25
THE WORST THOUGHT EXPERIMENT

John D. Norton

1 Introduction

When the great nineteenth-century physicist James Clerk Maxwell imagined (1871, 308) "a being whose faculties are so sharpened that he can follow every molecule in its course," he could not have foreseen the unlikely career his thought creation would have. A little over 50 years later, in a thought experiment devised by Leo Szilard and published in 1929, the demon was manipulating a single molecule and providing the grounding for a connection between information gathering or information processing and thermodynamic entropy. That 1929 thought experiment and its subsequent reception will be the principal subject of this chapter.

Part I will recount a standard narrative of the path to Szilard's thought experiment and its subsequent development. Sections 2 and 3 below sketch how Maxwell's original concerns evolved into Szilard's thought experiment. The thought experiment and its subsequent development will be described in Sections 4–7.

The narrative thus far will recapitulate, mostly, the standard, celebratory view of the thought experiment. In Part II, starting in Section 8, I will give a dissenting, less celebratory view of the thought experiment. It is, I will argue, a failed thought experiment and the locus and origin of enduring confusions in the subsequent literature. In Section 9, I will try to account for how a thought experiment like this can fail so badly. The failure, I will argue, derives from routinely accepted narrative conventions in thought experimenting. We give the thought experimenter extensive latitude in introducing idealizations, so that inessential distractions can be set aside. To avoid the clutter of needless generality, we grant the thought experimenter the presumption that the specific case developed is typical, so its behavior can stand in for a general result. The failure of the Szilard thought experiment results from misuse of both conventions.

PART I: RISE

2 Maxwell and his demon

Maxwell's overall project was to understand how large collections of molecules could manifest as familiar thermal systems, like gases. The natural approach was to devise a

dynamical theory that would track the courses of individual molecules, just as Newton's celestial mechanics tracked planets, moons and comet in their courses. Maxwell realized that there are too many molecules in a gas, all interacting in unimaginably complicated ways, for this dynamical approach to work. Instead, he must use statistical methods, such as employed in the study of mass populations. He could then arrive at a serviceable theory of molecular gases. The molecules are treated en masse only, through suitable probability distributions and statistical parameters.

An interesting consequence, Maxwell realized, is that the second law of thermodynamics depends essentially on the restriction to this statistical treatment of molecules. The law would be violated, he noted, if somehow we could manipulate molecules individually. To make his point, he imagined the being with sharpened faculties mentioned above. Such a being could, without the expenditure of work, cause a gas at uniform temperature to separate out into a hotter portion of faster moving molecules and a colder portion of slower moving molecules, in violation of the second law of thermodynamics. All that was needed was to confine the gas to a vessel with a dividing wall. The being would open and close a hole in the wall, when molecules approached it, so as to allow the faster molecules to accumulate on one side and the slower ones on the other.

Maxwell saw no special problem when he conceived his demon. He envisaged no threat to the second law in this thought experiment. The point was precisely that *we* do not have the powers of his imaginary being. We must treat molecules en masse, so we cannot overturn the second law.

3 Thermal fluctuations

Circumstances changed in the early twentieth century with the observation and analysis of thermal fluctuations. Einstein showed in 1905 that the jiggling motion of microscopically visible particles – "Brownian motion" – was due to the accumulated effect of many collisions of the particles with water molecules. Each time a microscopically visible particle was raised by these motions, we were seeing the full conversion of the heat energy of the water into work, in the form of the potential energy of height. This full conversion of heat into work is precisely a process prohibited by a standard statement of the law. It is a microscopic violation of the second law of thermodynamics.

The pressing question then was whether these microscopic violations could be accumulated into a macroscopic violation of the law. At that moment, Maxwell's demon had ceased to be a benign illustration of the statistical character of the second law. It had become a mortal threat to it. The threat was soon parried, decisively. The most thorough analysis came from Marian Smoluchowski (1912, 1914). He considered numerous mechanisms, each designed to accumulate these thermal fluctuations into a macroscopic violation of the second law. In each case, he showed that further fluctuations within the mechanisms themselves would on average defeat the intended accumulation. The demon must fail since the demon is itself a thermal system, rife with the very thermal fluctuations it sought to exploit. Momentary violations were possible, but they could not be accumulated. The second law was safe, as far as the average behavior of systems was concerned.

In the best known of many examples, Smoluchowski replaced Maxwell's cleverly opened hole in the gas chamber dividing wall with a simple, one-way flapper value, which later came to be known as the Smoluchowski trapdoor. The flapper or trapdoor is lightly

spring-loaded to keep it shut and is hinged so that it can open in one direction only. Collisions with molecules seeking to pass in that direction open it and they pass. Molecules seeking to pass in the reverse direction slam it shut. Over time, one part of the gas chamber becomes spontaneously pressurized, without any compensating process elsewhere, in violation of the second law.

This is a simple and apparently secure mechanical realization of Maxwell's imaginary being. It will fail in its purpose, Smoluchowski argued, since the trapdoor will have its own thermal energy. Since the trapdoor must be very light and lightly spring-loaded if a collision with a molecule can open it, that thermal energy will lead it to flap about wildly, allowing molecules to pass freely in both directions.

4 Intelligent intervention

A loophole remained. What if the mechanism that accumulated the fluctuations were operated intelligently by an animate being? The operation of that being would, in virtue of its vitality, lie outside the normal constraints of physics. This was the question put to Smoluchowski by Kaufmann at the conclusion of his lecture at the 84th *Naturforscherversammlung* (Meeting of Natural Scientists) in Münster in 1912. It would, Kaufmann suggested be (Smoluchowski 1912, 1018):

> a conclusion that one possibly could regard as proof, in the sense of the neo-vitalistic conception, that the physico-chemical laws alone are not sufficient for the explanation of biological and psychic occurrences.

Smoluchowski, taken somewhat aback by the question, granted the neo-vitalist presumption in his reply:

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

He then recovered and proceeded to suggest that perhaps even an intelligent being is constrained by normal physics, so that some neglected physical process would still protect the second law. Two years later, with the publication of a follow-up lecture, Smoluchowski was more certain that even an intelligent demon could not escape the confines of physics. He wrote then (1914, 397) of the demon:

> For the production of any physical effect through the operation of its sensory and also its motor nervous system is always connected with an energy cost, leaving aside the fact that its entire existence is bound up with continuing dissipation.

Smoluchowski concluded that successful operation even of an intelligent demon was thus "very doubtful, even though our ignorance of living processes precludes a definite answer."

5 Szilard and his one-molecule engine

The question of intervention by an intelligent demon languished until it was taken up by Leo Szilard in the next decade. His 1929 "On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings" became the founding paper of a new literature. Szilard explained his limited ambitions in the introductory section. He had just completed his doctoral dissertation on the topic of thermal fluctuations. Citing this work, he announced as established that no automatic machine or no strictly periodic intervention is able to exploit thermal fluctuations to produce a violation of second law. His goal was limited (Szilard 1929, 302): "[We] intend here only to consider the difficulties that occur when intelligent beings intervene in a system."

In pursuit of this goal, Szilard devised the simplest possible manifestation of thermal fluctuations, one that proved especially well-suited to easy analysis. For an ideal gas of many, say n, molecules, the probability that the gas fluctuates isothermally from the full vessel volume V to some smaller, lower entropy, volume $(V–\Delta V)$ is negligible, unless the fluctuation in volume ΔV is small with respect to V/n.[1] Since n is of the order of 10^{24} for ordinary samples of gases, volume fluctuations in them are minuscule – of the order of one part in 10^{24}. They become more prominent as n becomes smaller. The extreme case is $n = 1$. It is a gas of a single molecule whose density fluctuates wildly as the molecule bounces rapidly to and fro inside the confining chamber.

Described macroscopically, when the molecule has moved to one side of the chamber, its volume has fluctuated into that side and had momentarily, spontaneously compressed in volume. Szilard's ingenious idea was a mechanism that could lock in this spontaneous fluctuation to a lower entropy state: simply insert a partition into the chamber, dividing it into two parts. The molecule will be trapped on one side and, with it, we have locked in a volume fluctuation to a lower entropy state.

In the simplest case, we divide the chamber volume in half. Then inserting the partition halves the volume of space occupied by the one-molecule gas. It has been compressed from its initial volume V to $V/2$. Since the thermodynamic entropy varies with volume for an ideal, one-molecule gas according to k log (volume), where k is Boltzmann's constant, the process has reduced the thermodynamic entropy of the gas by:

$$\Delta S = k \log (V/2) - k \log V = - k \log 2$$

We assume some apparently benign idealizations: the partition can be inserted without friction and is massless so no work is done in moving it. Thus the sole effect of this step is to reduce the thermodynamic entropy of the gas, without compensating changes elsewhere, in violation of the second law.

What remains is to accumulate this molecular-scale violation of the second law by incorporating the uncompensated compression into a cycle that can be repeated indefinitely. Szilard devised a process that would accumulate these repeated entropy reductions in the entropy of an environmental heat bath that maintains the one-molecule gas at a constant temperature T.

To do this, the partition becomes a piston against which the confined one-molecule gas exerts a pressure, due to repeated impacts. The piston is coupled to a weight in such a way that, when the one-molecule gas expands isothermally, the work done by the gas on

the piston raises the weight. As the gas expands, it cools slightly due to the loss of energy transmitted as work to the raised weight. This lost energy is replaced by heat conducted to the one-molecule gas from the heat bath. Since the pressure exerted by the one-molecule gas is $P = kT/V$, the work W done by the gas in the expansion and the heat passed to the gas Q are given by[2]

$$Q = W = \int_{V/2}^{V} P dV = \int_{V/2}^{V} \frac{kT}{V} dV = kT \log \frac{V}{V/2} = kT \log 2$$

An essential condition on the cycle is that all processes be carried out in a thermodynamically reversible manner, else there will be unwanted thermodynamic entropy created in the process, undermining the goal. If heat $Q = kT \log 2$, passes reversibly from the heat bath, then its thermodynamic entropy is reduced by

$$\Delta S = -Q/T = -k \log 2$$

At the end of the cycle, the gas is restored to its initial state and the thermodynamic entropy of the heat bath has been reduced by $k \log 2$. This is a net reduction in the thermodynamic entropy of the universe, in violation of the second law. Merely repeating the cycle at will yields an arbitrarily large violation. It is a one-molecule heat engine that fully converts heat drawn from the heat sink into the work energy stored in the raised weight.

6 Szilard's principle: the entropy cost of measurement

A violation of the second law seems all but assured by this simple mechanism. How could it fail? Szilard had an answer. The account above neglects the presence of the agent operating the one-molecule engine. Successful operation requires that, upon insertion of the partition, the agent must discern on which side – left or right – the molecule is trapped. For that measurement is needed in determining which coupling to use for the pressure-driven raising of the weight. Does the gas expand to the left or the right?

Szilard's escape was to assume that this measurement requires the creation of entropy (1929, 303):

> We shall realize that the Second Law is not threatened as much by this entropy decrease as one would think, as soon as we see that the entropy decrease resulting from the intervention would be compensated completely in any event if the execution of such a measurement were, for instance, always accompanied by production of $k \log 2$ units of entropy.

The direct way of arriving at this entropy cost of measurement is simply to work backwards from the assumption that the second law is not violated, as Szilard notes (302):

> At first we calculate this production of entropy quite generally from the postulate that full compensation is made in the sense of the Second Law ...

This is an awkward moment. The demonstration of the necessary failure of the demon begins with the *assumption* of its failure? To escape the obvious circularity, Szilard devoted an extended part of his analysis to a specific, independent scheme for carrying out the measurements. The details of that scheme have proven to be opaque to later commentators and it has played no role in subsequent developments.[3] See Leff and Rex (2003, §1.3).

What survived in the subsequent literature was a much simpler version of Szilard's escape. John von Neumann, Szilard's fellow student in Berlin and Hungarian compatriot, included the simplified version in his authoritative and much admired text, *Mathematical Foundations of Quantum Mechanics* (1932). Having recounted the operation of Szilard's one-molecule engine, he stressed the importance of knowing the location of the trapped molecule for the cycle to be completed. He concluded (400): "That is, we have exchanged our knowledge for the entropy decrease $\kappa \ln 2$ [here, $k \log 2$]." To this sentence, von Neumann appended the footnote:

> L. Szilard has [reference] shown that one cannot get this "knowledge" without a compensating entropy increase $\kappa \ln 2$. In general, $\kappa \ln 2$ 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 to be invalid, although they may occasionally lead to very complicated automatic mechanisms.

Modern readers will see in this talk of "alternative of two cases" and the thermodynamic value of knowledge an immediate connection to Shannon and Weaver's mathematical theory of communication. There, information in a communication channel is quantified as information entropy, using the same formula in the probability calculus as can be used to compute the thermodynamic entropy of a thermal system. Shannon's theory, however, comes well after Szilard's work. Its founding paper was published in 1948 and, in a more popular introductory discussion to the canonical *The Mathematical Theory of Communication*, Weaver (1964, 3) cites both Szilard's and von Neumann's earlier work as precursors.

In the 1950s and 1960s, the idea of a connection between thermodynamic entropy and information entropy drew energetic attention.[4] The leading idea of this literature was subsequently labeled "Szilard's Principle": the gaining of information that allows us to discern among n equally likely states is associated with the creation of a minimum of $k \log n$ of thermodynamic entropy (Earman and Norton 1999, 5).

This was an era of many thought experiments illustrating how an entropy cost in information acquisition leads to the failure of a Maxwell's demon. A simple and popular example was provided by Leon Brillouin (1950), one of the founders of modern, solid state physics, in the Brillouin torch. It supplied the missing details in a quantum mechanical account of how the demon could locate a molecule: the demon would bounce a photon – a quantum of light – off it. The photon, however, must be sufficiently energetic for it to be visible above the thermal background. This condition forced so much entropy creation that it precluded the demon exploiting the information gained to violate the second law.

A celebratory headnote to the 1964 translation of Szilard's (1929) article in the journal *Behavioral Science* affirmed Szilard's priority:

> This is one of the earliest, if not the earliest paper, in which the relations of physical entropy to information (in the sense of modern mathematical theory of communication) were rigorously demonstrated and in which Maxwell's famous demon was successfully exorcised: a milestone in the integration of physical and cognitive concepts.

7 Landauer's principle: the entropy cost of erasure

This last explanation of the necessary failure of Maxwell's demon was short-lived. Within a few decades, it had been eclipsed by another exorcism. Bennett (1987, 108) reported this unexpected turn in a popular article in *Scientific American*:

> To protect the second law, physicists have proposed various reasons the demon cannot function as Maxwell described. Surprisingly, nearly all these proposals have been flawed. Often flaws arose because workers had been misled by advances in other fields of physics; many of them thought (incorrectly, as it turns out) that various limitations imposed by quantum theory invalidated Maxwell's demon.
>
> The correct answer – the real reason Maxwell's demon cannot violate the second law – has been uncovered only recently. It is the unexpected result of a very different line of research: research on the energy requirements of computers.

The new exorcism was founded on an idea proposed by Rolf Landauer (1961). When a computer memory is erased, heat is generated; and the heat generated creates entropy in the environment to which it passes. The amount created can be quantified under the rubric of "Landauer's principle": erasing a memory device with n states leads to the creation of $k \log n$ of thermodynamic entropy in the environment.

In the new exorcism, as elaborated in Bennett (1982, §V; 1987), the key fact was that the demon must record or remember the location of the molecule in operating a machine like the Szilard one-molecule engine. Left or right? One of the two states must recorded. Then the record must be erased if the demon is to complete the cycle and return to its initial state. That erasure, according to Landauer's principle, will create $k \log 2$ of thermodynamic entropy in the environment, cancelling precisely the $k \log 2$ entropy reduction of the earlier steps of the cycle.

The new exorcism moves the locus of creation of the compensating entropy from information acquisition to information erasure. Hence its cogency depends essentially on the possibility of processes that could acquire information without creating thermodynamic entropy. Szilard, von Neumann, Brillouin and others who worked in the earlier tradition gave powerful affirmations that this was impossible. Yet Bennett (1982, §V; 1987) now described several measurement devices that purportedly required no entropy creation. (Awkwardly, it was obvious to anyone who had absorbed the import of Smoluchowski's original analysis, that none of these devices could carry out dissipationless measurement. As noted in Earman and Norton (1999, 13–14), their delicate internal mechanisms would be fatally disrupted by thermal fluctuations.)

In any case, as the survey in Leff and Rex (2003) records, the new erasure-based exorcism now dominates a flourishing literature. In 2012, Bérut et al. (2012), in a letter to the

prestigious journal *Nature*, announced experimental confirmation of Landauer's principle and summarized its importance as:[5]

> Landauer's principle hence seems to be a central result that not only exorcizes Maxwell's demon, but also represents the fundamental physical limit of irreversible computation.

PART II: FALL

8 The worst thought experiment

The narrative so far has replicated the celebratory tone of standard accounts of the Szilard thought experiment. I do not share in the celebration. Rather I believe that this thought experiment is fatally flawed. It is the origin and an enduring stimulus for long-lived confusions in the subsequent literature. We have already seen a clue to the serious problems to be recounted below: the tradition of exorcism based on Szilard's principle simply collapsed when a new contender emerged. There was no new experimental result. There was merely an imperfectly supported pronouncement that decades of pronouncements by earlier, leading thinkers were wrong.

In its capacity to engender mischief and confusion, Szilard's thought experiment is unmatched.[6] It is the worst thought experiment I know in science. Let me count the ways it has misled us.

8.1 A worse solution to the Maxwell demon problem

The stated goal of Szilard's thought experiment was to treat the case not of an automated Maxwell's demon, but of an intelligent demon. We saw above that Smoluchowski had already dealt with that case quite effectively. If the intelligent demon has neo-vitalist powers that put it outside normal physics, then what it can do lies beyond what physical theory can determine. It can break physical laws since it is outside them. If however it is naturalized as a physical system,[7] then its proper functioning requires thermodynamic dissipation; and, if its mechanism is delicately balanced, thermal fluctuations will disrupt it. This basic idea is essentially correct. It can be developed into a much stronger exorcism, as shown in Subsections 8.2 and 8.5.

In its place, Szilard created a vague notion of an information gathering or measuring system, whose behavior will be all but impossible to quantify outside oversimplified and contrived cases like the one-molecule engine.

Almost immediately, this new approach became the only mode of exorcism in the literature. Information processing was the key, supposedly, to explaining why all Maxwell's demon must fail. Yet there are many proposals for Maxwell's demons to which this analysis cannot be applied. How can it explain the failure of the Smoluchowski trapdoor? How much information does the trapdoor gather? Where does this simple trapdoor mechanism store the information whose erasure is key to its failure? The information-based exorcism fails where Smoluchowski's simple observation of the disruptive effect of fluctuations succeeds.

8.2 Inadmissible idealizations that selectively neglect fluctuations

In the transition from the exorcism based on Szilard's principle to one based on Landauer's principle, the governing question was just which process is *the* ineliminable locus of dissipation. Measurement or erasure? The common assumption was that all the other processes could, in principle, be carried out without creating thermodynamic entropy and thus could be idealized as dissipationless processes.

This assumption is the most egregious of all assumptions in this literature. It is fatally and disastrously wrong. Thermal fluctuations are present in the systems of every step of the thought experiment; and dissipative, thermodynamic entropy creating processes are needed to suppress them and allow the step to complete. Since Szilard's one-molecule engine is driven by thermal fluctuations, to ignore them in some places but to depend on their presence in others is to render the whole analysis an inconsistent use of thermal physics.

We will see the need for dissipative processes to suppress fluctuations throughout the operation of the engine, first in an example and then in a quite general result. Consider the partition that is slid into the chamber to divide it. The partition is routinely assumed massless and to slide frictionlessly, so that no work is needed to move it. However it is just as much a component in the thermal system as is the molecule of the gas. Like the Smoluchowski trapdoor, it will have its own thermal energy as result of its thermal interactions with the environment. If it slides in one dimension, the equipartition theorem of classical statistical mechanics assigns it a mean thermal energy of $kT/2$. That will manifest as a jiggling motion, akin to Brownian motion. The partition can be slid into place and locked only by overcoming these motions, which requires the application of an unbalanced, dissipative force whose work energy is degraded into heat. Rendering the partition massless will make the operation more difficult. For its mean thermal energy $kT/2$ will equal its mean kinetic energy $mv^2/2$, where m is the partition mass and v its root mean square velocity. As we make the mass m small without limit, then the velocity v increases without limit, if the mean kinetic energy is to remain constant. That is, a near massless partition would be moving with near infinite speed!

This analysis could be continued for each step of the Szilard's cycle: the detection of the molecule; the coupling of the weight to the partition-piston; the expansion of the one-molecule gas and the raising of the weight; and the decoupling and the removal of the partition-piston.[8] In each case we would find that dissipative forces are needed to overcome fluctuations.

Fortunately, we do not need to labor through every case. There is a general result that covers them all. See Norton (2013, §9). On molecular scales, thermal fluctuations prevent assured completion of any process. There is always a chance that some fluctuation will undo the process. However a dissipative process that creates entropy ΔS can enhance the probability P_{fin} that the system will be in its desired final state, as compared to the probability P_{init} that the system has fluctuated back to its initial state. The three quantities are related by

$$\Delta S = k \log(P_{fin}/P_{init})$$

The quantities of entropy that this formula requires are large on molecular scales. If we ask only for a meager ratio favoring success of $P_{fin}/P_{init} = 20$, then the theorem tells us that, at

minimum, we must create $k \log 20 = 3k$ of thermodynamic entropy. This minimum quantity exceeds the $k \log 2 = 0.69k$ tracked in the Szilard's and Landauer's principle literature. This minimum quantity must be created not just once, but once for each of the many completing steps in the cycle.

8.3 The exorcisms rely on tendentious principles

Smoluchowski was right: fluctuations in the demon's mechanism will defeat it. Suppressing them will require creation of quantities of entropy that far exceed those tracked by Szilard's and Landauer's principle. This simple fact renders the principles insignificant in the analysis of Maxwell's demon, even if they are correct principles.

There are good reasons, however, to doubt their correctness. The most common derivation of Szilard's principle is to work backwards to arrive at the precise quantitative measure of entropy creation, $k \log n$, when discerning among n outcomes (see Earman and Norton 1999, §2.1). One assumes that the second law prevails in systems like the Szilard one-molecule engine and works backwards to the entropy that must be created in the measurement to protect the law. It is assumed – crucially – that measurement is the only step that must create entropy. Since this last crucial assumption fails, so does the derivation. Other, less precise demonstrations, such as the Brillouin torch, in effect exploit the general fact of Section 8.2 that any process on molecular scales that completes must be dissipative. That the process is a measurement is incidental to the dissipation inferred.

Landauer's principle, when it was first suggested in 1961, was an interesting speculation, founded on a loose plausibility argument, but in need of precise grounding. Over half a century later, in spite of considerable efforts, the principle remains at best loose speculation, grounded in many repetitions of the same misapplications of thermal physics. The details are too complex to be elaborated here. Discussion of these difficulties and an entrance in the broader literature, can be found in Norton (2011, 2013, §3.5).

8.4 The thought experiment legitimated a bad exemplar

Szilard's 1929 thought experiment introduced into the literature an exemplar for how thermodynamic analysis can be carried out on systems at molecular scales. It legitimated the idea that many processes can be effected reversibly, that is, dissipationlessly, including the insertion and removal of partitions into a one-molecule gas chamber and the reversible compression and expansion of the one-molecule gas. The literature, especially on Landauer's principle, is replete with manipulations of this type. A single molecule trapped in one or other side of a chamber has become the canonical example of a molecular-scale memory device; and it is manipulated by all the above processes. For a recent example, see Ladyman, Presnell, Short, and Groisman (2007); Ladyman, Presnell, and Short (2008); and for my critique, see Norton (2011).

The difficulty is that none of these processes can be carried out without dissipation. Hence the entire analytic regime is flawed, as is any result derived within it. The harm caused by Szilard's exemplar is not limited to the analysis of a one-molecule gas. The thermal properties of a one-molecule gas are analogous to those of other single component systems that may be used as memory or detection devices, such as a molecular-scale magnetic dipole in a thermal environment. The position of the molecule is analogous to the

direction of the dipole moment; and the compression of the one-molecule gas by a piston is analogous to the restriction of the direction of the dipole by an external magnetic field. The restriction is a compression in an abstract state space. It is routine to assume that the analogous operations on the dipole can be carried out dissipationlessly (see, for example, Bennett 1982, §5). These operations are, of course, equally disrupted by thermal fluctuations. No molecular-scale processes can be completed on them without dissipation.

8.5 Distraction from a far simpler exorcism: Liouville's theorem

The idea that information processing in some guise is the key to demonstrating the failure of Maxwell's demon has great popular appeal. It has, as a result, come to dominate virtually all writing on Maxwell's demon and has been responsible for an explosion of feckless analysis. It is only now becoming clear how thoroughly this seductive idea has misdirected us. Had we maintained Smoluchowski's focus on the mundane statistical physics, we might much earlier have hit upon a remarkably simple and quite general exorcism of Maxwell's demon. It requires nothing more than elementary notions in statistical physics and can be developed without even mentioning the sometimes troublesome notion of entropy. It turns out that a simple description of what a Maxwell's demon is required to do is incompatible with the Liouville theorem of Hamiltonian dynamics. Since this theorem is fundamental to classical statistical physics, it assures us that no Maxwell demon is possible. See Norton (2013, §4); and for a quantum theoretic version of the same exorcism, Norton (forthcoming).

9 How a thought experiment can fail

Were they not delivered through the medium of a thought experiment, I like to think that Szilard's misleading and troublesome speculations of 1929 would not have received assent. For his paper has no cogent demonstration that, as a general matter, an intelligent demon must fail because of entropy costs associated quite specifically with measurement, as quantified in his formulae. There is something about the medium of a thought experiment that induced this assent.

That something derives from the special narrative role that thought experiments play in science and especially physics. Results there are often quite complicated and hard to grasp, if displayed in precise terms and full generality. Here thought experiments have traditionally played a pedagogical role. They give us just the essentials in a form that is easy to visualize and easy to grasp. In return, we are willing to grant the thought experimenter considerable latitude. The failures of Szilard's thought experiment can be traced to a misuse of this latitude. It comes it two forms.

9.1 Hasty acceptance of idealizations

First, we allow many processes to be idealized away as nuisance distractions, so that we may focus just on the one that matters. Einstein (1952, Ch. XX; Peacock, this volume) asks us to image a large chest in a remote part of space with an observer inside. It is accelerated, he says, by "a 'being' (what kind is immaterial to us)," who pulls with constant force on a rope

attached to the lid, uniformly accelerating the chest. We suspend scepticism over whether such an idealized being really is possible. We accept Einstein's assurance that questioning this detail is an unhelpful distraction.

Szilard misuses this liberty in his thought experiment. It is simply assumed that most of the operations of his one-molecule engine can be carried out without dissipation. He briefly addresses the fluctuations ("agitations") that must be present when a piston, light enough to be raised by a collision with a single molecule, is repeatedly struck by one. He writes (1929, 304–5):

> It is best to imagine the mass of the piston as large and its speed sufficiently great, so that the thermal agitation of the piston at the temperature in question can be neglected.

This scheme would suppress the manifestation of molecular fluctuations only because a fast-moving, massive piston is in a state far from the equilibrium that non-dissipative processes demand.[9] Since thermal fluctuations are the primary subject, we should expect a more cogent defense of an idealization that eventually proves to be fatal. We are prompted by the narrative conventions of a thought experiment, however, not to press when the thought experimenter assures us that the idealization is benign.[10]

9.2 Hasty generalization

Instead of proving a result in all generality, a thought experiment may merely display typical behaviour. To enable a simple narrative, we accept that it is typical – a simplified surrogate for a complicated general statement.

The observer in Einstein's chest finds bodies inside it to fall just as if they are in a gravitational field. It is a striking coincidence and we are led immediately to connect acceleration and gravitation. The point of the thought experiment is that this is no mere coincidence, but a manifestation of a broader generality: all processes in a gravitational field proceed just as they do in the chest. Einstein immediately uses this generalization to infer that clocks deeper in a gravitational field are slowed, because clocks lower in the chest are slowed; and that light is bent by a gravitational field, since its propagation is bent inside the chest.

Without the thought experiment, we might well be reluctant to admit the generalization. Surely we should be more circumspect in jumping from the fall of ordinary bodies to the rates of clocks and the bending of light. However, we are carried along by the fictions in the narrative. Imagining ourselves with the observer in the chest, we agree that things are just as if we are in a gravitational field, as far as the motion of bodies is concerned. We have no means to know otherwise. How could we know that things are any different for other processes? Even if doubts linger, we conform with the narrative convention and grant that the thought experimenter knows which results are typical and thus support generalization.

Szilard's thought experiment misuses this latitude. The bare operation of the one-molecule engine leads to a reduction in thermodynamic entropy of $k \log 2$. We notice immediately, with von Neumann, that the two in the formula matches the count of alternatives of left and right available to the operating demon. Then, Shannon attaches $k \log 2$

of information entropy to a choice between two equally likely signals. Remarkably, the quantities of thermodynamic entropy and information entropy match.

This is a striking coincidence and seeing it is a memorable moment, when one first encounters the thought experiment. It is so striking that we readily accept the thought experimenter's suggestion that it is no mere coincidence. It is, we are to believe, a manifestation of a deeper generality. The entropy reduction produced by the machinations of Maxwell's demon will, in general, be compensated by entropy created in manipulating information, either acquiring it or erasing it. We readily assent because we presume that the thought experimenter is simplifying a deeper analysis, whose full details are suppressed by a narrative convention in the interests of preserving the simplicity of the thought experiment.

Alas, the latitude we accord to the thought experimenter in this case is misplaced. Aside from a few variant forms of the thought experiment with slight elaborations, there is no cogent general theory whose details are being suppressed for simplicity. We have been induced to make a faulty generalization.

10 Conclusion

In the standard view, Szilard's thought experiment was the initiating stimulus for a new tradition in physics. Until it drew attention to the role of information, we are supposed to believe, it was impossible to understand the necessity of failure of Maxwell's demon. The thought experiment illustrated how a quantitative relation is possible between information acquired or, later, erased and thermodynamic entropy. The principle that governs this relation, whichever it might be, forms the foundation of a new science of the thermodynamics of information or computation, with the exorcism of Maxwell's demon its signal achievement.

Alas, this standard view is a mirage and illusion. Szilard's novel response to Maxwell's demon added nothing useful to the superior analysis already given by Smoluchowski over a decade before. Instead, that older, better analysis was preempted by the popular appeal of the apparently paradoxical connection of information and thermodynamic entropy. What followed were expanding efforts to make precise ideas that are, at their heart, sufficiently vague and flawed as to admit no cogent development. The core of Szilard's thought experiment, his one-molecule engine, was an inconsistent muddle of improper idealizations. Yet it became the workhorse of new theorizing, ensuring that new ideas derived with its help, such as Landauer's principle, inherited its flaws.

The power of a thought experiment in physics lies in its capacity to focus on just the most essential. It can do that since the experiment is conducted purely in thought, where inessential distractions can be eradicated under the guise of idealizations; and one representative example can speak simply for the generality. For the activity to succeed, we must give the thought experimenter considerable latitude in deciding just which processes are the instructive, representative examples and just which can be idealized away as inessential. When that latitude is exercised well, we gain wonderful illumination. However that latitude can be misused. Through it, we may be induced to accept assumptions, hidden in the picturesque scenario, that we would never accept were they made explicit in a less picturesque environment. Szilard's thought

experiment is a powerful example of how this latitude can be misused. It is, in my view, unparalleled in science in the mischief it has caused. It is the worst thought experiment in science.

Notes

1 This follows since the probability of the fluctuation is $[(V\text{-}\Delta V)/V]^n \approx [1 - \Delta V/(V/n)]$ for large n and small $\Delta V/V$.
2 This equality Q =W follows since the internal energy of the one-molecule gas remains unchanged in the isothermal process; it is a function of temperature only.
3 Similarly, Szilard concluded that the k log 2 of entropy produced is an average only, so that the second law is preserved only in the averaging of many cycles.
4 For a convenient survey and compilation of papers, see the two editions Leff and Rex (1990 and 2003). Note that the earlier edition contains more material pertinent to this period than the later edition.
5 The claim of experimental confirmation is mistaken. See Norton (2013, §3.7).
6 Thought experiments in science are generally illuminating or, at the least, benign. It is not so with thought experiments in philosophy. They are a locus of misdirection and deception. We are supposed to derive important conclusions about fundamental matters from bizarre imaginings of zombies, who behave exactly like conscious humans, but are not conscious; or of substances that share exactly all the physical properties of water, but are not water. The narrative conventions of a thought experiment authorize us to contemplate hokum that would otherwise never survive scrutiny.
7 Szilard explicitly adopts this case, writing (1929, 302): "We may be sure that intelligent living beings – insofar as we are dealing with their intervention in a thermodynamic system – can be replaced by non-living devices whose 'biological phenomena' one could follow ..."
8 For detailed computation of these effects for the gas expansion, see Norton (2017).
9 For more discussion, see Norton (2013, §7).
10 This analysis conforms with El Skaf's (2016, ch. 6) account of how thought experiments can evolve. Tacit possibility claims in the background of the scenario are subsequently identified and brought to the foreground, where they are challenged.

References

Bennett, C. H. (1982) "The thermodynamics of computation – a review," *International Journal of Theoretical Physics* 21: 905–940.

Bennett, C. H. (1987) "Demons, engines and the second law," *Scientific American* 257: 108–116.

Bérut, A., Arakelyan, A., Petrosyan, A., Ciliberto, S., Dillenschneider, R. and Lutz, E. (2012) "Experimental verification of Landauer's principle linking information and thermodynamics," *Nature* 48: 187–190.

Brillouin, L. (1950) "Maxwell's demon cannot operate: Information and entropy I," in *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*, edited by H. S. Leff and A. Rex (pp. 120–123), Philadelphia: Institute of Physics Publishing.

Earman, J. and Norton, J. D. (1998, 1999) "Exorcist XIV: The wrath of Maxwell's demon," *Studies in the History and Philosophy of Modern Physics*, Part I "From Maxwell to Szilard," 29 (1998), 435–471; Part II: "From Szilard to Landauer and beyond," 30 (1999), 1–40.

Einstein, A. (1952) *Relativity: The Special and the General Theory*, New York: Bonanza.

El Skaf, R. (2016) *La Structure des Expériences de Pensée Scientifiques*, Ph.D. thesis, Université Paris I Panthéon-Sorbonne.

Ladyman, J., Presnell, S. and Short, A. J. (2008) "The use of the information-theoretic entropy in thermodynamics," *Studies in History and Philosophy of Modern Physics* 39: 315–324.

Ladyman, J., Presnell, S., Short, A. J. and Groisman, B. (2007) "The connection between logical and thermodynamic irreversibility," *Studies in the History and Philosophy of Modern Physics* 38: 58–79.

Landauer, R. (1961) "Irreversibility and heat generation in the computing process," *IBM Journal of Research and Development* 5: 183–191.
Leff, H. S. and Rex, A. (eds) (1990) *Maxwell's Demon: Entropy, Classical and Quantum Information, Computing*, Bristol: Adam Hilger.
Leff, H. S. and Rex, A. (eds) (2003) *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*, Philadelphia: Institute of Physics Publishing.
Maxwell, J. C. (1871) *A Theory of Heat*, London: Longmans, Green and Co.
Norton, J. D. (2011) "Waiting for Landauer," *Studies in History and Philosophy of Modern Physics* 42: 184–198.
Norton, J. D. (2013) "All shook up: Fluctuations, Maxwell's demon and the thermodynamics of computation," *Entropy* 15: 4432–4483.
Norton, J. D. (2017) "Thermodynamically reversible processes in statistical physics," *American Journal of Physics* 85: 135–145.
Norton, J. D. (forthcoming) "Maxwell's demon does not compute," in *Physical Perspectives on Computation, Computational Perspectives on Physics*, edited by M. E. Cuffaro and S. C. Fletcher, Cambridge: Cambridge University Press.
Shannon, C. E. and Weaver, W. (1964) *The Mathematical Theory of Communication*, Urbana, IL: University of Illinois Press.
Smoluchowski, M. (1912) "Experimentell nachweisbare, der üblichen Thermodynamik widersprechende Molekularphänomene," *Physikalische Zeitschrift* 13: 1069–1080.
Smoluchowski, M. (1914) "Gültigkeitsgrenzen des Zweiten Hauptsatzes der Warmetheorie," in *Vorträge über die Kinetische Theorie der Materie und der Elektrizität*, Berlin: B. G. Teubner; reprinted in *Oeuvres de Marie Smoluchowski*, Cracow: Jagellonian University Press (1927).
Szilard, L. (1929) "Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen," *Zeitschrift für Physik* 53: 840–56; translated as "On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings," *Behavioral Science* 9: 301–310, and *The Collected Works of Leo Szilard: Scientific Papers*, Cambridge: MIT Press (1972).
von Neumann, J. (1932) *Mathematische Grundlagen der Quantenmechanik*, Berlin: Julius Springer; translated by R. T. Beyer as *Mathematical Foundations of Quantum Mechanics*, Princeton: Princeton University Press.

THE ROUTLEDGE COMPANION TO THOUGHT EXPERIMENTS

Edited by
Michael T. Stuart, Yiftach Fehige
and James Robert Brown

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published in 2018
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book has been requested

ISBN: 978-0-415-73508-7 (hbk)
ISBN: 978-1-315-17502-7 (ebk)

Typeset in Goudy
by Book Now Ltd, London