What Chains Does Liouville's Theorem Put on Maxwell's Demon?*

Peter M. Ainsworth†‡

Recently Albert and Hemmo and Shenker have argued that, contrary to what is sometimes suggested, Liouville's theorem does not prohibit a Maxwellian demon from operating but merely places certain restrictions on its ability to operate. There are two main claims made in this article. First, that the restrictions Liouville's theorem places on Maxwell's demon's ability to operate depend on which notion of entropy one adopts. Second, that when one operates with the definition of entropy that is usual in this debate, the restrictions put on Maxwell's demon are not even as severe as Albert and Hemmo and Shenker argue.

1. Maxwell's Demon. Suppose we have a box divided into two compartments. Each compartment contains a gas. The gases are not at the same temperature. There is a small hole in the wall between the compartments and a massless shutter that can be moved over the hole. When the shutter is over the hole, the two compartments are thermally isolated.

Now suppose that the shutter is operated by a very observant demon. The demon follows the following procedure: normally, he leaves the shutter shut, so that the two gases remain isolated. But when he sees a molecule from the cooler gas whose speed is greater than the average speed of the molecules in the hotter gas coming toward the shutter, he opens the shutter to let the molecule through. Similarly, he opens the shutter if he sees a molecule from the hotter gas whose speed is less than the average speed of the molecules in the cooler gas coming toward the shutter. (If the speed

*Received April 2010; revised June 2010.

†To contact the author, please write to: Department of Philosophy, University of Bristol, 9 Woodland Road, Bristol BS8 1TB, United Kingdom; e-mail: plpma@bristol.ac.uk.

‡I would like to thank Meir Hemmo and an anonymous referee, both of whom provided many constructive comments on this article.

Philosophy of Science, 78 (January 2011) pp. 149–164. 0031-8248/2011/7801-0008$10.00

of the molecules is distributed according to the Maxwell-Boltzmann distribution, then there will be some such molecules.)

The effect of this will be to raise the temperature of the hotter gas and lower the temperature of the cooler gas. Since it seems that the demon need not perform any work, we apparently have here a process whose sole thermodynamic effect is a transfer of heat from a cooler to a hotter body, in violation of the second law of thermodynamics.[1]

Maxwell first summoned his demon in 1867 (in a letter to Peter Tait; see Earman and Norton 1998, 438). He took it to show that the second law is not, strictly speaking, true but only effectively true. This result may have seemed surprising at the time, but in the light of Loschmidt's reversibility objection to Boltzmann's H-theorem (Loschmidt 1876-77) and Poincaré's recurrence theorem (1890), this result is now generally accepted (see, e.g., Albert [2000], chap. 4, for a contemporary introduction to Loschmidt's reversibility objection and Poincaré's recurrence theorem).

Nonetheless, there is still a huge debate as to whether Maxwell's demon can operate in a way that achieves a net decrease of entropy. There seem to be two reasons for this. First, Maxwell's thought experiment seems to offer the prospect that we might, with sufficient ingenuity, be able to contrive net entropy decreases ourselves.[2] This would of course be tremendously useful. In fact, it sounds too good to be true. Second, there is an argument in the framework of the Boltzmann approach to statistical mechanics that Liouville's theorem prohibits Maxwell's demon from achieving a net decrease of entropy.[3] This second reason is the focus of this article.

2. Three Boltzmann Entropies. As noted, there is an argument in the framework of the Boltzmann approach to statistical mechanics that Liouville's theorem prohibits Maxwell's demon from achieving a net decrease

1. It is most transparently in (apparent) violation of Clausius's formulation of the second law, which states that no process whose sole thermodynamic consequence is the transfer of heat from a cooler body to a hotter body is possible. But of course it is also in (apparent) violation of the more usual modern formulation, which states that the entropy of an isolated system cannot decrease.

2. To some extent, so does Loschmidt's reversibility objection. But, intuitively, one expects that it would be harder to carry out a Loschmidt reversal than to carry out something like the operation performed by Maxwell's demon.

3. Even though the Gibbs's approach is more commonly used for practical purposes, the Boltzmann approach seems to be generally favored in foundational discussions of statistical mechanics (cf. Lavis 2005, 246). The question of whether it should be preferred is not addressed in this article. See Frigg (2008) for a discussion of the merits and problems of both approaches.

of entropy. The entropy that is referred to in this argument is the Boltzmann entropy, S_B, which is defined as follows:

$$S_B = k \ln(W).$$

In this equation, k is the Boltzmann constant. However, there is an ambiguity as to what exactly the Boltzmann entropy is because there are at least three interpretations of W in the literature: (i) W is the volume of phase space that is compatible with the macrostate of the system (e.g., Jaynes 1965/1983; Callender 2006; Hemmo and Shenker 2006), (ii) W is the number of arrangements compatible with the distribution of the system (e.g., Albert 2000; Ainsworth 2005), and (iii) W is the probability of the macrostate of the system (e.g., Swendsen 2008). It will be argued in the next three subsections that these three interpretations of W are not, in general, equivalent (although they are equivalent in special cases). Consequently, they give rise to three different Boltzmann entropies. In this article, these will be referred to as S_{B1}, S_{B2}, and S_{B3}, respectively. For many purposes it may not be necessary to decide which of these entropies is intended. But, as we shall see in sections 3 and 4, the type of restriction(s) (if any) that Liouville's theorem places on Maxwell's demon's ability to achieve a net decrease of entropy does depend on which entropy one has in mind.

2.1. W as the Volume of Phase Space That Is Compatible with the Macrostate of the System. The reference to "phase space" in this interpretation of W refers to the γ-space of the system. For an N particle system (where each particle has 3 degrees of freedom), this space has 6N dimensions (three for the x, y, and z coordinates of each particle and three for the x, y, and z components of the velocity of each particle). The microstate of the system is represented by a single point in γ-space. We can partition the γ-space into regions, such that each region contains all and only those points that correspond to microstates that are compatible with a given macrostate of the system, where each macrostate corresponds to a unique set of values for some selected macroscopic properties (e.g., temperature, pressure, volume). On this interpretation of W, the value of W for a system is the volume of the region of γ-space that contains all the points that correspond to microstates that are compatible with the actual macrostate of the system.

This interpretation of W is probably the one that is the most commonly used, both in the physics literature and in the philosophy literature. For example, Jaynes asks us to "recall Boltzmann's original conception of entropy as measuring the logarithm of phase volume associated with a macroscopic state" (1965/1983, 83), and Callender claims that "Boltzmann's great insight was to see that the thermodynamic entropy arguably 'reduced' to the volume in γ-space picked out by the macroscopic parameters of the system" (2006).

2.2. W as the Number of Arrangements Compatible with the Distribution of the System. The easiest way to understand this interpretation of W is to consider the μ-space of the system. For an N particle system (where each particle has 3 degrees of freedom), this space has six dimensions (three for the *x*, *y*, and *z* coordinates of the particles and three for the *x*, *y*, and *z* components of the velocity of the particles). The microstate of the system is represented by N points in μ-space (where each point represents the position and velocity of a particular particle). Suppose that the μ-space is "coarse grained," that is, partitioned into a number of equally sized cells of small but finite volume. The "distribution" of the system is given by specifying the number of particles in each cell. The "arrangement" of the system is given by specifying which particle is in which cell. On this interpretation of W, the value of W for a system is the number of arrangements that are compatible with the actual distribution of the system.

How does this interpretation of W compare to the first? Each arrangement of the system corresponds to a single coarse-grained cell in the system's γ-space. So each distribution corresponds to a number of such cells. So (in addition to the system's γ-space being divided into a number of coarse-grained cells: one for each possible arrangement) we can also think of it as being divided into a number of regions (one for each possible distribution), where each region contains all and only those cells that represent arrangements compatible with a particular distribution. So on this interpretation, the value of W for a system is essentially the volume of the region of γ-space that contains all the cells corresponding to the arrangements that are compatible with the actual distribution of the system (as well as being the number of arrangements that are compatible with the actual distribution of the system).

So there is a close link between these first two definitions of W. But the two definitions will only be equivalent if there is a one-to-one correspondence between the macrostates of the system and the distributions of the system, and, in general, this is not the case. For example, suppose we have an ideal gas (composed of identical molecules) whose temperature is not close to 0 K. And suppose we define the macrostate of the system by the values of the temperature, pressure, and volume of the system. The pressure, P, and temperature, T, are given by

$$P = c_1\left(\frac{N}{V}\right)m\langle v^2\rangle;$$

$$T = c_2 m\langle v^2\rangle.$$

In these equations, c_1 and c_2 are constants, N is the number of molecules in the gas, V is the volume of the gas, m is the mass of a single molecule of the gas, and $\langle v^2\rangle$ is the mean-squared velocity of the molecules of the gas. There are several distinct distributions that can give rise to the same values of T, P, and V (and thus the same macrostate). Here, for example, are two:

i) A distribution in which the molecules are evenly distributed over volume V, and all have velocity of magnitude u.
ii) A distribution in which the molecules are evenly distributed over volume V, and half of the molecules have velocity of magnitude w, such that $w^2 = 2u^2$, and half have velocity 0.

So, in general, the distribution of the system is not the same as its macrostate. This point is sometimes ignored or glossed over in the literature, which might lead one to mistakenly conclude that the first two interpretations of W are equivalent. For example, Ainsworth sloppily asserts that the distribution "is effectively the macrostate of the system" (2005, 631), and Albert sometimes uses the term "macrocondition" to refer to "distribution" (see, e.g., 2000, 50), but he introduces the term "macrocondition" as follows: "there is patently a possible *science* of these temperatures and pressures and volumes—a science (that is) of *macroconditions*" (23), which seems to suggest that he takes macroconditions to be equivalent to (what are here called) "macrostates."

2.3. W as the Probability of the Macrostate of the System. The third interpretation of W, as the probability of the macrostate of the system, is the least common. However, it has been urged by Swendsen as not only the interpretation that yields the most appropriate definition of entropy but also as the one that is closest to Boltzmann's own intentions: "The equation on Boltzmann's tombstone, S = klogW . . . does not refer to the logarithm of a volume in phase space. The equation was first written down by Max Planck, who correctly attributed the ideas behind it to Boltzmann [Planck 1901, 1906]. Planck also explicitly stated that the symbol 'W' stands for the German word 'Wahrscheinlichkeit' (which means probability) and refers to the probability of a macroscopic state" (Swendsen 2008, 16). Swendsen himself argues for an epistemic approach to the probabilities involved here (17), suggesting that the probability of a macrostate is the degree of belief we have (or perhaps should have) that we will find the system in that macrostate, other things being equal. But it seems that one could also offer an ontological interpretation of the probabilities involved here. For example, a frequentist might suggest that the probability of a macrostate is the limit (as time goes to infinity) of the ratio of the length of time the system is in that macrostate to the length of time the system has existed.

How does this definition of W relate to the first?[4] Clearly, if the volume of γ-space that represents the macrostate of a system is proportional to the probability of that macrostate, then this definition of W is essentially the same as the first. Advocates of S_{B1} suppose this to be the case (see, e.g., Albert 2000, 96). However, Swendsen claims that this supposition is not in general legitimate (so S_{B3} does not in general reduce to S_{B1}). In particular, he claims that it is not legitimate if we are dealing with a system of strongly interacting particles.[5] Intuitively, this seems right: if, for example, there is a strong attractive force between the particles, then a macrostate in which the particles are all bunched together seems to be more likely than a macrostate in which the particles are spread out over a large volume, notwithstanding the fact that, other things being equal, the latter macrostate corresponds to a larger region of γ-space. Consider, for example, a cloud of gas that collapses to form a star. We know that the gas particles are more densely concentrated in coordinate space after the collapse. So, on the face it, we would expect the "star" macrostate to correspond to a smaller region of γ-space than the "dispersed cloud of gas" macrostate. But presumably the collapse is not an entropy-decreasing process. One way to make sense of what is going on here is to assign a nonuniform probability distribution over γ-space (and take entropy to be a function of probability, not γ-space volume). Relatively small regions of γ-space corresponding to gas all clumped together can then be judged to be more probable (higher-entropy) states than large regions corresponding to the gas widely dispersed.

However, Albert (2000, 90) suggests that we can account for what is going on in this case without resorting to a nonuniform probability distribution over γ-space. He suggests that the "star" macrostate will correspond to a larger region of γ-space than the "dispersed cloud of gas" macrostate because, although the particles in a star are more densely concentrated in coordinate space, they will be more widely dispersed in momentum space. But while it seems plausible that the particles will be somewhat more dispersed in momentum space (simply because the mean-square velocity of the particles will be higher), it is not obvious that this

4. I will not explicitly address the issue of how this definition of W relates to the second. I hope that this will be clear enough from the following discussion and the preceding discussion of how the first two definitions of W are related to each other.

5. In fact, the idea is that it is not strictly legitimate unless we are dealing with noninteracting particles but that it is a reasonable approximation to make if we are dealing with weakly interacting particles (e.g., dilute gases).

will be sufficient to compensate for the fact that the particles are more densely concentrated in coordinate space. Moreover, it is not clear that this kind of answer will be at all plausible in other cases where we have a system of interacting particles.

3. What Chains Does Liouville's Theorem Put on Maxwell's Demon's Ability to Reduce S_{B1}? As noted, in classical statistical mechanics the microstate of a system at a given time is represented by a point in the system's γ-space. In general, the microstate of the system will change over time. The evolution of its microstate over time can be represented by a line in the system's γ-space. As the system is governed by the laws of classical mechanics, and these laws are deterministic, the line does not branch. Moreover, we can, in principle, determine exactly what the line will look like, if we know exactly what the initial microstate of the system is.

But suppose that we do not know exactly what the initial microstate of the system is (as is of course generally the case). Suppose we only know that the system is initially in one of several possible microstates. We can represent our knowledge of the system's microstate by a region of its γ-space: the region that contains all the points corresponding to the microstates that, for all we know, the system might initially be in. Call that region A. We can use the laws of classical mechanics to work out what the possible microstates of the system are at some later time by "evolving" all the points in region A. This will give us a new region, B. In general, B will not have the same shape as A. But Liouville's theorem states that, if the system is isolated, the volumes of regions A and B will be the same.

3.1. Iron Chains? Recall that S_{B1} is the Boltzmann entropy we get when we interpret W as the volume of γ-space that is compatible with the macrostate of the system. Liouville's theorem suggests an argument that Maxwell's demon cannot operate to decrease the S_{B1} of the gas without increasing his own S_{B1}. The argument runs as follows: first, let's consider what happens to the gas, ignoring what happens to the demon. The gas begins in some macrostate G_A and ends up in some macrostate of lower S_{B1}, G_B. Let's call the region of γ-space that contains all the points that represent microstates that are compatible with macrostate X "R(X)." As G_B is a macrostate of lower S_{B1} than G_A, $R(G_B)$ is smaller than $R(G_A)$. It seems that the initial microstate of the gas could be almost any of those that are compatible with G_A, and the demon would still be able to operate. So the point representing the microstate of the gas could start off from nearly anywhere in $R(G_A)$ and would end up somewhere in $R(G_B)$. So—as $R(G_B)$ is smaller than $R(G_A)$—the volume of the region that represents the possible microstates of the gas decreases (this is not forbidden by Liouville's theorem, as the gas is not an isolated system).

We know from Liouville's theorem that for an isolated system the volume of the region that represents the possible microstates of the system cannot decrease as the system evolves. The volume of the region that represents the possible microstates of a joint system is the product of the volumes of the regions that represent the possible microstates of the separate systems. By assumption, the gas + demon system is an isolated system. So if the volume of the region that represents the possible microstate of the gas decreases (and we saw above that it does), the volume of the region that represents the possible microstate of the demon must increase. So the volume of the region that represents the possible final microstates of the demon is greater than the volume of $R(D_A)$ (where D_A is the initial macrostate of the demon). Suppose the final macrostate of the demon is D_B. The region that represents the possible final microstates of the demon must be contained within $R(D_B)$.[6] So the volume of $R(D_B)$ must be greater than the volume of $R(D_A)$. So the S_{B1} of the demon must increase.[7]

So it seems that Maxwell's demon can only successfully decrease the S_{B1} of the gas at the cost of increasing his own S_{B1}. So it seems that Liouville's theorem prevents him from achieving a net decrease of S_{B1}. Much of the literature on Maxwell's demon can be thought of as trying to give a physical explanation as to why the demon's S_{B1} must increase. The main suggestion is that the demon's S_{B1} must increase because to carry out his job the demon must gain information about the position and velocities of the molecules in the gas and that there is an S_{B1} cost associated with either (i) acquiring this information or (ii) erasing this

6. But note that it need not be equal to $R(D_B)$. Which microstates are possible final microstates of the demon is determined not only by the final macrostate of the demon but also by his initial possible microstates and the laws of evolution governing the system.

7. Put a little more formally, let the volume of a region X be V(X). The volume of the region (in the γ-space of the gas) that represents the possible initial microstates of the gas is $V(R(G_A))$, and the volume of the region (in the γ-space of the demon) that represents the possible initial microstates of the demon is $V(R(D_A))$. So the volume of the region (in the γ-space of the joint system) that represents the possible initial microstates of the joint system is $V(R(G_A)) \times V(R(D_A))$. Similarly, the volume of the region (in the γ-space of the joint system) that represents the possible final microstates of the joint system is $V((G_B)) \times V(B)$ (where B is the region representing the final possible microstates of the demon). And by Liouville's theorem $V(R(G_A)) \times V(R(D_A)) = V(R(G_B)) \times V(B)$. So, as $V(R(G_B)) < V(R(G_A))$, $V(B) > V(R(D_A))$. The region that represents the possible final microstates of the demon must be contained within $R(D_B)$. So $V(B) \leq V(R(D_B))$. So $V(R(D_A)) < V(R(D_B))$. So the S_{B1} of the demon must increase.

https://doi.org/10.1086/658113 Published online by Cambridge University Press

information.[8] These claims have been criticized by Earman and Norton (1998, 1999) and Norton (2005).

3.2. Aluminium Chains? Albert (2000, chap. 5) and Hemmo and Shenker (2006) point out that there is a flaw in the argument presented in the previous subsection.[9] As we have seen, Liouville's theorem, together with the fact that the region of γ-space that represents the final possible microstates of the gas is smaller than the region that represents the initial possible microstates of the gas, shows that the region of γ-space that represents the final possible microstates of the demon must be larger than the region that represents the initial possible microstates of the demon. But it is only if we assume that the demon has only one possible final macrostate that we can conclude that the whole of this region must be contained within a single macrostate. If the demon has several possible final macrostates: D_B, D_C, and so on, then all we can conclude is that the union of $R(D_B)$, $R(D_C)$, and so on, must contain the whole of the region representing the final possible microstates of the demon. And of course it is possible that $R(D_B)$, $R(D_C)$, and so on, are all individually the same size or even smaller than $R(D_A)$. So the S_{B1} of the demon might remain the same or even decrease.

This shows that the chains placed on Maxwell's demon by Liouville's theorem are not as restrictive as we had initially thought. But note that it still seems that the theorem places limitations on the operation of the demon. It seems to show that if the S_{B1} of the demon does not increase, then there must be some uncertainty as to what the final macrostate of the demon will be. (Although, Hemmo and Shenker [2006] point out that in principle this uncertainty as to what the final macrostate of the demon

8. Even if there is an entropy cost associated with erasing information, it is not immediately obvious why that is relevant because it is not immediately obvious why the demon would need to erase the information. The claim is that either (i) he needs to do this because he must return to his initial state or (ii) he will have to do it sooner or later because he only has a finite memory.

9. It is reasonable to ask what definition of entropy Albert and Hemmo and Shenker work with. In the case of Hemmo and Shenker, it is clear that by entropy they mean S_{B1}: "S at time t is defined as the logarithm of the Lebesgue measure of the macrostate of S at time t," where "macrostates correspond to the values of some classical macroscopic *observables*" (2006). But, it is not so clear that this is the definition of entropy that Albert is working with. When he initially defines the Boltzmann entropy, he clearly intends S_{B2}, but, as noted, he goes on to refer to distributions as "macroconditions," which appear to be equivalent to what are here called "macrostates." This seems to suggest that he does not think there is any significant difference between S_{B1} and S_{B2}. However, as we shall see in sec. 4, the difference between them is significant in this context.

will be can be translated into uncertainty as to what the final macrostate of the environment will be.)

3.3. Paper Chains? Let us distinguish two types of Maxwellian demon: "brazen demons" and "subtle demons." A brazen demon wears its demonic powers on its sleeve, in the sense that everything that has the same macroscopic properties as a brazen demon is itself a (brazen) Maxwellian demon. Subtle demons do not wear their demonic powers on their sleeves, in the sense that not everything that has the same macroscopic properties as a subtle demon is itself a Maxwellian demon.

Intuitively, one might expect that any Maxwellian demon would be a subtle demon: if all we know about the (alleged) demon is really that he is in a certain macrostate (i.e., that he has certain values of temperature, pressure, volume, etc.), then it seems unlikely that we would know that he actually is capable of carrying out the operations required to reduce the S_{B1} of the gas. It seems that there could be lots of things that have the same macroscopic properties as the demon but cannot affect these operations. Or, to put it the other way round, if we know that the demon really is capable of carrying out the operations required to reduce the S_{B1} of the gas, then it seems that we know more about him than that he is in a certain macrostate. However, for present purposes there is no need to claim that a Maxwellian demon must be a subtle demon; it suffices to note that a Maxwellian demon could be a subtle demon.[10]

Albert and Hemmo and Shenker do not draw the distinction between brazen and subtle demons and consider (in effect) only brazen demons. The central claim put forward in this article is that subtle demons are not subject to even the limitation suggested by Albert and Hemmo and Shenker: it is contended here that it is possible for a subtle demon to decrease the S_{B1} of the gas and end up (with certainty) in the same macrostate that he began in. This state of affairs is, intuitively, what the original thought experiment seems to suggest is possible. The operations that the demon is required to perform by Maxwell do not appear to be the sort of operations that would in any way alter the macrostate of the demon (as characterized by his macroscopic properties: temperature, pressure, etc.). So the claim of Albert and Hemmo and Shenker that there must necessarily be some uncertainty as to the final macrostate of the demon seems rather surprising (as does the claim they refute: that the final macro-

10. A referee suggested that we could not build a subtle demon, as we are macroscopic beings. This seems to me wrong because we can manipulate systems at the microscopic level: this is what nanotechnologists do. However, it is certainly true that if we did build a subtle demon, then we would know more about it than is given by a specification of its macroscopic state.

state of the demon cannot be the same as his initial macrostate but must be a macrostate of higher S_{B1}).[11]

Consider a subtle demon (whose initial macrostate is D_A). We may or may not know that he is a subtle demon (see n. 10). But either way, given that he really is a subtle demon, his possible initial microstates are actually only a subset of those that are represented by all the points in $R(D_A)$. From Liouville's theorem (together with the fact that the volume of γ-space that represents the final possible microstates of the gas is smaller than the volume of γ-space that represents the initial possible microstates of the gas), we know that the volume of γ-space that represents the final possible microstates of the demon is bigger than the volume of γ-space that represents the initial possible microstates of the demon. But as the initial possible microstates of the demon are actually only a subset of those that are represented by points in $R(D_A)$, this does not imply that the volume of γ-space that represents the final possible microstates of the demon cannot wholly lie in $R(D_A)$. So the final macrostate of the demon could be (with certainty) the same as his initial macrostate.

Call the subregion of $R(D_A)$ representing the microstates that the demon could initially be in (given that he is in macrostate D_A and that he really is able to operate as a demon) A. Suppose the final macrostate of the demon will be (with certainty) the same as his initial macrostate (i.e., D_A). We know by Liouville's theorem and the fact that the gas is in a lower-entropy state at the end of the experiment that the subregion of $R(D_A)$ representing the microstates that the demon could be in at the end of the experiment is larger than A. So at the end of the experiment, the "demon" might not be in A. If he is not, he will not be able to repeat his demonic machinations (i.e., he will no longer be a demon) since, by hypothesis, A contains all the microstates (in region $R(D_A)$) from which he really is able to operate as a demon.

This is illustrated in figure 1 in which the dimensions representing the degrees of freedom of the gas are collapsed onto the *y*-axis, and the dimensions representing the degrees of freedom of the demon are collapsed

11. Perhaps it might be objected that this is taking the initial thought experiment too literally and that, since the demon is something very far removed from our everyday experiences, we cannot really trust our intuition that his macrostate will not be changed as he performs his operations. But imagine we just scale the whole thing up. Instead of a gas of atoms, suppose we have a "gas" of footballs. And instead of a demon, suppose we have a human being operating the massless shutter. Since all he has to do is watch carefully the footballs heading toward the shutter and estimate their speeds (he does not need to know them exactly; he will still achieve the desired result even if he only lets the balls through that are clearly moving fast enough/slow enough) and move a massless shutter up and down, we should surely not expect any change in his macroscopic properties.

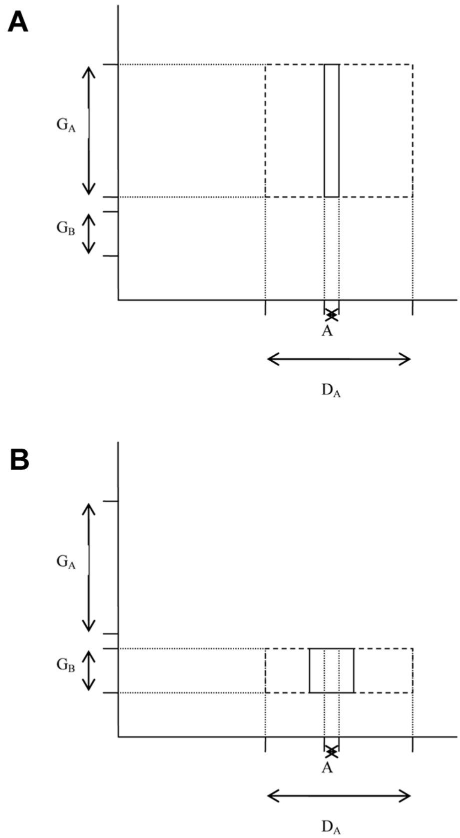

Figure 1. γ-space of the system at the start (*A*) and end (*B*) of the experiment.

onto the *x*-axis (cf. Hemmo and Shenker 2006). In figure 1*A*, the dashed lines enclose the region of phase space compatible with the initial macroscopic properties of the system (demon in state D_A and gas in state G_A). The solid lines enclose the region compatible with the macroscopic properties of the system and the fact that the demon is (initially) really a demon. In figure 1*B*, the dashed lines enclose the region of phase space compatible with the final macroscopic properties of the system (demon in state D_A and gas in state G_B). The solid lines enclose a region (compatible with the final macroscopic properties of the system) of the same volume as is enclosed by the solid lines in figure 1*A*. Evidently, there are some points in this region in which the "demon's" microstate does not lie in region A. If the system ends in one of these microstates, then the "demon" is no longer a genuine demon.

Suppose, for the sake of definiteness, that the region representing the final possible microstates of the demon is three times as big as region A (as in fig. 1). Then, other things being equal, in the best case scenario, the chance that the demon will not be able to repeat the operation is 2/3. But suppose that the demon can repeat the operation (i.e., he does end up in region A). We know that, second time around, he might end up in a microstate that does not lie in A (because, once again, the volume of γ-space representing the possible microstates of the gas has decreased, so the volume of γ-space representing the possible microstate of the demon must increase). So, even if he can repeat the operation once, he might not be able to repeat it a second time. And the chance that he will be able to carry out the operation N times tends to zero as N tends to infinity: sooner or later the demon will not be able to operate.

So, although Liouville's theorem does not prohibit the existence of a Maxwellian demon that (i) can cause a net decrease of S_{B1} and (ii) is guaranteed to end up in the same macrostate that he began in, it does imply that if the demon fulfills these conditions then there is no guarantee that he will be able to repeat the operation, and the chance that he will be able to carry out the operation N times tends to zero as N tends to infinity.

4. What Chains Does Liouville's Theorem Put on Maxwell's Demon's Ability to Reduce S_{B2} and S_{B3}? Recall that S_{B2} is the Boltzmann entropy we get when we interpret W as the number of arrangements compatible with the distribution of the system, and S_{B3} is the Boltzmann entropy we get when we interpret W as the probability of the macrostate of the system. In this section, we consider how the arguments of the previous section are affected if one changes one's concern from the demon's ability to achieve a net decrease of S_{B1} to its ability to achieve a net decrease of S_{B2} or S_{B3}.

How do things change if we substitute S_{B2} for S_{B1} in the foregoing arguments? The initial argument and the response of Albert and Hemmo and Shenker to this argument are essentially unchanged. However, the final argument, which purports to show that Liouville's theorem does not prohibit the existence of a Maxwellian demon that (i) can cause a net decrease of S_{B1} and (ii) is guaranteed to end up in the same macrostate that he began in, does not go through if we substitute S_{B2} for S_{B1}. This argument relied on the supposition that there could be a subtle demon, that is, that there could be a demon that could only operate from a special subset of the microstates compatible with its initial macrostate. The analogue of a subtle demon in the case of S_{B2} would be a demon that could only operate from a special subset of the arrangements compatible with its initial distribution. But it seems that there could be no such demon:

different arrangements compatible with a given distribution are permutations of one another, and it seems that if a demon is actually capable of carrying out the operations that Maxwell's demon carries out when he has one arrangement, then he should also be able to carry out those operations if he has a permutation of that arrangement. So, it seems that Liouville's theorem does prohibit the existence of a Maxwellian demon that (i) can cause a net decrease of S_{B2} and (ii) is guaranteed to end up with the same distribution that he began with.

If we substitute S_{B3} for S_{B1}, then the argument does not even get off the ground. Liouville's theorem puts restrictions on how volumes of γ-space evolve. In the case of S_{B3}, W is taken to be the probability of the macrostate of the system (and, in general, it is not assumed that there is a straightforward connection between probabilities and γ-space volumes). So, in general, Liouville's theorem puts no constraints on what happens to S_{B3}.

In particular, if we are taking entropy to mean S_{B3}, then the fact that the gas goes from a higher- to a lower-entropy state no longer means that the region representing the macrostate of gas at the end of the procedure is smaller than the region representing the macrostate of gas at the start of the procedure since W is now taken to be the probability of the macrostate of the system, not the volume of the region that represents it in γ-space. Moreover, as the volume of the region representing the possible microstates of the gas need not change over the course of the operation, the volume of the region representing the possible microstates of the demon need not change over the course of the operation. So, the region representing the final possible microstates of the demon could be exactly the same as the region representing the initial possible microstate of the demon. So, in this case the demon could continue to operate indefinitely.

However, if the demon operates on a gas of noninteracting particles (so the probability of a macrostate of the gas is essentially the same as the volume of the region that represents it in γ-space), the fact that the gas goes from a macrostate of higher S_{B3} to macrostate of lower S_{B3} does mean that the region representing the macrostate of the gas at the end of the procedure is smaller than the region representing the macrostate of gas at the start of the procedure after all. And this means that the region representing the final possible microstates of the demon must be larger than the region representing the initial possible microstates of the demon. So we have essentially the same results as in the case of S_{B1}.

5. Summary of Results. The main results of the article are (1) Liouville's theorem does not prohibit the existence of a Maxwellian demon that (i) can cause a net decrease of S_{B1} and (ii) is guaranteed to end up in the same macrostate that he began in, but it does imply that (iii) if the demon fulfills these conditions then there is no guarantee that he will be able to repeat the operation, and the chance that he will be able to carry out the operation N times tends to zero as N tends to infinity. (2) Liouville's theorem does not prohibit the existence of a Maxwellian demon that (i) can cause a net decrease of S_{B2}, but it does imply that (ii) if the demon fulfills this condition then there is no guarantee that he will end up with the same distribution that he began with. (3) Liouville's theorem does not prohibit the existence of a Maxwellian demon, operating on a gas of interacting particles, that (i) can cause a net decrease of S_{B3}, (ii) is guaranteed to end up in the same macrostate that he began in, and (iii) can repeat his machinations indefinitely. (4) Liouville's theorem does not prohibit the existence of a Maxwellian demon, operating on a gas of noninteracting particles, that (i) can cause a net decrease of S_{B3} and (ii) is guaranteed to end up in the same macrostate that he began in, but it does imply that (iii) if the demon fulfills these conditions then there is no guarantee that he will be able to repeat the operation, and the chance that he will be able to carry out the operation N times tends to zero as N tends to infinity.

REFERENCES

Ainsworth, Peter M. 2005. "The Spin-Echo Experiment and Statistical Mechanics." *Foundations of Physics Letters* 18 (7): 621–35.

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

Callender, Craig. 2006. "Thermodynamic Asymmetry in Time." In *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Stanford, CA: Stanford University, http://plato.stanford.edu/entries/time-thermo/.

Earman, John, and John D. Norton. 1998. "Exorcist XIV: The Wrath of Maxwell's Demon," pt. 1, "From Maxwell to Szilard." *Studies in History and Philosophy of Modern Physics* 29 (4): 435–71.

———. 1999. "Exorcist 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): 1–40.

Frigg, Roman P. 2008. "A Field Guide to Recent Work on the Foundations of Statistical Mechanics." In *The Ashgate Companion to Contemporary Philosophy of Physics*, ed. D. Rickles, 99–196. London: Ashgate.

Hemmo, Meir, and Orly Shenker. 2006. "Maxwell's Demon." Preprint, PhilSci Archive, http://philsci-archive.pitt.edu/archive/00003795/.

Jaynes, Edwin T. 1965/1983. "Gibbs vs Boltzmann Entropies." In *E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, ed. R. D. Rosenkrantz, 77–86. Repr. Dordrecht: Reidel.

Lavis, David A. 2005. "Boltzmann and Gibbs: An Attempted Reconciliation." *Studies in History and Philosophy of Modern Physics* 36:245–73.

Loschmidt, J. Josef. 1876–77. "Über die Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft." *Wiener Berichte* 73:128–42, 366–72; 75:287–98; 76:209–25.

Norton, John D. 2005. "Eaters of the Lotus: Landauer's Principle and the Return of Maxwell's Demon." *Studies in History and Philosophy of Modern Physics* 36:375–411.

Planck, Max. 1901. "Über das Gesetz der Energieverteilung im Normalspektrum." *Drudes Annalen* 309:553–62.

———. 1906. *Theorie der Wärmestrahlung*. Leipzig: Barth. Trans. M. Masius, *The Theory of Heat Radiation*, New York: Dover, 1991.

Poincaré, Henri. 1890. "Sur le Problème des Trois Corps et les Équations de la Dynamique." *Acta Mathematica* 13:1–270.
Swendsen, Robert H. 2008. "Gibbs' Paradox and the Definition of Entropy." *Entropy* 10: 15–18.