December 17, 2015. Rev. April 20, 2016.

Maxwell's Demon Does Not Compute

John D. Norton
Department of History and Philosophy of Science
University of Pittsburgh
http://www.pitt.edu/~jdnorton

Prepared for Michael E. Cuffaro and Samuel C. Fletcher, eds., *Physical Perspectives on Computation, Computational Perspectives on Physics*. Cambridge University Press.

1. Introduction

Is it instructive to model some physical process as a computational process or, more generally, as one that processes information? That it would be so is an hypothesis that needs to be tested case by case. Sometimes it will be very instructive. Shannon's information theory applied to communication channels is a striking success. There can be failures, however. This chapter will describe a lingering and striking failure.

A Maxwell's demon is a device that can reduce the thermodynamic entropy of a closed system, in violation of the Second Law of Thermodynamics, by means of molecular-scale manipulations. The received view since the mid-twentieth century is that such a device must fail for reasons most instructively captured by theories of information and computation. This received view of the demon's exorcism, I will argue here, is misdirected and mistaken.

First, there are many proposals for Maxwell's demons in which there is no obvious computation or information processing. As a result, the exorcism of the received view cannot be applied to them. It is no general exorcism.

Second, the received view depends variously on dubious principles, Szilard's Principle and Landauer's principle. They are at best interesting speculations in need of precise grounding; or, at worst, mistakes propped up by repeated misapplications of thermal physics.

Third, prior to the emergence of the received view, we already had a serviceable and generally applicable exorcism that made no use of notions of information or computation. In 1912, Smoluchowski had argued cogently that efforts to reverse the second law by manipulations at molecular scales will fail since they will be disturbed fatally by the very thermal fluctuations they seek to exploit.

Finally, I shall show here that the long-entrenched focus on information and computation-theoretic notions has distracted both supporters and opponents of the received view from a simpler exorcism, even stronger than Smoluchowski's arguments of 1912. A simple specification of what a Maxwell's demon must do turns out to be incompatible with the classical Liouville theorem of statistical physics or its quantum counterpart. Hence the demon must fail; and its failure is established without any recourse to notions of information or computation. The exorcism does not even require serious engagement with the notion of thermodynamic entropy.

The early Sections 2, 3 and 4 below will review Maxwell's invention of his demon, its naturalization with the discovery of fluctuation phenomena and Smoluchowski's argument that these same fluctuations defeat the demon. In Section 5, I will report on the appearance of the idea that an intelligent demon may need special accommodations. Sections 6 and 7 trace briefly how the ensuing idea of a naturalized, intelligent demon came to dominate the Maxwell's demon literature, with exorcisms focusing first on a supposed entropy cost in acquiring information and then in erasing it. This is, I will argue, a failing literature.

In Sections 8 and 9, I will report a new, stronger and simpler exorcism based on the contradiction between what the demon must do and Liouville's theorem of statistical physics. The exorcism reported is limited to classical physics. Sections 10, 11 and 12 will show a closely analogous exorcism using the quantum analog of Liouville's theorem.[1]

2. Maxwell's Fictional Demon

Maxwell [1871, pp. 308-309] unveiled his demon in print in 1871. He used it to make a point about the character of the second law of thermodynamics. We cannot reverse the second

---

[1] The content of Sections 11 and 12 can also be found in Norton [2014]. I thank Joshua Rosaler and Leah Henderson for helpful discussion of the quantum material.

law, Maxwell sought to establish, merely because we have no access to individual molecules. Instead we must treat molecular systems en masse. To make his point, he imagined a quite fictitious being who could access molecules individually. By carefully opening and closing a door in a dividing wall as the molecules of a gas approached it, this demonic being could accumulate slow molecules on one side and fast molecules on the other. The first side cools while the second warms, yet no work is done. The normal course of thermal processes is reversed, in contradiction with the second law.[2]

3. Fluctuations Bring Naturalized Demons

A major change in the demon's role came with the recognition in the early twentieth century that thermal fluctuations are microscopically observable. They could no longer be dismissed as an artifact of molecular theory of no practical import. They realize, it was concluded, a microscopic violation of the second law of thermodynamics, which could at best hold only for time-averaged quantities. The celebrated example is Einstein's [1905] analysis of Brownian motion. The larger movements of the Brownian particle arise through a transfer of the heat energy of the surrounding water into the particle's kinetic energy. It might then be converted to gravitational potential energy, a form of work energy, if the motion lifts the particle vertically. This is a momentary, microscopic violation of the second law of thermodynamics: ambient heat energy has been fully converted to work.

Maxwell had given no account of just how his demon might be constituted. Since the point was that his demon was fictional and intended to display vividly what we cannot do, there was no need for it. With the new recognition about thermal fluctuations, Maxwell's demon was moved from the realm of impossible fiction to a candidate physical possibility. If momentary, microscopic violations of the second law are possible, might we devise a real machine that can accumulate them and eventually lead to macroscopic violations of the second law? Such a machine would be a naturalized Maxwell's demon. That is, it would be one whose workings conform with the known natural laws of microscopic systems.

---

[2] For an account of Maxwell's original proposal and conception, see Myrvold [2011].

What followed were numerous proposals for naturalized Maxwell's demons of simple design. Some were intended to be realized in the laboratory. Such was Svedberg's [1907] colloid demon. In it, the Brownian motion of electrically charged colloid particles would lead them to radiate their thermal energy, which would be trapped in a carefully designed system of casings. The colloid would spontaneously cool, while the casing heated. Smolucholwski's [1912] paper contained a range of more schematic proposals. One was a one-way valve that would allow gas molecules to pass in one direction but not the other. This one-way transport was effected by a hole with a ring of hairs; or by a valve with a flapper.

This last proposal entered later literature in modified form as the Smoluchowski trapdoor. In his original thought experiment, Maxwell employed a fictional demon to open and close the door in the dividing wall of the chamber. Smoluchowski's trapdoor was an automatic device. It was lightly spring-loaded and configured so that molecules moving in one direction would flip it open and pass; whereas molecules moving in the opposite direction would slam it shut and be obstructed. For more discussion of these proposals, including what would later become Feynman's "ratchet and pawl" demon, see Norton [2013, §2].

4. Fluctuations Defeat Maxwell's Demon

The main point of Smoluchowski's analysis was that all these proposals for Maxwell's demons fail. For they are machines operating at molecular scales where fluctuation phenomena dominate. In each case, some fluctuation-driven process would reverse the normal course of thermal processes. The individual molecular collisions that flip open the valve flapper or the Smoluchowski trapdoor are pressure fluctuations in the gas. Smoluchowski then showed that, for each case, there was a second fluctuation process that undid the anti-entropic gains of the first. In the case of the Smoluchowski trapdoor, if the device is to operate as intended, the flapper must be so light that collisions with individual molecules can open it. But such a light flapper will have its own fluctuating thermal energy, which will lead it to flap about randomly, allowing molecules to pass in both directions. On average there is no accumulation of violations of the second law.

Smoluchowski made his case by examining many examples of candidate mechanisms and showing that they all failed in the same way. The analysis provided no principled proof of

the generalization that all demon proposals must fail this way. However once one sees the one mode of failure repeated again and again, in the range of examples treated by Smoluchowski, the generalization is hard to resist.

There is another way to see that fluctuations are a formidable obstacle to efforts to realize a Maxwell's demon. Such a demonic device will operate at molecular scales and will be composed of a series of steps, each of which must be brought to completion before the next can start. In recent work [Norton 2011, §7; 2013; Part II], I have shown that the completion of *any* single process at molecular scales, no matter how simple or complicated, intelligently directed or otherwise, involves dissipation. For any such process must overcome the thermal fluctuations that disrupt its orderly execution. They can only be overcome by the dissipative creation of entropy, if completion is to be assured, even just probabilistically. The quantities of entropy involved are great enough to swamp the entropy reduction envisaged in the operation of a Maxwell demon.

These considerations of fluctuations are not a deductive proof from first principles of the impossibility of a Maxwell's demon. However they make it quite plausible that a molecular-scale demon cannot overcome the disrupting effects of thermal fluctuations. They give us a simple and proven recipe for demonstrating the failure of any new proposal for a Maxwell's demon: look for the neglected effects of fluctuations.

5. The Distraction of Intelligent Intervention

Smoluchowski's 1912 verdict on the possibility of a naturalized Maxwell's demon provides a resolution that is still illuminating today. Naturalized demons will likely fail because thermal fluctuations will disrupt their intended operations. Smoluchowski's paper was delivered as a lecture at the 84th *Naturforscherversammlung* (Meeting of Natural Scientists) in Münster. The discussion that followed is reported at the end of the journal printing of Smoluchowski's lecture. In it, Kaufmann directed a quite awkward question to Smoluchowski:

> Kaufmann: The lecturer has indicated why presumably also no mathematical selection [among molecules of different speed] that contradicts the second law can be brought about by means of an automatic valve. The relations are otherwise for a valve with something like a sliding bar, whose motion requires no work in theory.

> Then there is an intelligence operating the valve and ensuring that the opening and closing is in the right moment; I believe that, for Brownian molecular motion, something like this is practically achievable. Then the second law would be violated by the participation of an intelligent creature. [This is] a conclusion that one possibly could regard as proof, in the sense of the neo-vitalistic conception, that the physico-chemical laws alone are not sufficient for the explanation of biological and psychic occurrences.

This is the sort of question any speaker dreads. Smoluchowski had just based his lecture on the presumption that a Maxwell's demon is naturalized, that is, it is subject to the normal physico-chemical laws. Then the demon will fail. Now he is asked to contemplate the case of a neo-vitalist demon; that is, an intelligence whose actions are not governed by those laws but is animated by some kind of vital force. It is even suggested that this might lead to an experiment that vindicates vitalism. The suggestion is far-fetched. If an intelligent organism—a human, for example—accumulates microscopic violations of the second law in Brownian motion in a real laboratory experiment, one must also account for the entropy created in the organism's metabolism. To ignore it through some vitalist commitment would make the vitalist interpretation of the experimental result circular.

Smoluchowski gives the best reply he can muster:

> Lecturer: What was said in the lecture certainly pertains only to automatic devices, and there is certainly no doubt that an intelligent being, for whom physical phenomena are transparent, could bring about processes that contradict the second law. Indeed Maxwell has already proven this with his demon.

This grants the tacit presumption of the question: that a vitalistic demon, were there such a thing, could succeed. However Smoluchowski then awkwardly reminds the questioner of the background assumption of Smoluchowski's entire analysis. He continued:

> However intelligence extends beyond the boundaries of physics. On the other hand, it is not to be ruled out that the activity of intelligence, the mechanical operation of the latter, is connected with the expenditure of work and the dissipation of energy and that perhaps after all a compensation still takes place.

Intelligence, presumably in the abstract, disembodied sense, is something that lies outside physics. But intelligence that can act in the world will do it through a physical system and this is

still a system that will be governed by the familiar laws. The wording is hesitant—it should not be ruled out. However I attribute the hesitancy merely to the politeness required to respond to a question clearly outside the scope of the speaker's talk.

6. Szilard's Principle

What happens if the intervening demon is an intelligence unconstrained by normal physico-chemical laws? This was a question best left to die quietly. If one allows such an intelligence, then no physical law is secure. If, however, the intelligence is embodied in a physical system, then Smoluchowski has already provided a quite serviceable answer: whether the system is intelligent or not, thermal fluctuations will likely preclude its operation. The question of an *intelligent* intervening demon is a distraction, since *all* demonic intervention will fail.

Unfortunately Leo Szilard was unable to resist the temptation of pursuing the distracting question. His 1929 "On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings" responded directly to Smoluchowski's work and quoted liberally from it. It initiated a decline in the literature on Maxwell's demon from which we have still to recover.

The details of Szilard's analysis are quite complicated and even obscure. See Earman and Norton [1998, §7] for a review. What survived into the ensuing literature were a few ideas in a form somewhat simpler than Szilard's formulation. The most important idea was that one need not provide physical details of the mechanism that animates the intelligent demon. All one needs to know is that its operation requires the gaining of information. The mere fact of gaining information, however it is done, creates enough entropy to defeat the demon.

To illustrate the point, Szilard introduced an ingeniously simplified arrangement in which the demon cyclically manipulates a one-molecule gas. Each cycle requires the demon to discern whether the molecule is trapped on the left or the right side of a partition. This discerning—in later literature the gaining of one bit of information—was, Szilard asserted, necessarily a dissipative process that creates entropy and protects the second law from violation.

How much entropy does this gaining of information create? If the second law is to be protected, then the process must create at least k log 2 of thermodynamic entropy for each bit of

information gained, where k is Boltzmann's constant. This principle was later called "Szilard's principle" [Earman and Norton, 1999]. That this amount suffices to protect the second law was assured by the expedient of working backwards. Assume that the second law is preserved and compute from that assumption how much entropy must be created. Szilard's principle ensues. While Szilard and others after him did try to justify the principle by examining particular detection processes, working backwards remained the simplest and most general justification.

The principle in this form supported a flourishing literature in the 1950s. It proclaimed a deep truth in the connection between information and thermodynamic entropy. This insight, it assured us, explains why a Maxwell demon must fail, even though its core claim of Szilard's principle was commonly derived by circular reasoning from the very presumption that a Maxwell's demon must fail.

For a synoptic discussion of this new literature and the ensuing literature in the thermodynamics of computation, and for reproductions of key papers, see Leff and Rex [2003].

7. Landauer's Principle

The success of this last exorcism was short-lived. It was replaced within a few decades by a modified version that drew on computational notions. The modified version retained the idea that one should abstract away all of the details of the demon's constitution excepting its treatment of information. But now the unavoidable dissipative step was not the acquiring of information. It was the erasure of information. To function, a demon must remember what it has learned. In the case of Szilard's example, the demon must remember that the molecule was trapped on the left or the right side of the partition; and that memory must be captured in some physical change in the demon. To complete the thermodynamic cycle, the demon's memory must be returned to its initial state. That return is the moment of dissipation. The erasure of this one bit of information is associated with k log 2 of thermodynamic entropy, which is just the amount needed to protect the second law. The statement of this erasure cost is "Landauer's principle," drawn from the work of Rolf Landauer [1961]. It is the central result of what soon came to be known as the "thermodynamics of computation."

The new computation-theoretic exorcism was laid out in Bennett [1982, §5]. In order to secure its primacy, the new exorcism needed to overturn the old exorcism. Its proponents, we

were now told, had simply erred in attaching the necessity of dissipation to information acquisition. All the clever arguments and manipulations of the old exorcism were deceptive mirages. Bennett [1982, §5, 1987] sketched new thought experiments in which information about the states of target systems could be gained by processes claimed to be thermodynamically reversible.

This computation-theoretic exorcism has now settled in as the standard in the literature. Although there have been amendments offered that draw on notions of complexity and quantum theory,[3] the basic ideas of the exorcism have survived with some stability. One might be excused for taking this stability as a sign of cogency. Alas, the computation-theoretic exorcism of the 1980s was no improvement on the fragile information-theoretic exorcism of the 1950s. It had merely rearranged some of its parts.

To begin, the essential problem remains. There are many proposal for Maxwell's demon in which there is no overt collection of information and no overt computation that employs a memory that must be erased. These processes, for example, are simply not present in the canonical Smoluchowski trapdoor or Feynman's ratchet and pawl demon. Therefore, neither information-theoretic nor computation-theoretic exorcism can touch them. However Smoluchowski's original, thermal fluctuation based exorcism applies to them and all the rest.

Second, the information-theoretic exorcism had been supported by ingenious thought experiments that illustrated how gaining information is thermodynamically costly. In a thought experiment reminiscent of the celebrated Heisenberg microscope of the quantum uncertainty principle, Brillouin [1950] had computed that dissipation compatible with Szilard's principle must occur, if a photon with energy above the thermal background is used to locate a particle. In spite of the luminaries of physics like Brillouin who had supported them, these thought experiments were all misleading and mistaken, we were now told. The trouble was that the thought experiments that replaced them were no better. Bennett's [1982, §5, 1987] illustrations of devices that could gain information dissipationlessly all required devices of delicate sensitivity. It takes only the most cursory of inspections to see that their operations would be fatally disrupted by thermal fluctuations, just as Smoluchowski envisaged. (See Norton, 2011, §7.3.) One defective set of thought experiments had merely been replaced by another.

---

[3] See Earman and Norton [1999] for further discussion.

Finally, the computation-theoretic exorcisms draw on Landauer's principle. When Landauer [1961] introduced the principle, it was little more than a promising speculation, supported by a sketchy plausibility argument. Over half a century later, one might imagine that this would be sufficient time to place the principle on a more secure foundation. This has not happened. It is not for want of trying. However, as I have documented in detail elsewhere (Norton, 2005, 2011 and summarized in Norton 2013, §3.5), the now burgeoning literature on Landauer's principle persists in committing repeatedly a small set of interconnected errors in thermal analysis.

8. Asking the Right Question

These failed traditions are driven by the belief that a successful exorcism of Maxwell's demon abstracts away all details of the demon's operation, other than its processing of information. As the discussion of the previous sections illustrates, this belief has presided over a descent into a feckless, convoluted and confused literature. As long as the attention of authors in the field, proponents and critics alike, remains focused on information processing, this descent is likely to continue. Here, ruefully and regretfully, I include much of my own writing over more than a decade on the topic. At best I have been able to show what does not work in exorcising the demon. What I should have asked is what does work.

Let us start again. Let us set aside information and computation-theoretic notions and take stock of what we know. We have known since Smoluchowski's work of 1912 that disruptions by fluctuations presents a formidable barrier to all efforts to realize a Maxwell's demon. We now also have strong empirical indications of the impossibility of such a demon. Nanotechnology has given us abilities to manipulate individual atoms far beyond anything Maxwell or Smoluchowski could have imagined. In 2013, scientists at IBM made a stop motion video of a stick figure boy playing with a ball.[4] The figures were drawn by lining up individual carbon monoxide molecules on a copper surface in a scanning tunneling microscope. Even with such prodigious capacities to manipulate individual molecules, no fully successful Maxwell's demon has been made. Rather all work at nanoscales struggles to overcome thermal fluctuations.

---

[4] http://www-03.ibm.com/press/us/en/pressrelease/40970.wss

They are the nemesis of nanoscience, just as Smoluchowski argued. The molecules of the IBM stop motion video were cooled to -268C to suppress fluctuations.

There have been other empirical clues. The biochemistry of a cell involves molecular processes of comparable refinement. The operation of a ribosome in a cell is a marvel of miniaturized molecular machinery. It was brought into being by the creative powers of evolution. Yet these same prodigious powers have failed to construct a demonic device in the cell, in spite of the obvious advantage to the cell of a process that converts ambient heat energy to useful work.

With some reasonable expectation that a Maxwell's demon is impossible, let us ask the question that has been neglected: is there a simpler way to demonstrate the impossibility of a Maxwell's demon that avoids the convolutions of the present literature?

9. A Better Exorcism

It came as a sobering surprise when I found recently [Norton, 2013, §4] that there is a very simple exorcism of Maxwell's demon that requires only elementary notions from statistical physics. There is no need for notions of information or computation or erasure, or tendentious principles like Szilard's or Landauer's. One need not even mention the ever-troublesome notion of entropy. The exorcism shows that a description of what a Maxwell's demon must do is incompatible with Liouville's theorem of statistical physics.

Here, in brief, is how it works. When presented with a target thermal system such as a gas in a vessel, a Maxwell's demon is presumed able to drive the system away from its normal state of thermal equilibrium into what would otherwise be judged a disequilibrated state, were there no interaction with the demon, and for the system state to remain so. For example, Maxwell's original demon or the Smoluchowski trapdoor takes a gas at uniform temperature and separates the hotter, faster molecules from the slower, colder ones. Once its work is done, we have the disequilibrated gas, with the hotter part on one side of a partition and the colder part on the other side. To ensure that there is no compensating hidden thermal dissipation or degradation in the demon itself or any supporting systems it uses, we require that the demon and these supporting systems are returned to their original states at the end of the process. Such a process reverses the second law of thermodynamics.

If we redescribe this process in the context of standard statistical physics, we quickly see that it is impossible. In that context, systems are presumed to be governed by Hamilton's equations, versions of which cover virtually all physical theories considered. The state of a system is fixed by determining a large number of generalized position and momentum variables. These variables are the coordinates of a space, known as a phase space. The state of a Hamiltonian system at one moment corresponds to a single point in the phase space. As the state changes, it traces a trajectory in the phase space.

A closed system will revert spontaneously to equilibrium states. For example, a gas confined in an isolated vessel will evolve to a state of uniform pressure, temperature and density. These equilibrium states occupy virtually all of the system's phase space. The non-equilbrium states with non-uniformities occupy only a tiny fraction of the volume of the phase space. This difference of volumes is the rough and ready explanation for why closed thermal systems revert to their equilibrium states. As the phase point of the system migrates in time through the phase space, it almost always ends up in the much larger part of phase space where equilibrium systems are found. The non-equilibrium states are mere temporary intermediates on the way to equilibrium.

When we couple a Maxwell's demon and its support systems to some target system in thermal equilibrium, we form a larger system with its own, larger phase space. If the demon operates as intended, the target system will evolve from an equilibrated to a disequilibrated, intermediate state, while the demon and its support systems revert to their original states. (Since the supposition of successful action of the demon upsets the normal notions of equilibrium and disequilibrium, henceforth these disequilibrated states will be labeled more neutrally "intermediate states.") This evolution is required to happen no matter which the equilibrium microstate of the target system; or at least for most of the equilibrium microstates of the target system. That is, the operation of the demon must compress the phase space volume of the target system down to a very much smaller volume, while leaving the phase space volume of the demon and supporting systems unchanged. The overall effect is that the successful operation of the demon must compress phase space of the combined system.

The combined system is governed by Hamilton's equations. An early and easily gained property of such systems is Liouville's theorem. It states that time evolution leaves phase space volumes unchanged. That is, if we select some set of states forming a volume in the phase space,

over time, as the systems evolves, the set of states occupied will move around the phase space. However the volume that they occupy in phase space remains unchanged.

In sum, the successful operation of a Maxwell's demon must compress phase space. Liouville's theorem of statistical physics asserts that this is impossible. Therefore a Maxwell's demon is impossible.

10. Classical or Quantum?

The exorcism just sketched informally was developed formally in Norton [2013, §4] and the main derivations will be reproduced again below. There is a weakness is this exorcism. The processes involved occur at molecular scales, where the quantum mechanical properties of systems can be important. Yet the exorcism employs classical physics.

The remaining analysis below rectifies this weakness. The bulk of the original analysis remains the same and an analogous result of comparable simplicity is recovered. 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 11 below. The following section will then list the premises of the classical exorcism along with their quantum counterparts.

11. 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	*Quantum Statistical Mechanics*
The state of a system is specified by 2n coordinates, the canonical momenta $p_1, \ldots, p_n$ and the canonical configuration space	The system state $\|\psi(t)\rangle$ is a vector in an n dimensional Hilbert space, with orthonormal basis vectors $\|e_1\rangle, \ldots, \|e_n\rangle$. 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:

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

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 defined on 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. Now ρ satisfies the equation of continuity:[5]

the system is governed by Schroedinger's equation:

$$i\hbar\frac{d}{dt}|\psi(t)> \; = H\,|\psi(t)> \qquad -i\hbar\frac{d}{dt}<\psi(t)| \; = <\psi(t)|\,H \tag{1b}$$

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:[6]

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

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] 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

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

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

Combining with the expression for the total derivative dρ/dt, we recover the classical Liouville equation

$$\frac{d\rho}{dt} = 0 \qquad (2a)$$

It asserts that the probability density in 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.

If the state vectors $|\psi_\alpha(t)\rangle$ evolve in time according to the Schroedinger equation (1b), the quantum Liouville equation follows:[7]

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

Alternatively, we can write the integral form of the Schroedinger equation with the unitary operator U(t) as

$$|\psi(t)\rangle = \exp(-iHt/\hbar)|\psi(0)\rangle = U(t)|\psi(0)\rangle,$$
$$\langle\psi(t)| = \langle\psi(0)|\exp(iHt/\hbar) = \langle\psi(0)|U^{-1}(t) \qquad (1c)$$

From it, we recover the integral form of the quantum Liouville equation:[8]

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

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

$$P = |e_1\rangle\langle e_1| + \ldots + |e_m\rangle\langle e_m|$$

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

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

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

[6] For a proof, see Nielsen and Chuang [2000, Section 2.4.2].

[7] Applying the Schroedinger equation to each $|\psi_\alpha\rangle\langle\psi_\alpha|$ in the expression for ρ yields

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

[8] $\rho(t) = \sum_\alpha p_\alpha |\psi_\alpha(t)\rangle\langle\psi_\alpha(t)| = \sum_\alpha p_\alpha U(t)|\psi_\alpha(0)\rangle\langle\psi_\alpha(0)|U^{-1}(t) = U(t)\rho(0)U^{-1}(t)$

Since the numbering of the basis vectors is arbitrary, the result holds for any subspace, which 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 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

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

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

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

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

---

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

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[9]

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

which entails that

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

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

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[10] 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

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

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 the Schroedinger equation.

---

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

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:[11]

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

12. 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 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 equilibrium state upon which the	(e') The equilibrium state upon which the

[11] The third equality uses the invariance of trace under cyclic permuation: Tr(ABC) = Tr(CAB). The fourth uses unitarity $U^{-1}U = I$.

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.	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)/(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)/(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

Bennett, Charles H. [1982] "The Thermodynamics of Computation—A Review." *International Journal of Theoretical Physics*, **21**, pp. 905-40.

Bennett, Charles, H. [1987] "Demons, Engines and the Second Law." *Scientific American*, **257**(5), pp.108-116.

Brillouin, Leon [1950] "Maxwell's Demon Cannot Operate: Information and Entropy I," pp.120-23 in Leff and Rex [2003].

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.

Einstein, Albert [1905] "On the Movement of Small Particles Suspended in a Stationary Liquid Demanded by the Molecular-Kinetic Theory of Heat," pp. 1-18 in *Investigations on the Theory of the Brownian Movement*. R. Fürth, ed., A. D. Cowper trans. Methuen, 1926; repr. New York: Dover, 1956.

Landauer, Rolf [1961] "Irreversibility and heat generation in the computing process", *IBM Journal of Research and Development*, **5**, pp. 183–191.

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.

Maxwell, James Clerk [1871] *Theory of Heat*. London: Longmans, Green and Co.

Myrvold, Wayne [2011] "Statistical mechanics and thermodynamics: A Maxwellian view," *Studies in History and Philosophy of Modern Physics*, **42** , pp. 237–243.

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.

Norton, John D. [2014] "The Simplest Exorcism of Maxwell's Demon: The Quantum Version." http://philsci-archive.pitt.edu/10572/

Smoluchowski, Marian [1912] "Experimentell nachweisbare, der üblichen Thermodynamik widersprechende Molekularphünomene," *Physikalische Zeitschrift*, **13**, pp. 1069–1080.

Svedberg, The. [1907] "Über die Bedeutung der Eigenbewegung der Teilchen in kolloidalen Lösungen für die Beurteilung der Gültigkeitsgrenzen des zweiten Haupsatzes der Thermodynamik," *Annalen der Physik*, **59**, pp. 451–458.

Szilard, Leo [1929] "On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings." pp.120–129 in *The Collected Works of Leo Szilard: Scientific Papers*. MIT Press: Cambridge, MA, 1972.