon, which stores the obtained information of "left" or "right," is itself a physical system, and the fluctuation of the measurement outcome implies an increase in the physical entropy of the memory. In fact, to decrease $k_B \ln 2$ of the physical entropy of the controlled system (i.e., the Szilard engine), at least the same amount of physical entropy must increases elsewhere corresponding to the obtained information, so that the second law of thermodynamics for the total system of the Szilard engine and demon's memory is not violated. This is a crucial observation made by Szilard. However, it was not yet so clear which process actually compensates for the excess work of $k_B T \ln 2$. This problem has been investigated by Brillouin, Landauer, and Bennett.

2.3 Brillouin's Argument

In 1951, Brillouin made an important argument on the problem of Maxwell's demon [7]. He considered that the excess work of $k_B T \ln 2$ is compensated for by the work that is needed for the measurement process by the demon.

He considered that the demon needs to shed a probe light, which is at least a single photon, to the molecule to detect its position. However, if the temperature of the heat bath is T, there must be the background radiation around the molecule. The energy of a photon of the background radiation is about $k_B T$. Therefore, to distinguish the probe photon from the background photons, the energy of the probe photon should be much greater than that of the background photons:

$$\hbar\omega_P \gg k_B T, \tag{2.3}$$

where ω_P is the frequency of the probe photon. Inequality (2.3) may imply

$$W_{\text{meas}} = \hbar\omega_P > k_B T \ln 2, \tag{2.4}$$

which means that the energy cost W_{meas} that is needed for the measurement should be larger than the excess work of $k_B T \ln 2$. Therefore, Brillouin considered that the energy cost for the measurement process compensates for the excess work, so that we cannot extract any positive work from the Szilard engine.

We note that, from the modern point of view, Brillouin's argument depends on a specific model to measure the position of the molecule.

2.4 Landauer's Principle

On the other hand, in his paper published in 1961 [8], Landauer considered the fundamental energy cost that is needed for the erasure of the obtained information from the memory. He propose an important observation, which is known as Landauer's principle today: to erase one bit ($= \ln 2$ nat) of information from the memory in the presence of a single heat bath at temperature T, at least $k_B T \ln 2$ of heat should be dissipated from the memory to the environment.

This statement can be understood as follows. Before the information erasure, the memory stores $\ln 2$ of information, which can be represented by "0" and "1." For example, as shown in Fig. 2.3, if the particle is in the left well, the memory stores the information of "0," while if the particle is in the right well, the memory stores information of "1." This information storage corresponds to $k_B \ln 2$ of entropy of the memory. After the information erasure, the state of the memory is reset to the standard state, say "0," with unit probability as shown in Fig. 2.3. The entropy of the memory then decreases by $k_B \ln 2$ during the information erasure. According to the conventional second law of thermodynamics, the decrease of the entropy in any isothermal process should be accompanied by the heat dissipation to the environment. Therefore, during the erasure process, at least $k_B T \ln 2$ of heat is dissipated from the memory to the heat bath, corresponding to the decrease of the entropy of $k_B \ln 2$. This is the physical origin of Landauer's principle, which is closely related to the second law of thermodynamics.

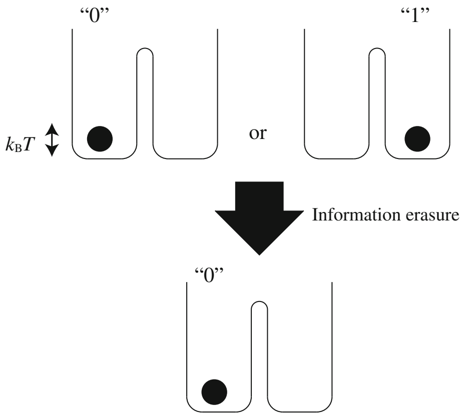

Fig. 2.3 Schematic of information erasure (reproduced from Ref. [25] with permission). Before the erasure, the memory stores information "0" or "1." After the erasure, the memory goes back to the standard state "0" with unit probability

If the internal energies of "0" and "1" are degenerate, we need the same amount of the work as the heat to compensates for the heat dissipation. Therefore, Landauer's principle can be also stated as

$$W_{\text{eras}} \geq k_B T \ln 2, \tag{2.5}$$

where W_{eras} is the work that is needed for the erasure process.

The argument by Landauer seems to be very general and model-independent, because it is a consequence of the second law of thermodynamics. However, the proof of Landauer's principle based on statistical mechanics has been given only for a special type of memories that is represented by the symmetric binary potential described in Fig. 2.3 [12, 15]. We note that Goto and his collaborators argued that there is a counter-example of Landauer's principle [14].

2.5 Bennett's Argument

In 1982, Bennett proposed an explicit example in which we do not need any energy cost to perform a measurement, which implies that there is a counter-example against Brillouin's argument [9]. Moreover, Bennett argued that, based on Landauer's principle (2.5), we always need the energy cost for information erasure from demon's memory, which compensates for the excess work of $k_B T \ln 2$ that is extracted from the Szilard engine by the demon.

His proposal of the resolution of the paradox of Maxwell's demon can be summarized as follows. To make the total system of the Szilard engine and demon's memory a thermodynamic cycle, we need to reset the memory's state which corresponds to information erasure. While we do not necessarily need for the work for the measurement, at least $k_B T \ln 2$ of work is always needed the work for the erasure.

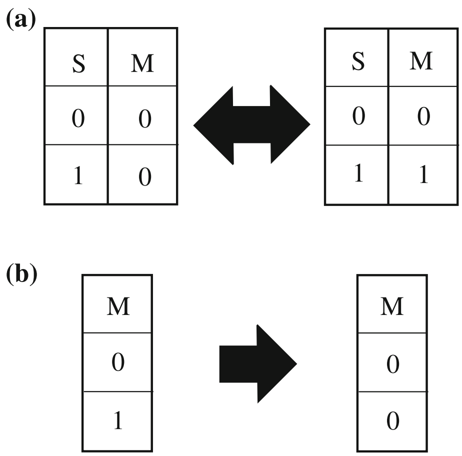

Fig. 2.4 Logical reversibility and irreversibility (reproduced from Ref. [25] with permission). **a** Logically reversible measurement process. **b** Logically irreversible erasure process

Therefore, the information erasure is the key to reconcile the demon with the second law of thermodynamics.

Bennett's argument is also related to the concept of logical reversibility of classical information processing. For example, the classical measurement process is logically reversible, while the erasure process is logically irreversible in classical information theory. To see this, let us consider a classical binary measured system **S** and a binary memory **M**. As shown in Fig. 2.4a, before the measurement, the state of **M** is in the standard state "0" with unit probability, while the state of **S** is in "0" or "1." After the measurement, the state of **M** changes according to the state of **S**, and the states of **M** and **S** are perfectly correlated. In terminology of theory of computation, this process corresponds to the C-NOT gate, where **M** is the target bit. We stress that there is a one-to-one correspondence of the pre-measurement and the post-measurement states of the total system of **M** and **S**, which implies that the measurement process is logically reversible.

On the other hand, in the erasure process, measured system **S** is detached from memory **M**, and the state of **M** returns to the standard state "0" with unit probability, irrespective of the pre-erasure state. Figure 2.4b shows this process. Clearly, there is no one-to-one correspondence between the pre-erasure and the post-erasure states. In other words, the erasure process is not bijective. Therefore, the information erasure is logically irreversible.

In the logically reversible process, we may conclude that the entropy of the total state of **S** and **M** does not change because the process is reversible. This is the main reason why Bennett considered we do not need any energy cost for the measurement process in principle. On the other hand, in the logically irreversible process, the entropy may always decrease, which means that there must be an entropy increase in the environment to be consistent with the second law of thermodynamics. In Landauer's argument, this entropy increase in the environment corresponds to the heat dissipation and the work requirement for the erasure process. Therefore, according to Bennett's argument, we always need the work for the erasure process, not for the measurement process, because of the second law of thermodynamics. This argument seems to be general and fundamental, which has been accepted as the resolution of the paradox of Maxwell's demon. However, we will discuss that the logical irreversibility is in fact irrespective to the heat dissipation, and the work is not necessarily needed for information erasure.

References

1. J.C. Maxwell, *Theory of Heat* (Appleton, London, 1871)
2. L. Szilard, Z. Phys. **53**, 840 (1929)
3. H.S. Leff, A.F. Rex (eds.), *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing* (Princeton University Press, New Jersey, 2003)
4. K. Maruyama, F. Nori, V. Vedral, Rev. Mod. Phys. **81**, 1 (2009)
5. O.J.E. Maroney, Information processing and thermodynamic entropy, in The Stanford Encyclopedia of Philosophy, ed. by E.N. Zalta (Fall 2009 Edition).
6. T. Sagawa, M. Ueda, Information thermodynamics: Maxwell's demon in nonequilibrium dynamics, in Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond, ed. by R. Klages, W. Just, C. Jarzynski (Wiley-VCH, Weinheim, 2012), arXiv:1111.5769 (2011) (To appear).
7. L. Brillouin, J. Appl. Phys. **22**, 334 (1951)
8. R. Landauer, IBM J. Res. Dev. **5**, 183 (1961)
9. C.H. Bennett, Int. J. Theor. Phys. **21**, 905 (1982)
10. W.H. Zurek, Nature **341**, 119 (1989)
11. W.H. Zurek, Phys. Rev. A **40**, 4731 (1989)
12. K. Shizume, Phys. Rev. E **52**, 3495 (1995)
13. R. Landauer, Science **272**, 1914 (1996)
14. H. Matsueda, E. Goto, K-F. Loe, RIMS Kôkyûroku **1013**, 187 (1997)
15. B. Piechocinska, Phys. Rev. A **61**, 062314 (2000)
16. C.H. Bennett, Stud. Hist. Philos. Mod. Phys. **34**, 501 (2003)
17. A.E. Allahverdyan, T.M. Nieuwenhuizen, Phys. Rev. E **64**, 0561171 (2001)
18. C. Horhammer, H. Buttner, J. Stat. Phys. **133**, 1161 (2008)
19. M.M. Barkeshli, arXiv:cond-mat/0504323 (2005).
20. J.D. Norton, Stud. Hist. Philos. Mod. Phys. **36**, 375 (2005)
21. O.J.E. Maroney, Phys. Rev. E **79**, 031105 (2009)
22. S. Turgut, Phys. Rev. E **79**, 041102 (2009)
23. T. Sagawa, M. Ueda, Phys. Rev. Lett. **102**, 250602 (2009)
24. T. Sagawa, M. Ueda, Phys. Rev. Lett. 106, 189901(E) (2011).
25. T. Sagawa, Prog. Theor. Phys. **127**, 1 (2012)

http://www.springer.com/978-4-431-54167-7

Thermodynamics of Information Processing in Small Systems
Sagawa, T.
2013, XVI, 120 p., Hardcover
ISBN: 978-4-431-54167-7