The Simplest Exorcism of Maxwell's Demon:

The Quantum Version

John D. Norton[1]

Department of History and Philosophy of Science

Center for Philosophy of Science

University of Pittsburgh

http://www.pitt.edu/~jdnorton

Norton (2013, Section 4) describes how a Maxwell's demon can be exorcised merely by considerations of the conservation of phase volume in classical physics. This exorcism is generalized here to quantum physics.

1. Introduction

A naturalized Maxwell's demon is a device that can manipulate individual molecules or other component at molecular scales in a way that leads to macroscopic violations of the second law of thermodynamics. The standard view in the present literature, as codified in Leff and Rex (2003), is that such a device must fail and that the reason for the failure is to be sought in a connection supposed between information and thermodynamic entropy. The demonic device is imagined to be structured like a computer. To complete a cycle of its operation, it must erase the information in its memory and, it is asserted, the entropic cost of this erasure outweighs any reduction in thermodynamic entropy achieved elsewhere in the cycle.

In papers written initially with John Earman, I have objected that this analysis is defective in many places. (Earman and Norton, 1998, 1999; Norton, 2005, 2011.) The connection between information and thermodynamic entropy remains unlikely speculation; and it is unclear that this analysis is sufficiently general to apply to the range of Maxwell's demons readily conceivable.

In a recent paper (Norton, 2013, Section 4), it is shown that there is a simple and general demonstration of the failure of a Maxwell's demon that makes no use of the connection supposed between information and thermodynamic entropy. When presented with a thermal system such as a gas in vessel, a Maxwell's demon is presumed able to drive the system away from its normal state of thermal equilibrium into a disequilibrated state and for it to remain there. Translating this behavior into the context of Hamiltonian systems in phase space, this behavior amounts to a time development in which a very large volume of phase space, corresponding to the many possible microstates of the equilibrium state, must evolve into a much smaller volume of phase space, corresponding to the far fewer microstates of the disequilibrated states. This time development violates Liouville's theorem of Hamiltonian mechanics, for that theorem requires that volumes of phase space remain unchanged under Hamiltonian time evolution.

There is no need to reproduce the full analysis of Norton (2013) here.[2] Rather, my purpose is to show that there is an analogous result of comparable simplicity if we assume that the systems at issue are quantum mechanical, not classical. The bulk of the analysis of Norton (2003) remains the same. All that is needed is to substitute quantum analogs for those parts of the argument that depend essentially on classical physics. The main substitution is to replace the conservation of phase volume of classical physics by its analog in quantum theory, the conservation of dimension of a subspace in a many-dimensional Hilbert space. This substitution will be described in Section 2 below. The following section will then list the premises of the classical exorcism along with their quantum counterparts.

2. Conservation of Volumes

The statistical treatment of thermal systems in classical and quantum contexts is sufficiently close for it to be possible to develop the relevant results in parallel, as in the two columns below. Corresponding results are matched roughly horizontally.

Classical Hamiltonian Dynamics
The state of a system is specified by 2n coordinates, the canonical momenta p_1, …, p_n and the canonical configuration space

Quantum Statistical Mechanics
The system state $|\psi(t)>$ is a vector in an n dimensional Hilbert space, with orthonormal basis vectors $|e_1>$, …, $|e_n>$. The time evolution of

coordinates $q_1, \ldots, q_n$ of the classical phase space Γ. The time evolution of the system is governed by Hamilton's equations:

$$\dot{p}_i = \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} \quad \dot{q}_i = \frac{dq_i}{dt} = \frac{\partial H}{\partial q_i} \quad i = 1, \ldots, n$$

(1a)

where $H(q_1, \ldots, q_n, p_1, \ldots, p_n)$ is the system's Hamiltonian.

Classical Liouville Equation

If $f(q_i, p_i, t)$ is a time dependent function over the phase space, then the total time derivative of f, taken along a trajectory $(q_i(t), p_i(t))$ that satisfies Hamilton's equations, is:

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \sum_{i=1}^{n} \left(\frac{\partial f}{\partial q_i} \frac{dq_i(t)}{dt} + \frac{\partial f}{\partial p_i} \frac{dp_i(t)}{dt} \right)$$

$$= \frac{\partial f}{\partial t} + \sum_{i=1}^{n} \left(\frac{\partial f}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial H}{\partial q_i} \right) = \frac{\partial f}{\partial t} + \{f, H\}$$

Set f equal to a probability density $\rho(q_i, p_i, t)$ that flows as a conserved fluid with the Hamiltonian trajectories. Thus ρ satisfies the equation of continuity:[3]

$$0 = \frac{\partial \rho}{\partial t} + \sum_{i=1}^{n} \left(\frac{\partial}{\partial q_i} (\rho \dot{q}_i) + \frac{\partial}{\partial p_i} (\rho \dot{p}_i) \right)$$

$$= \frac{\partial \rho}{\partial t} + \{\rho, H\}$$

Combining we recover the classical Liouville equation

$$\frac{d\rho}{dt} = 0$$

(2a)

It asserts that the probability density in

the system is governed by Schroedinger's equation:

$$i\hbar \frac{d}{dt} |\psi(t)> = H |\psi(t)>$$

$$-i\hbar \frac{d}{dt} <\psi(t)| = <\psi(t)| H$$

(1b)

where H is the system Hamiltonian.

Quantum Liouville Equation

In place of the classical probability density ρ, we have the density operator ρ, which is a positive, linear operator on the Hilbert space of unit trace. It may be written in general as:[4]

$$\rho(t) = \sum_{\alpha} p_{\alpha} |\psi_{\alpha}(t)> < \psi_{\alpha}(t)|$$

where $\sum_{\alpha} p_{\alpha} = 1$ for some set $\{|\psi_{\alpha}>\}$ of state vectors, which need not be orthogonal. This operator represents a "mixed state," that is a situation in which just one of the states in the set $\{|\psi_{\alpha}>\}$ is present, but we do not know which, and our uncertainty is expressed as the ignorance probability p_{α}.[5]

If the state vectors $|\psi_{\alpha}(t)>$ evolve in time according to the Schroedinger equation (1b), the quantum Liouville equation follows:[6]

$$i\hbar \frac{d\rho(t)}{dt} = H\rho(t) - \rho(t)H = [H, \rho(t)] \qquad 2(b)$$

Alternatively, we can write the integral form of the Schroedinger equation as

phase space evolves in time so that it remains constant as we move with a phase point along the trajectory determined by Hamilton's equations.

$$|\psi(t)> \;=\; \exp(-iHt/\hbar)|\psi(0)> \;=\; U(t)|\psi(0)>$$
$$<\psi(t)| \;=\; <\psi(0)|\exp(iHt/\hbar) \;=\; <\psi(0)|U^{-1}(t)$$

(1c)

From it, we recover the integral form of the quantum Liouville equation:[7]

$$\rho(t) = U(t)\rho(t)(0)U^{-1}(t) \qquad 2(c)$$

A quantum analog of classical phase space volume is the dimension of a subspace of the Hilbert space. It is measured by a trace operation. That is, the projection operator

$$P = |e_1><e_1| + \ldots + |e_m><e_m|$$

projects onto an m dimensional subspace of the n dimensional Hilbert space, spanned by the orthonormal basis vectors $|e_1>, \ldots , |e_m>$, where m<n. We can recover the dimension of the subspace as

$$Tr(P) = \sum_{i=1}^{n} <e_i|P|e_i> \;=\; \left(<e_1|e_1>\right)^2 + \ldots + \left(<e_m|e_m>\right)^2 = m$$

Since the numbering of the basis vectors is arbitrary, the result holds for any subspace that is closed under vector addition and scalar multiplication.

If the total dimension n of the Hilbert space is small, the dimension of a subspace is a coarse measure of size in comparison with the finer measurements provided by volume in a classical phase space. However, in the present application, the dimension of the Hilbert space is immense, with n at least the size of Avogadro's number, that is, at least 10^{24}. We need to assess the relative size of the thermal equilibrium states in the Hilbert space, in comparison with the non-equilibrium states. The equilibrium states are *vastly* more numerous than the non-equilibrium states. Our measure need only be able to capture this difference for the exorcism to proceed. While the dimension of the subspaces in which the equilibrium and non-equilibrium states are found is a coarse measure, it is but fully able to express the great difference in the size of the two.

We convert the forms (2a), (2b) and (2c) of the classical and quantum Liouville equation into expressions concerning conservation of volume by introducing analogous special cases of the probability density and density operator:

Classical Hamiltonian Dynamics

Consider a set of states that forms an integrable set S(0) in the phase space at time 0 of phase volume V(0). Under Hamiltonian evolution, it will evolve into a new set S(t). Define a probability density that is uniform over S(0) and zero elsewhere. That is

$$\rho_{S(0)}(q_i, p_i) = (1/V(0))\ I_{S(0)}(q_i, p_i)$$

where $I_S(q_i, p_i)$ is the indicator function that is unity for phase points in the set S and zero otherwise.

The classical Liouville equation (2a) tells us that the probability density remains constant in time along the trajectories of the time evolution. Hence if the initial probability density is a constant 1/V(0) everywhere inside the set S(0) and zero outside, the same will be true for the evolved set S(t). That is, the probability density will evolve to

$$\rho_{S(t)}(q_i, p_i) = (1/V(0))\ I_{S(t)}(q_i, p_i)$$

Since the new probability distribution must normalize to unity, we have[8]

$$1 = \int_\Gamma \rho_{S(t)}(q_i, p_i)\ d\gamma = \frac{1}{V(0)} \int_{S(t)} 1\ d\gamma = \frac{V(t)}{V(0)}$$

which entails that

$$V(t) = V(0) \qquad (3a)$$

Hence the phase volume of a set of points remains constant under Hamiltonian time evolution.

Quantum Statistical Mechanics

The projection operator $P_{S(0)}$ projects onto a closed subspace S(0) of the Hilbert space. Since $P_{S(0)}$ is a projection operator, it is idempotent

$$P_{S(0)} = P_{S(0)}\ P_{S(0)}$$

The dimension of the subspace onto which it projects is

$$V(0) = Tr(P_{S(0)})$$

The uniform density operator corresponding to $P_{S(0)}$ is

$$\rho_{S(0)} = (1/V(0))\ P_{S(0)}$$

Over time, using the quantum Liouville equation (2c), this density operator will evolve to a new density operator

$$\rho(t) = (1/V(0))\ U(t)\ P_{S(0)}\ U^{-1}(t)$$
$$= (1/V(0))\ P_{S(t)}$$

where $P_{S(t)} = U(t)\ P_{S(0)}\ U^{-1}(t)$ is the projection operator to which $P_{S(0)}$ evolves[9] after t. We confirm that $P_{S(t)}$ is idempotent since

$$P_{S(t)}\ P_{S(t)} = U(t)\ P_{S(0)}\ U^{-1}(t)\ U(t)\ P_{S(0)}\ U^{-1}(t)$$
$$= U(t)\ P_{S(0)}\ P_{S(0)}\ U^{-1}(t)$$
$$= U(t)\ P_{S(0)}\ U^{-1}(t)\ = P_{S(t)}$$

and define S(t) as the subspace onto which it projects. Hence we can write

$$\rho(t) = \rho_{S(t)}$$

Finally, density operators have unit trace, so

that

$$1 = \text{Tr}(\rho_{S(t)}) = (1/V(0)) \, \text{Tr}(P_{S(t)})$$

$$= V(t)/V(0)$$

where $V(t)$ is the dimension of $S(t)$. It follows that

$$V(t) = V(0) \qquad (3b)$$

Hence the dimension of a subspace remains constant as the states in it evolve over time under Schroedinger equation.

The derivation of the quantum result (3b) was carried out in a way that emphasizes the analogy with the classical case. The same result can be attained more compactly merely by noting that the trace of a projection operator is invariant under Schroedinger time evolution:[10]

$$V(t) = \text{Tr}(P_{S(t)}) = \text{Tr}(U(t) \, P_{S(0)} \, U^{-1}(t)) = \text{Tr}(U^{-1}(t)U(t) \, P_{S(0)}) = \text{Tr}(P_{S(0)}) = V(0)$$

3. Two Versions of the Exorcism

With the parallel results for the classical and quantum cases in hand, we can now restate the original assumptions of the classical exorcism, listed as (a)-(f) below. Quantum surrogates are needed only for (d)-(f) and are indicated on the right.

> (a) A Maxwell's demon is a device that, when coupled with a thermal system in its final equilibrium state, will, over time, assuredly or very likely lead the system to evolve to one of the intermediate states; and, when its operation is complete, the thermal system remains in the intermediate state.
>
> (b) The device returns to its initial state at the completion of the process; and it operates successfully for every microstate in that initial state.
>
> (c) The device and thermal system do not interact with any other systems.

(classical)	(quantum)
(d) The system evolves according to Hamilton's equations (1a) with a time-reversible, time-independent Hamiltonian.	(d) The system evolves according to the Schroedinger equation (1b), (1c), with a time-reversible, time-independent Hamiltonian.
(e) The final equilibrium state upon which the demon will act occupies all but a tiny portion α of the thermal system's phase space, V, where α is very close to zero.	(e) The final equilibrium state upon which the demon will act occupies all but a tiny subspace of dimension α of the thermal system's Hilbert space, where the dimension α is much smaller than the dimension of the thermal system's Hilbert space.
(f) The intermediate states to which the demon drives the thermal system are all within the small remaining volume of phase space, αV.	(f) The intermediate states to which the demon drives the thermal system are all within the small remaining subspace of Hilbert space of dimension α.

It is assumed in (e) that the Hilbert space of the thermal system and, tacitly, of the demon have a finite, discrete basis. This is the generic behavior of systems such as these that are energetically bound, such as a gas completely confined to a chamber.

The analysis now proceeds as in Norton (2013, Section 4). In brief, according to the behavior specified in (a)-(c), a demon is expected to take a thermal system that we would, under non-demonic conditions, consider to be in thermal equilibrium and evolve it to an intermediate state, that is, one which we would under non-demonic conditions consider to be a non-equilibrium state.

When coupled with the physical assumptions of (d)-(f) that behavior requires a massive compression of phase space volume or Hilbert space volume that contradicts the classical result of the conservation of phase space or the quantum analog for Hilbert subspace dimensions.

The key assumption is expressed in (e). A thermal system that has attained equilibrium under non-demonic conditions occupies one of many states that all but completely fill the phase

space or Hilbert space. The demon must operate successfully on all of these states, or nearly all of them. The intermediate states to which the demon should drive them must occupy the tiny, remaining part of the phase space or Hilbert space. Changes in the demon phase space or Hilbert space can be neglected, since the demon is assumed to return to its initial state.

References

Earman, John and Norton, John D. (1998, 1999) "Exorcist XIV: The Wrath of Maxwell's Demon." *Studies in the History and Philosophy of Modern Physics*, Part I "From Maxwell to Szilard" **29**(1998), pp.435-471; Part II: "From Szilard to Landauer and Beyond," **30**(1999), pp.1-40.

Leff, Harvey S. and Rex, Andrew (2003), eds., *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*. Bristol and Philadelphia: Institute of Physics Publishing.

Nielsen, Michael A. and Chuang, Isaac L. (2000) *Quantum Information and Quantum Computation*. Cambridge: Cambridge University Press.

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**, pp. 375–411.

Norton, John D. (2011) "Waiting for Landauer. *Studies in History and Philosophy of Modern Physics* **42**, pp. 184–198.

Norton, John D. (2013), "All Shook Up: Fluctuations, Maxwell's Demon and the Thermodynamics of Computation," *Entropy*, **15** pp. 4432-4483.

Notes

---

[1] I thank Joshua Rosaler and Leah Henderson for helpful discussion.

[2] The analysis is readily accessible since Norton (2103) is published in an open access journal, *Entropy*. There is also a second development of the result on my website as "The Simplest

Exorcism of Maxwell's Demon: No Information Needed,"

http://www.pitt.edu/~jdnorton/Goodies/exorcism_phase_vol/exorcism_phase_vol.html

[3] Since

$$\sum_{i=1}^{n}\left(\frac{\partial}{\partial q_i}(\rho\dot{q}_i)+\frac{\partial}{\partial p_i}(\rho\dot{p}_i)\right)=\rho\sum_{i=1}^{n}\left(\frac{\partial\dot{q}_i}{\partial q_i}+\frac{\partial\dot{p}_i}{\partial p_i}\right)+\sum_{i=1}^{n}\left(\frac{\partial\rho}{\partial q_i}\dot{q}_i+\frac{\partial\rho}{\partial p_i}\dot{p}_i\right)$$

Using Hamilton's equations (1a), the first term on the right vanishes since

$$\sum_{i=1}^{n}\left(\frac{\partial\dot{q}_i}{\partial q_i}+\frac{\partial\dot{p}_i}{\partial p_i}\right)=\sum_{i=1}^{n}\left(\frac{\partial^2 H}{\partial q_i\partial p_i}-\frac{\partial^2 H}{\partial p_i\partial q_i}\right)=0$$ and the second term is

$$\sum_{i=1}^{n}\left(\frac{\partial\rho}{\partial q_i}\dot{q}_i+\frac{\partial\rho}{\partial p_i}\dot{p}_i\right)=\sum_{i=1}^{n}\left(\frac{\partial\rho}{\partial q_i}\frac{\partial H}{\partial p_i}-\frac{\partial\rho}{\partial p_i}\frac{\partial H}{\partial q_i}\right)=\{\rho,H\}.$$

[4] For a proof, see Nielsen and Chuang (2000, Section 2.4.2).

[5] This representation of a mixed state by a density operator is an inadequate one that must be used cautiously. Distinct mixed states may be represented by the same density operator. The familiar example arises in an SU(2) spin space. The mixed states of equally likely spin-x up and spin-x down have the same density operator as the mixed states of equally likely spin-z up and spin-z:

$$\tfrac{1}{2}|x><x| + \tfrac{1}{2}|-x><-x| = \tfrac{1}{2}|z><z| + \tfrac{1}{2}|-z><-z|$$

The Born rule gives the probability of a measurement outcome $|\phi>$ to be $\text{Tr}(\rho\,P_\phi)$, where $P_\phi = |\phi><\phi|$. The probabilities p_α of the mixed state do not coincide with these Born rule probabilities for measurements of $|\psi_\alpha>$ if the states in $\{|\psi_\alpha>\}$ are not orthogonal.

[6] Applying the Schroedinger equation to each $|\psi_\alpha><\psi_\alpha|$ in the expression for ρ yields

$$i\hbar\frac{d}{dt}\sum_\alpha\left(|\psi_\alpha(t)><\psi_\alpha(t)|\right)=\sum_\alpha\left(H\psi_\alpha(t)>\right)<\psi_\alpha(t)|-|\psi_\alpha(t)>\left(<\psi_\alpha(t)|H\right)=H\rho-\rho H.$$

[7] $\rho(t)=\sum_\alpha p_\alpha|\psi_\alpha(t)><\psi_\alpha(t)|\ =\sum_\alpha p_\alpha U(t)|\psi_\alpha(0)><\psi_\alpha(0)|\ U^{-1}(t)=U(t)\rho(0)U^{-1}(t)$

[8] $d\gamma$ is the canonical phase space volume element $dq_1\ldots dq_n dp_1\ldots dp_n$.

[9] The derivation of this rule of time evolution closely parallels that of the density operator in (2c).

[10] The third equality uses the invariance of trace under cyclic permuation: $\text{Tr}(ABC) = \text{Tr}(CAB)$.