Entropy, information and Szilard's paradox

Jauch, Joseph-Maria; Báron, J.G.

How to cite

JAUCH, Joseph-Maria, BÁRON, J.G. Entropy, information and Szilard's paradox. In: Helvetica physica acta, 1972, vol. 45, n° 2, p. 220–232. doi: 10.5169/seals-114379

This publication URL: https://archive-ouverte.unige.ch/unige:162164

Publication DOI: 10.5169/seals-114379

Helvetica Physica Acta
Vol. 45, 1972

Birkhäuser Verlag Basel

Entropy, Information and Szilard's Paradox

by **J. M. Jauch**

Dept. of Theoretical Physics, University of Geneva,
and Dept. of Mathematics, University of Denver

and **J. G. Báron**

Rye, New York

(15. XII. 71)

This essay is presented in homage to Professor Markus Fierz, whose long-standing interest in statistical physics is well known, on the occasion of his 60th birthday.

Abstract. Entropy is defined as a general mathematical concept which has many physical applications. It is found useful in classical thermodynamics as well as in information theory. The similarity of the formal expressions in the two cases has misled many authors to identify entropy of information (as measured by the formula of Shannon) with negative physical entropy. The origin of the confusion is traced to a seemingly paradoxical thought experiment of Szilard, which we analyze herein. The result is that this experiment cannot be considered a justification for such identification and that there is no paradox.

1. Introduction

There is a widespread belief that the physical entropy used in thermodynamics is more or less closely related to the concept of information as used in communication theory.

This thesis has been made precise and explicit, primarily by Brillouin [1], who is of the opinion that both concepts should be united by identifying information (suitably normalized) by establishing an equivalence relation with negative physical entropy (called 'negentropy' by him), which then together satisfy a generalized principle of Clausius.

This point of view, however, is not universally accepted by those physicists who have thought about the question. We quote here as an example an explicit denial of such identification, by ter Haar [2], who writes in his textbook on statistical mechanics:

> 'The relationship between entropy and lack of information has led many authors, notably Shannon, to introduce "entropy" as a measure for the information transmitted by cables and so on, and in this way entropy has figured largely in recent discussions on information theory. It must be stressed here that the entropy introduced in information theory is *not* a thermodynamic quantity and that the use of the same term is rather misleading. It was probably introduced because of a rather loose use of the term "information".'

We want to elaborate ter Haar's point of view and discuss the reasons why we believe that the two concepts should not be identified.

One can trace the origin of this identification to a paper by Szilard [3], published in 1929, which discusses a particular version of Maxwell's demon and an apparent violation of the second law of thermodynamics.

The emphasis in that paper is on the intelligence of the 'demon', who, by utilizing the 'information' gained by observation of the detailed properties of a thermodynamic system, could use this information for the manipulation of a macroscopic gadget which could extract mechanical energy from the fluctuations of a thermodynamic system and thus produce a *perpetuum mobile* of the second kind.

Szilard based his version on a remark by Smoluchowski which was published in the latter's lectures on the kinetic theory of matter [4]. Smoluchowski said, 'As far as our present knowledge is concerned there does not exist a permanently working *automatic perpetuum mobile* in spite of molecular fluctuations, but such a contraption could function *if it were operated by intelligent beings in a convenient manner . . .'* (italics ours).

This statement seems to imply that the second law of thermodynamics could somehow be violated in the presence of intelligent beings and that this possible violation would be associated with the acquisition and retention of knowledge by such beings. It is with this in mind that Szilard contructed his thought experiment.

Although in a subsequent passage, Smoluchowski expressed considerable doubt ('recht zweifelhaft') about this possibility, Szilard proposed to elucidate the conjectured role of the intelligent being in creating the uncompensated entropy decrease. He described an idealized heat engine that seemingly functioned with continuous decrease of entropy. In order to save the second law, Szilard conjectured that the intelligent being (we shall call him 'the observer') must perform measurements in order to operate the engine, and that this process is in principle connected with a compensating increase of entropy.

We shall discuss Szilard's thought experiment in section 4 of this paper. Here we merely point out that this experiment provoked much discussion and, in our opinion, misinterpretation. We mention in particular the discussion by von Neumann [5], who transferred considerations of this kind into the realm of quantum mechanics with reference to the measuring process.

Many aspects of the measuring process in quantum mechanics are still controversial. Szilard's conjecture mentioned above has led many commentators [6] to believe that the measuring process in quantum mechanics is connected in an essential manner with the presence of a conscious observer who registers in his mind an effect, and that this conscious awareness is responsible for the oft-discussed, paradoxical 'reduction of the wave packet'.

We expect to show that the presence of a conscious observer in Szilard's experiment is not necessary; he can be replaced by an automatic device with no consciousness at all. Interestingly, Szilard noted this himself; toward the end of his paper, he concluded:

> 'As we have seen with this example, a simple, inanimate device can do exactly the same, as far as the essentials are concerned, as the intervention of an intelligent being would accomplish.'

It is strange that Szilard seemed not to realize that an automatic version of the intelligent observer contradicts the conclusion of Smoluchowski, according to which such mechanisms are not possible. As a matter of fact, the solution of the paradox in the case of the living observer is the same as that which Smoluchowski indicated for the explanation of the mechanical demon: The demon is himself subject to fluctuations, just as the system which he tries to control. To use a medical analogy, the demon who wants to operate the molecular trap is like a patient with a severe case of Parkinson's disease trying to thread a fast-vibrating needle!

We shall not question the analysis of the problem given by Smoluchowski; we shall consider this aspect of the problem as solved. From this it follows that Szilard's conjecture is not proven by his experiment.

2. The Classical Notion of Entropy

Entropy as a basic notion of science was introduced by Clausius to summarize thermal behavior of systems in equilibrium or changing in reversible fashion in the second principle of thermodynamics.

Boltzmann [8] and Gibbs [9] defined entropy of non-equilibrium states and entropy changes of irreversible processes in purely mechanical terms. Their theory was more general; it also explained how the same thermodynamic process can be irreversible from the phenomenological point of view—and completely reversible from the purely mechanical point of view. This paradoxical situation was cleared up by statistical interpretation of thermodynamic entropy.

Increase in generality resulted in some ambiguity of the notion of entropy. The reason for this is that in any statistical consideration a more or less arbitrary model must be used. Expressed differently, the system may be described at different levels (see H. Grad [7]).

We shall return to the significance of these ambiguities later in this section. First we briefly review some special features of the statistical interpretation of thermodynamic entropy.

In thermodynamics one may specify a homogeneous thermal system by a certain number of extensive variables $x_1, \dots x_n$ which usually have a simple physical interpretation (volume, surface, magnetic moment, etc.).

The generalized forces $y_1, \dots y_n$ associated with these variables are homogeneous functions of them, such that the element of work δA delivered by the system to the surrounding is related to the differentials dx_r $(r = 1, \dots n)$ by

$$\delta A = \sum_{r=1}^{n} y_r \, dx_r. \tag{1}$$

Mathematically, (1) is a differential form defined on an open region of $\mathbb{R}^n$, the Euclidean space of n dimensions.

The first principle of thermodynamics which expresses conservation of energy for a conservative system states that for an adiabatic system (that is, a thermally isolated system), this differential form is total. That means that there exists a function $U(x_1, \dots x_n)$ of the extensive variables x_r, which is itself extensive, such that

$$\delta A = -dU \tag{2}$$

Physically interpreted, this equation says that the work delivered to the outside by an adiabatic system is exactly compensated by the loss of internal energy.

If the system is not adiabatic, then equation (2) is no longer true and must be generalized to

$$\delta Q = dU + \delta A \tag{3}$$

where now δQ is the differential of the amount of heat added to the system in a reversible manner.

We may consider equation (3) as a new differential form in $n+1$ variables where $U = x_0$ may be defined as the new variable. Each of the generalized forms y_r is then a function of all the variables $x_0, x_1, \ldots, x_n$.

The differential form (3) is of a special kind which admits an integrating factor $T(x_0, x_1, \ldots x_n)$ such that the form

$$dS = \frac{\delta Q}{T} \tag{4}$$

is the total differential of an extensive function $S(x_0, x_1, \ldots x_n)$. This function is the entropy and the integrating factor (suitably normalized) is the absolute temperature of the system.[1])

The second principle of thermodynamics says that in a spontaneous evolution of a closed system not in equilibrium, the entropy always increases and attains its maximum value for the state of equilibrium.

Definition (4) determines the thermodynamic entropy only up to a constant of integration.

Boltzmann's statistical definition is given by the famous formula,

$$S = k \ln W \tag{5}$$

where W represents a probability for the system specified by the thermodynamic variables based on some appropriate statistical mode. This formula was apparently never written down by Boltzmann; yet it appears on his tombstone and indeed is one of the most important advances in statistical physics. For the practical application of this formula, one usually goes through the following procedure:

a) One assumes (explicitly or implicitly) an a priori probability. In phase space of a classical system it is given by a convenient selection of a volume element.

b) One then imposes constraints in agreement with a certain number of external parameters characterizing the thermodynamic state of the system.

c) One then calculates the probability of such constrained systems on the basis of the a priori probability field assumed.

d) Finally, one calculates a maximum value of this probability under the assumed constraints to obtain an expression for W.

The W thus calculated is in an arbitrary normalization. This arbitrariness corresponds to the constant of integration for the thermodynamic entropy S.

Boltzmann's theoretical interpretation of the entropy gives immediate insight into two important properties which are characteristic for the thermodynamic, and, as we shall see, for all other forms of entropy. They are:

a) *Extensity*

If there are two independent systems with their respective probabilities W_1 and W_2, then the joint system has a probability

$$W = W_1 W_2. \tag{6}$$

---

[1]) It seems not to be generally known that the existence of the integrating factor, hence the existence of the function entropy, is a consequence of the first principle of thermodynamics for conservative systems under reversible quasistatic variations. This was discovered by T. Ehrenfest [12]. A new proof of this statement will be given in a subsequent publication.

Hence

$$S = S_1 + S_2 = k \ln (W_1 W_2). \tag{7}$$

b) *Maximum property*

Any system outside the equilibrium state will have a probability $W < W_0$, the equilibrium probability.
Hence

$$S = k \ln W < k \ln W_0 = S_0, \tag{8}$$

since $\ln W$ is a monotonic function.

The arbitrariness in the definition of W, and thus the ambiguity of S, is brought out explicitly if we turn now to the definition for W used by Boltzmann and others in the derivation of the so-called H-theorem.

Here one considers the phase space Γ of a classical system endowed with a probability measure $\rho(P)$, $P \in \Gamma$. $\rho(P)$ is assumed to be a positive function, normalized by

$$\int_{\Gamma} \rho(P) d\Omega = 1 \tag{9}$$

and interpreted to represent the probability of finding the system at the point P in phase space. We have written $d\Omega$ for the volume element in phase space.

One can define a quantity $\eta = \ln \rho$ and its average

$$\sigma = \int_{\Gamma} \rho \ln \rho \, d\Omega = \bar{\eta} \tag{10}$$

and one can then show that this quantity reaches its maximum value under the subsidiary condition

$$\int \epsilon \rho d\Omega = E = \text{constant} \tag{11}$$

provided it has the form of the canonical distribution

$$\rho = e^{(\psi - \epsilon)/\theta}. \tag{12}$$

However, the quantity σ is a constant under the evolution in time. This is true for any σ of the form (10) for any ρ.

One obtains a more suitable statistical definition for the entropy if one uses the process of 'coarse-graining'. Physically, this corresponds to the process which one would use for describing a system, the state of which is incompletely known. It is carried out in the following manner:

One divides the phase space into a certain number of cells with volume Ω_i and defines

$$P_i = \frac{1}{\Omega_i} \int_{\Omega_i} \rho \, d\Omega.$$

It follows then that

$$\sum_i P_i \Omega_i = 1.$$

The coarse-grained density is defined as

$$P(x) = P_i \quad \text{if } x \in \Omega_i$$

and

$$\sum \equiv \sum_i P_i \ln P_i \Omega_i = \int_\Gamma P \ln P \, d\Omega.$$

Hence

$$\sum = \overline{\ln P}.$$

One can then prove that $\sum$ is a decreasing function in time, reaching its minimum value for the canonical distribution provided that the mean energy is kept constant.

This suggests that $-k\sum$ could be identified with the thermodynamic entropy when the system is away from equilibrium.

This result shows quite clearly that there are several ways of defining and interpreting statistical entropy. The arbitrariness is connected with both the assumed probability field and with the nature of constraints used in the coarse-graining process. Both Boltzmann [8] and Gibbs [9] were aware of these ambiguities in the statistical interpretation of thermodynamic variables. Boltzmann says, for example:

> 'I do not believe that one is justified to consider this result as final, at least not as long as one has not defined very precisely what one means by the most probable state distribution.'

Gibbs is still more explicit:

> 'It is evident that there may be more than one quantity defined for finite values of the degrees of freedom, which approach the same limiting form for infinitely many degrees of freedom. There may be, therefore, and there are, other quantities which may be thought to have some claim to be regarded as temperature and entropy with respect to systems of a finite number of freedoms.'

As an example of two different ways of interpreting entropy, we mention the mixing problem discussed by Gibbs.

If a blue and a red liquid are mixed, there will result after a while a purplish mixture which cannot be unmixed by further agitation.

In this case the statistical entropy increases if the coarse-graining size is much larger than the average size of the volume element that has a definite (unmixed) color. On the other hand, the statistical entropy remains constant if the volume element having a definite color is much larger than the coarse-graining size. Only on the molecular level, e.g., in the diffusion process of two different gases, can this entropy be related to thermodynamic entropy; it increases in the diffusion process by

$$S = R \log 2,$$

provided that the volumes of the two kinds of gases are equal. The so-called paradox of Gibbs results from the confusion of two different kinds of entropy, based on different statistical models.

All that we have said in this section has been said before. Nevertheless, we felt reiteration useful to emphasize the polymorphic nature of statistical entropy. It is precisely this non-uniqueness which makes the concept so versatile.

3. The Definition of Entropy as a Mathematical Concept

In this section we will define the notion of entropy in an abstract setting without reference to its interpretations. It will be seen that a suitable definition will immediately reveal the range of its applications to a variety of situations, including thermodynamic entropy and the notion of 'information'.

Before proceeding with the formal definition, let us give a heuristic description of what we seek in order to achieve sufficient motivation for the mathematical definitions.

We wish to establish between two measures a relationship which represents a quantitative expression for the variability of one with respect to the other. As an example, let us consider the probability of the outcome of one of two alternatives, head or tail in flipping a coin. If the coin is 'true' then the a priori probability for either of the two events is $\frac{1}{2}$. However, if the experiment is made and the coin has been observed, then this probability is changed and is now 0 for one and 1 for the other event. The observation, or the specification, has restricted the variability and as a result a certain quantity which we wish to define—entropy—has decreased.

This quantity should have the property that it behaves additively for independent events, so that if the coin is flipped twice, then the entropy for the two observations should be twice that for one. As we shall see, these properties determine the quantity almost uniquely.

Proceeding now to the formal definition, what we need first of all is a measure space; by this we mean the triplet $(X,\mathscr{S},\mu)$ where X is a non-empty set, $\mathscr{S}$ a collection of subsets of X, and μ a normalized finite measure on $\mathscr{S}$. We assume that the subsets of $\mathscr{S}$ are closed under the formation of complements and countable unions (hence also countable intersections), and we call such a collection a *field* of subsets.

The measure μ is a positive, countably additive set function defined for the sets of $\mathscr{S}$ such that

$$\mu(X)=1, \quad \mu(\varnothing)=0 \quad (\varnothing = \text{null set}).$$

Let $S \in \mathscr{S}$ be such that $\mu(S)=0$, then we say the set S has μ-measure zero. Two different measures μ and ν on $\mathscr{S}$ are said to be *equivalent* if they have exactly the same sets of measure zero. We write $\mu \sim \nu$ so that

$$\mu \sim \nu \Leftrightarrow \{\mu(S)=0 \Leftrightarrow \nu(S)=0\}.$$

A more general concept is *absolute continuity*. We say μ is absolutely continuous with respect to another measure ν if

$$\nu(R)=0 \Rightarrow R \text{ is } \mu\text{-measurable and } \mu(R)=0.$$

We then write

$$\mu \propto \nu.$$

The relation $\propto$ is a partial order relation between different measures since

1. $\mu \propto \mu$
2. $\mu \propto \nu$ and $\nu \propto \rho \Rightarrow \mu \propto \rho$.

Two measures are equivalent if and only if $\mu \propto \nu$ and $\nu \propto \mu$.

A function f on X is called $L^1(\nu)$ if it is measurable with respect to ν and if

$$\int_X |f|\, d\nu < \infty.$$

It is convenient to identify functions which differ only on a set of measure zero with respect to a measure ν. In this case we write for two such functions

$$f_1 = f_2 \text{ a.e. } [\nu].$$

where a.e. stands for 'almost everywhere'.

For any measure ν and $f \in L^1(\nu)$ ($f \geqslant 0$ a.e. $[\nu]$) we can define a measure μ_f by setting

$$\mu_f(R) = \int_R f \, d\nu \tag{13}$$

and it is easy to verify that $\mu_f < \nu$.

More interesting is the converse, which is the content of one of the Radon-Nikodym theorems [10].

If $\mu \propto \nu$ then there exists a uniquely defined $f \geqslant 0$ a.e. $[\nu]$ and $f \in L^1(\nu)$ such that

$$\mu = \mu_f$$

The function f so defined is called the Radon-Nikodym derivative and is often written as

$$f = \frac{d\mu}{d\nu}. \tag{14}$$

If X_1 and X_2 are two measure spaces with measures μ_1 and μ_2, respectively, then we can define the product measure μ_{12} on $X_1 \times X_2$ as follows: On every set of the form $S_1 \times S_2$ with $S_1 \in \mathscr{S}_1$ and $S_2 \in \mathscr{S}_2$ we set

$$\mu_{12}(S_1 \times S_2) = \mu_1(S_1)\,\mu_2(S_2).$$

There is then a unique continuation of this measure by additivity to the field of all sets generated by the sets of the form $S_1 \times S_2$. This is called the *product measure.*

If $\mu_1 \propto \nu_1$ and $\mu_2 \propto \nu_2$ then one can prove that

$$\mu_{12} < \nu_{12} \quad \text{and} \quad f_{12} = \frac{d\mu_{12}}{d\nu_{12}} = \frac{d\mu_1}{d\nu_1}\frac{d\mu_2}{d\nu_2} = f_1 f_2.$$

We have now all the concepts needed for the definition of *entropy*. If $\mu \propto \nu$ and $f = d\mu/d\nu$ we define entropy of μ with respect to ν by

$$H(\mu, \nu) = \int f \ln f \, d\nu. \tag{15}$$

It has the following properties

1. $H(\mu, \mu) = 0$.
2. $H(\mu, \nu) \geqslant 0$ for $\mu \propto \nu$.
3. $H(\mu_{12}, \nu_{12}) = H(\mu_1, \nu_1) + H(\mu_2, \nu_2)$.

We verify them as follows:

1. is obvious, since $f = 1$ for $\mu = \nu$,
2. we prove as follows: Let $f = 1 + \phi$. Since

$$\int_X d\mu = \int_X \frac{d\mu}{d\nu} \, d\nu = 1 = \int_X d\nu$$

it follows that $\int \phi \, d\nu = 0$. Therefore

$$0 = \int_X f \ln (f - \phi) d\nu = \int_X f \ln f \left(1 - \frac{\phi}{f}\right) d\nu$$

$$= \int_X f \ln f + \int_X f \ln \left(1 - \frac{\phi}{f}\right) d\nu$$

Here we use the inequality

$$\ln \left(1 - \frac{\phi}{f}\right) \leqslant \frac{-\phi}{f}$$

so that the second term is

$$\int_X f \ln \left(1 - \frac{\phi}{f}\right) d\nu \leqslant - \int \phi \, d\nu = 0.$$

Thus we have verified

$$0 \leqslant \int f \ln f \, d\nu$$

3. follows from the definition $f_{12} = f_1 f_2$ and Fubini's theorem:

$$\int f_1 f_2 \ln (f_1 f_2) \, d\nu_{12} = \int f_1 f_2 (\ln f_1 + \ln f_2) \, d\nu_1 \, d\nu_2$$

$$= \int f_1 \ln f_1 \, d\nu_1 + \int f_2 \ln f_2 \, d\nu_2$$

$$= H(\mu_1, \nu_1) + H(\mu_2, \nu_2).$$

We illustrate the foregoing with a few examples:

As a first example we consider a finite sample space X consisting of n objects. For the a priori measure ν we choose the value $1/n$ for each element of X. The family of sets $\mathscr{S}$ is the class of all subsets of X. The measure of any subset S with k elements is then given by $\nu(S) = k/n$.

For the measure μ we choose the value 1 on one particular element of X and zero for all others. Let A denote this element and denote by B any other element of X. We have then

$$\nu(B) = \frac{1}{n} \quad B \in X, \quad \mu(A) = 1$$

$$\mu(B) = 0 \text{ for } A \neq B.$$

If $S \in \mathscr{S}$ is any subset of X and if $\nu(S) = 0$ then it follows that $S = \phi$. Therefore $\mu(S) = 0$ also, and μ is absolutely continuous with respect to $\nu : \mu \propto \nu$. We can easily calculate the values of the Radon-Nikodym derivative $f = d\mu/d\nu$. It has the values

$$f(A) = n, \quad f(B) = 0 \quad \text{for } A \neq B$$

where A is that particular element for which $\nu(A) = 1$.

We obtain now for the entropy as we have defined it

$$H(\mu,\nu)=\int f\ln f\,d\nu=\sum_{B}f(B)\ln f(B)\cdot\frac{1}{n}=\ln n. \tag{16}$$

We observe that this is a very simple, special case of Boltzmann's formula, since n can be interpreted as the number of 'states' which are contained in a set X of n elements.

One can generalize this example a little to bring it closer to the use of 'entropy in information theory'. Let us assume as before that the a priori probability on the elements A_i of X has the value $\nu(A_i)=1/n$ $(i=1,\ldots,n)$. For the measure μ, however, we assume $\mu(A_i)=p_i$ where $p_i\geqslant 0$ and $\sum_i^n=1,\ p_i=1$. The positive number p_i represents the probability with which the element A_i appears for instance in a message with an alphabet of n letters. If $p_i=1/n$ the message is completely garbled; if it is $=1$ for one A_i and $=0$ for the others, it is clear. We consider now the intermediate case.

If $\nu(S)=0$ for some set $S\in\mathscr{S}$ then $S=\phi$; hence again as before $\mu(S)=0$ so that μ is absolutely continuous with respect to $\nu:\mu\propto\nu$.

Furthermore

$$\frac{du}{d\nu}(A_i)\equiv f(A_i)=np_i.$$

This follows from the formula

$$\int_S\frac{d\mu}{d\nu}\,d\nu=\mu(S).$$

When we calculate the entropy in this case, we obtain

$$H(\mu,\nu)=\sum_{i=1}^{n}np_i\ln(np_i)\cdot\frac{1}{n}=\ln n+\sum_{i=1}^{n}p_i\ln p_i. \tag{17}$$

This expression reaches its minimum value 0 for $p_i=1/n$ and its maximum value $\ln n$ for $p_k=1$ for $k=i$, $p_k=0$ for $k\neq i$.

In information theory [11] one uses the quantity $I=-\sum_{i=1}^{n}p_i\ln p_i$ as a measure of the information contained in the one-letter message and we may therefore write

$$H(\mu,\nu)=\ln n-I.$$

Our definition of entropy is therefore in this case, apart from a constant, $\ln n$ equal to the negative entropy of the measure μ with respect to ν.

The abstract mathematical concept of entropy introduced here is at the basis of numerous different applications of this notion in physics as well as in other fields. This was recently emphasized by Grad [7]. Mathematicians use this concept as a versatile research tool in mathematical probability theory.

As has been emphasized, the use of the word entropy should not lead to confusion of the mathematical concept defined here with the physical concept of thermodynamic entropy. The concept introduced here relates two measures, one of which is absolutely continuous with respect to the other and has at this stage of abstraction nothing to do with any particular physical system. The misleading use of the same word for mathematical and for physical entropy is well-entrenched; it is now unavoidable. We use it in the general sense; its special meanings should be clear from the context.

4. Szilard's Paradox

Many authors who have tried to identify thermodynamic entropy with information (or rather, its negative) refer explicitly to the thought experiment of Szilard which seemed to lead to a violation of the second principle of thermodynamics unless the loss of entropy in the hypothetical experiment, so it is alleged, is compensated by a gain of information by some observer. In this section we describe Szilard's thought experiment and present an analysis to show that it cannot be considered a basis for the alleged identity and interchangeability of these two kinds of entropy.

In our description of Szilard's thought experiment, we follow closely von Neumann's version. We make only minor, physically irrelevant changes in the experimental set-up to make it more convenient for analysis and to avoid extreme idealizations.

In the experiment, a rigid, hollow, heat-permeable cylinder, closed at both ends, is used. It is fitted with a freely moveable piston with a hole large enough for the molecule to pass easily through it. The hole can be closed from outside. All motions are considered reversible and frictionless. The cylinder is in contact with a very large heat reservoir to keep the temperature of the entire machine constant.

Within the cylinder is a gas consisting of a single molecule. At the beginning of the experiment, the piston is in the middle of the cylinder and its hole is open so that the molecule can move (almost) freely from one side of the piston to the other. The hole is then closed, trapping the molecule in one half of the cylinder.

The observer now determines the location of the molecule by a process called by Szilard 'Messung', meaning measurement. If it is found to the left of the piston (see figure), the observer attaches a weight to the piston with a string over a pulley so that the pressure is almost counterbalanced. He then moves the piston very slowly to the right, thereby raising the weight. When the piston reaches the end of the cylinder, the hole in the piston is opened and the piston is moved back to the middle of the cylinder, reversibly and without effect on the gas. At the end of this process the starting position has been reached, except that a certain amount of heat energy Q from the heat reservoir has been transformed into potential energy A of the weight lifted.

Although Szilard does not mention this, obviously the same procedure can be used if the molecule happens to be trapped on the other side of the cylinder after the closing of the piston's hole.

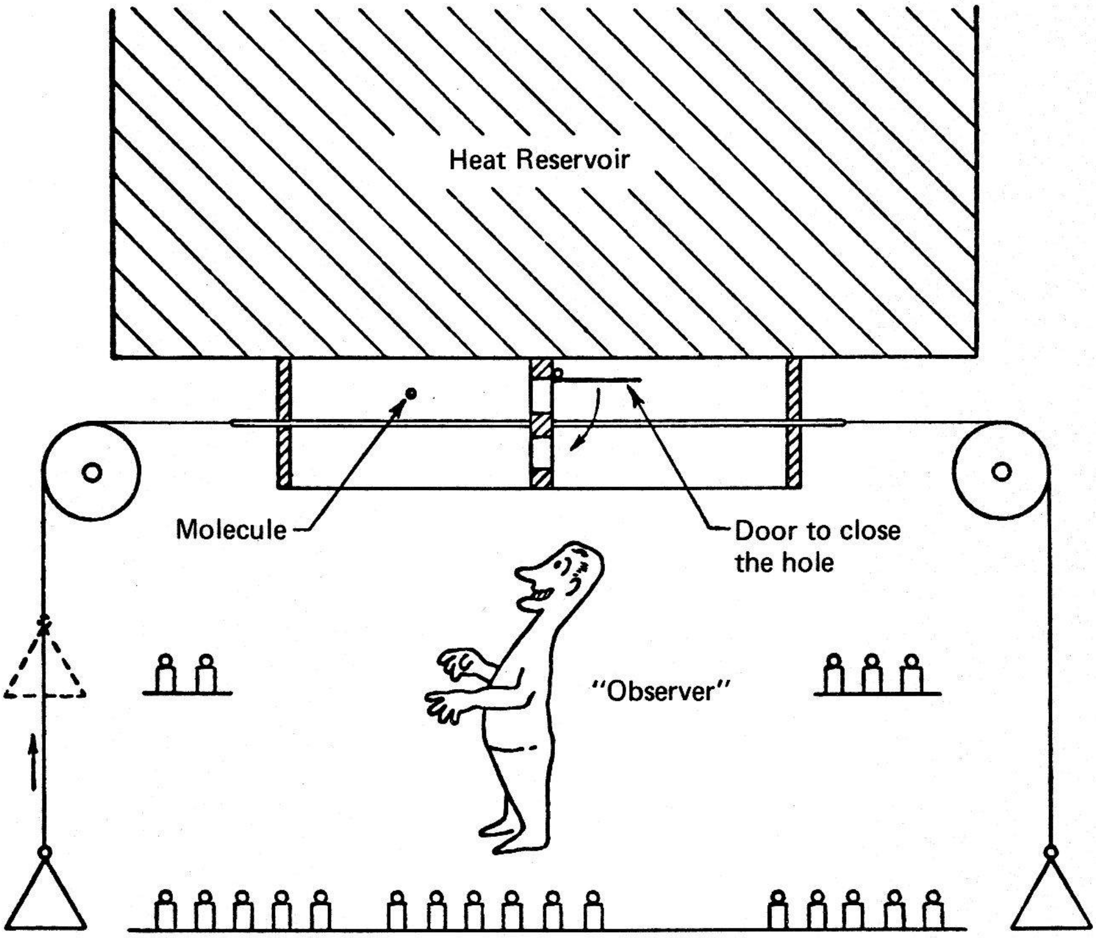
By repeating the process a large number of times (say, N), an arbitrarily large quantity of heat energy $Q = NA$ from the reservoir is transformed into potential energy without any other change in the system. This violates the second principle of thermodynamics.

In order to 'save' the second law, Szilard assumes that the observation of the molecule, for determining in which half of the cylinder it is contained, is in principle connected with an exactly compensating increase of entropy of the observer.

Before analyzing the experiment, a few remarks are in order concerning admissibility of procedures in idealized experiments. In this we are helped by a statement of von Neumann (p. 359 of his book, Ref. [5]): 'In phenomenological thermodynamics each conceivable process constitutes valid evidence, provided that it does not conflict with the two fundamental laws of thermodynamics.' To this we add: 'If an idealization is in conflict with a law that is basic to the second law, then it cannot be used as evidence for violation of the second law.'

Now this is precisely what is happening in the case of Szilard's experiment. Obviously, frictionless motion, reversible expansion and heat transfer, and infinitely large reservoir are all admissible under these criteria. Even the single-molecule gas is admissible so long as it satisfies the gas laws. However, 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 idealizations in Szilard's experiment are inadmissible in their actual context.

It is of further interest to analyze that phase of the experiment during which the hole in the piston is closed. This phase is almost a replica of the expansion phase of a Carnot cycle. Szilard believed that during this phase there is an uncompensated decrease in entropy, since during this interval the entropy of the heat reservoir decreases while at the same time the entropy of the gas increases by the same amount. The entropy of the entire system (piston, gas and reservoir) remains constant because the system is closed and the changes are reversible. Thus there is nothing to be compensated for by the alleged increase of entropy of the observer during observation.



Szilard's thought experiment

Finally, and perhaps most importantly, there is an unacceptable assumption in Szilard's interpretation of his experiment. He believed that the observer must *know* on which side of the piston the molecule is located in order that he may start the piston moving in the right direction. This knowledge is unnecessary, as is the pushing of the piston by the observer. The piston starts moving—under the idealized conditions of the experiment—by the pressure of the gas.

The automatic device, referred to in section 1, which can completely replace the observer, could work as follows:

Near the mid-plane of the cylinder and on both its sides are electrical contacts in its walls. When activated by the piston's motion along them, they operate mechanisms which attach a weight to the piston in whichever direction it moves. Thus a weight is lifted and the engine performs work, without interference by a conscious observer.

5. Summary and Concluding Remarks

Entropy is a fundamental mathematical concept, which relates two measures on a measure space in a certain manner.

The concept has many different applications, including thermodynamics (where it was first discovered) and information theory. It is also applicable in quantal systems, for example, and in random variables of any kind. The fact that entropy can be applied to many fields is no excuse for confusing its different meanings when applied to physical systems or mathematical constructions.

In particular, the identification of entropy of information (as defined by Shannon) as equivalent with negative thermodynamic entropy is unfounded and a source of much confusion. We have traced the origin of this confusion to the paradox of Szilard. Analysis of Szilard's paradox has shown specifically that:

1. Szilard's experiment is based on an inadmissible idealization; therefore it cannot be used for examining the principles of thermodynamics.
2. The observer needs no information about the location of the molecule at the beginning of the experiment.
3. There is no uncompensated entropy change in the system during the expansion phase of the experiment.
4. Thus, Szilard's thought experiment does not work; it is no paradox; and has nothing to do with information.

REFERENCES

[1] L. Brillouin, *Science and Information Theory* (Academic Press, New York, N.Y. 1956).
[2] D. ter Haar, *Elements of Statistical Mechanics* (Rinehart and Co., New York, N.Y. 1954, p. 161.
[3] L. Szilard, Z. Phys. *53*, 840 (1929).
[4] B. von Smoluchowski, *Vorträge über die kinetische Theorie der Materie und Elektrizität* (Leipzig 1914), esp. p. 89.
[5] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik*, p. 212 (1932).
[6] F. London and E. Bauer, *La Théorie de l'Observation en Mécanique Quantique*, Actual. scient. ind. *775* (Hermann, Paris 1939).
[7] H. Grad, *The Many Faces of Entropy*, Comm. Pure Appl. Math. *XIV*, 323 (1961).
[8] L. Boltzmann, Collected Papers, No. 42, p. 193.
[9] W. Gibbs, *Elementary Principles in Statistical Mechanics*, Dover Press, p. 169.
[10] S. K. Berberian, *Measure and Integration* (MacMillan & Co., New York, N.Y. 1962), esp. p. 160 seq.
[11] C. L. Shannon and W. Weaver, *Mathematical Theory of Communication*. (Univ. of Illinois Press, 1949).
[12] P. and T. Ehrenfest, *The Conceptual Foundations of the Statistical Approach in Mechanics*, transl. from the German by M. T. Moravcsik (Cornell Univ. Press, 1959), esp. Preface to the translation, by T. Ehrenfest-Afanassjewa. See also T. Ehrenfest, *Die Grundlagen der Thermodynamik* (Leiden 1956); A. Lande, *Axiomatische Begründung der Thermodynamik*, Handbuch der Physik *IX*; A. H. Wilson, *Thermodynamics and Statistical Mechanics* (Cambridge Univ. Press, 1960), esp. §2.5, p. 24 seq.