Pergamon

Stud. Hist. Phil. Mod. Phys., Vol. 29, No. 4, 435–471, 1998
© 1998 Elsevier Science Ltd. All rights reserved
Printed in Great Britain
1355-2198/98 $19.00 + 0.00

Exorcist XIV: The Wrath of Maxwell's Demon. Part I. From Maxwell to Szilard

*John Earman and John D. Norton**

In this first part of a two-part paper, we describe efforts in the early decades of this century to restrict the extent of violations of the Second Law of thermodynamics that were brought to light by the rise of the kinetic theory and the identification of fluctuation phenomena. We show how these efforts mutated into Szilard's (1929) proposal that Maxwell's Demon is exorcised by proper attention to the entropy costs associated with the Demon's memory and information acquisition. In the second part we will argue that the information theoretic exorcisms of the Demon provide largely illusory benefits. According to the case, they either return a presupposition that can be had without information theoretic consideration or they postulate a broader connection between information and entropy than can be sustained. © 1998 Elsevier Science Ltd. All rights reserved.

1. Introduction

Maxwell's Demon is the source of a literature that has no apparent end—the most prestigious scientific journals, including *Physical Review* and *Physical Review Letters*, continue to publish articles on this topic. It is a topic that is considered to be important enough to merit editorials in *Nature* and *New Scientist*.[1] From the point of view of philosophy of science, however, there are a number of peculiar characteristics of this literature, the most notable being the lack of any self-reflection on what the goals of the enterprise are and what the rules of the game are. In particular, why should one want to exorcise the

(*Received 15 July 1997; revised 18 December 1997*)
* Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A. (*e-mail*: jearman+@pitt.edu and jdnorton+@pitt.edu)
[1] See Brown (1990) and Maddox (1990).

PII: S1355-2198(98)00023-9

Demon? And what exactly would count as a legitimate and effective exorcism? Readers will search the literature in vain for explicit or tacit answers. Given the out-of-focus nature of the subject matter, it is hardly surprising that it has been viewed from a number of different perspectives and that new industries of knowledge have been eagerly seized upon as providing the keys to exorcism. Thus, the Demon has been seen as an information processor whose operations are circumscribed by results of information theory. Alternatively, the Demon has been analogised to a computer, and the key to corralling him has been seen to lie in considerations of memory storage and erasure.

The literature on Maxwell's Demon has produced much amusement and some good science—but, alas, in lopsided ratio. Where so many find exhilarating insight in the exorcisms of the latest Maxwell Demon literature, we find results that are both frustrating and obscure. We have been unable to reconcile ourselves to the notion, increasingly popular over the last half century, that information theory has provided the decisive exorcism of Maxwell's Demon. The deeper purpose of our critical review is to give sharp expression to our reservations. They will be formulated in Section 1 of the second part of this paper as a dilemma. The information theoretic exorcisms of Maxwell's Demon cannot be both sound and profound. In so far as information theory can protect the Second Law of thermodynamics from Maxwell's Demon by *sound* argumentation, it does so through the presumption that the Second Law must govern a naturalised Maxwell Demon. Thus the *sound* exorcism adds nothing of fundamental principle to the Second Law. It is at best a picturesque way to tease out some of its consequences. In so far as information theory provides a *profound* exorcism, it must do so by invoking hitherto neglected and novel physical principles. But what the exorcism literature has failed to present and, we believe, cannot present are compelling, independent reasons for accepting the new physical principles that connect information and thermodynamic entropy. In the present Maxwell Demon literature there seems to be no consensus as to which horn ought to be accepted; indeed it proceeds as if both could be accepted at once.

The first part of our paper will be devoted to understanding how a literature came to be that is so devoted to precarious constructions. Our answer will be provided in a historical review that traces how the role of Maxwell's Demon has changed since its earliest appearance. To begin, we shall see in Section 2 that Maxwell conceived of the Demon as a helpful spirit, assisting us to recognise most painlessly that the Second Law of thermodynamics can hold only with very high probability, apparently in the sense that there is a very small subclass of thermodynamic systems that assuredly reduce entropy. There may never have arisen an industry devoted to the Demon's exorcism had his function remained as Maxwell envisaged. However the Demon soon became entangled with another problem for the Second Law. After the turn of the century came the recognition that fluctuation phenomena could be observed in the laboratory. Brownian motion is the best known example. These phenomena, it was widely agreed as we shall see in Sections 3 and 4, constituted a microscopically visible violation of the Second Law of thermodynamics and the fear grew that these

microscopic violations might be convertible into macroscopic violations. They might not just permit momentary hesitations in entropy's rise; they might, under the right circumstances, turn back its course with inexorable certainty. All that was needed was some suitably constructed, adjunct device that could accumulate or amplify the fluctuations.

We shall see in Sections 5–9 how there arose a literature which sought to defeat this greater threat. Its principal strategy was to weaken the Second Law of thermodynamics in such a way that the law would continue to hold even allowing for fluctuation phenomena. In the work of Smoluchowski in the 1910s and Szilard in the 1920s, it became apparent that intelligent intervention by a Maxwell-type demonic intelligence was just the sort of adjunct needed to amplify fluctuation phenomena into macroscopic violations of the Second Law. This was a threat they worked hard to parry. Szilard's (1929) seminal paper on Maxwell's Demon came as the culmination of this tradition. He followed a hesitant Smoluchowski in presuming that any Demon must be a physical system itself subject to the Second Law of thermodynamics. From this assumption he inferred that a statistical form of the Second Law could be preserved as valid if we posited that there is a suitably large entropy cost associated with the Demon's information processing. With Szilard's work, the transmogrification of the Demon was all but complete. By the 1950s, the Demon was no longer a helpful spirit assisting us quite properly in mapping out the domain of validity of the Second Law. His exorcism had become an imperative in its own right, now quite divorced from the original threat of fluctuation phenomena. The Second Law had to be protected from him and the protection was to come from the principle of an entropy cost in information processing by the Demon. The latter principle became the foundation of the exorcism and it commonly became quite unclear whether that principle was merely a consequence of the supposition of the Second Law or an independent postulate. What also remained unclear was why any special effort would be needed to exorcise the Demon. In so far as he was subject to the weakened laws of thermodynamics, he could do nothing to violate them, so that his exorcism was a foregone conclusion—although the details might comprise an entertaining exercise. In so far as he lay beyond these laws, no exorcism was possible.

We conclude in Section 10 by listing a number of unresolved tensions in the literature that are warning signs of unclarities about both the aims and methods of exorcism. In an appendix we discuss Popper's attempt to save the phenomenological Second Law by reformulating it in a weakened form. We indicate how the recent work of Zhang and Zhang (1992) can be used to show that the standard version of classical statistical mechanics can underwrite a version of the Second Law very much in the spirit of Popper's formulation. But we also foreshadow a detailed example (Part II, Appendix 2) that displays a macroscopic dynamics that fails to conserve phase volume but is nonetheless energy conserving and time reversal invariant and can violate Popper's weakened Second Law. Attention to microdynamics is thus crucial to explaining and delimiting the validity of the Second Law.

In the second, forthcoming part of the paper we will focus on the main thrust of the recent Maxwell Demon literature in the latter half of this century. It is built around the idea that the key to exorcism is the entropy cost of information processing. In Section 1 we pose our sound vs profound dilemma designed to establish that the benefits of the information approach are largely illusory. In Section 2 we apply the dilemma to various attempts at exorcism that appeal either to the Szilard principle, which specifies a minimal entropy cost for the acquisition of a bit of information, or Landauer's principle, which posits a minimal entropy cost for erasing a bit of information. All of these attempts, we claim, are impaled on one or the other of the horns of our dilemma. Section 3 offers some brief comments on the notion, which appears sporadically in the literature, that quantum mechanics holds the key to exorcism. Concluding remarks are offered in Section 4.

2. The Birth and Childhood of Maxwell's Demon[2]

The Demon made its first appearance in a letter of 11 December 1867 from Maxwell to Tait, although Maxwell himself refers to the creature not as a Demon but as a 'very observant and neat-fingered being'.[3] We are asked to imagine a container of gas separated into two sections, A and B, by a diaphragm.

> Now conceive a finite being who knows the paths and velocities of all the molecules by simple inspection but who can do no work except open and close a hole in the diaphragm by means of a slide without mass. Let him first observe the molecules in A and when he sees one coming the square of whose velocity is less than the mean sq. vel. of the molecules in B let him open the hole and let it go into B. Next let him watch for a molecule of B, the square of whose velocity is greater than the mean sq. vel. in A, and when it comes to the hole let him draw the slide and let it go into A, keeping the slide shut for all other molecules (Knott, 1911, p. 214).[4]

As a result of these operations, 'the hot system has got hotter and the cold system colder and yet no work has been done, only the intelligence of a very observant and neat-fingered being has been employed'. The upshot is that the Demon has produced a violation of the Second Law of thermodynamics (see Appendix 1).

[2] Readers who want a more detailed history of Maxwell's Demon may consult the review articles by Collier (1990), Daub (1970), and Heimann (1970). Much valuable information is also to be found in Brush (1976) and Leff and Rex (1990).

[3] It was William Thomson (later Lord Kelvin) who christened these creatures Demons. In an undated letter to Tait, Maxwell wrote: 'Concerning Demons: 1. Who gave them this name? Thomson' (Knott, 1911, p. 214). See also Thomson (1874).

[4] According to Maxwell's velocity distribution law, the temperature of a gas in equilibrium is proportional to the average value of the square of the molecular velocities.

Maxwell realised that the Second Law could fall prey to what was later called Loschmidt's reversibility objection;[5] namely, if a system whose microdynamics is governed by deterministic time reversible laws exhibits thermodynamic behaviour, then anti-thermodynamic behaviour can be produced by reversing the velocities of microconstituents. But as Maxwell wrote to John William Strutt (later Lord Rayleigh) on 6 December 1870, 'the possibility of executing this experiment is doubtful'. Besides, he thought that a violation could be more easily produced simply by employing 'a doorkeeper, very intelligent and exceedingly quick, with microscopic eyes'. The moral drawn was that '[t]he 2nd law of thermodynamics has the same degree of truth as the statement that if you throw a tumblerful of water into the sea, you cannot get the same tumblerful of water out again' (Strutt, 1968, p. 47).

The allied moral drawn by Maxwell was that the 'learned Germans', i.e. Boltzmann and company, were deluding themselves in trying to derive the Second Law from mechanics:

> But it is rare sport to see those learned Germans contending for the priority in the discovery that the second law of $\theta\Delta^{cs}$ is the Hamiltonsche Princip. [...] The Hamiltonsche Princip, the while, soars along in a region unvexed by statistical considerations while the German Icari flap their waxen wings in nephelococcygia, amid those cloudy forms which the ignorance and finitude of human science have invested with the incommunicable attributes of the invisible Queen of Heaven (Knott, 1911, pp. 115–116).

Many painful years of struggle were in store for Boltzmann before he was forced to concede the point.

A third related moral was that it is possible to reverse the dissipation of energy, and more fundamentally, that the distinction between dissipated energy and energy available for work depends on our state of knowledge. In his 1878 *Encyclopedia Britannica* article on 'Diffusion' Maxwell (1952, p. 646) wrote:

> It follows [...] that the idea of dissipation of energy depends on the extent of our knowledge. Available energy is the energy which we can direct into any desired channel. Dissipated energy is energy which we cannot lay hold of and direct at our pleasure, such as the energy of the confused agitation of molecules which we call heat. Now, confusion, like the correlative term order, is not a property of material things themselves, but only in relation to the mind which perceives them. [...] Similarly the notion of dissipated energy would not occur to a being who could not turn any of the energies of nature to his own account, or to one who could trace the motion of every molecule and seize it at the right moment. It is only to a being in the intermediate stage, who can lay hold of some forms of energy while others elude his grasp, that energy appears to be passing inevitably from the available to the dissipated state.

[5] Brush (1976, p. 602) indicates that Tait, or possibly William Thomson, deserves priority for formulating the reversibility objection.

It is hard to disagree with the morals Maxwell drew from the Demon, although later developments will present many obstacles for the arguments that lead from the Demon to these morals.[6] But the bald statement of these morals leaves matters in an unsatisfactory state. Granted, the Second Law has only statistical validity. But what exactly is the nature and source of that validity? Granted, the assumption of a deterministic dynamics promotes the idea that randomness cannot be a property of material things themselves but is merely a reflection of our ignorance of the values of relevant variables. Nevertheless, does not the fact that the overwhelmingly vast majority of macroscopic systems behave according to the dictates of thermodynamics suggest that the probabilities are not purely subjective? There are two ways to approach these questions. One is to work very hard on statistical mechanics. The other is to wrestle with the Demon, looking for ways to patch, hedge, or protect the Second Law. The results of the first approach have not been entirely satisfactory, as attested by the contentious and controversial nature of the foundations of statistical mechanics.[7] Nevertheless, we contend (but will not argue here) that this approach has yielded much valuable understanding. On the other hand, we contend (and will argue here) that while the second approach has yielded much amusement, it has produced little in the way of enlightenment.

3. The Threat of Fluctuations: the Demon Within

Only a little reflection is needed to realise that the Demon is made possible by the form of the velocity distribution law which Maxwell announced in 1860 for a gas in equilibrium (see Maxwell, 1860). For no matter how sharply peaked this bell-shaped distribution is, it has tails which extend infinitely far. The Demon operates by sorting out molecules whose velocities lie sufficiently far out in the tails from those whose velocities lie in the main hump.

We need not await the intervention of Maxwell's fictitious Demon to realise threats to the Second Law. For thermal systems carry their own natural Demons, derived directly from the probabilistic distribution of their properties. A gas left to itself and undisturbed by an external agency will spontaneously exhibit fluctuations away from equilibrium, and such fluctuations can produce violations of the Second Law. The origin of this threat to the Second Law lies deep within the microdynamics of thermodynamic systems. Suppose that the gas

[6] Maxwell does not make clear whether he conceives the Demon as an entity outside or within thermodynamics. In the former case, Maxwell's argument would become opaque. Why should we care if a gas coupled with a non-thermodynamic Demon disobeys the Second Law? By supposition, the Demon is already beyond the reach of the Second Law. In the latter case we would probably conceive of the Demon as composed of many molecules in some kind of thermal equilibrium. Later developments will show that we cannot automatically presume that such a Demon can perform as intended and successfully sort fast and slow molecules.

[7] For a recent overview, see Sklar (1993).

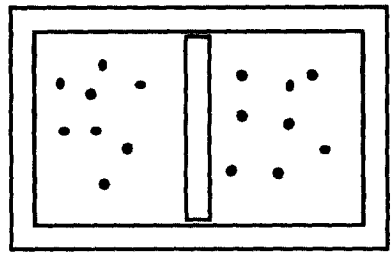

Fig. 1. *Poincaré machine.*

or fluid in question can be treated as a deterministic dynamical system whose phase volume is preserved under the dynamical flow (see Appendix 2). For any such system we have Poincaré's recurrence theorem: the system returns, almost surely, arbitrarily closely to its starting state (see Appendix 2 for a precise statement). It follows that if the system is ever in a low entropy state, it will almost surely eventually return to a low entropy state. For ergodic systems, the antecedent of the previous statement can be affirmed since almost every phase orbit will pass arbitrarily near to any chosen phase point (see Appendix 2 for a precise statement). In particular, density fluctuations of however great a magnitude will occur spontaneously without the need for a clever Demon to intervene.

To illustrate how fluctuations can violate the Second Law without the help of a Demon, consider what we will call a Poincaré machine (see Fig. 1). The working substance is a gas of hard spheres. This gas is divided equally between the two chambers of a cylinder with a movable piston such that each half is initially at the same temperature, pressure and volume. The system is thermally isolated from its surroundings. On average over time, the gas will exert equal pressures on either side of the piston. But, at any moment, the pressure will fluctuate away from its mean value. If we ask after the system's state at some particular moment after its initial set up, then very probably the fluctuations are small. There is, however, an extremely small probability of fluctuations so large that the gas has been spontaneously compressed to, say, half its volume in one side of the cylinder and correspondingly expanded in the other.[8] The system is now in a lower entropy state than its initial state. While improbable, this spontaneous compression is not impossible, Thus this machine embodies a violation of the Second Law of thermodynamics, for that law prohibits any reduction in the entropy of an isolated system.

If one is innocent of how the Maxwell's Demon literature develops, one will be tempted to add extra components to the Poincaré machine. We will refrain from such additions since they all prove to open new lines of debate that cloud the simple fact that this machine violates the Second Law. Perhaps one might like to

[8] The minuteness of this probability exceeds hyperbole. To see such fluctuations in a typical macroscopic gas of 10^{24} molecules, one would need to allow times greater than the current estimates of the age of the universe.

add a mechanism that would lock the piston in its compressed state. But how is such a mechanism to operate? Does it dissipate energy and thus perhaps supply a compensating increase in entropy? Might we be able to add machinery that could convert this spontaneous reduction in entropy into usable work? But then we may need to know when to lock the piston, where the piston is and how to tap usefully into the pressure differential no matter the side on which it may lie. All these requirements may involve the dissipation of work and thus perhaps an increase in entropy. They need not distract us yet.

These complications do, however, allow us to disentangle two senses in which the Second Law may be violated:

> (Straight Violation): it is possible for the entropy of an isolated system to decrease.
> (Embellished Violation): these decreases in entropy can be exploited reliably to provide work.

4. The Demon Within Made Manifest

At the turn of the century, the threat of these types of violations of the Second Law was a very abstract one. Indeed the threat depended on the presumption of the atomic theory of matter. The famous atomic debates were still unresolved. The anti-atomism of Ostwald and Mach was driven by a positivist demand for conceptual parsimony in physical theory. Thus their debates with atomists such as Boltzmann presumed that no direct experimental verification of the existence of atoms would emerge. With the new century, that assumption failed. Its failure derived from ever deepening investigations of a multitude of fluctuation phenomena by Einstein, Smoluchowski, Perrin and others. Most famously, the phenomenon of Brownian motion came to be accepted as a directly observable manifestation of molecular motions. Almost as well known were observable consequences of density fluctuations in fluids. In fluids near their critical states, these fluctuations are sufficiently large to produce an opalescence observable under the microscope. That air is made of molecules and is not a continuous fluid gives air a granular character akin to density fluctuations. Indeed this granularity proved to be responsible for the blueness of the sky since it scattered blue frequencies out of sunlight.[9] The pace of change was very rapid. Ostwald's famous retraction of his opposition to atomism had come in 1909.[10] Smoluchowski (1914, p. 80), reflecting a little over a decade into the new century, already found anti-atomism fading into history.[11]

[9] For further discussion see Brush (1983, Chap. II), Brush (1976, Vol. 2, Chaps 14–15), Nye (1972), Perrin (1921).

[10] See Brush (1983, Vol. 2, p. 699).

[11] He wrote: 'Today it is no longer easy for us to think back on the agreement that prevailed towards the end of the last century. Then indeed the scientific leaders of Germany and France — with few exceptions — were convinced that the atomic kinetic theory had played out its role'. And then (p. 90): 'Today the situation is the reverse of twenty years ago. Atomism is generally accepted as the foundation of current physics'.

With the new understanding of fluctuation phenomena and Brownian motion, physicists no longer needed to invent a fictitious Demon or wait eons with a Poincaré machine to see violations of the Second Law—then commonly known as Carnot's Principle. Such a spectacle was as close as a microscope trained on pollen grains, exhibiting the endless dance of Brownian motion. Or so it was reported by Poincaré (1904, p. 287):

> [...] we see under our eyes now motion transformed into heat by friction, now heat changed inversely into motion, and that without loss since the movement lasts forever. This is the contrary of the principle of Carnot.

Einstein's (1905, p. 549) judgment in his famous 1905 paper on Brownian motion was the same:

> If it is really possible to observe the motion to be discussed here, along with the laws it is expected to obey, then classical thermodynamics can no longer be viewed as strictly valid even for microscopically distinguishable spaces and an exact determination of the real size of atoms becomes possible.

These violations of the Second Law are the 'straight violations' we described earlier in the context of the Poincaré machine. They are the same violations—just made smaller and consequently much more frequent. But no usable work is recovered from them; these fluctuations lift no weights for us and wind no springs. Might it be possible to convert these violations into the 'embellished violations'? That is, might it be possible to harness fluctuations, such as those of Brownian motion, in such a way that we can extract usable work from them? Could we conceive such a machine?

Gouy in 1888 had already imagined just such a machine based on Brownian motion: a ratchet wheel would be wound by a thread attached to particles undergoing Brownian motion, thus allowing the conversion of the thermal energy of the suspending fluid into work drawn from the wheel.[12] Responding to the rapid pace of advance of atomic theory in the early part of the century Svedberg (1907) continued with further proposals. The basic idea was the same; one somehow needed to couple a fluctuating system to an energy store so its thermal energy could be extracted. But he looked to colloids, his area of experimental expertise, and to couplings more sophisticated than a simple thread. He offered two devices, each described with elaborate concern to ensure that no unforeseen contingency might compromise their operation.[13]

Svedberg's first engine exploited the result that the particles of a colloid carry an electric charge. Since these particles' thermal energy is manifest in random

[12] For an account, see Laymon (1991, Section IV). Poincaré (1905, p. 179) in reflecting on Brownian motion and Gouy's work made the connection to Maxwell's Demon: 'One can almost see Maxwell's demon at work', he exclaimed.

[13] We learn in the concluding pages of the paper that Svedberg takes as unproven the existence of molecules 'as discrete particles' (p. 457). His constructions are based on the presumption that the thermal energy of colloids resides in their kinetic energy.

fluctuations in their motion, they were accelerating and, it now followed from electrodynamic theory, radiating their thermal energy in electromagnetic waves. In Svedberg's device, the colloid in a water jacket was surrounded by a vacuum layer and then a lead casing whose size was carefully determined so that the radiated electromagnetic waves would be fully absorbed. Thus the thermal energy of the colloidal particles would be transferred to the lead casing, spontaneously cooling the colloid and heating the lead casing.[14] The resulting temperature difference could then be exploited in the usual ways to generate work and restoring the original thermal equilibrium. This cycle could be repeated at will, converting the thermal energy of the colloid into work. It would thereby constitute a perpetual motion machine of the second kind: one whose net effect is solely the conversion of heat into work.

Svedberg's second engine was a slight modification of the first. In the second, the thermal motions of a colloid are used to set in motion a small spiral of conducting wire, with the spiral and colloid concentration sized to maximise such motion. The entire apparatus sits in an external magnetic field, so that the result of the spiral's motion is an induced electric current. Thus the thermal energy of the colloid is transferred to the thermal motion of the spiral and then into the energy of the induced electric current. The colloid, spiral and a water jacket are in a silvered container in a vacuum. The wires run out of this container through the vacuum to an enclosing water jacket which is in turn bounded by another vacuum layer with all walls silvered. The wires connect to resistance coils in the outer water layer where the energy of the induced current reverts to heat. The overall effect is a transfer of heat energy of the inner colloid to the outer water jacket; the colloid cools and the jacket heats spontaneously. The resulting temperature difference can then by utilised to generate work. We once again have a perpetual motion machine of the second kind.

5. Smoluchowski and the Rescue of the Second Law of Thermodynamics

While accepting the heuristic power of the atomic doctrine, anti-atomists such as Ostwald had proposed that the doctrine would ultimately be superseded by a purely phenomenological thermodynamics. They were wrong; the triumph of atomism was now threatening to overturn thermodynamics itself. The only real question was how complete would be the defeat of thermodynamics. This was the question that Marian Smoluchowski (1912) investigated in a conference paper that was provocatively entitled 'Experimentally Demonstrable Molecular Phenomena that Contradict Ordinary Thermodynamics'. The bulk of Smoluchowski's paper was given over to an account of the experimentally

[14] Svedberg went to further pains to ensure the expected operation. Recognising that a warmed lead casing would reradiate heat back to the cooler colloid, he interposed a water jacket between the two in the middle of the vacuum layer for the purpose of absorbing this reradiated heat. A sufficiently massive water jacket, he reported, would prevent the heat radiation of the lead reaching the colloid.

demonstrable fluctuation phenomena that had played so decisive a role in this triumph of atomism. Its concluding Part III, however, turned to the question of whether these fluctuation phenomena would permit the construction of a perpetual motion machine of the second kind. Smoluchowski (p. 1078) cited contemporary opinion, including Svedberg (1907) that such a construction would be possible. He continued to describe some ways that fluctuation phenomena could apparently be used to build perpetual motion machines of the second kind.

The first example was a microscopic hole in a dividing wall that would pass emulsion particles only in one direction since the hole was equipped with a one-way valve or a ring of little elastic hairs. The effect would be a sustained difference of pressure over the wall that could be continuously exploited to produce work. Or one might envisage a toothed wheel with a catch that would allow it to turn in one direction only. Fluctuations in pressure would enable this wheel to wind a torsion spring. In all these cases, the thermal energy of fluctuations would be converted completely into work without a corresponding discharge of heat to a cool reservoir—in violation of the Second Law of thermodynamics.

Given Smoluchowski's bold assertion that fluctuation phenomena do contradict ordinary thermodynamics, we might well expect Smoluchowski to affirm that the devices would violate the Second Law. What else could he mean in the title of his paper when he asserted that these phenomena contradict thermodynamics? The theory, as then formulated, had two laws and some background assumptions. Fluctuation phenomena do not contradict the First Law, the law of conservation of energy. Poincaré had already allowed that they do violate the Second and Smoluchowski opened his closing discussion with the statement of his 'question of principle': 'How do things stand with the Second Law?' (p. 1078).

But Smoluchowski balked (p. 1078):

> In spite of all this I do not believe that in this way we obtain a perpetual motion machine that continuously *produces work*; for right in the constitution of the one-way valve, the ratchet, there is an impossibility of principle, in so far as the considerations of statistical mechanics are correct [Smoluchowski's emphasis].

His worry was soon clear. The endless agitation of Brownian motion amounted to a direct violation of the Second Law—the 'straight violation' above. But Smoluchowski was unable to see success in any of the proposed devices intended to harness these motions and effect the unlimited conversion of heat energy into work—the 'embellished violation'. In the case of the one-way valve, for example, the valve's flapper would need to be restrained by a spring of sufficiently weak force to enable the one-way passage of particles. But if this spring is weak, then the flapper would attain thermal energy itself and its resulting motion would cause it to open and allow the reverse passage of particles. He concluded (p. 1078):

> Therefore a perpetual motion machine would only be possible if one could construct a one-way valve of quite another kind, without a tendency to molecular fluctuations, and for that today we see no possibility.

This failure of the one-way valve has been much cited in later literature. For Smoluchowski it was but an example. He urged that similar considerations would defeat any attempt to construct a perpetual motion machine that continuously produces work from heat on the basis of fluctuation phenomena. As an illustration, he gave another example of such a device and its defeat. He imagined a charged condenser that used air as its dielectric, and one of whose plates is connected to ground by a conductor. Its supposed anti-entropic behaviour depended on fluctuations in air pressure. These cause fluctuations in the capacitance of the condenser which are in turn reflected in alternating currents in the conductor. Because of it resistance, these currents heat the conductor, whose temperature will rise above that of its surroundings. Overall, the device draws heat energy from air—the source of the pressure fluctuations—and converts it to heat at a higher temperature, in violation of the Second Law. Once again, further consideration of fluctuations defeat this device's demonic behaviour. There are, in any case, fluctuations in the potential of a condenser.[15] These fluctuations produce mechanical effects that move the surrounding air, heating it at the expense of the energy of the conductor. The combined effect is simply an interchange of energy between the air and the conductor. Exactly this type of compensating interaction would defeat another simpler perpetual motion machine. Smoluchowski mentioned attempts to use the energy of the Brownian motion of particles converted by friction to thermal energy through thread coupling as a means of heating just one part of a fluid. He concluded (p. 1079):

> Therefore, in spite of our current knowledge of those fluctuations, it does not appear to be possible as a result to bring about by any kind of device of this type a continual concentration of heat in a medium in equilibrium; and it seems that the construction at present of a perpetual motion machine that produces continuous work is to be excluded, not through purely technical difficulties, but through matters of principle.

Smoluchowski did not explicitly address Svedberg's engines beyond the mere citation of Svedberg's paper mentioned above. However this conclusion clearly applied to Svedberg's devices.[16] Svedberg, it must be presumed, had neglected some further processes that would counteract and defeat the devices' operation. In the case of Svedberg's first engine, there is a familiar mechanism that returns heat from the lead casing to the colloid. The colloid radiates its thermal energy to an electromagnetic radiation field that in turn transmits the heat to the lead casing. This radiation field would itself be a system of heat radiation with its

[15] Fluctuations in the potential of a capacitor had already been investigated by Einstein (1907).

[16] Recent historical research has found that Svedberg and Smoluchowski were corresponding in 1907 concerning their common interest in Brownian motion, although this correspondence seems to throw no direct light on their disagreement over the validity of the Second Law. See Sredniawa (1991a, 1991b). For appreciation of Smoluchowski's work, see Ingarden (1986).

own temperature and fluctuations. These fluctuations would transmit a thermal agitation back to the colloid. If the colloid temperature drops below that of the radiation field, there would be a net transfer of heat from the radiation to the colloid; and since the lead casing and radiation field would be exchanging heat energy by the same mechanism, there would be a similar compensating transfer of heat from the lead casing to the cooled radiation field. The overall effect would be thermal equilibrium between colloid, radiation field and lead casing and not an uncontrolled cooling of the colloid. Any cooling of one component would be immediately compensated by energy transfers from the others.

This type of dynamic equilibrium between thermally agitated bodies and a field of heat radiation had an important place in the physics literature of the early part of the century. Einstein's (1909a, pp. 189–190; 1909b, pp. 496–497) now famous thought experiment that established wave-particle duality for quantised electromagnetic fields considered a thermally agitated mirror that transferred its thermal energy to the radiation field by radiation damping; and in turn, fluctuations in the radiation field returned that thermal energy to the mirror. This return mechanism allowed thermal equilibrium to be established and prevented by the uncontrolled cooling of the mirror.[17] Just prior to 1900, Max Planck had prepared the ground for early work in the old quantum theory of black body radiation with his detailed, classical analysis of the dynamic equilibrium between an oscillating charge and the electromagnetic field (see Kuhn, 1978, Chapter III).

Smoluchowski's conclusion—that fluctuation phenomena could not be exploited to produce a perpetual motion machine—was deemed by him to be of the deepest significance. He decided to make it the basis of his rescue of thermodynamics. While fluctuation phenomena were a threat to the theory, it could be protected by a modification of the Second Law. He wrote (p. 1079):

> Molecular fluctuation phenomena today give us no reason to overturn completely the Second Law of thermodynamics, as we have so many other dogmas of physics. They compel us only to a weakened formulation, if we demand universal validity for the laws of thermodynamics. Perhaps an apparently quite minor extension of the wording suffices, in so far as one says: 'There can be no automatic device that would produce *continuously* usable work at the expense of the lowest temperature'. The brief version [of the Second Law] 'impossibility of a perpetual motion machine of the second kind' is even sufficient, for one has transferred the difficulty into the explication of the latter concept [Smoluchowski's emphasis].

The crucial modification lies in the word Smoluchowski emphasised: *continuously*. While some device may convert heat entirely to work, the success would be

[17] See Norton (1991) for further discussion.

temporary.[18] In the long run, it would not be sustained. In the limit of infinite time, the rate of conversion of heat of the surroundings into work by such devices would drop to zero. Smoluchowski (1912, p. 1079; 1914, p. 117) gave this result formal expression in the equation

$$\lim_{t \to \infty} \frac{A}{t} = 0$$

where A is the work converted to heat in time t. We shall call this weakened version of the Second Law the 'time averaged Second Law of thermodynamics'.

6. Smoluchowski and the Naturalisation of Maxwell's Demon

The devices considered by Gouy, Svedberg and Smoluchowski were physical constructs designed to bring about the same effect as Maxwell's Demon. They were conceived as mechanised surrogates of the Demon—as Smoluchowski explained (1912, p. 1078) in introducing them:

> Indeed we need no Maxwell demon at all, for instead of him we can employ an automatic device, for observable fluctuations of pressure and motion phenomena appear comfortably already in the domain of the microscopic, visible parts of space (indeed even in that visible with the naked eye) [...].

But what if we consider the intervention of an intelligent being in the operation of these devices? In his later paper (1914), Smoluchowski repeated the viewpoints laid out in his 1912 lecture, including his time averaged form of the Second Law of thermodynamics. He also addressed the question of the intervention of intelligent beings.

His remarks came in the context of two further proposals for perpetual motion machines that exploit fluctuations. The first (§18, pp. 117–118) was a simple variant of Gouy's machine. Smoluchowski imagined a particle of gamboge—a substance used in the experimental investigation of Brownian motion—that is raised in a fluid against its own weight by Brownian motion. Smoluchowski proposed that the particle be coupled with a device that would permit only upward motion of the particle. He suggested devices such as a ratchet and spring-loaded pawl. Thus the particle would slowly but inexorably ascend, converting its thermal energy into the potential energy of height. The error in the analysis, Smoluchowski continued, was that fluctuations must also arise in the ratchet and pawl arrangement so that it would fail to enforce unidirectional motion.[19]

[18] Short term violations were even to be expected. 'One does not need any device at all', he wrote (p. 1079). 'One must only just wait until it happens by itself in accord with the laws of chance, that is, until a correspondingly great deviation from the normal state takes place'. This is our Poincaré machine.

[19] Smoluchowski's analysis is brief and somewhat cryptic. Presumably his view is similar to the lengthier treatment given in Feynman *et al.* (1963, §46-1, 2). The thermal motion of the pawl causes it to jump off the ratchet and no longer reliably prevents reverse motion.

The second device (§18, pp. 118–119) was based on a very simple notion that would prove of immense importance to the work of Szilard that follows. If two bodies are in thermal contact, even at equilibrium, fluctuations will lead to continual but slight imbalance in thermal energy between the two bodies. If they are separated, we arrest these fluctuations at whatever might be their present state. One body will be slightly colder, the other slightly warmer. Smoluchowski's proposal was to exploit this effect to build up a difference of temperature without the corresponding expenditure of work. He envisaged two bodies A and C that were to act as heat sinks and a third body B that would shuttle between them. They would all start at the same temperature. The following cycle would be repeated as shown in Fig. 2:

(i) B would be brought into contact with A.
(ii) When B's temperature fluctuated to one higher than C, B would then be separated from A, locking it in its higher temperature.
(iii) B would be brought into contact with C.
(iv) When B's temperature fluctuated to one lower that A, B would be separated from C. (Return to (i)).

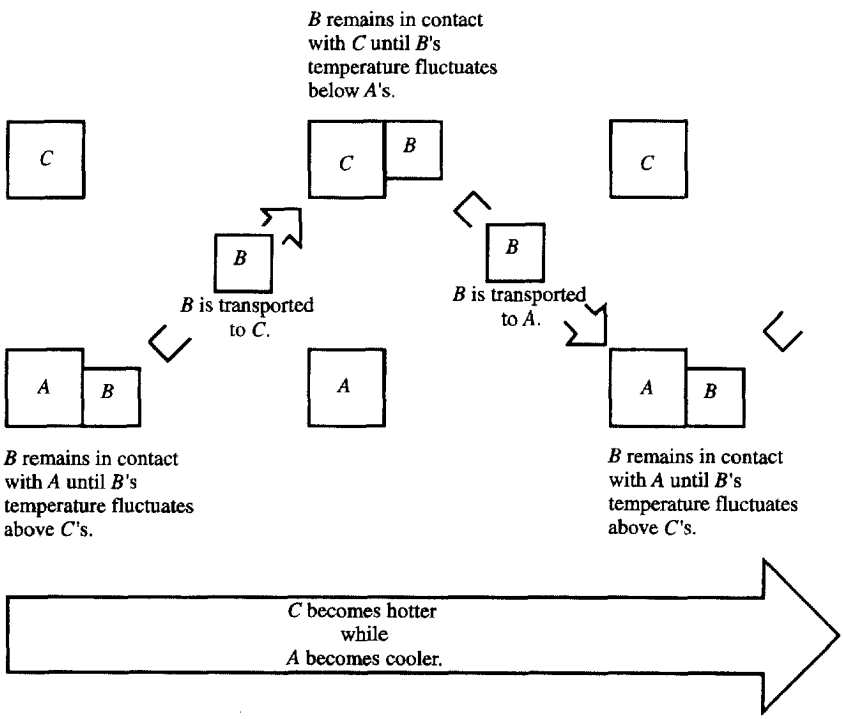

Fig. 2. *Smoluchowski's thermal fluctuation machine.*

The overall effect of the cycle is to draw heat from A and transport it to C, so that A cools as C heats. The resulting temperature difference could be used to generate work. Smoluchowski imagined that these steps would all be carried out by some mechanical device equipped with a thermometer to measure temperature. He presumed that the amount of work needed to operate this device could be made arbitrarily small in comparison to the work recoverable from the temperature difference between A and C.

As before, Smoluchowski declared that neglected fluctuation phenomena would defeat the operation of this automatic device. The added mechanical devices would in turn be subject to fluctuations. He concluded all too briefly that 'indeed—this is here the essential point—mechanical fluctuations are not coordinated with thermal [fluctuations]. Thereby automatic operation is rendered impossible' (p. 119).

Smoluchowski now turned in §20 (pp. 119–120) to the intervention of intelligent beings. While thermal fluctuations had defeated both proposals for automatically operated perpetual motion machines, they would succeed were they operated by intelligent beings. Such a being would just know when to move the shuttle B between bodies A and C in the thermal fluctuation engine. It could also enforce the elevation of the gamboge particle by raising a massless floor each time the gamboge particle happened to ascend. He continued:

> Therefore a perpetuum mobile is possible in case one considers, in accord with the customary methods of physics, the experimenting person as a kind of 'Deus ex machina' that is informed continually and exactly of the momentary state of nature and can set in motion or interrupt microscopic natural processes at any moment without expenditure of work. Thus he would not at all need to possess the capacity of a Maxwell Demon to intercept individual molecules, but he would definitely still be distinct from real living beings in the above points. For, the production of some physical effect through operation of the sensory and also motor nervous systems is always associated with an energy cost, independent of the fact that its entire existence is tied up with a continuous dissipation of the same.
>
> Therefore, in view of these circumstances, that real living beings could produce work continuously, or at least in a regular fashion, at the expense of the heat of the lowest temperature, appears truly doubtful, although our knowledge of living processes excludes a definite answer.

Smoluchowski had now described the strategy that would become standard for defeating Maxwell's Demon. He is to be naturalised: in so far as Maxwell's intelligent being could be realised as a real living being, that being must be subject to the ordinary restrictions of physical systems, including energy costs for effects they produce. While Smoluchowski does not say so explicitly, this escape surely assumes that, just as fluctuations in the operation of automatic mechanisms prevent their realising a perpetual motion machine, fluctuations within the living being's system would do the same for a naturalised Demon.

While Smoluchowski describes this escape, his endorsement of it is equivocal. He will not exclude the possibility that living beings are free of the energetic restrictions imposed by thermodynamics on ordinary processes. This was not

a concession to a straw man. For in the discussion following Smoluchowski's 1912 paper (p. 1080), Kaufmann has urged that an intelligent being could operate a perpetual motion machine of the second kind using a sliding shutter as a one-way valve and that the success of its operation 'could possibly be seen in the sense of the neovitalist conception as proof that the physico-chemical laws do not extend to the explanation of biological and psychological occurrences'. Smoluchowski's (1912, p. 1080) reply was equivocal, once again leaving unanswered the question of whether naturalisation defeats Maxwell's Demon:

> [...] there is certainly no doubt that an intelligent being to whom physical phenomena are transparent could bring about processes that contradict the Second Law. Indeed Maxwell has already proven this with his Demon. However intelligence extends beyond the boundaries of physics. On the other hand, it is not to be excluded 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.

7. Szilard and the Thermodynamic Analysis of Fluctuations

Smoluchowski's proposal for a weakened Second Law seems, at first blush, to be unworkable and to offer us no useful information. His time averaged form of the law no longer tells us what happens in any particular cycle; it tells us only about the accumulated behaviour of many cycles. That a weakened law still enables interesting consequences to be drawn was shown by Leo Szilard in a 1922 dissertation at the University of Berlin, whose results were published as Szilard (1925). Szilard's apparently impossible task was to show (as he put it in the abstract (p. 70) how '[...] purely phenomenological thermodynamic considerations can lead to an understanding of the laws governing fluctuation phenomena [...]'. The task seems impossible since traditional phenomenological thermodynamics has no place for fluctuations: a given system in thermal equilibrium at temperature T has a certain fixed energy $E(T)$. How can a theory based on that supposition yield results concerning fluctuations in the energy E? And how can the laws of phenomenological thermodynamics cohere with fluctuations when large fluctuations constitute direct violations of the laws?

The answers are, of course, that Szilard did a little less than his abstract promised. The existence of fluctuations is postulated, not inferred, and it is supposed that the admissible energies of a system are governed by a probabilistic law. If a body gains heat Q_i as a result of some cyclic process, the probability that this heat lies in Q to $Q+dQ$ is $W_i(Q)dQ$, where the functional form of W_i is to be determined. 'The claim that some probabilistic law applies', noted Szilard in a footnote (p. 71) 'is an assumption on which this work is based'. Moreover the version of the Second Law invoked is not the standard version but a version weakened analogously to Smoluchowski's version; it prohibits, in effect, cyclic processes whose expected outcome—the averaged behaviour in the limit of infinitely many repeated cycles—would violate the ordinary version of the

Second Law. To be precise, Szilard imagined a cyclic process that exchanged heat with many heat reservoirs. If the mean value of the heat absorbed by all reservoirs but one is zero (the Q_i below) for the cycle, then the mean value of the heat absorbed by the remaining reservoir (Q below) cannot be negative. He wrote[20] (p. 71):

> We now postulate: for all cyclic processes the amount of heat taken up on average by one reservoir must satisfy
>
> $$\bar{Q} = \int_{-\infty}^{\infty} QW(Q)dQ \geqslant 0 \qquad [(1)]$$
>
> if the other reservoirs on average take up no heat, so that for them
>
> $$\bar{Q}_i = \int_{-\infty}^{\infty} QW_i(Q)dQ = 0. \qquad [(2)]$$

We shall call this Szilard's 'statistical form of the Second Law of thermodynamics'. One quickly sees that it is properly named. If it is violated, then on average in each cycle some quantity of heat will be lost by the reservoirs. Conservation of energy dictates that this heat must have been converted to work, so that very many probabilistically independent repetitions of the cycle will very probably produce an unlimited conversion of heat into work.[21]

We refrain from further analysis of Szilard's paper since it has been analysed in revealing detail by Martin (1996).[22] For our purposes it will be sufficient to give a flavour of Szilard's work by applying it to the simple example of Smoluchowski's thermal fluctuation machine described above. While there is no mention of this machine in Szilard's paper, we can be sure that Szilard knew of this proposal for his Szilard (1929) cites Smoluchowski (1914) and gives the same quote from it as we have in the last section.[23] Szilard (1929) was written about

[20] We translate here directly from the German since the English translation (p. 71) is misleading. We also correct the first equation which reads $\bar{Q} \geqslant \int_{-\infty}^{\infty} QW(Q)dQ \geqslant 0$ in the original.

[21] That Szilard overstated his project may explain why, or Szilard's (1978), pp. 9–11) report, Einstein was first skeptical of the project but then quickly convinced: 'He [Einstein] said, "Well, what have you been doing?" And I told him what I had done. And Einstein said "That's impossible. This is something that cannot be done". And I said,"Well yes, but I did it". So he said, "How did you do it?" It didn't take him five or ten minutes to see, and he liked this very much'. Presumably Einstein rapidly saw past Szilard's hyperbole and that the modest reality of his project was quite sound. Interestingly, in a paper to which Szilard (p. 89) draws attention, Einstein (1914) had also used thermodynamic reasoning to infer to the laws governing the probabilistic behaviour of systems at the molecular level. Einstein's methods were quite different from Szilard's, however. Einstein ingeniously exploited an analogy between the entropy of mixing of different chemical species and the entropy of a system of molecules with different energies.

[22] We would also like to thank Chris Martin for helpful discussion of Szilard's work.

[23] There is an unfortunate error in the translation of Szilard (1929). The English version (Collected Works, p. 121; Leff and Rex, 1990, p. 125) closes the quotation marks too early, so that the major part of the quote from Smoluchowski in presented as Szilard's own text. The original German text (*Collected Works*, pp. 104–105) is correct, although its closing sentence ('Doch führen die [...]') actually is the introductory sentence of a new section, §21.

six months after Szilard (1925).[24] Since they deal with closely related material, it is likely that Szilard already knew of Smoluchowski's work at the time of writing Szilard (1925). Comparing Smoluchowski's proposal with Szilard's system, it seems very likely that Szilard's analysis arose as a response and elaboration of Smoluchowski's analysis.

First we recall Szilard's notion of a normal distribution. Assume that some body has come to thermal equilibrium with an infinite heat reservoir at temperature T. Now assume that this contact is broken. In ordinary thermodynamics, the body will now have a definite energy content dependent on its constitution. In Szilard's probabilistic scheme, it may have many different energies according to a characteristic probability law.[25] This Szilard calls the 'normal distribution at temperature T'. According to it, in Szilard's notation, the probability of an energy between u and $u + du$ is

$$W^*(u; T)du. \tag{3}$$

The principal burden of Szilard's work is to show that this normal distribution has the form

$$W^*(u; T) = C(T)g(u)\exp \varphi(T)u, \tag{4}$$

where $g(u)$ is a weight function characteristic of the body, $\varphi(T)$ is a universal function of T, the same for all bodies, and $C(T)$ a normalisation constant. If we set $\varphi(T) = -1/kT$, for k Boltzmann's constant, readers will immediately recognise this W^* as a form of the Maxwell–Boltzmann distribution—a core result of statistical mechanics. With this distribution in hand, Szilard will have no problems inferring many further results in the statistical domain, including the fluctuations entailed by this distribution.

Szilard's strategy for deriving this result is to imagine cyclic operations that involve bodies shuttling energy between heat reservoirs. By sufficiently ingenious selection of these cycles, the requirement that their outcomes agree with his statistical version of the Second Law places strong restrictions on the functional form of the normal distribution and yields the result above. Smoluchowski's thermal fluctuation machine operates on a very much simpler cycle that places few restrictions on W^* but can illustrate Szilard's methods. Imagine in particular that the machine cycle is executed without the problematic conditions of (ii) and (iv). That is, the shuttle B proceeds from A to C and back without any test for whether its thermal energy has fluctuated to favourable levels at each separation. It proceeds with whatever energy chance fluctuations give it. So that there is some interest in the cycle, imagine that C is initially hotter than A. We also assume that both A and C are infinite heat sinks. Then, on average, in

[24] According to his *Collected Works*, p. 32.

[25] Presumably we are to imagine that the equilibrium is dynamic, with the energy ebbing and flowing between reservoir and object. Removing the object simply halts this exchange, fixing the body's energy at whatever its current value may be. Presumably we are to imagine this only, since such dynamics form no part of Szilard's theory.

accord with the statistical form of the Second Law of thermodynamics, we expect that the effect of each cycle will be to transfer some amount of heat from the hotter body C to the cooler body A.

Szilard's ingenious idea was to translate such expectations into mathematical conditions on the distribution W^*. To see how this occurs, assume the hotter body C has temperature T_h and the cooler body A has temperature T_c. After step (i), the shuttle will have an energy u_c, distributed according to $W^*(u_c, T_c)$. After step (iii), that energy will have become u_h, distributed according to $W^*(u_h, T_h)$. Thus the average effect of many repeated cycles is to transfer energy[26] $\overline{u_h - u_c} = \overline{u_h} - \overline{u_c}$ from the hot reservoir to the cold in each cycle, where the expectation is given as

$$\bar{u} = \int_0^\infty uW^*(u; T)du. \tag{5}$$

The statistical form of the Second Law requires that this process should not, on average, transfer heat from cold to hot; that is, it requires

$$\overline{u_h - u_c} = \overline{u_h} - \overline{u_c} \geqslant 0. \tag{6}$$

This condition in turn places constraints on the functional form of W^*. They are not powerful — but also not unimportant. They require that W^* be such that $\bar{u}$ is an everywhere non-decreasing function of T.

This simple cycle illustrates Szilard's project.[27] It is also sufficient to reveal a potentially fatal defect. If we imagine the cycle operated as Smoluchowski first suggested, then its overall effect would be no constraint on the distribution W^* but a violation of the statistical form of the Second Law. We might expect Szilard to exclude such a possibility by recalling Smoluchowski's well rehearsed escape. Indeed he takes up exactly the consideration to which Smoluchowski (1914) proceeded after his proposal of the thermal fluctuation machine. To operate anti-entropically, these types of cycles require some sort of intelligent intervention, able to decide, for example, whether the temperature of the shuttle exceeds some nominated bound. In a footnote, Szilard promises to treat this problem of intelligent intervention elsewhere; for the present paper he will assume all actions are mechanical — tacitly assuming that this type of intelligent action is beyond a machine. The footnote foreshadows Szilard's famous 1929 paper:[28]

> It is plausible to express an objection that originates from Maxwell against the above form of the Second Law, which also claims strict validity in the face of fluctuation phenomena:

[26] Notice that this equality tacitly assumes that the distributions of u_h and u_c are independent.

[27] Szilard's work was not a dead end. His statistical approach to thermodynamics has been carried on by such researchers as Mandelbrot (1964).

[28] This translation is a version of the English translation (p. 73) corrected slightly against the original German (pp. 38–39).

> If we had some demon at our disposal who could accurately guess the instantaneous values of the parameters and take the appropriate action, then it would certainly be possible to construct a perpetuum mobile of the second kind, if this demon were willing to help. We humans cannot guess these parameters, but we can measure them and could take appropriate action. This raises the question whether we do not arrive in this way at a contradiction with the dogmatically exact form of the Second Law. We hope to give a satisfactory answer in a soon forthcoming paper and avoid in the present paper this difficulty by here not coupling the actions taken in our 'Gedanken'-experiments with the fluctuations. Instead we restrict ourselves to actions that could be carried out equally well by periodically functioning machines.

In short, the answer supplied in that paper is that there is a hidden entropy cost in cycles of the type just sketched. We assume that an operator can detect whether the shuttle temperature exceeds a threshold. This measurement process may require entropy generation and, Szilard postulated, this amount must be sufficient to save the statistical form of the Second Law. The escape looks rather like Smoluchowski's. But it was not. Szilard had sought to defeat the Demon by naturalising him. Szilard now located the hidden entropy that was to defect this naturalised Demon in the processes he used for measurement. Szilard's analysis had taken a turn that would redirect the literature on Maxwell's Demon.

8. Szilard and the Entropy Cost of Information

Szilard's (1929) 'On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings' is a paper that has proved enormously influential, although it is conceded that details of its argumentation are obscure.[29] While parts of Szilard's arguments remain unclear to us as well, this overall project is quite explicit. Any device that employs fluctuations in an attempt to violate the Second Law of thermodynamics will fail, he urges, since there is an inevitable hidden entropy cost in the acquisition of information needed to run the device. This entropy cost of information is to be computed from the supposition of the Second Law of thermodynamics; it is not an independent postulate that then proves able to save the Second Law (p. 124):

> We show that it is a sort of a memory faculty, manifested by a system where measurements occur, that might cause a permanent decrease of entropy and thus a violation of the Second Law of Thermodynamics, were it not for the fact that measurements themselves are necessarily accompanied by a production of entropy. At first,[30] we calculate this production of entropy quite generally *from the postulate*

[29] e.g. Leff and Rex (1990, p. 16).

[30] The second part of the paper, Szilard explains, is a check on the result derived in the first by checking the entropy production of a particular measuring device. He continues the passage quoted as:

> Second, by using an inanimate device able to make measurements—however under continual entropy production—we shall calculate the resulting quantity of entropy. We find that it is exactly as great as is necessary for full compensation.

> *that full compensation is made in the sense of the Second Law (Equation (1))* [our emphasis].

That is, Maxwell's Demon cannot succeed in so far as he is subject to ordinary thermodynamic law. Moreover the sense of the Second Law is clearly the statistical sense. This is entirely in keeping, of course, with his analysis in Szilard (1925) of fluctuations. Szilard's (1929, p. 125) explication of perpetual motion machines of the second kind clearly addresses only the long term, average effect of a cycle repeated indefinitely often. Any individual execution of the cycle may well violate the law:

> [...] in a system left to itself no 'perpetuum mobile' (perpetual motion machine) of the second kind (more exactly, no 'automatic machine of *continual* finite work-yield which uses heat at the lowest temperature') can operate in spite of the fluctuation phenomena. A perpetuum mobile would have to be a machine which *in the long run* could lift a weight at the expense of the heat of a reservoir. In other words, if we want to use the fluctuation phenomena in order to gain energy at the expense of heat, we are in the same position as playing a game of chance, in which we may win certain amounts now and then, although the expectation value of the winnings is zero or negative [our emphasis].

The core result of Szilard's paper concerns the case of measurement of a system parameter that varies according to a probabilistic law and can have two values. The entropies $\bar{S}_1$ and $\bar{S}_2$ are the entropies produced by measurement when the outcomes are the first or second values respectively. The lower bound for these two entropies is given by

$$\exp(-\bar{S}_1/k) + \exp(-\bar{S}_2/k) \leqslant 1. \tag{7}$$

This core result is unfamiliar. However a short manipulation will at least show a natural connection to the later literature. If we assume that the probabilities of the two outcomes are w_1 and w_2, then, invoking later perspectives, we might assign a lower bound to the entropies associated with each outcome:

$$\bar{S}_1 \geqslant -k \log w_1, \quad \bar{S}_2 \geqslant -k \log w_2, \tag{8}$$

which are equivalent to the inequalities

$$w_1 \geqslant \exp(-\bar{S}_1/k), \quad w_2 \geqslant \exp(-\bar{S}_2/k). \tag{9}$$

Now the probabilities w_1 and w_2 satisfy

$$w_1 + w_2 = 1. \tag{10}$$

If we substitute for w_1 and w_2 using the above inequalities, we recover the core result (7) as a weaker consequence of the more familiar (8).[31]

[31] The consequence (7) is weaker since, as we shall see below, (7) can be satisfied by values for $\bar{S}_1$ and $\bar{S}_2$ that do not satisfy (8). For example, whatever the values of w_1 and w_2, (7) can be satisfied by $\bar{S}_1 = \bar{S}_2 = k \log 2$.

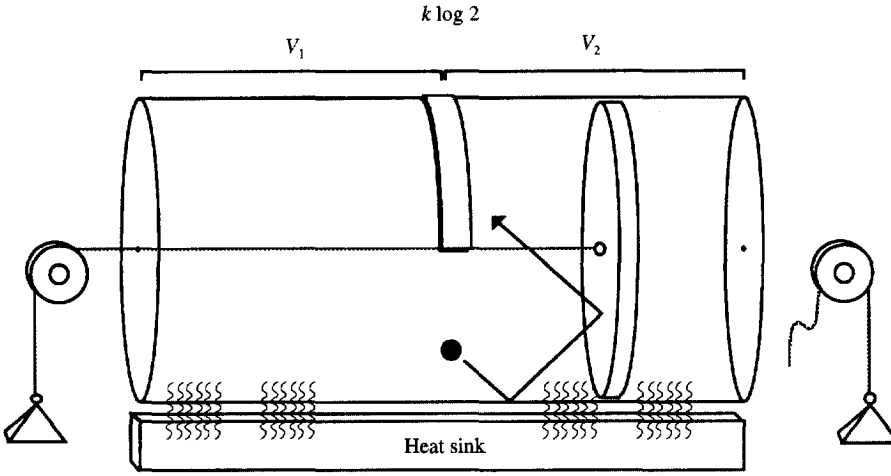

Fig. 3. *Szilard's one-molecule engine.*

Szilard's first and most famous application of this result is his celebrated one-molecule engine. The engine consists of a cylinder containing a gas of a single molecule, as illustrated in Fig. 3.[32] A partition may be inserted into the cylinder so that the cylinder is divided into two volumes V_1 and V_2. This partition can be used as a piston that expands under the pressure of the single-molecule gas. The work recovered during expansion is stored in a raised weight. A heat reservoir maintains the gas at a temperature T. The engine operates under the following cycle. The partition is inserted, trapping the molecule on one side of the partition. The operator ascertains which side contains the molecule and then allows an isothermal, reversible expansion to proceed until the piston has reached the end of the cylinder. Work is derived from the expansion and the temperature of the gas is maintained by the transfer of heat from the heat reservoir. The piston partition is removed, completing the cycle. If we assume that the molecule happens to be trapped in volume V_1 when the partition is inserted, then the heat absorbed during the expansion and the work recovered are both equal to $kT \log((V_1 + V_2)/V_1)$. If we ignore any entropy costs associated with the manipulation of the piston partition, the net effect of the cycle is to convert this amount of heat into work. This corresponds to a net reduction of

$$k \log((V_1 + V_2)/V_1) \tag{11}$$

in the entropy of the heat reservoir, with no compensating entropy increase elsewhere. (Since the one-molecule gas is returned to its initial state, its entropy has not changed.) In virtually all later discussion, it is assumed that $V_1 = V_2$ so

[32] We differ from Szilard's description in aligning the cylinder axis horizontally rather than vertically to remove the distraction of the work needed to raise and lower the piston against gravity.

that the reduction in entropy per cycle is the simpler expression

$$k \log 2. \tag{12}$$

Szilard's analysis is intended to defeat this apparent violation of the Second Law. His concern is not whether the engine allows a decrease in entropy in a single cycle of operation. Rather it is whether there is such a decrease in the average of many repeated cycles of operation. Here he shows that the *average* entropy change per cycle is non-negative if one factors in the average entropy production due to measurement. This average production in the weighted sum of the entropy costs of measurements $\bar{S}_1$ and $\bar{S}_2$ corresponding to the two cases in which the molecule is measured to be in one or other part of the cylinder. The probabilities of these two outcomes are w_1 and w_2. Thus the average entropy cost of measurement per cycle is

$$\bar{S} = w_1\bar{S}_1 + w_2\bar{S}_2 \tag{13}$$

Since $\bar{S}_1$ and $\bar{S}_2$ are allowed any values that satisfy (7), Szilard adopts an *Ansatz* apparently merely for convenience of calculation that

$$\bar{S}_1 = \bar{S}_2 = k \log 2 \tag{14}$$

These values are convenient in so far as they happen to satisfy the inequality (7) and give a value of $\bar{S} = k \log 2$ that is independent of w_1 and w_2. These values also happen to give the minimum value for $\bar{S}$ if $\bar{S}_1$ and $\bar{S}_2$ are defined by (8). On average, in each cycle, this measurement-based entropy production is no less than the entropy decrease from the conversion of heat to work. The Second Law is saved—but clearly only in its statistical form—for the calculation gives no assurance as to the outcome of any single cycle.

Szilard then proceeds to what he calls a derivation of his core result (7). It arises within an analysis of a cycle of separation and recombination of a molecular gas of two differing molecular species by means of semi-permeable membranes. One step even presupposes an intelligent semi-permeable membrane that passes or obstructs molecules *not* according to their present species, but according to their species in a previous step. Thus the whole process requires an intelligent being to measure the species of the molecules and remember the results for each molecule, no matter how its species may alter in future steps. The semi-permeable membrane can then base its function on this memory in a future step. Szilard's exposition is quite brief and the details of his analysis obscure. Fortunately Leff and Rex (1994) have been able to supply a plausible reconstruction that gives many of the details omitted by Szilard.[33] For our purposes what matters is that the cycle would be able to bring about a net entropy reduction, a violation of the Second Law of thermodynamics, unless it is supposed that there is a hidden entropy cost in the intelligent intervention. Szilard locates this

[33] Even though Leff and Rex (1994, p. 997) have been charitable in their interpretation, they still find places where Szilard's model is weak. In one step, Szilard imposes a reversibility condition unnecessarily. More seriously, they find that Szilard fails to account for the entropy cost of intervention by an intelligent being in another step.

cost as arising in the measurement process that distinguishes the two molecular species. If $\bar{S}_1$ and $\bar{S}_2$ are the entropies generated by measurements that identify the first and the second species, then the average entropy cost of each measurement in the process is $\bar{S} = w_1\bar{S}_1 + w_2\bar{S}_2$, where there are frequencies w_1 of molecules of the first species and w_2 of the second. Szilard then shows that the total entropy of the process is non-negative for any value of w_1 and w_2 provided that his core result (7) sets a lower bound for the values of $\bar{S}_1$ and $\bar{S}_2$. Szilard's final calculation is to devise a particular example of a device that performs a measurement and to show that the entropy it produces in the course of measurement is in accord with (7).

9. From the Statistical to the Absolute

Szilard's project was well defined. From the assumption of the Second Law of thermodynamics in a statistical form, he will derive a lower bound for the entropy cost of information acquisition. This cost will preclude naturalised Maxwell Demons from violating this Second Law. Szilard's framework limits how much he can achieve. Since he proceeded from a statistical form of the Second Law, then the best he can infer is a statistical result about the entropy cost of information. He can assign a definite value to the cost only for the long term averages of many repeated measurements. He has no way to preclude the possibility that the entropy costs of individual measurements are subject to fluctuation and may be greater or smaller in repetitions of the same measurement. That this is all he intended is suggested by his calculations of the overall entropy production in a cycle: they are always for the expected entropy production, the average of many repeated cycles. Moreover Szilard's notation of $\bar{S}_1$ and $\bar{S}_2$ as the entropy costs for measurements suggest that these two costs individually are only expected entropies — the overhead bar is the standard symbol in probability for an expectation. However Szilard's verbal descriptions of $\bar{S}_1$ and $\bar{S}_2$ do not seem to allow such a statistical construal. Here is his description of his core result (7):[34]

> One may reasonably assume that a measurement procedure is fundamentally associated with a certain definite average entropy production, and that this restores concordance with the Second Law. The amount of entropy generated by the measurement may, of course, always be greater than this fundamental amount, but not smaller. *To put it precisely: we have to distinguish here between two entropy values* [Szilard's emphasis]. One of them, $\bar{S}_1$, is produced when during the measurement y assumes the value 1, and the other, $\bar{S}_2$, when y assumes the value -1. We cannot expect to get general information about $\bar{S}_1$ or $\bar{S}_2$ separately, but we shall see that, from the assumption that the amount of entropy produced by the 'measurement' must compensate in the sense of the second law the entropy

[34] The standard translation reproduced in Leff and Rex (1990, p. 127) has been corrected slightly against the original German.

decrease affected by utilization, there follows quite generally the relation

$$\exp(-\bar{S}_1/k) + \exp(-\bar{S}_2 k) \leqslant 1. \tag{15}$$

If Szilard's 'average entropy production' denotes the average $\bar{S} = w_1\bar{S}_1 + w_2\bar{S}_2$, then nothing he says unequivocally attributes $\bar{S}_1$ and $\bar{S}_2$ statistical character; they are treated as if they are not statistical.

Whatever Szilard may have intended, he was almost immediately understood as asserting that the quantities $\bar{S}_1$ and $\bar{S}_2$ were absolute values. In considering the entropy cost in ascertaining the location of the molecule in Szilard's one-molecule engine, von Neumann (1932, p. 400), for example, described Szilard's result as:

> L. Szilard has (see reference [Szilard, 1929]) shown that one cannot get this 'knowledge' without a compensating entropy increase $k \ln 2$. In general $k \ln 2$ is the 'thermodynamic value' of the knowledge, which consists of an alternative of two cases. All attempts to carry out the process described above without the knowledge of the half of the container in which the molecule is located, can be proved invalid, although they may occasionally lead to very complicated automatic mechanisms.

Notice also that Szilard's (7) has been replaced by the specific values, $k \log 2$, he chose merely as conveniences.[35]

By 1950, Szilard was reported as having secured the absolute validity of the Second Law of thermodynamics on the basis of his discovery of the entropy cost of information and his core result (7) reduced to the catchier formula $k \log 2$. Brillouin (1951, p. 136) summarised it as:

> [... Szilard] inferred that the process of physical measurement, by which the information could be obtained, must involve an increase of entropy so that the whole process would satisfy Carnot's principle.

And Gabor (1951, p. 148) has Szilard as

> [...] showing that a simple observation, which amounts to a selection from n equally likely possibilities, enables the observer to decrease the entropy of the system observed by a maximum of
>
> $$k \log n. \tag{16}$$
>
> Hence, in order to save the Second Principle, it must be assumed that such an observation could not be made by any 'demon', intelligent or mechanical, without an entropy increase of *at least* this amount. Szilard proved this in detail in one example [...] [Gabor's emphasis].

We must wonder how this transformation has occurred. Szilard's project was to protect a statistical form of the Second Law of thermodynamics from demonic violation by naturalising the Demon; he is now reported as succeeding so well that his discoveries protect not just the law in its statistical form, but also in its

[35] Has von Neumann misinterpreted Szilard's intentions? Or does Szilard (1929) not give a faithful rendering of them? What complicates the decision is that von Neumann was in Berlin with Szilard at this time and had the opportunity to discuss these matters with Szilard.

absolute form. Possibly the transformation is connected with the separation of Szilard's exorcism of the Demon and fluctuation phenomena. While Szilard's analysis was motivated originally by the need to prevent a Demon exploiting fluctuation phenomena, explicit consideration of fluctuation phenomena disappears from later analyses. In so far as it is natural to associate the statistical character of the Second law with fluctuation phenomena only, then it would be tempting to presume that this statistical character is not relevant to Maxwell's Demon. That, of course, would be a mistake. The operation of the Demon depends on phenomena that are equivalent to fluctuation phenomena. The recompression of the single-molecule gas in Szilard's engine amounts to exploiting a huge pressure fluctuation, for example.

What is puzzling is that von Neumann, Brillouin and Gabor surely knew that the Second Law could hold at most statistically and that this was of immediate relevance to Maxwell's Demon. On at least some occasions, Brillouin does report the result as holding merely statistically. He writes for example in Brillouin (1953, p. 1153):

> Any experiment by which an [sic] information is obtained about a physical system corresponds *in average* to an increase of entropy in the system or in its surroundings. This average increase is always larger than (or equal to) the amount of information obtained [Brillouin's emphasis].

A little later (p. 1155) the origin of the crucial 'in average' qualification is made clear:

> [...] *the second principle holds only in average*[36] [... it] is always limited by the possibility of unpredictable *fluctuations*. It may happen that one particular observation could be made at an exceptionally low cost, but we have no way to foresee when and how this may happen. Only averages may be safely predicted [Brillouin's emphasis].

This explicit recognition of the statistical character of the result is exceptional; it is usually absent. Thus we must conclude that Brillouin and his colleagues do not literally mean what they say in their reports of an absolute entropy cost of information acquisition. But if they do not say what they mean, what hope is there that their readers will understand them and that the literature can develop without confusion?

10. Conclusion: Tensions in the Maxwell Demon Literature

By 1950, with the assimilation of Szilard (1929), the literature on Maxwell's Demon had become a rich repository of suggestive ideas. Beneath the alluring shine of its clever thought experiments were several tensions whose presence left the basic principles of the new literature unclear and its direction of

[36] The 'second principle' refers to Brillouin's modified version of the Second Law of thermodynamics which requires that a quantity 'entropy minus information' is non-decreasing.

development unsure. The modern literature on Maxwell's Demon has been built on these uncertain foundations. The continued presence of these tensions is, we believe, responsible for the discomfort we feel with present day literature on Maxwell's Demon.

(I) Does Maxwell's Demon afford a means of circumscribing the domain of validity of the Second Law? Or is it a threat against which the Second Law must be protected?

Maxwell chose in favour of affirming the first and this seems to us still to be the correct choice. By 1950, however, the literature had come to treat Maxwell's Demon as a threat to the Second Law that must be contained or parried. The thermodynamics literature had seen one successful deflection of such a threat, that due to fluctuation phenomena. It succeeded by weakening the Second Law into a statistical form no longer contradicted by fluctuation phenomena. Szilard (1929) attempted to maintain this success in the face of intelligently manipulated fluctuations. Just as the Second Law had been protected from unmanipulated fluctuations by postulating a weakened version, it also seemed that the Second Law could be protected against intelligent intervention by a further postulate, this time of an unnoticed entropy cost in measurement. Commonly recognised fluctuation phenomena do fall under the statistical thermodynamics generated by the weakened form of the Second Law. But can we be so sure that all Maxwell Demons will be similarly brought into accord with the Second Law by some such general postulate? The range of potential Demons at issue is great, with many contrived exactly to maximise their incompatibility with the Second Law. Since these Demons remain the playthings of thought without precise rules as to which idealisations and liberties are permissible, one might well wonder how any general postulate can hope to tame them all. Should we require, for example, that the microdynamics of a Demon be Hamiltonian? To do so is arbitrary. To fail to do so, as we shall see in the next part, invites inexorable Demons. The more secure approach, we urge, is to look to the microphysics that underpins the various Demons' operation and determine from that which accord with this or that version of the Second Law and which violate it.

Related to the first tension is a confusion of purpose in the exorcism literature:

(II) Are the exorcisms of Maxwell's Demon aimed at protecting an absolute form of the Second Law or merely a statistical form?

In so far as the exorcisms are built on the work of Smoluchowski and Szilard, one could only expect protection of statistical validity of the Law, for that is the basic presumption of both their analyses. They sought not to protect thermodynamics from fluctuation phenomena but from intelligent intervention that might allow fluctuations to be accumulated into macroscopic violations of the Second Law. But then how are we to explain that this crucial statistical qualification is so often omitted in the exorcism literature?

Jauch and Baron's (1972, Section 4) assault on the mainstream viewpoint in the later Demon literature allows us to see just how completely fluctuations

have been decoupled from the operation of the Demon. They complain that the Szilard one-molecule engine employs an illegitimate idealisation in so far as the door closing operation violates the gas law—the equation of state presumed in the phenomenological thermodynamics of gases:

> [...] at the exact moment when the piston is in the middle of the cylinder and the opening is closed, the gas violates the law of Gay–Lussac because the gas is compressed to half its volume without expenditure of energy. We therefore conclude that the idealisations in Szilard's experiment are inadmissible in their actual context.

Of course they have missed the point of Szilard's original problem completely. The variations in gas density and pressure as the single molecule moves about the piston are simply large fluctuations in the gas's density and pressure. The point of the Szilard engine is to determine whether an intelligent Demon can exploit these fluctuations systematically. Since the equation of state for a gas reports only on mean densities and pressures, we should not expect fluctuations to be reflected in the equation. To decry fluctuations as inadmissible is to deny the whole project. The establishment response seems also to have lost the connection to fluctuations. Costa de Beauregard and Tribus' (1974) response insists on the legitimacy of the door closing but without mentioning fluctuations. More curiously, Zurek's (1984, p. 250) analysis seeks to show that 'the apparent inconsistency pointed out by Jauch and Baron is avoided by quantum treatment', thereby conceding the viability of Jauch and Baron's objection for the classical case. Worse, on the same page, Zurek treats fluctuations as a separate problem that can be evaded, allowing that their effect can be reduced by considering many linked one-cylinder engines. We return to Zurek's analysis below in Part II, Section 3.

There is a variant form of this last unclarity of purpose.

(III) Are the exorcisms aimed at protecting the Second Law against straight violations or against embellished violations in which useful work is continuously extracted from a macroscopic system?

The literature often gives the impression that the first, stronger aim is operative. But not untypically it turns out that the prospects of success are nil unless the second, weaker aim is adopted.

There are also unclarities in the means of exorcism:

(IV) Once Maxwell's Demon is naturalised, is he exorcised by allowing that he is a thermal system himself subject to fluctuations that defeat his purposes, or by allowing that he is an information system and that there are hidden costs associated with information processing?

The first alternative follows Smoluchowski; the second Szilard. Are they mutually exclusive alternatives? Or, as Szilard seemed to suggest in this analysis, is his approach merely a picturesque representation of Smoluchowski's?

(V) Is the entropy cost of information acquisition and processing a result that is independently postulated or is it derived from the supposition of the Second Law in some suitable form?

Szilard clearly advocated the latter. Yet his assertion of the entropy cost of information acquisition is now often treated as an independent postulate.

These tensions have left their mark on the modern literature on Maxwell's Demon as it seeks to come to terms with the Demon within and the Demon without.

Appendix 1: The Second Law of Thermodynamics

Among the many formulations of the Second Law, there are three to which we will have occasion to refer.

1. No cycle is possible whose sole result is the removal of heat from a reservoir at one temperature and the absorption of an equal quantity of heat by a reservoir at a higher temperature.
2. No cycle is possible whose sole result is the abstraction of heat from a single reservoir and the performance of an equivalent amount of work (perpetual motion machine of the second kind[37]).
3. The entropy of a closed system cannot decrease with time.

Appendix 2: Dynamical Systems

A *dynamical system* is a triple (X, T_t, μ). Each point $\xi \in X$ of the state space or phase space corresponds to a possible instantaneous state of the system. The $T_t : X \to X$, $t \in R$, form a one-parameter group of transformations. ($T_0 = \mathrm{id}$, $T_{-t} = T_t^{-1}$, and $T_{t_1+t_2} = T_{t_2} \circ T_{t_1}$). T_t represents temporal evolution, i.e. if the state at t_0 is ξ, then the state at $t_0 + t$ is $T_{t_0+t}(\xi)$. μ is a normed measure on X. It is often assumed that μ is invariant under the phase flow, which means that for any measurable set $A \subseteq X$, $\mu(T_t(A)) = \mu(A)$ for all t. In the familar case of Hamiltonian dynamics of a system of N point particles, X is $6N$-dimensional (three coordinates for the position of each particle, three for momentum). Conservation of phase volume under the Hamiltonian flow is known as *Liouville's theorem.*

For any dynamical system with an invariant μ, *Poincaré's theorem* shows that the system returns, almost surely, arbitrarily closely to its starting state. More precisely, this theorem shows that for any measurable set A, $\mu(F) = 0$, where $F = \{\xi \in A : T_t(\xi) \notin A \text{ for all } t > 0\}$.

A dynamical system with an invariant μ such that $\mu(X) < \infty$ is said to be *ergodic* if and only if almost every phase orbit passes arbitrarily closely to any

[37] A perpetual motion machine of the first kind is a machine which violates the First Law, which states the conservation of energy.

chosen phase point. More precisely, ergodicity means that for almost every phase point ξ and for any measurable set A such that $\mu(A) > 0$, there is a t such that $T_t(\xi) \cap A \neq \emptyset$. If (X, T_t, μ) is an ergodic dynamical system and μ' is any invariant measure that is absolutely continuous with respect to μ (i.e. for any measurable set $A \subseteq X$, $\mu'(A) = 0$ implies that $\mu(A) = 0$), then $\mu' = \mu$. This fact has been taken to justify the use of the so-called microcanonical distribution in the case of Hamiltonian dynamics where the state of the system is confined to a constant energy surface (see Malament and Zabell (1980); but see also Earman and Rédei (1995)). A dynamical system is *mixing* if and only if for any measurable sets A and B, $\lim_{t \to \pm\infty} \mu((\varphi_t A) \cap B) = \mu(A) \cdot \mu(B)$. Mixing (which is stronger than ergodicity) guarantees a coarse-grained approach to equilibrium: for any integrable observable $f: X \to R$ and for any normed density ρ, the expectation value of f approaches its equilibrium value, i.e. $\lim_{t \to \pm\infty} \int_X f \rho_t d\mu = \int_X f d\mu$, where $\rho_t(x) = \rho(T_t(x))$.

Appendix 3. Exorcising the Demon Within — Fluctuations: Popper's Reformulation of the Second Law

Like Smoluchowski, Popper (1957) took fluctuation phenomena to provide violations of the classical Second Law of thermodynamics. And although he faulted Smoluchowski's attempted reformulation of the Second Law, Popper shared Smoluchowski's goal of saving the Second Law by weakening it. Popper's reformulation reads (p. 153)

> A satisfactory phenomenological formulation of the entropy law [...] appears to be the following:
>
> > (F) A gas or liquid in a closed circular tube, immersed in a heat bath of any temperature and fitted with a *one-way valve*, does not constantly circulate through the tube, however slowly.

Popper took (F) to be equivalent to (*ibid.*)

> > (F′) There does not exist a semi-permeable membrane with an asymmetric structure (like a one-way valve) such that the probabilities of passing through are not equal in both directions.

If we take Popper's project at face value, then these laws ought to stand on their own, without support of a particular micropicture. The crucial qualification 'does not *constantly* circulate' allows the possibility of short term violations in which circulation does occur. Without some guide from a micropicture as to how long this short term may be, Popper's (F) makes no prediction about the behaviour of thermodynamic systems over finite periods of time. This is no trivial quibble. We can readily concoct scenarios in which such short term violations may persist for so long that they are the law of practical interest. Consider, for example, a gas that has spontaneously expanded to fill a vessel. Poincaré's recurrence lemma assures us that, aside from cases of measure zero

probability, the gas cannot constantly maintain its expanded state. Of course, in the short term it may maintain its expanded state. As it turns out, this short term extends into eons of time for macroscopic gas systems, so that the short term violation becomes the useful prediction.

Moreover, as long as any particular picture of microdynamics is renounced, the formulations (F) and (F') are not equivalent. To see this we have to take account of the fact that Popper's 'probability of passage' is ambiguous; for example, the probability of a left-to-right passage could mean either $\Pr(L \to R)$ or $\Pr(L \to R) \times \Pr(L)$. If the latter reading is chosen, it is possible to have systems in which (F) holds but (F') fails, whereas on the former reading it is possible to violate (F) without violating (F'). As an example of the first kind imagine a semipermeable membrane in a gas-filled closed circular tube and imagine that the membrane will pass faster moving molecules only in the right-to-left direction and slower moving molecules in the left-to-right direction. While the membrane passes molecules asymmetrically in violation of (F'), it is not necessary that (F) be violated. In designing the membrane, one might adjust the probability of passage in the two directions so that the average momentum flow in each direction through the membrane is equal. Presumably the adjustment would require a lesser probability of passage for the faster molecules and a greater probability of passage for the slower molecules. Also, the adjustment might only succeed in balancing the momentum flows for one particular temperature. But that is sufficient. At that temperature, (F') will be violated, but there will be no net circulation in the tube, in violation of (F).[38]

As an example of the second kind, once again imagine a gas-filled ring, this time without membrane—the case of the vacuous membrane! The vacuous membrane clearly satisfies (F') for the vacuum passes molecules with equal ease in all directions. It will be possible, however, to select the geometry of the tube and the initial position and velocities of the molecules so that they do not collide, their trajectories are re-entrant and they all circulate in the same direction. Of course such an arrangement is likely to be pseudo-stable: the slightest perturbation will destroy it. But that is not our concern. The example shows that (F) can fail without violation of (F'). In ordinary statistical mechanics, this arrangement is dismissed as a case of measure zero probability. This escape is not open to Popper's (F) and (F') since they make no disclaimer about measure zero cases. And adding such a disclaimer would not help. For such a disclaimer would raise the issue of the origin of the probability measure invoked.

Popper was pessimistic about grounding his weakened version of the Second Law in statistical mechanics: 'I have little doubt that my formulae (F) and (F')

[38] To see that there will be no net circulation, recall that for a tube of sufficient size filled with a gas of sufficiently great density, both faster and slower molecules that pass the membrane will rapidly lose their special 'fast' and 'slow' characteristic through collision with other molecules. At some distance from the membrane on either side, the molecular velocity distributions will be the same and there will be no net flow in either direction.

cannot be derived from any of the versions of statistical mechanics, as they exist at present, except of course, if the non-existence of Maxwell's demon is simply assumed *ad hoc* [...]' (p. 154).

The recent work of Zhang and Zhang (1992) can be seen as an attempt to supply the derivation of which Popper despaired. Replace (F) with the kindred

> (F'') Postulate of no *Spontaneous Momentum Flow* (SMF): an isolated mechanical system does not admit of a sustaining and robust momentum flow.

Zhang and Zhang offer a 'proof' that (F'') is equivalent to the non-existence of a perpetual motion machine of the second kind. To show that the non-existence of perpetual motion mechines of the second kind rules out an SMF, suppose that an isolated system admitted an SMF. We could then extract work from the system by means of a paddle wheel. The energy lost could be replenished by heat from a heat bath in contact with the system. The result would be a perpetual motion machine of the second kind. In the other direction, if a perpetual motion machine of the second kind existed, we could use the work output to run a paddle wheel in a liquid, which would set up an SMF. Any heat generated by the paddle wheel could be fed back to the heat bath.

To derive any precise results about the existence or non-existence of an SMF, one needs to make suppositions about the micropicture and to provide a technical characterisation of an SMF introduced informally in (F''). For a system of particles whose dynamics make it an instance of an abstract dynamical system (see Appendix 2) one can define the momentum $J_V(\xi)$ associated with a spatial volume V and a phase trajectory determined by the phase point ξ. The long-term average $\overline{J_V(\xi)}$ is defined as $\lim_{\tau\to\infty} 1/\tau \int_0^\tau J_V(T_t(\xi))\,\mathrm{d}t$, where T_t is the time evolution operator. An SMF is said to exist if and only if there is a V such that $\overline{J_V(\xi)}$ has the same non-zero value for almost every ξ on the energy surface. Zhang and Zhang (1992) prove:

> *Lemma*: If the energy is symmetric under momentum reversal and the phase volume is conserved under time evolution, then the system does not admit an SMF.

Now, normal time reversal invariant Hamiltonian dynamics satisfies the conditions of the Lemma, which should then mean, by the above reasoning, that any such system does not permit the operation of a perpetual motion machine of the second kind. We would then have a proof that such a dynamical system implies the full validity of the unweakened Second Law. This is too good to be true: normal classical statistical mechanics typically assumes phase volume conserving and time reversal invariant dynamics; but, of course, it implies the existence of fluctuation phenomena that violate the Second Law.

What has gone wrong is that the above 'proof' is this: it is a fallacy that the existence of a perpetual motion machine of the second kind implies an SMF. A perpetual motion machine of the second kind need not *continuously* produce

work that could be used to set up an SMF. Such a machine is, in effect, any system that violates the Second Law, even if it does it just once, or extremely sporadically or with vastly small probability. Consider, for example, a large and vastly improbable momentum fluctuation in a kinetic gas. If it just happens to raise a heavy weight, then the heat energy of the gas has been converted into the work needed to raise the weight without discharge of waste heat, in violation of the Second Law. Zhang and Zhang try to exclude such cases by saying that the perpetual motion machine is supposed to perform robustly. But there is nothing about robust performance in the classical statement of the Second Law. The law prohibits any case of conversion of heat to work without discharge of waste heat to a cooler reservoir.

However, Zhang and Zhang's lemma is valid, and since a violation of Popper's (*F*) would seem to involve an SMF, we have a proof — Popper's pessimism to the contrary — that a standard version of classical statistical mechanics using time reversal invariant Hamiltonian dynamics does imply Popper's weakened version of the Second Law (allowing exceptions of measure zero). This much Popper would have liked. But Zhang and Zhang's analysis contains other results that Popper would not have found congenial.

They give concrete examples of dynamical systems that conserve energy and are time reversal invariant but have a dynamics that alters phase volume. These examples are shown to admit an SMF. Since the implication from an SMF to a perpetual motion machine of the second kind is more solid, these examples gave a violation of the classical Second Law. Such examples also violate the spirit if not the letter of Popper's (*F*).[39]

We conclude that whether or not the Demon can be exorcised along the lines Popper suggested depends on contingent features of nature. And the relevant features must be investigated at the microlevel; no clever phenomenological reformulation will save the Second Law if the relevant microdynamical features are unfavourable.

Acknowledgement—We are grateful to Frank Arntzenius, Al Janis, Chris Martin and anonymous referees of this journal for helpful discussion and comments.

References

Brillouin, L. (1951) 'Maxwell's Demon Cannot Operate: Information and Entropy. I', *Journal of Applied Physics* **22,** 334–337, reprinted in Leff and Rex (1990), pp. 134–137; 'Maxwell's Demon Cannot Operate: Information and Entropy. II', *Journal of Applied Physics* **22,** 338–343.

Brillouin, L. (1953) 'The Negentropy Principle of Information', *Journal of Applied Physics* **24,** 1152–1163.

Brown, N. (1990) 'A Demon Blow to the Second Law of Thermodynamics?' *New Scientist* (14 July) 35.

[39] In Part II, Appendix 1 we discuss a modified version of one of Zhang and Zhang's models.

Brush, S. G. (1976) *The Kind of Motion We Call Heat: A History of the Kinetic Theory of Gases in the 19th Century*, 2 Vols (Amsterdam: North-Holland).

Brush, S. G. (1983) *Statistical Physics and the Atomic Theory of Matter, from Boyle and Newton to Landau and Onsager* (Princeton: Princeton University Press).

Collier, T. D. (1990) 'Two Faces of Maxwell's Demon Reveal the Nature of Irreversibility', *Studies in the History and Philosophy of Science* **21,** 257–268.

Costa de Beauregard, O. and Tribus, M. (1974) 'Information Theory and Entropy', *Helvetica Physica Acta* **47,** 238–247. Reprinted in Leff and Rex (1990), pp. 173–182.

Daub, E. E. (1970) 'Maxwell's Demon', *Studies in the History and Philosophy of Science* **1,** 213–227.

Earman, J. and Rédei, M. (1995) 'Why Ergodic Theory Does Not Explain the Success of Equilibrium Statistical Mechanics', *British Journal for the Philosophy of Science* **47,** 63–78.

Einstein, A. (1905) 'Über die von der molecularkinetischen Theorie der Wärme gefordete Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen', *Annalen der Physik* **17,** 549–560. Translation from A. Beck (transl.) P. Havas (consultant), *The Collected Papers of Albert Einstein*, Vol. 2 (Princeton: Princeton University Press, 1989).

Einstein, A. (1907) 'Über die Gültigkeitsgrenze des Satzes vom thermodynamisshen Gleichgewicht und über die Möglichkeit einer neuen Bestimmung der Elementarquanta', *Annalen der Physik* **22,** 569–572.

Einstein, A. (1909a) 'Zum gegenwärtigen Stand des Strahlungsproblems', *Physikalische Zeitschrift* **10,** 185–193.

Einstein, A. (1909b) 'Über die Entwicklung unserer Anschauungen über das Wesen und die Konstitution der Strahlung', *Deutsche Physikalische Gesellschaft, Verhandlungen* **7,** 482–500.

Einstein, A. (1914) 'Beiträge zur Quantentheorie', *Deutsche Physikalische Gesellschaft, Verhandlungen* **16,** 820–828.

Feynman, R., Leighton, R. and Sands, M. (1963) *The Feynman Lectures on Physics*, Vol. 1 (Reading, MA: Addison–Wesley).

Gabor, D. (1951) '§5 A Further Paradox: "A Perpetuum Mobile of the Second Kind"', in Leff and Rex (1990), pp. 148–159.

Heimann, P. M. (1970) 'Molecular Forces, Statistical Representation and Maxwell's Demon', *Studies in the History and Philosophy of Science* **1,** 189–211.

Horowitz, T. and Massey, G. J. (eds) (1991) *Thought Experiments in Science and Philosophy* (Savage, MD: Rowman and Littlefield).

Ingarden, R. S. (1986) *Marian Smoluchowski: His Life and Scientific Work* (Warsaw: Polish Scientific Publishers).

Jauch, J. M. and Baron, J. G. (1972) 'Entropy, Information and Szilard's Paradox', *Helvetica Physica Acta* **45,** 220–232. Reprinted in Leff and Rex (1990), pp. 160–172.

Knott, C. G. (1911) *Life and Scientific Work of Peter Gutherie Tait* (Cambridge: Cambridge University Press).

Kuhn, T. (1978) *Black Body Theory and the Quantum Discontinuity* (Oxford: Clarendon Press).

Laymon, R. (1991) 'Thought Experiments by Stevin, Mach and Gouy: Thought Experiments as Ideal Limits', in Horowitz and Massey (1991) pp. 167–191.

Leff, H. S. and Rex, A. F. (1990) *Maxwell's Demon: Entropy, Information, Computing* (Princeton: Princeton University Press).

Leff, H. S. and Rex, A. F. (1994) 'Entropy of Measurement and Erasure: Szilard's Membrane Model Revisited', *American Journal of Physics* **62,** 994–1000.

Maddox, J. (1990) 'Maxwell's Demon Flourishes', *Nature* **345,** 109.

Malament, D. and Zabell, S. (1980) 'Why Gibbs Phase Space Averages Work: The Role of Ergodic Theory', *Philosophy of Science* **47,** 339–349.

Mandelbrot, B. (1964) 'On the Derivation of Statistical Thermodynamics from Purely Phenomenological Principles', *Journal of Mathematical Physics* **5,** 164–171.

Martin, C. (1996) 'A Phenomenological Basis for Statistical Thermodynamics: Leo Szilard's 1925, "On the Extension of Phenomenological Thermodynamics of Fluctuation Phenomena"', preprint.

Maxwell, J. C. (1860) 'Illustration of the Dynamical Theory of Gases', *Philosophical Magazine* **19,** 19–32; **20,** 21–37. Reprinted in Maxwell (1952), pp. 377–409.

Maxwell, J. C. (1878) 'Diffusion', *Encyclopedia Britannica* (9th edn) **7,** 214. Reprinted in Maxwell (1952), pp. 625–646.

Maxwell, J. C. (1952) *Scientific Papers of James Clark Maxwell*, edited by W. D. Niven (New York: Dover).

Neumann, J. von (1932) *Mathematical Foundations of Quantum Mechanics* (Princeton: Princeton University Press, 1955).

Norton, J. D. (1991) 'Thought Experiments in Einstein's Work', in Horowitz and Massey (1991), pp. 129–148.

Nye, M. J. (1972) *Molecular Reality: A Perspective on the Scientific Work of Jean Perrin* (London: MacDonald).

Perrin, J. (1921) *Atoms*, transl. D. Ll. Hammick (Woodbridge, Connecticut: Ox Bow Press, 1990).

Poincaré, H. (1904) 'The Principles of Mathematical Physics', in *Physics for a New Century: Papers Presented at the 1904 St. Louis Congress* (Los Angeles: Tomash, 1986), pp. 281–299.

Poincaré, H. (1905) *Science and Hypothesis* (New York: Dover, 1952).

Popper, K. (1957) 'Irreversibility, or Entropy Since 1905', *British Journal for the Philosophy of Science* **8,** 151–155.

Sklar, L. (1993) *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics* (Cambridge: Cambridge University Press).

Smoluchowski, M. (1912) 'Experimentell nachweisbare, der üblichen Thermodynamik widersprechende Molekularphänomene', *Physikalische Zeitschrift* **13,** 1069–1080.

Smoluchowski, M. (1914) 'Gültigkeitsgrenzen des zweiten Hauptsatzes der Wärmetheorie', in *Vorträge über die kinetische Theorie der Materie und der Elektrizität* (Leipzig: Teubner).

Sredniawa, B. (1991a) 'Collaboration of Martin Smoluchowski and Theodor Svedberg in the Investigations of Brownian Motion and Density Fluctuations', preprint, Department of Theoretical Physics, Jagellonian University, Cracow.

Sredniawa, B. (1991b) 'Martin Smoluchowski's Collaboration with Experimentalists in the Investigations of Brownian Motion and Density Fluctuations', in B. Sredniawa (ed.), *Essays Devoted To Scientific and Didactic Work of Marian Smoluchowski (1872–1917)* (Universitatis Iagellonicae, Folia Physica, Fasciculus XXXIII, Krakow), pp. 9–46.

Strutt, R. J. (1968) *The Life of John William Strutt* (Madison: University of Wisconsin Press).

Svedberg, T. (1907) 'Über die Bedeutung der Eigenbewegung der Teilchen in Kolloidal-Lösungen für die Beurteilung der Gültigkeitsgrenzen des zweiten Hauptsatzes der Thermodynamik', *Zeitschrift für Physikalische Chemie* **59,** 451–458.

Szilard, L. (1925) 'On the Extension of Phenomenological Thermodynamics to Fluctuation Phenomenon', in Szilard (1972), pp. 70–102.

Szilard, L. (1929) 'On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings', Szilard (1972), pp. 120–129. Reprinted in Leff and Rex (1990), pp. 124–133.

Szilard, L. (1972) *The Collected Works of Leo Szilard: Scientific Papers* (Boston, MA: MIT Press).

Szilard, L. (1978) *Leo Szilard: His Version of the Facts: Selected Recollections and Correspondence*, edited by S. R. Weart and G. W. Szilard (Boston, MA: MIT Press).

Thomson, W. (1874) 'The Kinetic Theory of the Dissipation of Energy', *Nature* **9**, 441–444.

Zhang, K. and Zhang, K. (1992) 'Mechanical Models of Maxwell's Demon with Non-invariant Phase Volume', *Physical Review* **A46**, 4598–4605.

Zurek, W. H. (1984) 'Maxwell's Demon, Szilard's Engine and Quantum Measurements', in G. T. Moore and M. O. Scully (eds), *Frontiers of Nonequilibrium Statistical Mechanics* (New York: Plenum). Reprinted in Leff and Rex (1990), pp. 249–259.