# Szilard's Perpetuum Mobile*

Meir Hemmo and Orly Shenker†‡

In a previous article, we have demonstrated by a general phase space argument that a Maxwellian Demon is compatible with statistical mechanics. In this article, we show how this idea can be put to work in the prevalent model of the Demon, namely, a particle-in-a-box, used, for example, by Szilard and Bennett. In the literature, this model is used in order to show that a Demon is incompatible with statistical mechanics, either classical or quantum. However, we show that a detailed phase space analysis of this model illustrates that a Maxwellian Demon is compatible with statistical mechanics.

**1. Introduction.** In a previous article (Hemmo and Shenker 2011), we have demonstrated by a general phase space argument that a Maxwellian Demon is compatible with statistical mechanics. In this article, we show how this idea can be put to work in the prevalent model of the Demon, namely, a particle-in-a-box, used, for example, by Szilard (1929) and Bennett (1982). In the literature, this model is used in order to show that a Demon is incompatible with mechanics, either classical or quantum (see Leff and Rex 2003).

There are different theories that go under the name of statistical mechanics. In this article, we work in the framework of the so-called Boltzmannian statistical mechanics. For an overview of the different theories, see, for example, Sklar (1993), Callender (1999), Uffink (2007), and Frigg (2008).

The aim of this article is to give a detailed stage-by-stage analysis of Szilard's particle-in-a-box experiment in the context of Boltzmannian classical statistical mechanics. We will show by a detailed phase space analysis

*Received March 2010; revised June 2010.

†To contact the authors, please write to: Meir Hemmo, Philosophy Department, University of Haifa, Haifa 31905, Israel; e-mail: meir@research.haifa.ac.il. Orly Shenker, Program in the History and Philosophy of Science, the Hebrew University, Jerusalem 91905; e-mail: orlyshenker@pluto.mscc.huji.ac.il.

‡This research is supported by the Israel Science Foundation, grant 240/06.

Philosophy of Science, 78 (April 2011) pp. 264–283. 0031-8248/2011/7802-0005$10.00

of this model that a Maxwellian Demon is compatible with statistical mechanics. Our analysis points in a schematic way to some general features of the type of Hamiltonians that can implement the Demonic evolution in Szilard's set up. As far as we are aware, there are no no-go theorems in mechanics that preclude Hamiltonians with these features. Albert's (2000, chap. 5) and our (Hemmo and Shenker 2011) discussions of Maxwell's Demon are in the same spirit.

We do not attempt to write down in this article a detailed Hamiltonian of a demonic set up since the question we wish to consider is whether one can rule out a demonic evolution on the basis of theorems in classical (statistical) mechanics (e.g., Liouville's theorem, conservation of energy). We stress that in statistical mechanics there are no general theorems that are probabilistic counterparts of the second law of thermodynamics, as is well known. Indeed, if a demonic evolution is consistent with statistical mechanics (as we will show), there can be no such general theorems. Of course, theorems assuming very specific conditions, for example, Lanford's theorems (see Uffink 2007), are compatible with our proof of the possibility of a Demon. In the literature (e.g., Leff and Rex 2003), there are numerous discussions of specific Hamiltonians associated with putative Demons. It may be interesting for further research to examine such specific Hamiltonians in the light of our possibility argument, but this is not our aim here.

It is important to distinguish between the question of whether Maxwellian Demons are possible by theorems of statistical mechanics and the question of whether Demons are feasible in practice. J. C. Maxwell thought that his Demon is possible, given the laws of mechanics, but perhaps not feasible. If, as we argue, Maxwellian Demons are possible but not feasible, it becomes an extremely interesting question in physics why that is so. We leave this question to further research.

**2. Macrostates.** Consider the experiment in figure 1. A particle G is placed in a box of volume $V$, which is initially thermally isolated (the set up is adiabatic up to stage f; see below). Particle G can be treated as an ideal gas obeying the equation $pV = kT$, where $p$ is the pressure, $T$ is the temperature, and $k$ is Boltzmann's constant. At stage a, a device D that can measure whether G is on the left- or the right-hand side of the box is prepared in some standard ready state S, while G is free to move around in the entire volume of the box. Figure 1 illustrates the two possible evolutions of the experiment, although only one of the two evolutions is realized at each time we carry out the experiment.[1] In classical mechanics,

1. In fig. 1, we follow the illustration of the experiment given in Bennett (1982), with some significant changes.

https://doi.org/10.1086/659231 Published online by Cambridge University Press

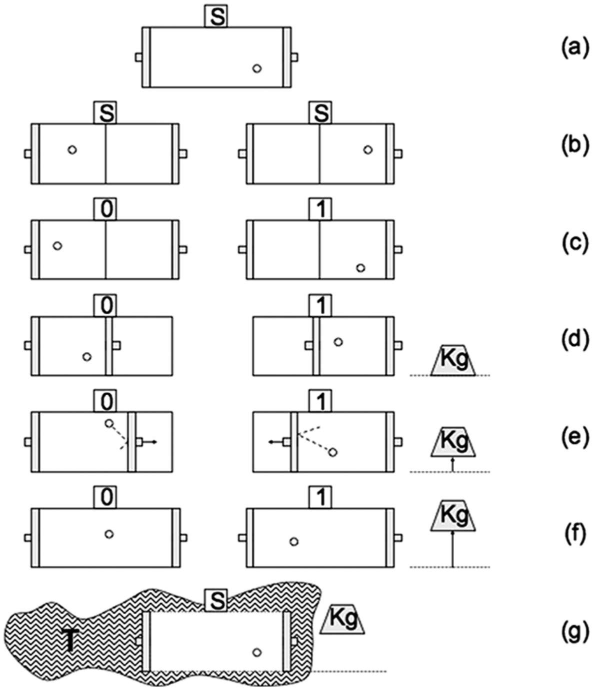


Figure 1. Szilard's particle-in-a-box experiment.

the fundamental assumption is that such an experiment can be completely described by a trajectory, which is a series of microstates determined by the equations of motion.

In statistical mechanics, the phase space description of this stage is as follows (see fig. 2). The system consists of three sets of degrees of freedom, two sets belonging to G and one to D. The horizontal axis in figure 2 stands for the position $x$ of G in the box. We ignore the directions $y$ and $z$, since they are unchanged throughout the experiment. Since in later parts of the experiments we shall be interested in whether G is on the right-hand side of the box or the left-hand side of it, we divide the accessible region on $x_G$ into two regions corresponding to the position of G, and denote these regions by [L] and [R]. In Boltzmannian statistical mechanics, these regions are macrostates. Formally, macrostates are equivalence classes of microstates with respect to some phase space function. Such classes are physically significant when this function corresponds to physical observables, such as the thermodynamic magnitudes. A famous example is the one given by Boltzmann in which the microstates in any

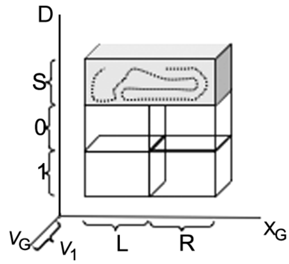


Figure 2. Phase space description of stage a of the experiment.

given macrostate differ only by permutations of particles (see Ehrenfest and Ehrenfest 1912/1990; Uffink 2007). We write macrostates in square brackets and other regions in the phase space in curly brackets.

The axis perpendicular to the page stands for G's velocity $v$, which is determined by the total kinetic energy of G ($E = mv^2/2$). Since the projections of the velocity on its three spatial directions can vary (even when the total kinetic energy $E$ is constant) due to the collisions with the box's walls, for any given $E$ the velocity in each spatial direction ranges over a region between ($-v$, $+v$) for $v = (2E/m)^{1/2}$. Therefore, we represent the velocity macrostate at stage a as the range on the axis perpendicular to the page.

The vertical axis corresponds to D's memory state, which is divided into three macrostates [S], [0], and [1]. We assume for simplicity (as in Bennett 1982) that D's macrostates are of equal Lebesgue measure. As we shall see later, this assumption is natural but not necessary.

At stage a of the experiment depicted in figure 1, the actual microstate of D + G is in the region corresponding to the macrostate [S, L + R, $v_1$], as illustrated in figure 2. As long as the external constraints are not changed (i.e., the volume and the total energy are kept constant), the trajectory of D + G evolves inside this region.

Another useful way to treat stage a is by saying that stage a describes the outcome of a preparation measurement of G in the macrostate [L + R, $v_1$] and similarly for the state [S] of D. More generally, stage a describes a preparation measurement in which D + G is brought into the joint macrostate [S, L + R, $v_1$]. Of course, this preparation means that suitable constraints have been placed on D + G and that other possible macrostates have been ruled out by some observation, which is not described here. We shall come back to the notion of preparation later on in section 6.

**3. Trajectories.** At stage b of the experiment (see fig. 1), a partition is placed (with negligible investment of work) in the middle of the box so

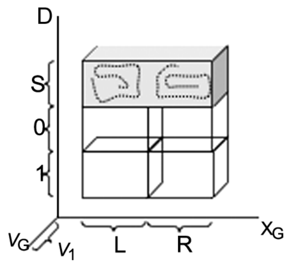


Figure 3. Phase space description of stage b.

that G is trapped on either the left- or the right-hand side of the box. This means that the trajectories of D + G are now confined to either [L] or [R] and no longer pass between the left- and right-hand sides of the box.

It is important to distinguish, at stage b, between the structure of the trajectories of D + G in the phase space and the macrostate of D + G (see fig. 3). In figure 3, this fact is expressed as follows. The macrostate is depicted by the shaded area, within which there are two disconnected parts of the bundle of trajectories represented by the dotted lines. This means that while the trajectories' structure changes in the transformation from a to b, the macrostate remains unchanged. This is because whereas at stage a the dynamical structure of the phase space may be topologically connected, at stage b the region occupied by [L + R] is necessarily topologically disconnected. The classical dynamics does not allow a transformation, which makes connected trajectories disconnected (or vice versa); this is possible only for projections of trajectories on some subset of the degrees of freedom of the system in question. And so the transformation from a to b necessarily involves the intervention of some degrees of freedom not indicated in figure 1, such as the partition itself and the automata that manipulate it. In our description of the experiment, these additional degrees of freedom are treated as external constraints, and they impose limitations on possible evolutions of D + G. By distinguishing between the degrees of freedom of D + G and the degrees of freedom of the external constraints, it is possible to account for notions such as preparation, which will turn out to be important later (see secs. 6 and 12).

Despite the change in the structure of the trajectories, the macrostate of D + G at stage b is still [S, L + R, $v_1$], as in stage a.[2] The reason is

2. One might say that the macrostate of D + G at this stage is either [S, L, $v_1$] or [S, R, $v_1$]. This depends on the way one understands the notion of macrostates. On this view, the entropy of D + G decreases already at this stage rather than at stage c, as we argue below.

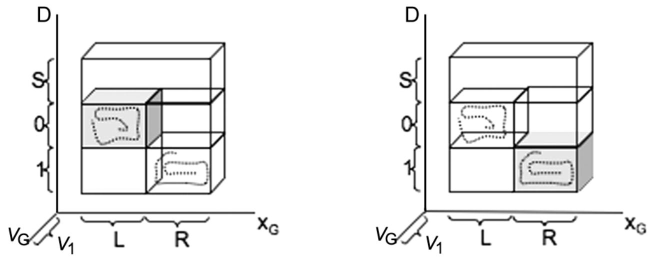


Figure 4. Phase space description of stage c.

that there are no correlations between the macrostates [L] and [R] of G and the macrostate of D (which is still S), and so by looking at the macrostate of D it is impossible to infer the macrostate of G. In figure 3, we present (in dotted lines) trajectories that are permanently trapped in the subregions [S, L, $v_1$] and [S, R, $v_1$] of the phase space, while the macrostate at stage b is exactly the same as in stage a. This is an example of the crucial distinction between macrostates and the dynamics in statistical mechanics. This distinction plays a central role in our analysis below.

**4. Thermodynamics and Kinematics.** The transition from a to b is this. The volume of the box accessible to G is reduced from $V$ to $V/2$. By inserting the partition adiabatically, no energy is invested in G, and therefore G's velocity is unchanged. Consequently, the pressure on the walls of the box is doubled from $p$ to $2p$. This increase in $p$ is brought about solely by the increase in the frequency of the collisions of G with the box's walls due to the reduced volume. If we take G to be an ideal gas, this entails that G's temperature is unchanged.

**5. Measurement in Statistical Mechanics.** The measurement of the location of G by D is described in this experiment by the transition from stage b in figure 3 to stage c in figure 4. The interaction Hamiltonian brings about correlations between the macrostates of D and the macrostates of G, such that trajectories that start out in [S, L, $v_1$] end up in [0, L, $v_1$] and trajectories that start out in [S, R, $v_1$] end up in [1, R, $v_1$] (where [0, L, $v_1$] is the macrostate in which D registers the outcome 0, G is located in the left side of the box, and its velocity is in the velocity range $v_1$, and similarly for [1, R, $v_1$]). The macroscopic evolution of D + G is expressed by the shaded areas in figures 3 and 4. It is crucial to note that the measurement of the location of G by D has a single outcome, that is, either [L] or [R]. To account for this fact, we need two phase space descriptions of the universe, both appearing in figure 4. The phase space on the left side of the figure represents the outcome [L], while the phase space on right side represents the outcome [R]. It would be wrong to superimpose these two descriptions in a single phase space since this would imply, mistakenly, that the outcome of the measurement is both [L] and [R]. At this point, our description of the measurement becomes crucially different from Bennett's (1982) description. To understand this transition, it is essential in statistical mechanics to distinguish between the following two aspects of the evolution:

i) The dynamical behavior of the bundle of trajectories of D + G.
ii) The macroscopic evolution of D + G.

We consider now these two aspects in turn. (i) The first thing to note is that the description of the behavior of the bundle of trajectories in the transition from stage b to c refers to the set of all possible (or counterfactual) trajectories of D + G that start out in the region {S, L + R, $v_1$}, which coincides (due to the preparation in stage a) with the initial macrostate [S, L + R, $v_1$]. We follow the trajectories of D + G that start out in the initial region {S, L + R, $v_1$} and note the regions of the phase space to which they arrive at each time. The set of end points, to which the trajectories bundle arrives at any given time $t$, we call the *dynamical blob* (or blob, for short) at $t$.[3]

At stage b, the dynamical blob coincides with the region {S, L + R, $v_1$}. We then follow the evolution of this blob. One subset of the blob evolves to the region {0, L, $v_1$}, and the other subset evolves to the region {1, R, $v_1$}. In other words, at stage c the blob coincides with the union of the regions {0, L, $v_1$} and {1, R, $v_1$}. In figure 4, this evolution is illustrated by the trajectories represented by the dotted lines: the dotted lines that have been confined to {S, L + R, $v_1$} in figure 3 are now confined to both regions {0, L, $v_1$} and {1, R, $v_1$}. Since the description in terms of the blob applies to the actual trajectory as well as the counterfactual trajectories of D + G, the evolution of the blob is identical in the two possible histories (or possible measurement outcomes). We understand Bennett's description of this set up using a single phase space as applying to the evolution of the blob (rather than the evolution in terms of macrostates). For this reason, the dotted lines representing the entire blob cover both regions {0, L, $v_1$} and {1, R, $v_1$} and are copied in both phase spaces in figure 4.

The evolution of the blob satisfies Liouville's theorem since the Lebesgue measure of the region {S, L + R, $v_1$} is equal to the Lebesgue

3. We use here an idea similar to that of Poincare sections, which are hyperplanes crossed by trajectories and often used to study degree of periodicity.

https://doi.org/10.1086/659231 Published online by Cambridge University Press

measure of the union of the regions {0, L, $v_1$} and {1, R, $v_1$}. This is the case in each of the two parts of figure 4. It is important to note that this conclusion about Liouville's theorem holds regardless of the fact that the regions {0, L, $v_1$} and {1, R, $v_1$} coincide with the corresponding macrostates.

(ii) We now consider the macroscopic evolution of D + G as it is described in terms of the partition to macrostates. Here it is crucial to note the interplay between the evolution of the blob and the change in the macrostate of D + G. The macrostates in stage c of the experiment are denoted in figure 4 by the shaded regions, which stand for the two possible macrostates corresponding to the two possible outcomes of the measurement. In our setting, the phase space is partitioned into macrostates in such a way that the three macrostates of D (i.e., [S], [0], and [1]) are pairwise equal in Lebesgue measure, and so the macrostates at stage c exactly overlap with the blob at this stage. Liouville's theorem applies to the behavior of the blob only. It states that the measure of the blob over time is conserved, but it does not put any constraint on the way in which the blob can spread over the macrostates of D + G, nor is it relevant at all in determining the measure of each macrostate of D + G. The only constraint that follows from Liouville's theorem with respect to the change in the macrostates of D + G during the evolution is that the total Lebesgue measure of (the union of) the macrostates into which the blob evolves cannot be smaller than the Lebesgue measure of the blob. That is all. Thus, we can see that our setting satisfies this constraint since the two macrostates at stage c exactly overlap with the blob.

Note, however, that this exact overlap is not necessary: macrostates are sets of microstates that are observationally indistinguishable, and this indistinguishability is contingent on the physical structure of the observer in question. Figure 5 illustrates a scenario in which the Lebesgue measures of the macrostates [S, L + R, $v_1$], [0, L, $v_1$], and [1, R, $v_1$] are pairwise equal. Consequently, the blob does not fill up the regions occupied by the postmeasurement macrostates of D + G. However, Liouville's theorem is satisfied also in this scenario precisely because the blob at stage c does not fill up the regions occupied by the macrostates [0, L, $v_1$] and [1, R, $v_1$] but is confined to half of each of them. We focus on the partition in figure 4 for simplicity of illustration only. But the argument in this article goes through for the case of figure 5 as well, and in fact for any partition to macrostates, as we shall see below.

As we said above, in our set up of the experiment, illustrated in figure 4, the postmeasurement macrostates of D + G are pairwise equal in Lebesgue measure. At the initial stage of the experiment, the blob, by construction, coincides with the initially prepared macrostate [S, L + R, $v_1$]. In the subsequent stages, the blob evolves in accordance with the

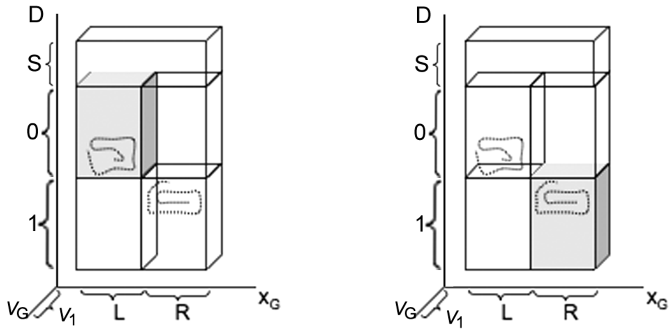


Figure 5. Phase space description of stage c, with alternative partitioning to macrostates.

Hamiltonian. At stage b, the blob still exactly coincides with the macrostate of D + G. Note that in figure 1, which describes the physical set up of the experiment, there are two possibilities at stage b. But the description of these two possibilities is given by figure 3 in a single phase space since the macrostate of D + G at stage b is the same for both possibilities. By contrast, at stage c of the experiment, since the measurement has a single outcome, that is, either the macrostate [0, L, $v_1$] or the macrostate [1, R, $v_1$], a correct description of the phase space depicts only one of these macrostates. In figure 4, we present the two possible outcomes with two different phase spaces in order to avoid the mistaken idea that both possibilities are equally real. In Boltzmannian statistical mechanics, entropy is defined by the logarithm of the Lebesgue measure of the actual macrostate of the system (more on entropy in sec. 7). Therefore, the entropy of D + G after the measurement is given by the Lebesgue measure of either [0, L, $v_1$] or [1, R, $v_1$], depending on the actual outcome. This does not imply, of course, a violation of Liouville's theorem since the measure of the blob is conserved in each of the two possibilities depicted in figure 4.

**6. Preparation.** Here is another way to understand the measurement as described above in the framework of Boltmannian statistical mechanics. On the one hand, the trajectories' bundle that started out at stage a overlaps with both macrostates at stage c. By Liouville's theorem, the volume of the bundle never changes. On the other hand, the measurement is intended to reveal the actual macrostate of G. This means that once we discover whether G is in [L] or [R], in practice we follow thereafter only the corresponding subset of the bundle. The other subset is not cut off but only practically ignored.

The above account of measurement is exactly the way in which one explains the preparation of a system in any given macrostate and in particular the preparation of the initial macrostate at stage a. Unless we take it that one of the two outcomes of the measurement is the newly prepared macrostate, and unless we follow this macrostate only (and ignore the other outcome), we cannot explain the preparation of the initial stage a of the experiment and, in particular, why we follow only the blob that starts out in the macrostate [S, L + R, $v_1$] rather than a blob that starts out in the union of this macrostate and some other macrostates that were possible at the initial time. These considerations lead to the conclusion that at the end of the measurement, at stage c, the macrostate of D + G is not given by the entire blob of the trajectories that started out in the macrostate [S, L + R, $v_1$] but rather by either one of the macrostates [0, L, $v_1$] or [1, R, $v_1$].[4]

**7. Entropy.** The above treatment of measurement and preparation has some crucial implications in statistical mechanics concerning entropy. In Boltzmannian statistical mechanics, the entropy of a system at a moment of time is given by $S = k \ln w$, where $w$ is the Lebesgue measure of the macrostate in which the microstate of the system happens to be at that moment. The entropy of D + G is reduced during measurement in this setting since the volume of each of [0, L, $v_1$], [1, R, $v_1$] is smaller than the volume of [S, L + R, $v_1$], and so the Lebesgue measure of the macrostate along each evolution decreases in the transition from b to c.

Of course, as illustrated in figure 5, whether entropy is reduced in measurement is contingent and depends on the relative Lebesgue measures of the macrostates in question, in our experiment, on the relative volumes of [S], [1], and [0].[5] The fact that entropy is reduced in our setting is absolutely compatible with the laws of classical mechanics. This was pointed out by Albert (2000, chap. 5) and in Hemmo and Shenker (2011). Here we see a concrete example of these general arguments. Figure 5 is a case in which the entropy is conserved during measurement, and if the

4. The idea of retaining all possible evolutions of the system perhaps comes from quantum mechanics. But in the context of classical mechanics, this idea is flawed for the following reasons. First, tracking all possible evolutions immediately leads to a problem in accounting for measurement since there is no way to express in this approach the empirical fact that a measurement has a single outcome. Second, while in quantum mechanics there might be some arguments for retaining the evolution of all possible branches of the universal quantum state based on the possibility of reinterference, such arguments are inapplicable in classical mechanics.

5. In general, the relative size of the macrostates and the behavior of entropy over time are fixed neither a priori (in whatever sense) nor conventionally but rather objectively by the structure of the universe.

https://doi.org/10.1086/659231 Published online by Cambridge University Press

macrostates [0, L, $v_1$], [1, R, $v_1$] were larger in measure, then the entropy would even increase.

If this scenario is a Maxwellian Demon, then it might well be that we are amply surrounded by Demons. Since the second law of thermodynamics seems to be violated here (at first sight; we shall come back to this point later), one may be tempted to postulate that in such a partitioning of the phase space entropy must increase by at least the same amount elsewhere in the universe. However, in the context of the Demon question, such a move would be circular (see Earman and Norton 1998, 1999). Moreover, the foregoing argument is based only on mechanical considerations that, as far as we see, do not support such a postulate. In particular, we emphasize once again that Liouville's theorem is irrelevant in this context since it applies to the evolution of the blob only, which is indeed measure conserving. The decrease of entropy during measurement is a decrease in the measure of macrostates only. This decrease does not mean that the measure of the blob has decreased but only that from that time onward, we follow the corresponding subset of the blob, in accordance with the outcome of the measurement.

**8. Szilard on the Entropy of Measurement.** Our account of the experiment so far is different from Szilard's (1929) account (see also Earman and Norton 1998).[6] Szilard argued that in measurement the entropy of D + G decreases, and therefore there must be an increase of entropy elsewhere in the universe. We have just shown that the total entropy of D + G may decrease in measurement in a way that is consistent with the classical dynamics, in particular with Liouville's theorem. Therefore, unless one presupposes that the entropy of a closed system cannot decrease as stated by the second law of thermodynamics (thus begging the question of Maxwell's Demon), no compensating increase of entropy is required. It seems to us that Szilard was in the right direction in arguing that the entropy of D + G decreases during the measurement (assuming the same partition of the phase space into macrostates). But he failed to realize the full implications of his idea since he did not pursue a purely mechanical analysis to the end.

**9. Work in Thermodynamics.** We now move on to the transition starting at d through e and ending at stage f of the experiment. In our set up, this process is carried out adiabatically. At stage d, we press a piston against a vacuum (in accordance with the outcome of the previous measurement), remove the partition without investing any work, and then, at

6. We do not address arguments here (e.g., Bennett 1982) in the framework of Gibbsian statistical mechanics.

stage e, we release the piston quasi-statically allowing the pressure exerted by G to produce work, until at stage f the particle G is again free to move throughout the volume $V$ of the box. The work produced by G on the piston is stored in some external degree of freedom, say a weight that is lifted outside the box (see fig. 1). This quasi-static expansion of an ideal gas is the paradigmatic case of a reversible (i.e., entropy-conserving) process in thermodynamics. Particle G exerts work on the piston, thereby transferring energy to the weight outside the box. Consequently the free energy of the weight increases, by the amount $mgh$, where $m$ is the mass of the weight, $g$ the gravitational acceleration, and $h$ the height to which the weight is lifted. Conservation of energy entails that the internal energy of G decreases by an amount equal to the increase of the free energy of the weight. (We focus on an adiabatic process here, but a similar argument can be run by considering an isothermal process; see below.)

Now, it follows from the ideal gas law that the increase in the volume accessible to G (from either L or R to L + R) is accompanied by changes in both the pressure and the temperature. The temperature decreases because the temperature of an ideal gas is proportional to its internal energy, which has decreased due to the transfer of energy from G to the weight. The pressure decreases during this transition, both because of the decrease in the average kinetic energy (and thus the velocity) of G and because of the increase in the distance between the box's walls.

Note that since the weight here ends up in a certain fixed height $h$, which is essentially determined by the momentum it gains from G, it can be held in place at its maximal altitude with negligible investment of work. The mechanism places the weight in some standard position at height $h$, which does not depend on whether G is in [L] or [R], and then returns to its ready state. Of course, the entire evolution of the mechanism will depend on the memory macrostate of D. But the set up is such that once the weight is in its final place and the mechanism is back in its ready state, the only traces of whether G was in [L] or in [R] are in D's memory. We will come back to the issue of D's memory in section 12.

**10. Work in Statistical Mechanics.** In terms of phase space, the transition from c to f is as follows (see fig. 6; for simplicity we do not draw the phase space at stages d and e). From stage c to stage f, there is an increase in the Lebesgue measure along the $x_G$ axis from either [L] or [R] to [L + R] and a simultaneous decrease of the measure along the $v_G$ axis from $v_1$ to $v_2 = v_1/2$, in such a way that the total Lebesgue measure of the Poincare section at all times is conserved in accordance with Liouville's theorem. By the end of stage f, the macrostate of D + G is either [0, L + R, $v_2$] or [1, L + R, $v_2$], as represented by the two parts of figure 6. During this transition, the low entropy of D + G is conserved, while the

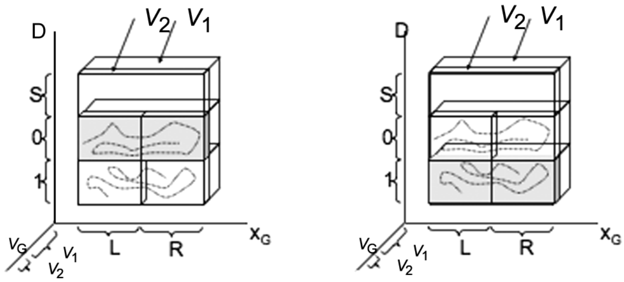


Figure 6. Phase space description of stage f.

weight is reversibly lifted. The free energy of the weight can be used now in order to compress another ideal gas G′ in a reversible and entropy-conserving way.

Here we have (once again) followed the assumption in the literature that the partition of the phase space to macrostates matches exactly the spread of the blob. This assumption is not necessary (see fig. 5), but it makes the discussion simple, and it has the consequence that the total entropy of G (and of course of D + G) is conserved during this evolution from stage c to stage f. If the partition is coarser than the spread of the blob, such that the measure of $v_2$ is greater than the measure of the blob along the $v_G$ axis, then the entropy in $v$ will increase and ipso facto so will the entropy of D + G. However, the partition may also be finer (i.e., $v_2$ may be smaller) than the spread of the blob, in which case the entropy in $v$ will decrease (in this case, the blob will partially overlap with more than one macrostate, and probability considerations will come into play). The crucial point that needs be stressed here is that there is no general argument based on mechanics (or statistical mechanics) to the effect that the entropy in $v$ must increase during the transition from c to f, so as to result in an increase of the total entropy of D + G.

On balance, therefore, in our setting, the total entropy of D + G has decreased in the measurement transition from b to c and then remained unchanged throughout until stage f, and then in the transition from stage f to stage g, as we will show shortly, it increases back to its initial value. The decrease in entropy during the measurement is used in our setting to transform some of the heat energy of G into free mechanical energy in the external weight, which can be used, for instance, to compress another ideal gas G′ in a box. Therefore, what we have shown here is that entropy decreases in measurement on the assumption that the partition matches exactly the spread of the blob.

**11. The Second Law of Thermodynamics.** In the context of Boltzmannian statistical mechanics, the second law of thermodynamics applies to the dynamical evolutions of blobs. In this approach, the second law says roughly what follows. Take an initial set of microstates (e.g., compatible with some macrostate) and follow the trajectories of the system, which start out in this set as they evolve according to the Hamiltonian. In the course of this evolution, the blob overlaps with various macrostates at each moment of time. The statistical counterpart of the second law in this framework then says that most (in the sense of the Lebesgue measure) of these trajectories will arrive into macrostates with an increasingly larger Lebesgue measure until equilibrium is reached. (We discuss the question of completing the cycle of operation in the next section.) Our analysis of measurement above violates this understanding of the second law. It shows that the second law cannot be true for all partitions of the phase space into macrostates. Whether the partition to macrostates in the scenario of figures 2–4 is true about the world is a question of fact. It seems to us that this partition is natural, and if so this statistical mechanical version of the second law of thermodynamics cannot be in general true in statistical mechanics.

For this reason, our conclusion that entropy might decrease in measurement is compatible with our analysis of Maxwell's Demon in Hemmo and Shenker (2011). In this analysis, although we did not stress explicitly the special role of measurements, the decrease of the total entropy of D + G is a straightforward consequence of both the partition to macrostates and the selection of a single final macrostate among several possible ones.

**12. Completing the Cycle.** We saw that we can extract work from heat using the decrease of the entropy of D + G at stage c. In order to decide whether we have here a Maxwellian Demon, what still remains to be examined is whether the cycle of operation can be completed. Let us explain what this means. We shall say that the cycle is completed if the following conditions are satisfied: (i) D + G returns to its initial macrostate, (ii) there are no macroscopic traces of the previous history of the universe (e.g., concerning the outcome of the measurement, the memory of D) in either D + G or the environment Q, and (iii) the free energy of the weight at the final state is the same as at stage f.[7]

We treat the environment Q as a mechanical system that may consist of any number of degrees of freedom. For simplicity, we take Q to consist of a single particle. Note that up to now we did not refer to Q since we carried out an adiabatic process in which Q is irrelevant. However, as we

7. Of course, there are always microscopic traces because of the determinism of classical mechanics.

shall see now, the environment plays a crucial role in completing the cycle. This very idea is prevalent: for example, Szilard (1929) and Bennett (1982) argue that completing the cycle of operation involves dissipation in Q, and therefore Q's final macrostate is a fortiori different from its initial macrostate. For them, not only the macrostate of Q changes, but the entropy of Q increases. We will now show how to complete the cycle such that the above three conditions are satisfied. We will then examine the question of the entropy of Q in the course of the cycle and consider the implications with respect to the question of Maxwell's Demon.

Here is a way to complete the cycle. At stage f, the blob describing D + G's evolution covers the entire region {0 + 1, L + R, $v_2$} (see the dotted lines in fig. 6). We now construct an evolution that maps the points in this blob back to the initial macrostate [S, L + R, $v_1$] in figure 2. Here, it is important to keep in mind that figure 6 describes two possible histories, only one of which is the actual history of the universe, but in both the full blob is depicted. The measure of the blob {0 + 1, L + R, $v_2$} in figure 6 is equal to the measure of the final macrostate [S, L + R, $v_1$] in figure 2: while the measure of the blob along the D degree of freedom decreases by half, from the union of [1] and [0] to [S], the measure of the blob along the $v_G$ degree of freedom is doubled since the measure of $v_1$ is twice the measure of $v_2$. Therefore, Liouville's theorem is satisfied.

Recall that inserting the partition at stage b resulted in the splitting of the phase space into two disconnected regions [S, L] and [S, R]. However, during the expansion stages from c to f, the gradual changes in the constraints translate into a topological change in the trajectories of G, namely, that the region [L + R] at stage f is connected again. A similar topological change occurs in D, so that by the end of stage f, the phase space is connected as indicated in figure 2. Note further that in this transition the entropy of D + G is restored to its initial value; that is, it is doubled. The entropy of G is doubled in this transition, while the entropy of D remains unchanged. The weight remains in its lifted state.

The evolution just described in the transition from f to g requires a source of energy since it involves an increase in the Lebesgue measure of the blob along the $v_G$ degree of freedom corresponding to the increase in G's speed. In other words, in restoring G's initial macrostate, its internal energy increases (by the amount corresponding to the change from $v_2$ to $v_1$). Here, Q comes into play. Consider figure 7. Up to now, in figures 2–6 we depicted three sets of degrees of freedom. We now need two more, namely, the velocity $v_Q$ and position $x_Q$ of Q. We describe them separately in figure 7, but of course the dimensions of figure 7 belong to the same phase space of figures 2–6. This means, in particular, according to Liouville's theorem that the total volumes of the blobs in figures 2–6 and 7 must be conserved; that is, the total volume of the blob describing D +

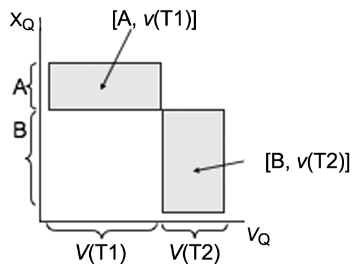


Figure 7. Phase space description of the environment at stage g.

G + Q must be conserved. Since in the transition from stage f to stage g the volume of the blob in the D + G degrees of freedom is conserved, and since the total volume of the blob describing D + G + Q must be conserved, it follows that we need to show now that the volume of the blob describing Q alone must be conserved during the interaction of Q with D + G. Here is a way to achieve this.

Throughout stages a to f, Q is in some fixed macrostate [A, *v*(T1)], where A is the position macrostate of Q and *v*(T1) is its velocity macrostate corresponding to the initial temperature T1 of Q. As a result of the interaction of G with Q, energy flows from Q to G, and Q's blob evolves from the region overlapping with the macrostate [A, *v*(T1)] to the region overlapping with the macrostate [B, *v*(T2)], where T2 is the temperature of Q at the final state. As in the previous stages, for simplicity, we constructed the macrostates such that they overlap exactly with the blob. The final temperature T2 of Q is lower than its initial temperature T1 in exactly the amount of energy transferred to G. Therefore, the volume of the blob in figure 7 along the $v_Q$ degree of freedom decreases. Liouville's theorem dictates that (since we required that the total volume of the blob of D + G remains fixed) the volume along $v_Q$ must increase by the same amount. This increase in the volume of $x_Q$ is expressed by the transition to the region overlapping with macrostate [B, *v*(T2)] in which the spread of the blob over the $x_Q$ degree of freedom is larger than in the macrostate [A, *v*(T1)]. Here we have shown that Liouville's theorem is satisfied by the transition from f to g.

If we think of Q as a single particle, then increasing the spread of Q along $x_Q$ just means that we have less information about the position of Q after the interaction (whereas we have more information about Q's velocity). If Q consists of many particles, the analysis of the spread along $x_Q$ is a simple generalization of this point. We reiterate that the increase in the spread of the blob here only means a decrease of information about where Q is. In classical mechanics, nothing forbids such an increase in $x_Q$, nor are there any mechanical preconditions concerning the rest of the universe that need be satisfied in order for this evolution to take place.[8]

The transition from figure stage f to stage g completes the erasure of any traces of the actual position of G at stage b and any record in D of the outcome of the measurement at the end of stage c. Since it is impossible to reconstruct the earlier macrostate of either D or G from the final macrostate of D + G + Q in stage g, this is a macroscopic erasure. Of course, in classical mechanics there is no microscopic erasure since from any given microstate one can reconstruct the full history of the system. However, for any given partition, one can construct a macroscopic erasure by mapping macrostates or (subsets of the macrostates) into other macrostates such that macroscopic retrodiction would become impossible (another example of macroscopic erasure is proposed in Hemmo and Shenker [2011]). In our setting, given the partition to macrostates as described throughout our experiment, subsets of the macrostate [S, L + R, $v_1$] are indistinguishable. In particular, one cannot distinguish between subsets that originate in the two macrostates [0, L + R, $v_2$] and [1, L + R, $v_2$] in figure 6. Consequently, one cannot retrodict the macrostate in figure 6, given the macrostate of stage g in figure 2. We stress here that a macroscopic erasure is relative to a given partition of the phase space into macrostates, and there is no universal erasure in classical statistical mechanics that would fit in advance all possible partitions.

**13. Maxwell's Demon.** Let's consider the changes in the entropy of D + G + Q during the transition from stage f to g. At the end of the cycle, D + G returns to its initial macrostate, and the weight remains lifted. Environment Q evolved from [A, $v$(T1)] to [B, $v$(T2)]. This set up is a Maxwellian Demon, provided the entropy change of Q is not higher than the equivalence in entropy of the free energy of the weight. The net change in the total volume of the blob of D + G + Q as a result of the evolution from stage f to stage g is zero, in accordance with Liouville's theorem: D + G is back in its initial state of stage a, and so the cycle is completed. Since the weight has a higher free energy than it did at stage a, and since the net effect is that Q has less energy by exactly the same amount, one may say that this experiment illustrates how to transform into work heat energy from a single heat source. We have here a bona fide Maxwellian Demon that is consistent with the laws of statistical mechanics.[9]

8. In particular, this evolution does not require any pressure difference between Q and the rest of the universe at any time.

9. We do not say that the Demon strictly violates Kelvin's formulation of the second law since, as we said, Q is not a heat bath.

Note that we assumed above that the interaction of Q with G resulted in a transfer of energy from Q to G rather than the other way around. In mechanics, of course, these two evolutions are velocity reversals of one another. And if one is possible given a certain Hamiltonian, so is the other. Which evolution takes place depends on the initial conditions. Throughout the experiment, we have assumed that the microscopic initial conditions in the initial macrostate of D + G + Q at stage a are such that the velocities induce a process from a to g as we described. If we reverse the velocities of D + G + Q, then the process would go in the reversed order of macrostates from g to a, contrary to our assumption. In this case, the entropy of the universe (including the weight) would increase rather than decrease. The question of whether Maxwell's Demon in this case is possible boils down to whether D + G + Q can be prepared in the initial macrostate that would contain microconditions with velocities in one direction only. If this cannot be done, the efficiency of the Demonic evolution would decrease (see Hemmo and Shenker 2011), but any degree of efficiency greater than the standard predictions of statistical mechanics would be a Demon. We are not aware of any no-go theorems that preclude such preparations that lead to a Demonic evolution given the right Hamiltonian.

In general, the question as to the entropy of Q during this transition depends on the partition of the phase space of Q into macrostates. Our choice of macrostates in figure 7 is natural but once again not necessary (see sec. 14): the entropy of Q may increase or decrease depending on the partition to macrostates. The macrostate of Q could be larger (a case that would be similar to fig. 5) or smaller, by, for example, dividing the region overlapping with [B, T2] into two macrostates. In this latter case, the spread in $x_Q$ will be exactly the same as in [A, T1], despite the decrease in the spread along $x_\nu$. The crucial point is that such construction does not seem to violate any known law of mechanics; in particular, it obeys Liouville's theorem.[10] Since the macroscopic erasure is relative to a given set of macrostates, the change of entropy during the erasure varies according to the measure of the relevant macrostates. In the case brought here, the transition from f to g is entropy conserving.

**14. Repeating the Cycle.** As we saw already, the first cycle of our experiment is a Maxwellian Demon. The question now is whether we can repeat the cycle in order to transform more of the heat energy of Q into work. Since at the end of the first cycle the possible position of Q is more

10. Note that here the environment Q is not a heat bath in the thermodynamic sense. It is a consequence of energy conservation that in our set up above Q's temperature varies.

https://doi.org/10.1086/659231 Published online by Cambridge University Press

dispersed, the operation of the second cycle may be less efficient. However, as we showed above, despite this possible decrease in efficiency, the entropy of Q is conserved.

On repetitions of the experiment, the outcome of the measurement at stage c is unpredictable. However, the decrease of entropy and the extraction of work are completely predictable, independently of whether the microstate of G happens to be in region [L] or in region [R] after the insertion of the partition at stage b. The fact that the expansion from stage c to stage f requires different operations depending on whether the particle is found in [L] or [R] at stage c has no bearing on the question of Maxwell's Demon.

**15. Does the Demon Depend on the Partition to Macrostates?** In the Boltzmannian approach, entropy depends on the Lebesgue measure of the macrostate of the system. Therefore, a system in a given microstate may be assigned different values of entropy depending on the partition to macrostates. In particular, in our experiment the entropy changes in, say, figures 2–4 are different from the entropy changes in figure 5 (in the latter, the entropy of D + G in the measurement is conserved). However, the crucial point in our argument is that, since the cycle is completed, the particular partition in the D + G degrees of freedom makes no difference to the question of the total entropy balance by the end of the cycle. However, the final total entropy balance depends on the entropy change in Q, which does depend on the partition to macrostates in Q's degrees of freedom. The macrostates of Q in figure 7 are contingent on the specific environment, but they are consistent with the principles of classical mechanics. We do not know of any no-go theorem in mechanics that rules out such macrostates of Q. Indeed the core of our argument is precisely that a Demon is consistent with classical mechanics.

In order to show that a Demon is impossible and save the second law, one needs to show that the partition to the macrostates of Q is necessarily (by the principles of mechanics) such that the entropy of Q increases by more than the work gained in the weight. Until then, we join Lebowitz in his conclusion: "I do not know if anyone is making bets on the eventual resolution of the apparent paradoxes relating to the co-existence, in the description of the same phenomena, of both determinism and randomness, reversibility and time a-symmetry, and so on. If there are people betting, however, I would be very happy to be the banker and keep the money until everyone has agreed on the matter" (quoted in Sklar 1993, 420).

**16. Conclusion.** In our previous article (Hemmo and Shenker 2011) on this topic, we gave a general argument to the effect that a Maxwellian Demon is compatible with Boltzmannian statistical mechanics. We further showed that the Demon's operation cycle can be completed. In the current article, we illustrated this argument by focusing on Szilard's particle-in-a-box thought experiment. The analysis here uses purely Boltzmannian statistical mechanical concepts. In particular, we did not appeal to ideas and concepts from thermodynamics. Indeed, we believe that this was Maxwell's original intention in putting forward his Demon. Like Maxwell's, our experiment is ideal, and we do not argue that an actual construction of a Demon is feasible. The reasons, however, are pragmatic and concern the complexity of real physical systems and the difficulty to control multiple degrees of freedom.

## REFERENCES

Albert, David. 2000. *Time and Chance*. Cambridge, MA: Harvard University Press.

Bennett, Charles. 1982. "The Thermodynamics of Computation: A Review." In Leff and Rex 2003, 283–318.

Callender, Craig. 1999. "Reducing Thermodynamics to Statistical Mechanics: The Case of Entropy." *Journal of Philosophy* 96 (7): 348–73.

Earman, John, and John D. Norton. 1998. "Exocist XIV: The Wrath of Maxwell's Demon." Pt. 1, "From Maxwell to Szilard." *Studies in History and Philosophy of Modern Physics* 29:435–71.

———. 1999. "Exocist XIV: The Wrath of Maxwell's Demon." Pt. 2, "From Szilard to Landauer and Beyond." *Studies in History and Philosophy of Modern Physics* 30:1–40.

Ehrenfest, Paul, and Tatiana Ehrenfest. 1912/1990. *The Conceptual Foundations of the Statistical Approach in Mechanics*. New York: Dover.

Frigg, Roman. 2008. "A Field Guide to Recent Work on the Foundations of Statistical Mechanics." In *The Ashgate Companion to Contemporary Philosophy of Physics*, ed. Dean Rickles, 99–106. London: Ashgate.

Hemmo, Meir, and Orly Shenker. 2011. "Maxwell's Demon." *Journal of Philosophy*, forthcoming.

Leff, Harvey, and Andrew Rex. 2003. *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*. London: Institute of Physics.

Sklar, Lawrence. 1993. *Physics and Chance*. Cambridge: Cambridge University Press.

Szilard, Leo. 1929. "On the Decrease of Entropy of a Thermodynamic System by the Intervention of an Intelligent Being." In Leff and Rex 2003, 110–19.

Uffink, Jos. 2007. "Compendium to the Foundations of Classical Statistical Physics." In *Handbook for the Philosophy of Physics*, pt. B, ed. Jeremy Butterfield and John Earman, 923–1074. Amsterdam: Elsevier.