International Journal of Theoretical Physics, Vol. 29, No. 9, 1990

Szilard's Demon Revisited

Jorge Berger[1]

Received February 26, 1990

We show that piston fluctuations are of crucial importance in the analysis of a Szilard engine. Some engines which do not require information in order to perform a Szilard cycle actually do not work. We pinpoint the mechanism and stages which require work investment when a measuring instrument is reset.

1. SZILARD'S DEMON

In a classical paper, Szilard (1929) presented several examples which show that additional information about a system yields a decrease in the entropy of that system. His simplest and most popular example is an engine which will be described in the following. (The present engine is equivalent to Szilard's when piston fluctuations are ignored, but is neater when they are considered.)

Figure 1a represents a cylinder divided by a rigid wall into two cells of maximal volume v each, which is in thermal contact with a heat reservoir

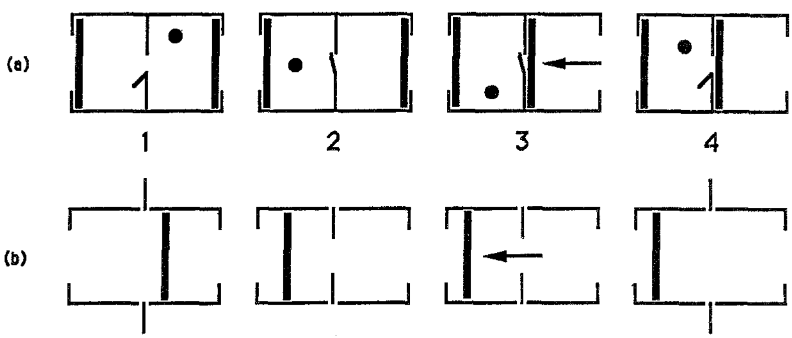

Fig. 1. Two Szilard engines.

[1]Department of Physics and Mathematics, Oranim—School of Education of the Kibbutz Movement, Tivon 36910, Israel.

0020-7748/90/0900-0985$06.00/0

at temperature T. A single molecule is located in the cylinder and a valve may either permit or impede the molecule from crossing the wall. The cylinder is bounded by two movable pistons. At state 1 the valve is open and the molecule can move through the whole volume $2v$. At state 2 the valve is closed and a measurement is performed, so that the engine knows in which cell the molecule is located. At stage $[2\to 3]$ the empty cell is compressed by an infinitesimal applied force. Since there is no internal pressure to oppose compression, this stage requires no work. At stage $[3\to 4]$ the valve is opened and at stage $[4\to 1]$ the internal pressure pushes the piston outward, performing positive work. The work performed in the entire cycle is positive, in apparent contradiction with the second law of thermodynamics.

The mechanism which leads to this entropy decrease has been called "Szilard's demon" and many physicists have claimed to have "exorcised" it. However, different authors use different explanations to make this thought-experiment comply with the accepted result that overall entropy cannot decrease. Jauch and Báron (1972) claimed that information gain should not be identified with entropy decrease. This view was rebutted by Costa de Beauregard and Tribus (1974). Brillouin (1962) and Gabor (1961) claimed that there is an entropy cost for acquiring information. Landauer (1961, 1971, 1988, 1989) and Bennett (1982, 1987) interpret measurement as copying the state of the measured system into the corresponding state of the measuring instrument, which must initially be in a standard state. They devised a procedure for dissipationless measurement and therefore concluded that there is a price for erasing information. Lubkin (1987) and Zurek (1986) discuss quantum aspects of Szilard's problem.

At present, it seems that the view of Landauer and of Bennett has become the orthodox view of the physics community and the conundrum has come to its end. Landauer (1989) writes:

> Szilard reaffirmed belief in the second law, and that the measurement process, in some overall sense, requires energy dissipation. Szilard, however, did not pin down the exact source of the dissipation, within a measurement cycle. We now know, from analysis of the computational process, that resetting of the meter requires energy dissipation, and for Maxwell's demon this is enough to save the second law. Authors following Szilard, however, did not understand that, and looked for dissipation in the step in which information is transferred from the object to be measured to the meter. Brillouin, Gabor, 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.

In spite of this categorical statement, Landauer and Bennett betray their own philosophy when they jump to the conclusion that there is an entropy cost for erasing information: they nowhere provide an explicit argument which shows that mechanical energy has to be invested in erasure, but rather use the premise that, since entropy cannot decrease, this cost has to exist.

In an attempt to clarify things, I discuss some aspects which have not been considered previously. I question whether it is essential to perform measurements in order to operate Szilard's engine. I evaluate exactly the influence of fluctuations and propose an engine which is simpler than Szilard's. For this simple engine it is clearly seen why resetting requires work.

2. DISSIPATIONLESS ERASURE?

The following argument is transcribed from Bennett's (1982) proof that a measurement can be reversible and we could expect it to be as valid here as it was in its original context. Let us consider erasure of a one-bit movable element which is forced to assume one of the two possibilities which are favored by a bistable potential (two minima separated by a barrier considerably higher than kT). A physical example could be a single-domain ferroelectric particle which has a preferred axis due to either geometric or crystalline anisotropy. This particle is represented in Figure 2a and the corresponding bistable potential as a function of the orientation angle is drawn in Figure 2b.

The element is initially at position 1 and its polarization is arbitrary. To perform the erasure operation, this element is slowly brought into a region where a standard field exists. During the first part of the operation (positions 1 to 4), the element is forced to move parallel to its preferred axis (possibly along a narrow tube). During this part of the operation, the element passes through a region where a transverse field is found, which is strong enough to render the orientation potential unistable. This situation is depicted by position 2. From then on, the initial polarization of the

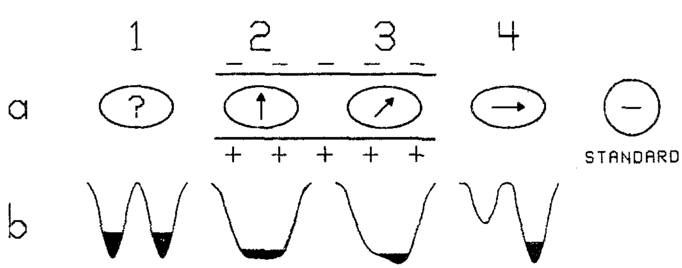

Fig. 2. Erasure of a movable bit. The bit is moved from 1 to 4. (a) In this figure a negative charge is the source of the standard field. The work performed on the bit and its final polarization do not depend on its initial polarization. (b) For the respective places, the dipolar energy and probability density as functions of polarization direction.

element becomes irrelevant. As the movable bit leaves the region of strong transverse field, the orientation potential is biased by the standard field (position 3). The contribution of the standard field to the orientation potential is considerably greater than kT when the bit reaches position 3, though negligible before it reaches position 2. (This can be achieved if the transverse-field region is long enough.) Therefore, when the bit leaves the transverse-field region (position 4), it is almost certain that it has the standard polarization. Finally, the bit is brought back from position 4 to position 1, but this time along a path which avoids the transverse-field region.

Summing up, the polarization of the bit, which was initially arbitrary and possibly unknown, has been set into a controllable standard direction. In order to perform this operation, we have to apply an external force which slightly overcomes the forces exerted by the reservoir and by the electric fields. The interaction with the reservoir is essential in order to stabilize the polarization direction, but we assume that it does not exert a net translational force on the movable element. Since electrostatic forces are conservative and since for quasistatic motion the force acting on the element is, in principle, a predictable function of its position only, the entire operation can be performed by investing an arbitrarily small amount of work.

We shall criticize this conclusion in our last section.

3. A SZILARD CYCLE WITHOUT MEASUREMENT

Following Bennett (1982, 1987) and our previous section, we might be tempted to think that the measuring stage is in fact unessential and try to devise a Szilard engine in which this stage does not exist.

This seems to be feasible since, in order to perform the stage $[2 \to 3]$ in Figure 1a, it appears to be unessential to know in which cell the molecule is located. We could as well apply a slight pressure (as compared to kT/v) at *both* sides of the cylinder. The empty cell would then collapse completely, whereas the cell with the molecule in it would just contract to a new volume $v' \approx v$. The work performed by this engine would be $kT \ln(v'/v)$ in stage $[2 \to 3]$, $kT \ln(2v/v')$ in stage $[4 \to 1]$, and $kT \ln 2$ in the entire cycle, as in the original engine described in section 1.

An additional possibility is provided by Popper's engine (Rothstein, 1964). We immediately see from Figure 3 that the shaft will be driven counterclockwise and could be used to raise a weight, regardless of the side at which the molecule is trapped.

In this way we have gotten rid of measurement, registering, erasure, and all their philosophical burden; the contradiction to the second law is now neat and plain. However, a more careful analysis shows that the conclusions of this section are incorrect.

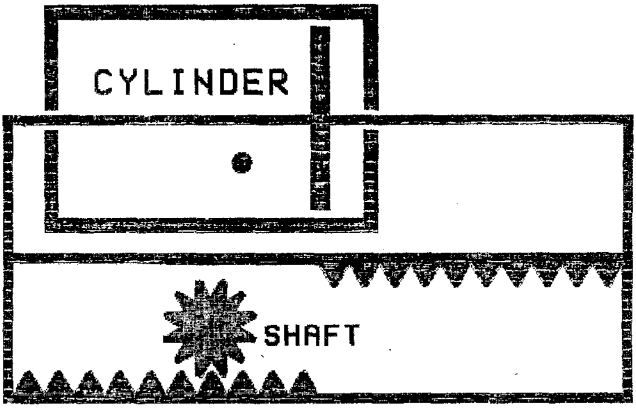

Fig. 3. Popper's engine. The cylinder is fixed in its place and its temperature is constant. The piston, as usual, comprises the entire movable complex. It should also have a "valve" (not shown in the figure) to permit skipping the molecule during the appropriate stage. The shaft can rotate around its fixed axis.

4. THE EFFECT OF FLUCTUATIONS

The descriptions and calculations of Sections 1 and 3 ignore the fluctuations of the pistons, which will now be considered.

The force applied on the pistons (external pressure) can be precisely controlled. For instance, we could assume that the pistons are attached to electric charges which are pulled by infinitesimal charged layers which are brought to some appropriate place. However, the positions of the pistons cannot be precisely controlled. The pistons must be free to move, otherwise they will not be able to recoil and absorb energy from the molecule. Note that in this context the concept "piston" includes all the objects which are rigidly attached to it for the purpose of translating its motion into mechanical work. The piston–molecule collisions are precisely what is meant by thermal contact. Therefore, the piston will be at temperature T and, for any given external pressure, its position will fluctuate accordingly. Once a piston is allowed to move, we lose the information on its position (unless we measure it, and this is not easier than measuring the position of the molecule).

Taking these position fluctuations into account, we shall now calculate the average work performed by the engines in the previous section. When we use the word "work," we mean the work performed against the external pressure, which can be stored as potential energy.

Let us begin with the symmetric-pressure engine of Figure 1a. Consider first the stage $[2 \to 3]$ with the valve closed. Let us denote by v_N the volume of the cell which contains N molecules. (In the present example, $N = 0$

or 1, but generalization is quite simple.) The work performed by the cell (including its piston) when its volume increases by dv_N under applied pressure p is $đw_N = p\,dv_N$. Taking the average over many cycles (at the same stage and pressure), we write $\overline{đw_N} = p\,d\bar{v}_N$. The volume v_N is not a function of p, since it fluctuates; however, assuming that averaging over the appropriate Gibbs ensemble is equivalent to averaging over many cycles, $\bar{v}_N$ may be considered a function of p. Defining the dimensionless parameters $y_N = \bar{v}_N/v$ and $x = pv/kT$, we write

$$\overline{đw_N} = kTx\,dy_N = kTx\frac{dy_N}{dx}\,dx \tag{1}$$

Since the temperature and the pressure are controlled, the probabilities for every possible volume v_N are distributed according to the isobaric-isothermal ensemble. In dimensionless parameters, the probability $\rho_N(y)\,dy$ to have $y \leq v_N/v \leq y+dy$ is given by

$$\rho_N(y) \propto y^N e^{-xy} \tag{2}$$

with the symbol $\propto$ meaning "proportional to." An alternative way to understand the distribution (2) is to regard the pressure as a force field and use the canonical distribution. The kinetic energy is independent of y and the potential energy (of the piston) is $kTxy$. The volume in phase space is proportional to y^N and Boltzmann's factor to e^{-xy}. Since v_N has to be in the range $[0, v]$, it follows from (2) and the definition of y_N that

$$y_N(x) = \int_0^1 y^{N+1} e^{-xy}\,dy \bigg/ \int_0^1 y^N e^{-xy}\,dy \tag{3}$$

It may be interesting to note that the ideal gas law, $pv_N = NkT$, gives the most probable volume according to (2). The average volume is given by (3) and is somewhat different.

Defining

$$z_N(x) = \int_0^1 y^N e^{-xy}\,dy = \frac{N!}{x^{N+1}}\left(1 - e^{-x}\sum_{i=0}^{N}\frac{x^i}{i!}\right) \tag{4}$$

we can express (3) in the form

$$y_N = z_{N+1}/z_N = -d(\ln z_N)/dx \tag{5}$$

The average work $\bar{w}_N(x_1 \to x_2)$ performed by a cell when the pressure changes from kTx_1/v to kTx_2/v is found by integrating (1) by parts and using (5):

$$\bar{w}_N(x_1 \to x_2) = kT\{x_2 y_N(x_2) - x_1 y_N(x_1) + \ln[z_N(x_2)/z_N(x_1)]\} \tag{6}$$

Let us consider now the stage [4→1] with the valve open. Defining $V_1 = v_0 + v_1$ and $Y_1 = \bar{V}_1/v$, we have that the work performed by the engine is now given by

$$\overline{đW_1} = kTx\,dY_1 = kTx\frac{dY_1}{dx}\,dx \tag{7}$$

The probability $P_1(Y)\,dY$ to have $Y \leq V_1/v \leq Y + dY$ is given by

$$P_1(Y) \propto l(Y)\rho_1(Y) \tag{8}$$

where $l(Y)$ is the distance which both pistons are free to move as a whole, without changing the total volume $v_0 + v_1 = vY$. Explicitly,

$$P_1(Y) \propto \begin{cases} Y^2\,e^{-xY} & \text{for} \quad 0 \leq Y \leq 1 \\ Y(2-Y)\,e^{-xY} & \text{for} \quad 1 \leq Y \leq 2 \end{cases} \tag{9}$$

Following the steps which led to (3)-(5), we now have

$$Y_1 = -d(\ln Z_1)/dx \tag{10}$$

with

$$\begin{aligned} Z_N(x) &= \int_0^1 Y^{N+1}\,e^{-xY}\,dY + \int_1^2 (2-Y)\,Y^N\,e^{-xY}\,dY \\ &= 2[z_{N+1}(x) - z_N(x)] - 2^{N+2}[z_{N+1}(2x) - z_N(2x)] \end{aligned} \tag{11}$$

A direct calculation shows that $Z_1(x) = 2z_0(x)z_1(x)$. From here, (5), and (10), it follows that

$$Y_1(x) = y_0(x) + y_1(x) \tag{12}$$

Namely, for any exerted pressure, the average volume of the cylinder does not depend on whether the valve is open or closed. This statement can be made much stronger: it is a trivial exercise to show that, if we do not know at which side the molecule is, then the probability density to have volumes vy_L at the left and vy_R at the right is proportional to

$$(y_L + y_R)\,\exp[-x(y_L + y_R)]$$

both in the case that the valve is open or closed. Finally, it follows from (1), (7), and (12) that the work delivered by the engine in the stage [4→1] is exactly balanced by the work invested in the stage [2→3].

In summary, the work $kT\ln 2$ per cycle, which could be expected to be obtained by naively following Szilard's first example, is precisely canceled by volume fluctuations.

The same kind of analysis may be applied to Popper's engine. Its piston should have a valve for the purpose of making it either transparent or impermeable to the molecule. We denote by vy the volume enclosed between the piston and the middle of the cylinder (y is defined positive, no matter at which side the piston lies; for each value of y there are two possible positions, but a unique potential energy). When the valve is open, the molecule is irrelevant and the probability density for a value y is proportional to $\exp(-xy)$ (x is determined by the torque applied to the shaft). When the valve is closed, the probability for having the value y and the piston at the same side as the molecule is proportional to $(1-y)\exp(-xy)$; the probability for this same y, but with the piston at the opposite side, is $(1+y)\exp(-xy)$. Summing the probabilities of both sides, we obtain that the probability density for y is proportional to $\exp(-xy)$. Again, the average work performed by quasistatically changing the value of x does not depend on whether the valve is open or closed.

5. A SIMPLIFIED SZILARD ENGINE

We want to engage in devising an engine whose average volume does depend on whether the valve is open or closed. But we first note that, when pistons are allowed to fluctuate, there is no essential difference between a piston and a molecule. Consequently, Szilard's engine contains more movable parts than necessary or, in other words, is more complicated than necessary. The natural candidate is the engine described in Figure 1b. It consists of the same cylinder as in Figure 1a, but this time there is no molecule and only one piston. The rigid wall at the middle of the cylinder has been replaced by a set of pins which act as a "valve," permitting or impeding the piston to cross it. At state 1 the valve is open. At state 2 the valve is closed and a measurement is performed, so that the engine knows at which side the piston is located. At stage $[2\to 3]$ a quasistatically increasing force is applied toward the side where the piston is located. At stage $[3\to 4]$ the valve is opened and at stage $[4\to 1]$ the force is quasistatically released.

Defining the volume of the engine as the volume enclosed between the piston and the end of the cylinder toward which the force is applied, the formalism developed in the previous section remains appropriate. While the valve is closed, y_0 and $\bar{w}_0$ are still given by equations (3)-(6). When the valve is open, we denote the average volume of the engine by $v\mathscr{Y}_0$, with $\mathscr{Y}_0 = -d(\ln \mathscr{Z}_0)/dx$ and

$$\mathscr{Z}_0(x) = \int_0^2 e^{-x\mathscr{Y}}\, d\mathscr{Y} = (1+e^{-x})z_0(x) \tag{13}$$

Using (6), we find that the average work performed by the engine in the stages $[2\to 3]$ and $[4\to 1]$ is

$$\bar{w}_0(0\to x) + \bar{\mathscr{W}}_0(x\to 0) = kT\left\{x[y_0(x) - \mathscr{Y}_0(x)] + \ln\frac{z_0(x)\mathscr{Z}_0(0)}{z_0(0)\mathscr{Z}_0(x)}\right\} \tag{14}$$

Since $y_0(x) \neq \mathscr{Y}_0(x)$, additional work $kTx[\mathscr{Y}_0(x) - y_0(x)]$ is performed in the stage $[3 \to 4]$, when the average volume increases without changing the external pressure, due to the opening of the valve. From here, (14), and (13), the average work in the entire cycle is

$$\text{work per cycle (informed)} = kT \ln[2/(1+e^{-x})] \tag{15}$$

For $x > 0$, this work is always *positive*. For $x \gg 1$, it approaches $kT \ln 2$. We see that the same volume fluctuations which canceled the work expected from the previous engines are the source of the work supply in the present case.

What happens if we forgo the measurement and do not know at which side the piston gets trapped? In order to answer this question, we have to calculate the work per cycle when the force is applied in "the wrong" sense (toward the middle of the cylinder). The only necessary change in the analysis is the replacement of $y_0(x)$ by $1 + y_0(x)$. This has no effect on the work described by equation (14), which is sensitive only to the derivative of y_0. However, the work performed in the stage $[3 \to 4]$ will now be $kTx[\mathcal{Y}_0(x) - y_0(x) - 1]$. If no measurement is performed at state 2, then the force will be applied equally often either in the "right" or in the "wrong" sense. Therefore, the average work per cycle will be (after some rearrangement)

$$\text{work per cycle (uncorrelated)} = -kT \ln(\cosh x/2) \tag{16}$$

This work is negative for $x > 0$. A negative work was expected, since in the stage $[3 \to 4]$ a constraint is released while an unbalanced force is present, and this is an irreversible step.

The conclusion of this and the previous section is that information is indeed the key ingredient for the performance of a "demon."

6. DISCUSSION

In the previous section we showed that the work performed by the engine in Figure 1b will be positive if and only if we know at which side the piston is located in state 2. One way to know where the piston lies is to perform a measurement. Another way is to "prepare" the engine properly. The way to do this is hinted by section 2. Let us assume that the piston is charged. Then, we have to take the engine, with its valve open, to a region in which an electric field exists. For an appropriate electric field, the piston will have different potential energies for different sides of the cylinder. If this energy difference is considerably larger than kT, we can know almost with certainty at which side the piston is located. We then close the valve and withdraw the engine from the electric field.

However, bringing the engine into an electric field is equivalent to bringing the sources of the field close to the engine. Therefore, "preparation" of the engine amounts to application of an increasing force with the valve open, followed by release of this force with the valve closed. But this is precisely the almost cyclic process $[1 \to 4 \to 3 \to 2]$ (see Figure 1b). Obviously, the average work which we have to invest in this process is nothing else than the work we expect to obtain from operation of the engine $[2 \to 3 \to 4 \to 1]$.

The erasure process described in Section 2 involves the same features contained in the preparation process we have discussed. It is not true that the force acting on the movable element depends on its position only, since the polarization direction fluctuates. Moving the element through the transverse field is analogous to having the valve open, and skipping that field is analogous to having the valve closed.

Thus, we arrive naturally at the following picture for resetting a measuring instrument from an unknown initial state: first, the instrument is set in interaction with a standard field. The instrument has to be in thermal contact with a heat bath, so as to permit damping of the initial state and assume the desired standard state (which depends on the standard field only). During this stage, we must allow passage between the different possible states of the instrument and the desired state. Second, the instrument–field interaction is withdrawn. At this stage, we must forbid passage to other possible states. Since the second stage has more constraints than the first, fluctuations have a different effect on the average work performed during each stage. As a result, these works do not cancel out; a net work has to be invested in the resetting process.

So far, the second law wins. However, if we accept Bennett's (1982) results that both measuring and uncopying are dissipationless, the demon still has a chance. (Measuring and uncopying differ qualitatively from erasure of an unknown state, since in these cases we know how to build a symmetric process such that the work vanishes separately for the constrained and the unconstrained stage.) All we have to do is to equip the "mind" of a Szilard engine with a couple of fixed bits (one "right" and one "left") for the purpose of uncopying. Once a measurement is performed, its information is used for a compound action: first, the empty cell is compressed, and second, the register is brought toward the appropriate fixed bit and uncopied. In this way we can complete the cycle without paying the price of erasure. It still remains to be asked whether our notions of thermodynamics are really contradicted by the existence of a macroscopic-like process that yields just a microscopic amount of work.

REFERENCES

Bennett, C. H. (1982). *International Journal of Theoretical Physics*, **21**, 905-940.

Bennett, C. H. (1987). *Scientific American*, **257**(11), 88.

Brillouin, L. (1962). *Science and Information Theory*, 2nd ed. Academic, New York.

Costa de Beauregard, O., and Tribus, M. (1974). *Helvetica Physica Acta*, **47**, 238-247.

Gabor, D. (1961). In *Progress in Optics*, Vol. 1, E. Wolf, ed., North-Holland, Amsterdam, pp. 109-153.

Jauch, J. M., and Báron, J. G. (1972). *Helvetica Physica Acta*, **45**, 220-232.

Landauer, R. (1961). *IBM Journal of Research and Development*, **5**, 183-191.

Landauer, R. (1971). *Ferroelectrics*, **2**, 47-55.

Landauer, R. (1988). *Nature*, **335**, 779-784.

Landauer, R. (1989). In *Selected Topics in Signal Processing*, S. Haykin, ed., Prentice Hall, Englewood Cliffs, New Jersey, p. 18.

Lubkin, E. (1987). *International Journal of Theoretical Physics*, **26**, 523-535.

Rothstein, J. (1964). *Bulletin of the American Physical Society II*, **9**, 98.

Szilard, L. (1929). *Zeitschrift für Physik*, **53**, 840-856 [Translated and reprinted in *Quantum Theory and Measurement*, J. A. Wheeler and W. H. Zurek, eds., Princeton University Press, Princeton, New Jersey (1983)].

Zurek, W. H. (1986). In *Frontiers of Nonequilibrium Statistical Physics*, G. T. Moore and M. O. Scully, eds., Plenum Press, New York, pp. 151-161.