Article

On the Precise Link Between Energy and Information

Cameron Witkowski * , Stephen Brown , Kevin Truong

Posted Date: 10 January 2024

doi: 10.20944/preprints202401.0761.v1

Keywords: Maxwell's Demon; Landauer's Principle; Szilard's Engine; Erasure; Cost; Energy; Information; Measurement

Article

On the Precise Link between Energy and Information

Cameron Witkowski *,†, Stephen Brown and Kevin Truong

University of Toronto; prof.brown@utoronto.ca; kevin.truong@utoronto.ca

* Correspondence: cameron.witkowski@mail.utoronto.ca; Tel.: +1-905-809-1696

† Current address: 6 King's College Road, Toronto, ON, Canada, M5S 3H5.

Abstract: We present a modified version of the Szilard Engine, demonstrating that an explicit measurement procedure is entirely unnecessary for its operation. By considering our modified engine, we are able to provide a new interpretation of Landauer's original argument for the cost of erasure. From this view, we demonstrate that a reset operation is strictly impossible in a dynamical system with only conservative forces. Then, we prove that to approach a reset yields an unavoidable instability at the reset point. Finally, we present an original proof of Landauer's principle that is completely independent from the Second Law of Thermodynamics.

Keywords: maxwell's demon; landauer's principle; szilard's engine; erasure; cost; energy; information; measurement

1. Introduction

Since the inception of thermodynamics, a delicate tension between physics and information has been unfolding. On the one hand, it is generally believed that knowledge of a system's evolution will not, by itself, change that evolution. Simultaneously, what an observer can do with a system (ie. extract work or decrease entropy) does depend upon the knowledge they possess. Since the Second Law of thermodynamics, roughly speaking, requires that the thermodynamic entropy of a closed system can only increase, a paradox emerged: can an intelligent being circumvent the laws of thermodynamics?

The first recognition of this paradox was by Maxwell, who described how the entropy of a gas could be decreased by "the intelligence of a very observant and neat-fingered being" [1]. In a thought experiment, Maxwell imagined this being opening and closing a massless shutter between two vessels of gas at equilibrium. With knowledge of the paths and velocities of all the molecules, the intelligent being can selectively let fast-moving molecules pass to one side and slow-moving molecules to the other. As a temperature difference grows between the two vessels, the entropy of the system decreases. This intelligent being became known as Maxwell's Demon.

Since the Second Law of thermodynamics forbids such decreases of entropy in closed systems, there must be a way of accounting for the Demon's information about the system. Such was the thought of Leo Szilard, who in 1929 created an engine that permits easier analysis of the connection between information and thermodynamics [2]. A depiction of Szilard's engine is presented in Figure 1.

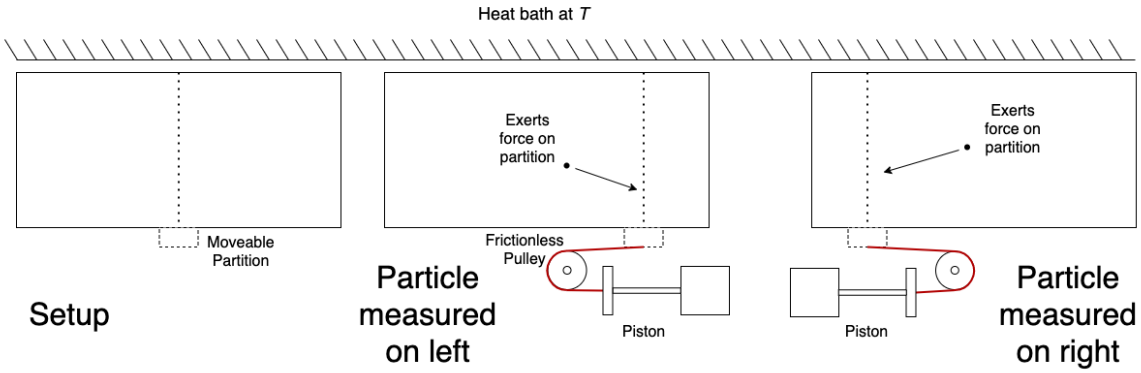

Figure 1. A depiction of the classic Szilard engine.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2024 doi:10.20944/preprints202401.0761.v1

In contrast to the Maxwell's Demon thought experiment, Szilard's engine contains only one particle in a closed vessel kept at temperature T_b. A movable partition is inserted in the centre of the vessel, creating two sub-chambers, which we take here to be equal volumes $V_l = V_r = \frac{1}{2}V_{total}$. The partition also confines the particle to one side of the vessel. Several assumptions are made in the analysis of the Szilard engine:

1. The partition can be inserted or removed from the chamber at a fixed position with zero energy cost.
2. When the partition is removed from the chamber, it can be slid left and right with zero energy cost.
3. The heat bath at temperature T is infinitely large.
4. The practical difficulties (ie. constructing a particular mechanical assembly) of extracting work from a single particle may be ignored.
5. During expansion, the partition can be moved slowly enough to be considered quasi-static, so nonequilibrium and transitory effects may be ignored.
6. The pulleys exert no force in equilibrium other than to redirect the tension of the string.

To justify assumptions 1 and 2, one may note that when the partition is not in contact with the particle, the partition may be moved by conservative forces alone (ie. any kinetic energy transferred to the partition may be recovered when slowing it to a halt). Assumptions 3, 4, and 5 are, strictly speaking, idealizations. Assumption 6 is weaker than assuming that the pulleys are massless and frictionless (typical for dynamics problems), and is hardly a step from their real behavior. Szilard made assumptions 1-5 either implicitly or explicitly, and here we add assumption 6 for our analysis [2].

Following Szilard, we start with the partition at the midpoint of the chamber. If the piston is positioned correctly, then work can be extracted from this engine by a quasi-static isothermal expansion. For a single particle, this work is given in Joules by:

$$W = \int_{V_i}^{V_f} P\, dV \tag{1}$$

$$= \int_{V_i}^{V_f} \frac{NkT}{V}\, dV \tag{2}$$

$$= NkT \ln \frac{V_f}{V_i} \tag{3}$$

$$= (1)kT \ln \frac{V_{total}}{\frac{1}{2}V_{total}} \tag{4}$$

$$= kT \ln 2 \tag{5}$$

where N is the number of particles (in this case 1), k is the Boltzmann constant, and T is the temperature in degrees Kelvin.

In order to position the piston correctly, however, a measurement must be made to determine which side of the partition the particle occupies. Thus, Szilard argued, we must associate $k \ln 2$ units of entropy with the measurement, in order to account for the work we are able to extract as a result. Szilard writes:

> If we do not wish to admit that the Second Law has been violated, we must conclude that the intervention which establishes the coupling between y and x, the measurement of x by y, must be accompanied by a production of entropy [2].

Since these words were put down in 1929, the story has remained much the same. The only major change was made by Landauer, who suggested that the *erasure* of information was specifically what generated heat. In particular, Landauer wrote that the energy cost we must pay when erasing this

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2024 doi:10.20944/preprints202401.0761.v1

measurument equals or surpasses $kT \ln 2$ [3]. Thus, the cost of erasing our measurement ultimately saves the Second Law from the Demon's wiles.

Surprisingly, the question of whether measurement is necessary at all to operate Szilard's engine seems completely absent from the literature. This consideration does not appear to have crossed Szilard's mind, or the minds of any subsequent authors. While we would be delighted to find out we overlooked an analysis somewhere, our search through the literature did not reveal any previous discussion of this question. We present our modified engine to demonstrate one way the engine could work without us measuring.

2. Modified Szilard Engine

In Figure 2, the modified Szilard engine is shown. The only difference between the setups in Figures 1 and 2 is the positioning of the piston, and the use of a second pulley. Importantly, the piston does not have to be moved a different location to extract work from the engine in Figure 2, regardless of the side the particle is on. Thus, since the side the particle is on does not matter to action of the engine, the measurement is superfluous.

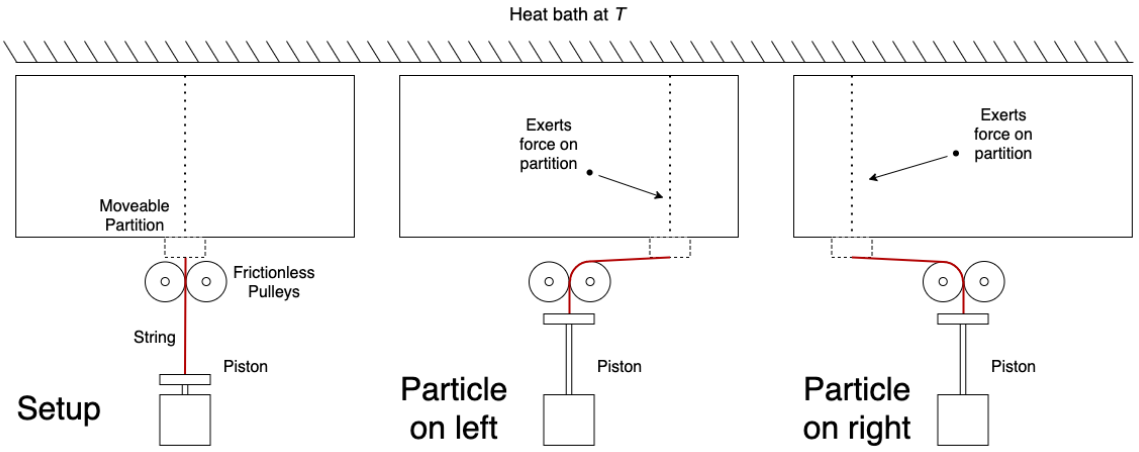

Figure 2. Our modified Szilard engine.

2.1. Work Extraction Protocol

The most likely objection to our modified engine in Figure 2 is that work cannot actually be extracted by it; work can only be extracted in a directed manner. Since the modified engine does not allow for knowledge of which way the partition should move, no sort of directed expansion is possible. Note, however, that the necessity of directing the expansion (thus the necessity of measuring) is exactly what is under question to begin with. We cannot assume *a priori* that this is impossible simply because it is unfamiliar.

To shed some more light on the analysis of work extraction, consider the following common description of quasi-static compression and expansion. Imagine a pile of sand placed on top of a piston against which gas is compressed. By adding a single grain of sand to the pile, the gas compresses slightly and reaches a new equilibrium. Grain-by-grain, the gas can be compressed to any desired amount. Likewise, grains can be removed one-by-one and the pile of sand will rise to find a new equilibrium. Assuming a constant temperature, the work done on the sand during this compression or expansion is given as:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2024 doi:10.20944/preprints202401.0761.v1

$$W = \int_{x_i}^{x_f} F \cdot dx \tag{6}$$
$$= \int_{x_i}^{x_f} -m(x)g \cdot dx \tag{7}$$
$$= \int_{x_i}^{x_f} P(x)A \cdot dx \tag{8}$$
$$= \int_{V_i}^{V_f} \frac{NkT}{V}\, dV \tag{9}$$
$$= NkT \ln \frac{V_f}{V_i} \tag{10}$$

where x is the piston's displacement, F is the force on the gas, and m(x) is the mass of the sand pile as a function of displacement. In equation 8, since the system is in equilibrium, we may use $P(x)A = -m(x)g$. In equation 9, we use the fact that $A \cdot dx$ is a change in volume dV. Unsurprisingly, the final expression in equation 10 is equivalent to equation 3. Thus, as long as we may remove grains of sand one-by-one from a piston, we may extract work in a quasi-static manner.

Can grains of sand be placed on the piston in Figure 2 as easily as they could for Szilard's engine? Upon close inspection, we see nothing that would prevent this. Sure, the gravitational force from a single grain is orders of magnitude greater than the average pressure from a single particle, but the same challenge is faced by Szilard's engine. For both cases, in principle, nothing prevents the design of a piston with enough mechanical advantage that the average force exerted by the particle will reach equilibrium with the gravitational force of a reasonably sized pile of sand. Besides, we made assumption 4 to secure us against such practical challenges. Thus, we conclude that work can be extracted by quasi-static expansion of the engine shown in Figure 2.

To be fully explicit about the cycle we imagine for Figure 2, we specify the following four steps, beginning with the partition at the midpoint of the chamber:

1. 'Grains of sand' are placed on the piston.
2. The partition is inserted into the chamber (with no energy cost, per assumption 1).
3. 'Grains of sand' are removed yielding a quasi-static expansion.
4. The partition is removed from the chamber and brought back to the midpoint (with no energy cost, per assumption 2).

2.2. Considering Information

At this point, it is natural to wonder what happened to the information. It seems to have played no role thus far—and precisely characterizing its role was our motivation from the start. Is it encoded in the engine somehow?

Upon closer inspection, we find that the position of the partition (or equivalently, the position of the string), carries the information about the particle's original position. Let x represent the (horizontal) position of the partition, with the starting position being $x = 0$, and the positive direction being to the right. After one expansion, if the particle started on the left, then we will have $x > 0$, and if the particle started on the right, then we will have $x < 0$. Thus, the sign of x, taking two possible values, can be treated as a bit of memory which stores the measurement of particle's initial side.

The reader may feel some unease with interpreting the partition's position as a 'measurement,' for this is certainly an unfamiliar way of thinking about measurement. However, consider Szilard's description of measurement in his 1929 paper:

> For brevity we shall talk about a "measurement," if we succeed in coupling the value of a parameter y_s (for instance the position co-ordinate of a pointer of a measuring instrument) at one moment with the simultaneous value of a fluctuating parameter x_s of the system, in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2024 doi:10.20944/preprints202401.0761.v1

> such a way that, from the value y_s, we can draw conclusions about the value that x_s had at the moment of the "measurement."[1] [2]

We contend this description accords exactly with the common intuition of what a measurement is: a coupling between one variable and another, such that the one informs an observer of the other. So, by letting $y_s = \text{sign}(x)$, and letting x_s represent the original side of the particle, the value of x_s can be concluded from the value of y_s. Thus, the description justifies the interpretation of the partition's location as representing a measurement.

At face value, this reinterpretation seems to offer little value, as it appears we are in the same position as with Szilard's original engine. Namely, our work extraction protocol generates information which must be accounted for in the analysis. However, we are in fact at a great advantage, since now informational concepts are on the same playing field as the dynamics; we can analyze this information strictly using the tools of physics. In doing so, we will find a better reason for the link between energy and information than simply not wanting to admit that the Second Law has been violated.

3. Landauer's Original Argument

Landauer's principle states that the act of erasing one bit of information necessarily carries an energy cost of $kT \ln 2$. With our modified engine, we are now in a position to fully explain the reason for this cost, pinpoint its source, and demonstrate its generality. But before turning attention to the reset operation (step 4) of our modified engine in Figure 2, it will be most helpful to remind ourselves of Landauer's argument for why erasure is necessarily dissipative. He considers a single particle in a bistable potential well, then asks whether we can reset the particle to the ONE state with a single time-varying force. He writes:

> Since the system is conservative, its whole history can be reversed in time, and we will still have a system satisfying the laws of motion. In the time-reversed system we then have the possibility that for a single initial condition (position in the ONE state, zero velocity) we can end up in at least two places: the ZERO state or the ONE state. This, however, is impossible. The laws of mechanics are completely deterministic and a trajectory is determined by an initial position and velocity. (An initially unstable position can, in a sense, constitute an exception. We can roll away from the unstable point in one of at least two directions. Our initial point ONE is, however, a point of stable equilibrium.) Reverting to the original direction of time development, we see then that it is not possible to invent a single F(t) which causes the particle to arrive at ONE regardless of its initial state [3].

Landauer's first point is that for a conservative system, the history can be reversed in time. A classical mechanical system is conservative if there exists a potential function V such that

$$F(x,t) = -\nabla V(x) \tag{11}$$

where F is the net force vector, x is position, and t is time [4]. In such a system, Newton's equations are time reversal invariant, since the forces depend only on position and not time. Thus, $F(x,v,t) = F(x,-v,-t)$. Recognizing this fact is critical to the rest of the argument.

The dynamics of such a system are described by the second order ordinary differential equation:

$$\ddot{x} = -\frac{\nabla V(x)}{m} \tag{12}$$

---

[1] The s subscripts were added to distinguish Szilard's notation from ours.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2024 doi:10.20944/preprints202401.0761.v1

where m is the mass.[2] With such dynamics in mind, Landauer then states that, in the time reversed system, for a single initial condition we can end up in two places, which is impossible. This fact can be seen as a direct consequence of the Existence and Uniqueness Theorem for Ordinary Differential Equations, also known as Picard-Lindelöf Theorem [5].

Theorem 1 (The Existence and Uniqueness Theorem; Picard-Lindelöf)**.** *Let $R \subseteq \mathbb{R} \times \mathbb{R}^n$ be a closed rectangle with $(t_0, \mathbf{x}_0) \in R$. Let $f : R \to \mathbb{R}^n$ be continuous in t and Lipschitz continous in $\mathbf{x}$. Then, there exists some $\varepsilon > 0$ such that the initial value problem*

$$\dot{\mathbf{x}}(t) = f(t, \mathbf{x}(t)), \quad \mathbf{x}(t_0) = \mathbf{x}_0 \tag{13}$$

has a unique solution, $y(t)$ on the interval $[t_0 - \varepsilon, t_0 + \varepsilon]$.

To apply the theorem to the dynamics in equation 12, we set

$$\mathbf{x} = \begin{bmatrix} x \\ v \end{bmatrix} = \begin{bmatrix} x \\ \dot{x} \end{bmatrix} \tag{14}$$

$$f(t, \mathbf{x}(t)) = \begin{bmatrix} v(t) \\ -\nabla V(x)/m \end{bmatrix} \tag{15}$$

then it follows that, so long as $\nabla V(x)$ is Lipschitz continuous, then a unique solution $\mathbf{x}(t)$ is guaranteed to exist on some interval including t_0. If we set $t = t_0$ at the moment of reset, then the reverse dynamics of the reset operation will yield two nonunique solutions to the same initial value problem. Thus, if we allow reset under conservative dynamics, we violate The Existence and Uniqueness Theorem. This is another crucial fact to recognize for the argument.

Landauer then notes that an unstable equilibrium constitutes an exception in some sense. This point is actually quite nuanced, and we will treat it comprehensively in the following analysis. For now we simply mention that it will play an instrumental role in proving the cost-of-erasure bound, and will constitute the precise location where this cost is paid.

Finally, again considering the possibility of a reset operation, Landauer writes "if, however, we permit the potential well to be lossy, this becomes easy" [3]. Here, lossy may be taken as a synonym for non-conservative. Thus, the seeds of a rigorous argument are laid: a reset operation is not possible under conservative dynamics due to the Existence and Uniqueness Theorem and therefore it must involve nonconservative dynamics resulting in an energy cost.

What remains is to explicitly demonstrate that the cost of erasing one bit has a particular lower bound, namely $kT \ln 2$. Landauer's approach was to include this bit in the thermodynamical state space and conclude that its erasure decreased the system's entropy by $k \ln 2$, thus generating $kT \ln 2$ J of heat. While satisfying to some, the validity and generality of his conclusions remain highly controversial to this day. In the section 5, we will prove this lower bound directly by mechanical and statistical considerations alone, providing what we hope is a satisfying and definitive conclusion to this controversy.

4. Reset Operations with Conservative Forces

We now shift our gaze to step 4 of our modified Szilard's engine cycle: removing the partition from the chamber and returning it to the midpoint. At the end of step 3, the partition can be in one of two places: the right side of the chamber, or the left side. In step 4, we hope to bring the partition back

[2] Equation 12 and the following arguments are written for a one-dimensional system for the sake of simplicity, although extending them to multiple dimensions would be relatively straightforward.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2024 doi:10.20944/preprints202401.0761.v1

to the midpoint regardless of which side it was on. Thus, if we look closely at step 4, we should expect to catch the act of erasure on full display, ready to be subjected to our scrutiny.

4.1. Approaching Reset

In section 3, we stated that a reset operation under conservative dynamics is impossible, and in this section, we will analyze what happens when we get close. Recall that a system is conservative if all forces can be expressed as the gradient of a potential function, and consider the potential function in Figure 3.

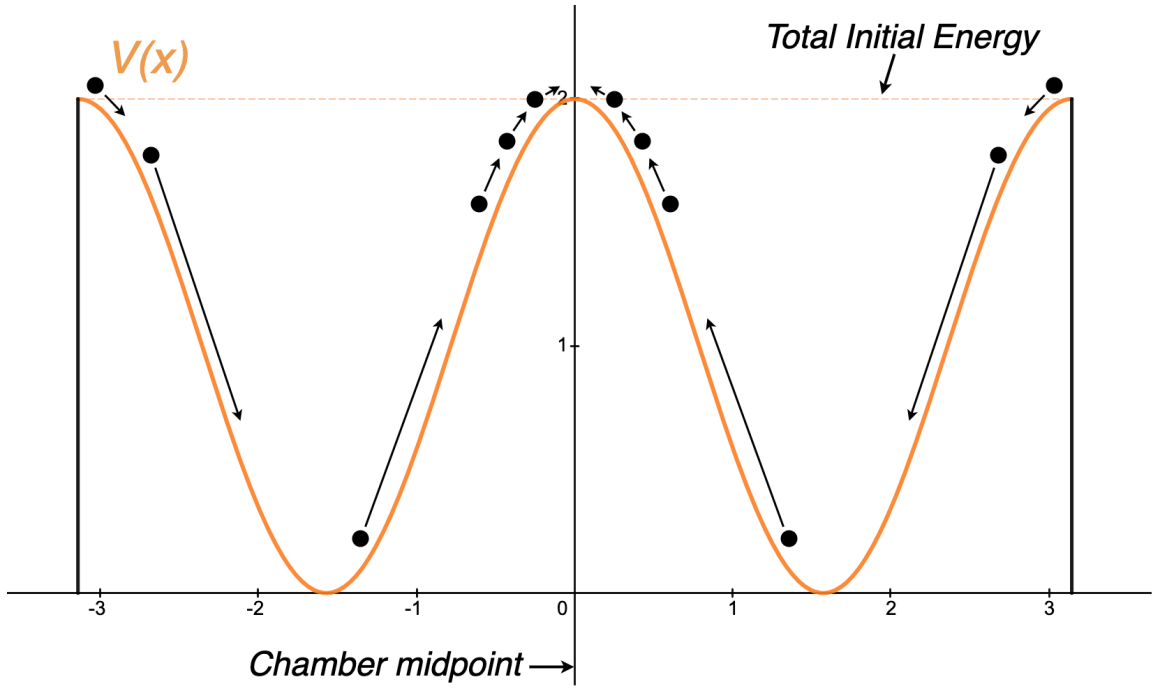

Figure 3. A potential energy function, $V(x)$, one might use to attempt a reset procedure using conservative forces.

The ball represents the partition. The arrows showcase how the partition would be brought back to the midpoint if it started on the left and the right. We find that when the partition comes to rest at $x = 0$, it will be at an unstable equilibrium point. We now see in greater detail why reset in a conservative system is impossible. If the partition starts *exactly* at $x = 0$, then it will stay at $x = 0$ as long as there are no disturbances. If the partition starts anywhere else, it will never come to rest at $x = 0$. This can be seen as another consequence of the time reversal invariance property and the Existence and Uniqueness Theorem, presented in section 3.

The presence of an unstable equilibrium at $x = 0$ is no coincidence, and will play an important role. It turns out that every system approaching a reset operation with conservative forces will result in an unstable equilibrium at the reset point. We present a proof of this fact next.

4.2. General Proof of Instability

First we define a parameter h which measures how close we are to executing a reset. To be precise, consider two trajectories $x_1(t)$ and $x_2(t)$, and some equilibrium point x_e which we will treat as our reset state. We characterize these trajectories as follows:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2024 doi:10.20944/preprints202401.0761.v1

$$||x_1(0) - x_2(0)|| > 0 \tag{16}$$
$$||x_1(\tau) - x_e|| \leq h \tag{17}$$
$$||x_2(\tau) - x_e|| \leq h \tag{18}$$
$$||v_1(\tau)|| \leq h \tag{19}$$
$$||v_2(\tau)|| \leq h \tag{20}$$
$$\nabla V(x_e) = 0 \tag{21}$$

where $\tau > 0$ is some elapsed time. Equation 16 says that the two trajectories start in different places, while equations 17-20 specify how close our trajectories are to being 'merged.' Our goal is to investigate what happens as $h \to 0$. We will prove that, for any conservative system under these conditions, the reset state is an unstable equilibrium.

Definition 1 (Lyapunov Stability). *Consider an autonomous dynamical system given by*

$$\dot{\mathbf{x}} = f(\mathbf{x}(t)), \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{22}$$

where $\mathbf{x}(t) \in \mathcal{D} \subseteq \mathbb{R}^n$ denotes the system state vector, $\mathcal{D}$ is an open set containing the origin, and $f : \mathcal{D} \to \mathbb{R}^n$ is a continuous vector field on $\mathcal{D}$. Suppose f has an equilibrium at x_e such that $f(\mathbf{x}_e) = 0$.

This equilibrium is said to be Lyapunov stable, if, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that, if $\|\mathbf{x}(0) - \mathbf{x}_e\| < \delta$, then for every $t \geq 0$ we have $\|\mathbf{x}(t) - \mathbf{x}_e\| < \varepsilon$ [6].

Definition 2 (Instability). *The equilibrium point $\mathbf{x}_e$ is defined to be unstable if it is not Lyapunov Stable.*

We write out our conservative system from equation 12 as follows:

$$\mathbf{x}(t) = \begin{bmatrix} x(t) \\ v(t) \end{bmatrix} \tag{23}$$
$$f(\mathbf{x}(t)) = \begin{bmatrix} v(t) \\ -\nabla V(x)/m \end{bmatrix} \tag{24}$$
$$\dot{\mathbf{x}}(t) = \begin{bmatrix} \dot{x}(t) \\ \dot{v}(t) \end{bmatrix} = f(\mathbf{x}(t)) \tag{25}$$

where $v = \dot{x}$ is the velocity.

Theorem 2 (Instability of Conservative Reset). *Let $x_1(t)$ and $x_2(t)$ be trajectories of a conservative system and let x_e be a point. If $x_1(t)$, $x_2(t)$, and x_e satisfy equations 16-21, then in the limit as $h \to 0$, x_e is an unstable equilibrium.*

Proof. We must show that it is not the case that for every $\varepsilon > 0$, there exists a $\delta > 0$ such that, if $||\mathbf{x}(0) - \mathbf{x}_e|| < \delta$, then for every $t \geq 0$ we have $||\mathbf{x}(t) - \mathbf{x}_e|| < \varepsilon$. Equivalently, we will show that there exists an $\varepsilon > 0$ such that for every $\delta > 0$, there exists a $t \geq 0$ and $\mathbf{x}(0)$ satisfying $||\mathbf{x}(0) - \mathbf{x}_e|| < \delta$ such that $||\mathbf{x}(t) - \mathbf{x}_e|| \geq \varepsilon$.

Let $\mathbf{x}_1(t) = \begin{bmatrix} x_1(t) \\ v_1(t) \end{bmatrix}$, $\mathbf{x}_2(t) = \begin{bmatrix} x_2(t) \\ v_2(t) \end{bmatrix}$, and $\mathbf{x}_e = \begin{bmatrix} x_e \\ 0 \end{bmatrix}$. We then set

$$\varepsilon = \max\left(||\mathbf{x}_1(0) - \mathbf{x}_e||, ||\mathbf{x}_2(0) - \mathbf{x}_e||\right) \tag{26}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2024 doi:10.20944/preprints202401.0761.v1

We may have that $\mathbf{x}_1(0) = \mathbf{x}_e$ or $\mathbf{x}_2(0) = \mathbf{x}_e$, but these two conditions cannot both be true, as this would violate equation 16. Thus, our selection for ε always yields $\varepsilon > 0$. Consider the reverse dynamics.

Case 1: if $||\mathbf{x}_1(0) - \mathbf{x}_e|| > 0$ then set $\mathbf{x}(0) = \begin{bmatrix} x_1(\tau) \\ -v_1(\tau) \end{bmatrix}$. Then, $\mathbf{x}(\tau) = \mathbf{x}_1(0)$ and $\lim_{h\to 0} ||\mathbf{x}(0) - \mathbf{x}_e|| \leq \lim_{h\to 0} h < \delta$ for all $\delta > 0$. Thus, for every $\delta > 0$ there exists a $t \geq 0$ such that

$$||\mathbf{x}(t) - \mathbf{x}_e|| \geq \max\left(||\mathbf{x}_1(0) - \mathbf{x}_e||, ||\mathbf{x}_2(0) - \mathbf{x}_e||\right) = \varepsilon \tag{27}$$

Case 2: if $||\mathbf{x}_2(0) - \mathbf{x}_e|| > 0$ then set $\mathbf{x}(0) = \begin{bmatrix} x_2(\tau) \\ -v_2(\tau) \end{bmatrix}$. Then, $\mathbf{x}(\tau) = \mathbf{x}_2(0)$ and $\lim_{h\to 0} ||\mathbf{x}(0) - \mathbf{x}_e|| \leq \lim_{h\to 0} h < \delta$ for all $\delta > 0$. Thus, for every $\delta > 0$ there exists a $t \geq 0$ such that

$$||\mathbf{x}(t) - \mathbf{x}_e|| \geq \max\left(||\mathbf{x}_1(0) - \mathbf{x}_e||, ||\mathbf{x}_2(0) - \mathbf{x}_e||\right) = \varepsilon \tag{28}$$

□

Thus, we have demonstrated that any equilibrium point at which two trajectories merge in a conservative classical mechanical system is necessarily unstable.[3] This result can easily be generalized to trajectories that merge (anywhere) away from equilibrium, simply by viewing the trajectories in the proper inertial or non-inertial frame of reference (such that the merge point is an equilibrium in that frame). Moreover, we did not require any assumption that either $\mathbf{x}_1(0) \neq \mathbf{x}_e$ or $\mathbf{x}_2(0) \neq \mathbf{x}_e$. As a result, even though the reset state in Figure 3 is distinct, our proof covers the case of 'reset to ONE' which Landauer originally discussed [3]. To conclude, without any loss of generality, we can view Figure 3 as stereotypical of any scheme to erase information without spending energy.

5. Proof of Landauer's Principle

In section 4.2, we showed that performing a reset operation with only conservative forces is not only impossible, but to even approach it we create an unavoidable instability at the reset point. Fortunately, we can overcome both these difficulties if we are just willing to spend a little energy. To determine how much energy we need to spend, consider Figure 4 below, which we will analyze in detail.

The system in Figure 4 is no longer conservative: we have placed a friction force, labelled 'Brake,' at the $x = 0$ location to dissipate some small quantity of energy and ensure the partition does not spontaneously slide away. Our intention with the brake is to 'trap' the partition at the reset point. The quantity of energy we dissipate is labelled by ϵ.

Our ultimate question is: what is the minimum value of ϵ such that we can reliably perform a reset? At first glance it appears that our brake will have this desired effect for any $\epsilon > 0$. In other words, we can 'trap' the partition at $x = 0$ as long as we dissipate nonzero energy; we imagine that once the partition falls into our trap, it simply will not have the energy to spontaneously jump back out.

This conclusion is compelling, and it would be true if the partition was at absolute zero. If the partition has any significant thermal energy, however, it will constantly be undergoing vibrations. We immediately see that if we make ϵ too small the partition may actually vibrate out of our trap. Fortunately for Landauer's principle, these vibrations place a lower limit on ϵ, meaning it cannot be arbitrarily close to zero. In our system, the chamber is in thermal contact with a heat bath at

[3] Note that, in a nonconservative system, the preceding argument fails, for the time-reversal property played a necessary role in setting $\mathbf{x}(0)$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2024 doi:10.20944/preprints202401.0761.v1

temperature T. Thus, unless we pretend there are other energy sources or sinks, we should find the partition at temperature T also.

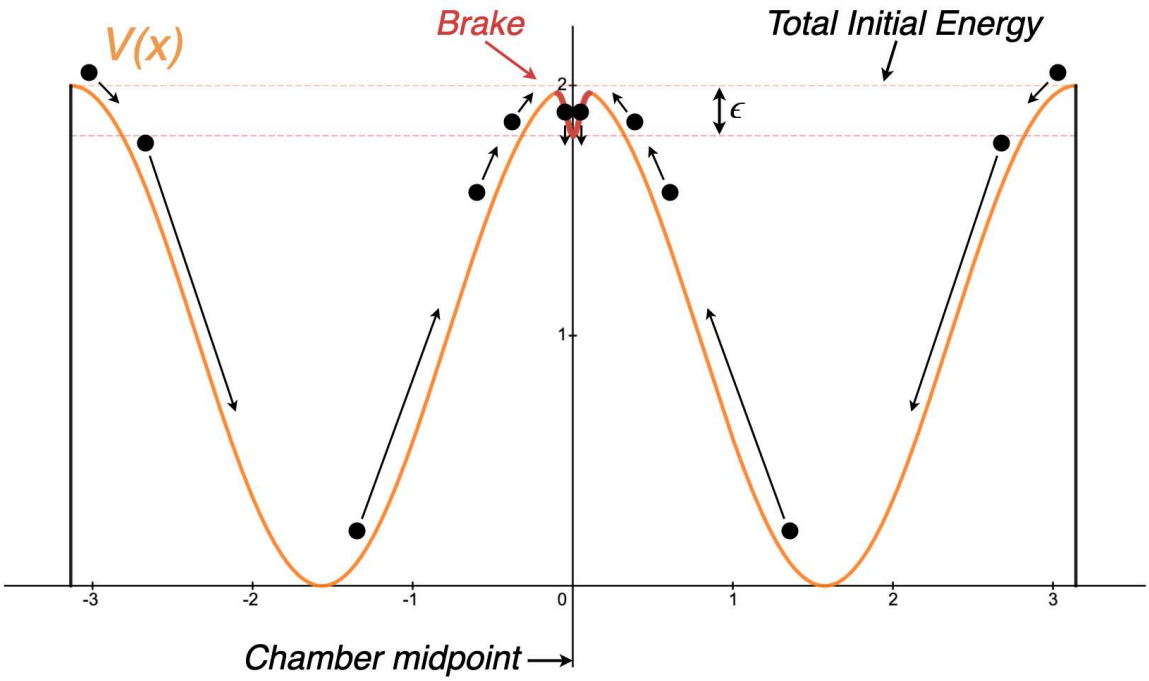

Figure 4. An energy landscape one might implement to perform a reset with minimal energy loss.

When we consider the possibility of the partition vibrating out of our trap in the context of our engine cycle for Figure 2, we face a startling and beautiful realization: the entire engine cycle could work in reverse. In particular, consider the following alternate steps, recalling that the partition starts at the midpoint:

1. The partition jumps away from the midpoint and comes to rest at either the right or left of the chamber, then is inserted into the chamber.
2. 'Grains of sand' are placed on the piston, yielding a quasi-static compression.
3. The partition is removed from the chamber.
4. The grains of sand are removed from the piston.

Thus, we see that for a given value of ϵ, there will be some probability of the forward cycle and some probability of the reverse cycle. Fundamentally, this means that the measurement that got made may instead be un-made, and the work done on the sand (by the gas) may instead be done on the gas (by the sand). Here we are reminded of the ratchet and pawl thought experiment, beautifully analyzed by Feynman [7]. The ratchet and pawl appear more likely to proceed in one direction than another, but are ultimately found to be in equilibrium. We will prove Landauer's principle by a similar approach to the argument Feynman makes.

Let $\mathcal{X}$ denote an autonomous physical system in contact with a heat bath at temperature T. Let x_L, x_R, and x_e be memoryless states of $\mathcal{X}$, representing the ZERO, ONE, and RESET states. Let $x(t)$ represent the system's trajectory through these states over time. Also, let E_L, E_R, and E_e represent the energy of states x_L, x_R, and x_e respectively, with $E_L = E_R$. Finally, define $E_L - E_e = E_R - E_e = \epsilon$ to be the energy cost of reset. We define these terms in full generality, applying to any system, though it may be helpful to imagine x_L corresponding to the partition at the left, x_R to the partition at the right, and x_e to the partition at the midpoint.

Consider some time interval $[t_i, t_f]$. Let

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2024 doi:10.20944/preprints202401.0761.v1

$$P(x(t_f) = x_e \mid x(t_i) = x_L) = P(x(t_f) = x_e \mid x(t_i) = x_R) = p \in (0,1) \tag{29}$$
$$P(x(t_f) = x_L \mid x(t_i) = x_e) = P(x(t_f) = x_R \mid x(t_i) = x_L) = q \in (0,1) \tag{30}$$
$$P(x(t_f) = x_L \mid x(t_i) = x_L) = P(x(t_f) = x_R \mid x(t_i) = x_R) = r \in (0,1) \tag{31}$$

These transition relations are represented graphically in Figure 5.

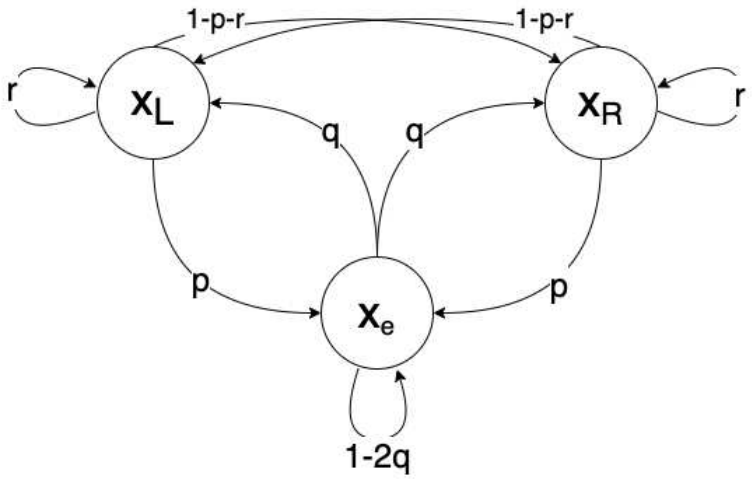

Figure 5. A graphical representation of the transition probabilities described by equations 29-31.

To perform a reset, we should like the probability that the system goes *into* the reset state to be greater than the probability that it *leaves* the reset state. Observe that if the system is in x_L or x_R, the probability that it will move to x_e (performing the reset) is p. On the other hand, if the system is in x_e, the probability that it will move to x_L or x_R (undoing the reset) is $2q$. We say $\mathcal{X}$ implements a reset if the former case is more probable than the latter. Precisely, $\mathcal{X}$ implements a reset if

$$p > 2q \tag{32}$$

When applied to our engine cycle, this constraint would enforce that the forward cycle is more likely than the reverse.

Theorem 3 (Landauer's Principle)**.** *If $\mathcal{X}$ implements a reset then $\epsilon > kT \ln 2$.*

Proof. Since x_L, x_R, and x_e are memoryless states and $\mathcal{X}$ is autonomous, the transition probabilities described by equations 29-31 generate a Markov Chain. Since $p \in (0,1), q \in (0,1)$, and $r \in (0,1)$, it is easily verified that this chain is aperiodic and irreducible, and thus has a stationary distribution. Let $P(x_L)$, $P(x_R)$, and $P(x_e)$ be the probabilities of each state in the stationary distribution, which we can also consider as a statistical ensemble.

For the stationary distribution, we will have:

$$P(x_e)(2q) = P(x_L)(p) + P(x_R)(p) \tag{33}$$
$$P(x_L)(p) + P(x_L)(1-p-r) = P(x_e)(q) + P(x_R)(1-p-r) \tag{34}$$
$$P(x_R)(p) + P(x_R)(1-p-r) = P(x_e)(q) + P(x_L)(1-p-r) \tag{35}$$

Subtracting equation 35 from 34, we obtain

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2024 doi:10.20944/preprints202401.0761.v1

$$(P(x_L) - P(x_R))(1-r) = (P(x_R) - P(x_L))(1-p-r) \tag{36}$$
$$(P(x_L) - P(x_R))(p) = 0 \tag{37}$$
$$P(x_L) = P(x_R) \tag{38}$$

Applying equation 38 to 33, we obtain

$$P(x_e)(2q) = 2P(x_L)(p) \tag{39}$$
$$P(x_e)q = P(x_L)p \tag{40}$$

Now, recalling we must have $p > 2q$ if $\mathcal{X}$ implements a reset, we obtain

$$P(x_e)q > P(x_L)(2q) \tag{41}$$
$$P(x_e) > 2P(x_L) \tag{42}$$

Equation 42 was the key relation we needed from the analysis of the Markov Chain. Now, we will seek to write the stationary probability of states in terms of their energy. First observe that the expected energy of the statistical ensemble is given by:

$$\langle E \rangle = P(x_L)E_L + P(x_R)E_R + P(x_e)E_e \tag{43}$$

If the distribution over states is stationary, the energy of the statistical ensemble will be constant. Then, there can be no net flow of thermal energy between $\mathcal{X}$ and the heat bath. Thus, the stationary distribution is in thermal equilibrium with the heat bath.

Since the stationary distribution is a statistical ensemble in thermal equilibrium with a heat bath, it is exactly the canonical ensemble [8]. The probability distribution over states as a function of energy (measured in Joules) is thus given by:

$$P(x_i) = \frac{e^{-\frac{1}{kT}E_i}}{\sum_j e^{-\frac{1}{kT}E_j}} \tag{44}$$

where k is Boltzmann's constant, and T is the temperature in Kelvin. We then continue from equation 42:

$$\frac{e^{-\frac{1}{kT}E_e}}{\sum_j e^{-\frac{1}{kT}E_j}} > 2\frac{e^{-\frac{1}{kT}E_L}}{\sum_j e^{-\frac{1}{kT}E_j}} \tag{45}$$
$$e^{-\frac{1}{kT}E_e} > 2e^{-\frac{1}{kT}E_L} \tag{46}$$
$$e^{\frac{1}{kT}(E_L - E_e)} > 2 \tag{47}$$
$$e^{\frac{\epsilon}{kT}} > 2 \tag{48}$$
$$\frac{\epsilon}{kT} > \ln 2 \tag{49}$$
$$\epsilon > kT \ln 2 \tag{50}$$

□

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2024 doi:10.20944/preprints202401.0761.v1

6. Discussion

The result in equation 50 is quite general. It is not limited to particles in boxes, but applies to any autonomous system in contact with a heat bath. Naturally, it is trivial to extend the argument for the cost of erasure to any other logically irreversible function or 'merging of computational paths.' Also, for systems of multiple bits, the bound scales exactly as expected. For instance, imagine the engine in Figure 2 were divided into four quadrants, rather than two chambers, thus generating a 'measurement' of two bits rather than one. An isothermal expansion to four times the volume, by the same calculations as equations 1-5, gives $W = kT \ln 4$. The two bits would occupy four states that merge into one, thus equation 32 would become $p > 4q$. With this, it is easy to recompute the bound as $\epsilon > kT \ln 4 = 2kT \ln 2$. By extension, the cost to erase n bits has a lower bound of $nkT \ln 2$.

Interestingly, the case of equality ($\epsilon = kT \ln 2$) corresponds to the reset process having equal likelihood of working forward or backward. In the context of our engine from Figure 2, the forward cycle will be equally as likely as the backward cycle. This result should not be surprising, since a nearly identical consideration is used to demonstrate that the ratchet and pawl cannot produce work at equilibrium [7].

With regard to the heat generated by erasure, we may now observe exactly where it comes from. In the reset scheme of Figure 4 for instance, we see that the mechanical energy of the partition had to be dissipated. In general, the source of heat will depend on the memory device used, but it will be whatever form of energy facilitated the switch to the reset state; this energy must be spent or else the same energy could facilitate a switch back.

We may gain a deeper intuition of this idea by the following analogy, with regard to the reverse dynamics. Imagine balancing on a nearly unstable equilibrium, such as that of Figure 4 with $\epsilon = kT \ln 2$. If we stay perfectly atop, our total energy will not change. In the presence of thermal vibrations, however, eventually a disturbance will push us along one trajectory or another. This 'push' is actually a small quantity of heat that (by starting our motion) gets converted to mechanical energy, in accordance with the conservation of energy. As a result, we can view the entire backward cycle as an isothermal compression used to cool the partition. Each cycle the engine operates in reverse, $kT \ln 2$ work is performed on the particle, and $kT \ln 2$ heat is removed from the partition. In the forward direction then, we see in great detail why the mechanical energy must be converted to heat.

7. Conclusion

In conclusion, we offer a definitive exorcism of Maxwell's Demon by clarifying the necessity of measurement in Szilard's engine and presenting a proof of Landauer's principle. Remarkably, our proof is entirely independent of the Second Law. Nowhere did we require any assumption that the Second Law is true or that it holds for our engine. Instead, we compute the energy cost of erasure directly by mechanical and statistical means alone. Our result instills greater confidence in the Second Law, as it sheds light on independent reasons why perpetual motion machines are impossible even for Maxwell's Demon.

We summarize our conclusions as follows. We showed that an explicit measurement procedure is unnecessary to operate Szilard's engine if we instead interpret the partition's location as bearing information. This reinterpretation shed light on how information can be analyzed strictly using the tools of physics—dynamical systems theory in particular. Using these tools, it follows that a reset operation in a conservative system is strictly impossible due to the Existence and Uniqueness theorem for ordinary differential equations. Worse, to even approach a reset operation produces an unavoidable instability (in the sense of Lyapunov) at the reset point. Practically, thermal vibrations at this instability allow the reset operation to proceed in reverse, which becomes more likely as ϵ decreases. Finally, we showed that when a reset operation is more likely to proceed forward than backwards, we must have $\epsilon > kT \ln 2$. Finally, to the question of whether an intelligent being can circumvent the Second Law by gathering and exploiting information, we answer no.

Author Contributions: Conceptualization, C.W.; formal analysis, C.W.; investigation, C.W.; writing, C.W.; supervision, S.B. and K.T.; All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Sciences and Engineering Research Council of Canada [#RGPIN-2020-07118 to S.B., #RGPIN-2019-04183 to K.T.]; and the Canadian Institutes of Health Research [#PJT-156317 to K.T.].

Acknowledgments: This work has greatly benefited from the insightful and fruitful discussions with Artemy Kolchinsky, whose valuable input was instrumental to gaining an appreciation for the current wisdom in informational physics and analyses of the Szilard engine. Immense gratitude is extended to Saiyam Patel for sanity checking the initial concerns with Szilard's engine, and for employing his discerning mind to thoroughly vet the proofs presented in this paper. Sincere appreciation extends to Professor Frank Kschischang, who graciously served as a soundboard for several preliminary ideas on the subject, offering his wisdom and expertise. Lastly, heartfelt thanks go to Simone Descary, whose unwavering support and encouragement have been a cornerstone of this endeavor.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals

References

1. Knott, C.G. Quote from undated letter from Maxwell to Tait. *Life and Scientific Work of Peter Guthrie Tait. Cambridge University Press* **1911**, p. 215.
2. Szilard, L. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Behavioral Science* **1964**, *9*, 301–310.
3. Landauer, R. Irreversibility and Heat Generation in the Computing Process. *IBM Journal of Research and Development* **1961**, *5*, 183–191. doi:10.1147/rd.53.0183.
4. Arnol'd, V.I. *Mathematical methods of classical mechanics*; Vol. 60, Springer Science & Business Media, 2013; p. 22.
5. Coddington, E.A.; Levinson, N.; Teichmann, T. Theory of ordinary differential equations, 1956.
6. Lyapunov, A.M. The general problem of motion stability. *Annals of Mathematics Studies* **1892**, *17*.
7. Feynman, R.P. *Feynman lectures on physics*; California Institute of Technology, 1967; chapter 46.
8. Gibbs, J.W. *Elementary Principles of Statistical Mechanics*; Charles Scribner's Sons, 1902.