Preprints.org

Thermostatistics, Information, Subjectivity, Why Is This Association So Disturbing?

Didier Lairez [*]

Article

Thermostatistics, Information, Subjectivity, Why Is This Association so Disturbing?

Didier Lairez [iD]

Laboratoire des solides irradiés, École polytechnique, CEA, CNRS, IPP, 91128 Palaiseau, France
didier.lairez@polytechnique.edu

Abstract: Although information theory resolves inconsistencies (known under the form of famous enigmas) of the traditional approach of thermostatistics, its place in the corresponding literature is not what it deserves. This is interpreted as being mainly due to epistemological rather than scientific reasons: the subjectivity introduced into physics is perceived as a problem. This paper attempts to expose and clarify where exactly this subjectivity lies: in the representation of the reality and in probabilistic inference. Two aspects which have been integrated into the practice of science for a long time and which should no longer frighten anyone, but which become explicit with information theory.

Keywords: thermodynamics; statistical mechanics; information theory

---

With great success, statistical mechanics provides *"the rational foundations of thermodynamics"* (Gibbs [1]) which thus becomes thermostatistics. However, *"[it] is notorious for conceptual problems to which it is difficult to give a convincing answer"* (Penrose [2]). Since the origin, these problems are illustrated by famous enigmas which do not prevent the theory from advancing, but are like pebbles in the shoe. I think about the two Gibbs' paradoxes (related to the mixing of two volumes of gas), the Poincaré-Zermelo paradox (related to the recurrence of dynamical systems), the Loschmidt's paradox (related to the reversibility of the equations of mechanics) and its demonic version, the Maxwell's demon (a ratchet-pawl mechanism at the scale of particles). These enigmas all have one thing in common: they are all concerned with the second law of thermodynamics and entropy, the concept that was invented by Clausius [3] to account for the irreversibility of energy exchanges, linked by Boltzmann [4], Planck [5] and Gibbs [1] to probabilities and finally "enlightened" by Shannon's information theory [6].

"Enlightened" is wishful thinking because, although information theory resolves the inconsistencies raised by these enigmas (as will be shown), this contribution is far from being unanimously recognized. In recent textbooks, apart from one exception [7], the information theory is either just mentioned but not really used [8] or totally ignored [9–12]. The situation is also ambiguous in recent research literature.

The contribution of information theory to thermostatistics is two folds that must be well identified. The first one is linked to the encoding significance of entropy [13] and the relation it gives between energy and the information needed to reproduce the system as it appears to our senses, that is to say make a representation of it. The second fold, which is no less fundamental, is related to the "maximum entropy principle" that legitimates an inductive probabilistic inference based on our partial knowledge of the system [14–16] to describe its state of equilibrium. It legitimates prior probabilities (the first meaning of probability seen as a degree of belief) as opposed to *a posteriori* probabilities on which a frequentist inference could be done.

This dual contribution allows for very efficient shortcuts of thought and for resolving inconsistencies in the theory, such as those illustrated by the enigmas mentioned above. This is not new and these advantages are sometime recognized even by those who do not defend the viewpoint of information theory and virulently combat it. For instance: *"Although information theory is more comprehensive than is statistical mechanics, this very comprehensiveness gives rise to objectionable when it is applied in physics and chemistry"* (Denbigh & Denbigh [17] p.117). So, if information theory is not

more widely adopted, it is either because its benefits are poorly understood (which we will also try to remedy), or because they are fully understood but rejected for epistemological rather than scientific reasons. In fact, the stumbling block is that information theory is seen as introducing subjectivity into physics, that is classified as a "hard science" practiced with rigor and objectivity.

The aim of this paper is, in a first part, to clarify how exactly the subjectivity is brought to thermostatistics via entropy by two means: the encoding of a representation and the probabilistic inference, which are linked to the two features already evoked. Also, the last point will be compared to the alternative frequentist (objectivist) inference, namely the ergodic hypothesis.

The second part addresses the above puzzles and highlights the inconsistencies they raise in relation to the "objectivist" position. These inconsistencies are all removed with information theory (namely the "subjectivist" position) in a concise manner.

The last part tries to put the debate at an epistemological level: objective versus subjective conception of entropy. It aims to extricate things, to show where the arbitrariness lies precisely and to answer the question: why is subjective entropy such a disturbing concept? A particular focus will be done on the filiation of the ideas behind the approach of information theory once applied to thermostatistics. This filiation corresponds ultimately to a conception about what is science that originates from the Plato's allegory of the cave and develops where modern representationalism (or indirect-realism), empiricism, falsificationism and Bayesianism meet. That being exposed, everyone can decide whether this conception is natural or worrying, weigh it against the advantages provided by information theory and make a choice.

1. Information

1.1. Encoding, Information, Uncertainty

In everyday life, the question "How much information does this newspaper contain?" is understood in terms of the novelty of the meaning (the substance). With the advent of communication and computer sciences an alternative signification concerns the minimum quantity of bits that would be needed to transmit, store and reproduce it later (the form). But, the form is the physical support of the substance and the novelty may lie in the form. Also, an entirely predictable source of "information" does not require storing data to be reproduced. So that both acceptations of the term "quantity of information" are linked. However, the latter has the advantage to be much more manageable. This was the approach of Shannon: *"[The] semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. [Transmission and storage devices] must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design."* [6]. With Shannon, the message becomes a random variable to be lossless encoded and stored.

Consider a source (a thermodynamic system) that sequentially emits a random message $\mathbf{x} \in \Gamma$ (e.g. adopts a given microstate, Γ is the phase space) according to a fixed probability distribution $p(\mathbf{x})$ (the system is at equilibrium). We plan to perform a lossless recording of the sequence to reproduce it exactly, for instance to study and describe it later. Whatever the nature of the random events, if the number W of their possibilities (the cardinality of Γ or its "volume") is finite, we can establish a one-to-one correspondence table m (a mapping) that assigns to each event $\mathbf{x}$ an integer $n = m(\mathbf{x})$ ranging from 1 to W:

$$m: \quad \mathbf{x} \in \Gamma \quad \longmapsto \quad n \in \{1, ..W\} \tag{1}$$

So that recording the system behavior (the source emission) would start by recording the correspondence table (the meta-data), then continue with recording the sequential outcomes of the random integer-variable n (the data). This passes through the encoding of the latter, say a binary representation. Thus the question rises: what minimum number H of bits per outcome should be

provided for storage or transmission with a given bandwidth? By quantity of information emitted by the source, we mean this minimum number of bits for lossless recording of an outcome.

1.1.1. Fixed-Length Encoding

The central point is to seek the most economical encoding rule [18]. A first answer is to plan a fixed-length per outcome (per word). Of course, for the decoder to be able to discriminate the end of a word from the start of the next, this conventional fixed-length must be recorded with the meta-data. The length must be large enough to store the largest integer $\mathcal{W}$ that is expected to outcome. Since $\mathcal{W} = 2^{\log_2(\mathcal{W})}$, up to a rounding-error:

$$H = \log_2(\mathcal{W}) \tag{2}$$

The greater the number $\mathcal{W}$ of possibilities, the greater the uncertainty on a given outcome and the greater the minimum length. This last equation allows us to consider H, either as the *a posteriori* average number of bits per outcome, or as the *a priori* expected number of bits that should be scheduled to record upcoming events in case we have absolutely no idea about their actual probability distribution. In the former case, H is a measure of the quantity of information that has been emitted by the source, whereas in the latter case it is for the observer a measure of the uncertainty about the outcomes. These two facets are found in the usual meaning of probability.

1.1.2. Variable-Length Encoding

Equation 2 is not an optimal solution for storing data, because small numbers, that only require a few bits, take up the same storage space as large numbers. A variable-length encoding that uses just enough space for each outcome, i.e. $\log_2(m(\mathbf{x}))$ bits for the outcome $\mathbf{x}$, is better. Of course, this supposes a special encoding, named prefix-code, making it possible for decoding to identify the end of a given word and the beginning of the next. For instance, this can be done by using a delimiter. This also supposes that the encoding rules are recorded with the rest of the metadata, the size of which, however, will be assumed to be negligible compared to that of the data (which is legitimate for a long sequence of recordings). With a variable-length encoding, the average number of bits per outcome is:

$$H = \sum_{\mathbf{x} \in \Gamma} p(\mathbf{x}) \log_2(m(\mathbf{x})) \tag{3}$$

where $p(\mathbf{x})$ is the probability of the outcome $\mathbf{x}$ to which was assigned the integer $m(\mathbf{x})$ according to the mapping.

Let us first examine the special case where $p(\mathbf{x})$ takes the constant value $1/\mathcal{W}$ whatever $\mathbf{x}$. Eq.3 gives:

$$H = \frac{1}{\mathcal{W}} \sum_{\mathbf{x} \in \Gamma} \log_2(m(\mathbf{x})) \tag{4}$$

The smallest values for the series $n = m(\mathbf{x})$ are obtained by starting from $n_1 = 1$ and then by applying the rule $n_i = n_{i-1} + 1$. This sequence is that of natural numbers up to $\mathcal{W}$, so one gets:

$$H = \frac{1}{\mathcal{W}} \sum_{i=1}^{\mathcal{W}} \log_2(i) = \frac{1}{\mathcal{W}} \log_2(\mathcal{W}!) \tag{5}$$

For large $\mathcal{W}$, the Stirling formula leads to:

$$H = \log_2(\mathcal{W}) + o(1) \tag{6}$$

which is asymptotically the same as Eq.2. Note that the same result is obtained either if no *a priori* information is known about the outcomes except it is bounded (Eq.2), or if we know in advance that outcomes obey a uniform probability distribution (Eq.6).

Variable-length encoding does not make economy of storage-space for a uniform probability distribution. But it remains others. For non-uniform distributions, it is possible to choose $m(\mathbf{x})$ as being an increasing function of improbability $1/p(\mathbf{x})$, so that the rarest events are mapped to the values that require the most storage space. Different rules of assignment can be applied. For instance, according to the median, split Γ into two sub-sets labeled 0 and 1. The first encoding bit for $m(\mathbf{x})$ is the label to which subset $\mathbf{x}$ belongs and the others bits are obtained by subsequent similar recursive dichotomies. This procedure, named Fano encoding [19], gives a near optimal encoding length. Shannon [6] showed that in no case can the average length per outcome be less than

$$H = \sum_{\mathbf{x} \in \Gamma} p(\mathbf{x}) \log_2(1/p(\mathbf{x})) \tag{7}$$

Note that the uniform distribution is a special case of this last equation.

1.1.3. Information Encoding and Energy

To a factor $\ln 2$, one recognizes in equations 2 and 6 the formula for the Boltzmann entropy of an isolated system (microcanonical), and in equation 7 that of the Gibbs entropy of a closed system (canonical). In both cases one can write:

$$S = H \ln 2 \tag{8}$$

that is called Shannon entropy [20] (in this paper temperature is in Joule, so that entropy is dimensionless). The Boltzmann-Gibbs entropies are in reality special cases of that of Shannon for which the random events would be the different microstates that a thermodynamic system can adopt.

Here, let us recall that the formula for the Gibb's entropy was obtained from the canonical distribution of energy levels and from an identification with certain thermodynamic equalities involving Clausius entropy [21]. The derivation of the Gibbs entropy formula is entirely dependent on thermodynamics. The Shannon entropy, for its part, is obtained independently of thermodynamics, but with the aim of optimizing the encoding size. Aim which is reminiscent of the idea that Gibbs entropy is maximum at thermodynamic equilibrium. Equality of the formula and of the idea behind are likely not coincidental.

The Shannon entropy of the distribution of microstates and Boltzmann-Gibbs entropy are the same quantity. As the latter is the same as Clausius entropy, the three are one and the same quantity. Hence the connection between energy on one side and information/uncertainty on the other side.

To be more precise about this connection, let us recall some thermodynamics. Consider a system with internal energy U, for any quasistatic process it undergoes, one can write:

$$\Delta U = Q + W, \tag{9}$$

where Q is the heat exchanged with the surroundings and W is the work defined as the complementary part of Q in virtue of the conservation law (1st law of thermodynamics).

The 2nd law of thermodynamics is two folds: the first defines the entropy S as a state quantity linked to heat exchanged for a reversible process, whereas the second is the Clausius inequality that concerns the general irreversible case:

2nd law of thermodynamics (§1):
There exists a state quantity S which variation for a reversible process is such as $Q = T\Delta S$, where T is the temperature.

2nd law of thermodynamics (§2.1):
Clausius inequality: in all cases

$$Q \le T\Delta S \tag{10}$$

The energy exchanges can be seen as a dissipation (Q) of an energy cost (W). Then, the Clausius inequality says that the energy cost to achieve the process is always greater than $-T\Delta S$.

The twin of Clausius inequality in terms of the quantity of information emitted by the source, or equivalently in terms of the uncertainty about its emission, is obtained with Eq.8, leading to : $W \geq -T\Delta H \ln 2$. The second part of the 2nd law (§2) can thus be rewritten as

2nd law of thermodynamics (§2.2):
The energy cost W to vary by ΔH the uncertainty about the microstate of a system is always such as

$$W \geq -T\Delta H \ln 2 \tag{11}$$

The acquisition of data about the system, via measurement of certain properties (for instance the boundary of the phase space) is a way to reduce the uncertainty about the outcomes. So that the previous equation can be expressed per bit ($\Delta H = -1$) of acquired data:

$$W_{\text{acq/bit}} \geq T \ln 2 \tag{12}$$

Equations 11 and 12 are the key equations for linking energy and information. There is nothing more than that.

The two above versions of the 2nd law (§2.1 and §2.2) express exactly the same thing but in different ways. The first speaks of heat dissipation and entropy, whereas the second prefers to speak of work and uncertainty. It is thus legitimate to question the real usefulness of the notion of information encoding in thermostatistics. The first answer is that a link between two fields of knowledge is part of what we call understanding something. The second answer is that the link between information encoding and energy allows us to express certain ideas in a more concise and consistent way. In particular, it provides shortcuts in solving thermostatistic enigmas. So that if a theory is ultimately an economy of thought [22–24], this is undoubtedly a progress. The third answer is that the entropy so defined as an uncertainty forms a package with the probabilistic induction that will be discussed in the following section.

1.1.4. Stability of Equilibrium

The 2nd law of thermodynamics is often expressed for an isolated system to which neither heat nor work is exchanged with the surroundings ($Q = W = 0$). So that equations 10 becomes

$$T\Delta S \geq 0 \tag{13}$$

This leads to another version of the second law:

2nd law of thermodynamics (§2.3):
The entropy of an isolated system cannot spontaneously decrease.

Or in terms of information:

2nd law of thermodynamics (§2.4):
The uncertainty (the quantity of information) about the microstate of an isolated system cannot spontaneously decrease.

With the Boltzmann-Gibbs entropy, probabilities come into play for the description of the thermodynamic equilibrium. Therefore, an isolated system can now fluctuate and deviate slightly from equilibrium. This notion was absent in phenomenological thermodynamics, so that we are now faced with this problem: How can the equilibrium be stable? What is the restoring force of the system when it deviates from equilibrium? To ensure the stability of equilibrium, we need an additional

ingredient under the form of a postulate or a definition of the nature of equilibrium. This definition can be the following:

Definition of equilibrium (v1):
The equilibrium of an isolated system is the state of maximum entropy.

Or equivalently

Definition of equilibrium (v2):
The equilibrium of an isolated system is the state of maximum uncertainty about its microstate.

With these definitions, the restoring force comes from the 2nd law: due to fluctuations, the system may deviate from equilibrium but will return to it spontaneously.

1.2. Inductive Probabilistic Inference

The first programme of statistical mechanics is to calculate certain observables macroscopic quantities of a thermodynamic system at equilibrium, from the average of certain random variables which are relevant at a microscopic level. For instance, calculating the temperature from the average kinetic energy of particles. These averages are computed over a probability distribution.

The central point is thus to determine which random variable to consider and what probability distribution it is supposed to obey. That is to say make a statistical inference.

1.2.1. Subjective Versus Objective Probabilities

Two different types of statistical inferences are traditionally distinguished: "probabilistic inference" and "frequentist inference", which depends on how probabilities are defined.

1. subjective (or prior) probabilities are reasonable expectations or degrees of belief that one thing or another will happen. They are subjective in that they depend on our knowledge of the system.
2. objective relative frequencies of occurrence (over an ensemble) of one thing or another that actually happened (or *a posteriori* probabilities). They are supposed to be a tangible property of the system.

Subjective probabilities are general and can always apply, so they are *de facto* the most common on which to base a decision. But their arbitrary nature poses a problem without a rational criterion to assign a value to them. They appear illegitimate and may turn out to be false *a posteriori*. In contrast, frequencies are reliable provided the corresponding measurement has been carried out. But this is often impossible or at least not possible before making a decision. Their use is conditioned on the existence of an ensemble, or at least on the hypothesis of its existence provided it can be done in a consistent manner.

Consider the game of die. The die is cubic and offer six possible outcomes. Also from a symmetry argument there is no reason to believe that one is more likely than another. Prior to any toss of die, we can reasonably assign to outcomes a uniform discrete probability distribution lying from 1 to 6. This reasoning, which accords with common sense [25], is called "Laplace's principle of insufficient reason" (or "principle of indifference" [26]). It is a typical exampe of probabilistic inference.

An interesting point is that the most reasonable decision for a bet would be exactly the same if we know in advance that the die is loaded (we know in advance that the distribution is not uniform), but do not know which number is favored. The first assignment of probabilities is ultimately based on a criterion that seems much more arbitrary ("there is no reason to believe otherwise") than waiting for a few tosses of die and estimating the frequency distribution from the sampling of outcomes (make a frequentist inference). The decision is based on prior probabilities which do not seem legitimate, but are the only ones available.

In the problem of information processing, faced with an unknown source that we want to record, we must begin by using one or other of the encoding rules and then eventually use an adaptive

procedure to reconsider (to update) the encoding according to the observations. For a lossless recording, the best choice to begin is a fixed-length encoding (with an overestimated number of possible outcomes if it is unknown). This choice maximizes the compatibility of the encoding procedure with future incoming data. In section 1.1 we saw that this choice amounts to assigning a uniform prior distribution for the outcomes and is therefore in agreement with the principle of insufficient reason. We know in advance that this prior distribution is certainly not the true one, but this does not prevent it from being chosen rationally. This choice is the best we can make, any other would be judged irrational.

The problem of assigning a prior distribution is ultimately reduced to the search for an optimal compromise between two contradictory goals: 1) avoid any loss of information; 2) avoid unnecessary volume of storage. This problem of optimization is formalized and generalized with another aspect of the Shannon's information theory that essentially aims to optimize the use of our prior knowledge of the source for the statistical inference.

1.2.2. Maximum Entropy Probabilistic Inference

Suppose we are dealing with a source that emits outcomes i about which we have only a partial knowledge of the true probability distribution p_i, so that different distributions can potentially satisfy these constraints. To start recording the source, we need to assign a prior distribution that *"agrees with what is known, but expresses a 'maximum uncertainty' with respect to all other matters"* (Jaynes [15]) and thus leaves a maximum chance of compatibility with subsequent data. The problem is reduced to maximizing the uncertainty, which therefore remains to be measured.

In addition to the definition of the entropy as the optimum length to encode the outcomes of a source, Shannon [6] introduces another idea. He asks *"Can we find a measure [...] of how uncertain we are of the outcome?"*, and continues *"If there is such a measure, say $H(p_1, p_2, ...)$, it is reasonable to require of it the following properties:"* 1) being continuous in p_i; 2) being increasing in $W = 1/p_i$ for a uniform distribution; 3) being additive over different independent sources of uncertainty. Then, Shannon demonstrated that the only function (the only measure of uncertainty) that fulfills these requirements is to a factor $\sum p_i \ln(1/p_i)$, i.e. the Shannon entropy. Hence the theorem:

Maximum-entropy theorem: *the best prior distribution $p(x)$ that maximizes the uncertainty on x while being consistent with our knowledge, is the one that maximizes Shannon entropy.*

The validity of this theorem is ultimately determined by what is supposed to be required for a measure of uncertainty. These requirements play the role of starting axioms to its demonstration. As natural as they seem, some may find them arbitrary, wondering why not prefer others, leading to another function to be maximized. Shore and Johnson [16] start with a completely different requirement, a consistency axiom that can be stated like this: *"if the problem of assigning a prior distribution can be solved in more than one way of taking the same information into account (for instance in different coordinate systems), the results should be consistent"*. On which everyone should agree. Then, Shore and Johnson [16] prove that the only procedure satisfying this requirement of uniqueness is that of maximizing Shannon entropy. Given a random variable $\mathbf{x}$, the maximum entropy theorem provides a legitimate method to determine the prior probability distribution $p(\mathbf{x})$.

However, to complete the programme of statistical mechanics, it remains the first point risen in the beginning of this section: which random variable $\mathbf{x}$ to consider. For instance, imagine a random variable $x \in [0, \pi]$ with a uniform distribution, $\sin(x)$ is also a random variable which distribution is not uniform but has a maximum for $x = \pi/2$. Thus, using the maximum-entropy criterion directly on x or on $\sin(x)$ lead to different distributions *in fine* for x. However, *"among all these distributions there is one particular one, corresponding to the absolute maximum of entropy, which represents absolutely stable equilibrium."* (Planck [5] p.32). Jaynes [27] outlines a crucial point. The assumption of uniqueness of equilibrium actually automatically brings to our knowledge others crucial informations: the solution is not supposed to depend on the orientation of the observer (invariance under rotation), nor on its position (invariance under translation), nor on the scale it is considered (invariance under scaling).

Among all the possible variables describing a system, considering only those whose distributions are invariant in form under similarities avoids inconsistent results (lead to the same result). Hence the definition of equilibrium:

Definition of equilibrium (v3):
The equilibrium of a system is the only state that maximizes the uncertainty on variables whose distributions are similarity-invariant in form.

This definition plus the maximum-entropy theorem, can be expressed all-in-one under the form of the so called maximum-entropy principle that is actually also an alternate definition of the equilibrium:

Definition of equilibrium (v4):
The equilibrium of a system is the only state that maximizes Shannon entropy of variables whose distributions are similarity-invariant in form.

Two classical examples of similarity-invariant distributions mentioned in the above definitions are those of Boltzmann and Gibbs:

- Boltzmann[4] considers the phase-space of one single particle ($\Gamma_1 \subset \mathbb{R}^6$). For N particles, provided that their phase spaces are identical, the probability distribution $p(\mathbf{x})$ he considers is that of finding one particle in a given elementary "volume" $\mathbf{x}$ ($p(\mathbf{x})$ is the particle density). The H-function he defines is the Shannons's entropy of $p(\mathbf{x})$.
- Gibbs[1], considers directly the phase space of N particles ($\Gamma \subset \mathbb{R}^{6N}$). $p(\mathbf{x})$ is the probability for the system to be in microstate $\mathbf{x}$. The corresponding Shannon's entropy is the same as that defined by Gibbs who shows that it is also equal to the Clausius entropy of the system.

Maximizing the entropy of either distribution yields two consistent descriptions of equilibrium. In practice, the maximization procedure is a variational calculus taking into account the constraints imposed by our knowledge about the system[6]. For instance, suppose that the only thing we know about $p(\mathbf{x})$ is that it has a finite support (the minimum required for a discrete distribution to be properly normalized), then the best distribution is uniform (this is the microcanonical distribution for an isolated system). If $p(\mathbf{x})$ is only known to have a positive support allowing $\mathbf{x}$ to have a finite average value, then the best distribution is the exponential decay (this is the canonical distribution of energy levels for a closed thermalized system).

In these last two definitions of equilibrium, the statement of uniqueness should not be controversial. The core of the controversy lies in the rest. One may consider that there is absolutely no reason why the system would actually maximize the uncertainty we, observers, have about its microstate. But making another inference would be neither optimal nor reasonable. Moreover, the similitude between the first definitions of equilibrium (v1 and v2) that were given in section 1.1.4 and these last two is noticeable. Whereas the first definitions were based on thermodynamical considerations and from the need for the theory to define the equilibrium as stable, the second are more general (not only concerned with microstates) and emerge from a reasoning totally free from thermodynamics and free from any considerations regarding stability. This reasoning is an inductive probabilistic inference, that will be called "subjective" in contrast with the alternative, called "frequentist", that is believed to be more objective.

1.2.3. Alternative "Frequentist" Inference

Statistical mechanics did not wait for information theory to infer distributions at equilibrium. Alternatives approaches focus on the distribution of microstates. In addition, the problem lies uniquely in deciding what is the distribution for an isolated system. Because, this being determined, the distribution for any closed subpart can be deduced[21]. These alternative approaches are essentially of two kinds.

The first is based on the already mentioned principle of insufficient reason, which is renamed for the circumstance "fundamental postulate of statistical mechanics" [28]. In fact, it is nothing other than a less formalized and general expression of the maximum entropy principle [29].

We will only focus on the second alternative approach, that of "frequentists", which intends to adopt an objective point of view. To compensate for the lack of knowledge of the system, the idea is to make a strong hypothesis, that of "ergodicity". This hypothesis essentially aims to fulfill the prerequisite for the definition of probabilities as relative frequencies: the existence of an ensemble on which to calculate them.

Consider an isolated thermodynamic system, its phase space Γ is the set (ensemble) of all possible microstates $\mathbf{x}$ of probability $p(\mathbf{x})$ under which copies of the system can be found. Take one of this copy. It is dynamical, that is to say it continuously undergoes over time a transformation $F : \Gamma \mapsto \Gamma$, allowing to define, from the initial condition $\mathbf{x}_0$, a trajectory (an orbit) as the set of points in the phase space $\mathcal{T}(t) = \{\mathbf{x}_0, F(\mathbf{x}_0), F^2(\mathbf{x}_0), ...F^t(\mathbf{x}_0)\}$. The ergodic hypothesis is that this trajectory will finally pass recurrently through all points of the phase space at the frequency p, that is therefore supposed to remain unchanged over time (the transformation F preserves the volume of the phase space).

For our concern to determine which probability distribution is that of microstates, one consequence of the ergodic hypothesis is that "volume preserving" transformation means for any point $\mathbf{x}$ of the trajectory

$$p(\mathbf{x}) = p(F^{-1}(\mathbf{x})) \tag{14}$$

That is, the probability of a given consequent microstate is that of its antecedent. The probability distribution is therefore uniform and unchanged throughout the trajectory which will ultimately cover the entire phase space. We thus obtain the microcanonical distribution we were looking for.

For comparison with the maximum entropy approach, it is interesting to express the hypothesis of ergodicity in the same form as previously, i.e in the form of a definition of equilibrium:

Definition of equilibrium (v5):
The equilibrium of a system is the only state where the distribution of microstates is the same over its time-transformation as over an ensemble of its copies.

It immediately appears that this definition, contrary to the previous ones, does not contain any warranty of stability of the equilibrium. There is no restoring force for the equilibrium, which so defined is not an attractor for the system. This point is one of the source of the inconsistencies that will be discussed in the following.

At the basis of the hypothesis of ergodicity is the fact that deterministic Hamiltonian systems, (according to the Liouville's theorem) are volume preserving and thus ergodic. In this context, the trajectory of the system in the phase space is a chain of causality. So that Eq.14 is more particularly interpreted as: the probability of a given consequent microstate is that of its "cause".

Still for the sake of comparison, the detailed logical steps of the inductive reasoning for the ergodic hypothesis, which makes it natural to us, are the following: 1) a set of atoms (a thermodynamic system) is analogous to a set of colliding rigid spheres (in the time of Maxwell and Boltzmann, as the notion of atom itself, this was far from obvious); 2) usually, a set of colliding rigid spheres comes under classical mechanics; 3) usually, the equations of motion of mechanics alone determine the future state of classic mechanical systems (this is generally the meaning of the word "deterministic", but there exist exceptions of non-deterministic classic mechanical systems, e.g. the "Norton's dome" [30]). It follows that: 4) given a microstate $\mathbf{x}$ adopted by the system, the equations of motion of classical mechanics alone determine the future microstate $F(\mathbf{x})$ (Eq.14).

In short, there is no definitive proof for the system to be ergodic as measurements is not possible, but we consider this very likely. The reasoning is ultimately not that far from a probabilistic inference, but much less explicit on this point than that of maximum-entropy.

2. Enigmas

Information theory introduces subjectivity in thermostatistics by two different and related manners: 1) the encoding of a representation of the system and its link with energy; 2) a probabilistic inference. In what follows, these two features are used to resolves the inconsistencies raised by the thermostatistic enigmas quoted in the introduction.

Here, these enigmas are classified into two categories, paradoxes and demons, which have not the same level of importance. Paradoxes raise inconsistencies that cannot be removed by classical statistical mechanics without information theory. As for demons and the devices they drive, they do not actually introduce inconsistencies, in so far as they are physical systems that obey to the second law of thermodynamics, as do all others. But they can rather be considered as the first evidences of the link between information and energy that were given by Maxwell, Zermello and Szilard well before Shannon.

2.1. Gibbs Paradoxes

2.1.1. The Problems

Consider a system D (for disjoined) made of two adiabatic containers, denoted A and B, of same volume V, each containing the same quantity N of an ideal gas (of the same species or not) with the same degree of freedom and at the same temperature T (see Figure 1). So that, the entropy of the two sub-systems are equal $S = S_A = S_B$. From the additivity of entropy:

$$S_D = 2S \tag{15}$$

Joining these two volumes by removing the partition between them results in another system J (for joined). The process is never accompanied by any observable thermodynamic effect: neither heat nor work is exchanged with the surroundings. To determine whether or not this is accompanied by an increase of Clausius entropy, we must go back to the disjoined state (close the cycle) by a reversible process and measure heat exchange. Here, two paradoxes arise and continue to be debated for 150 years (for a review of the debate see e.g. the papers in [31,32]).
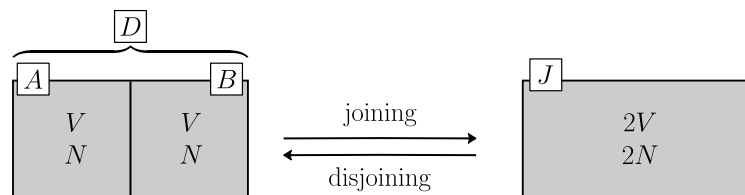

Figure 1. Usual image of the joining-disjoining thermodynamic cycle of two volumes of gas: removing the partition joins the two volumes and putting it back restores the system to its initial state with no work nor heat exchanged with the surroundings.

If the two gases are identical, according to thermodynamic phenomenology replacing the partition between the two containers allows the system to return to its initial state. This is done without heat exchange leading to the conclusion that the two states have the same Clausius entropy.

$$\Delta S = S_J - S_D = 0 \tag{16}$$

However, if initially the two gases differ, removing the partition mixes them and just putting it back is insufficient to separate them again. The separation can be done by two isothermal compressions against two pistons equipped with different semi-permeable membranes [33]. The first piston is only

able to compress one species (from $2V$ to V) and the second only the other species (by the same ratio). The total work is thus equal to $W = -Q = T\Delta S = T \times 2N \ln 2$. Leading to

$$\Delta S = S_J - S_D = 2N \ln 2 \tag{17}$$

In Eq.16 and 17, ΔS is called entropy of mixing.

The first paradox mentioned by Gibbs [34] is that ΔS is a bivalued discontinuous step function of the dissimilarity of the gases, whereas it can legitimately be expected to be continuous like other property-variations of the system (density, refractive index...).

The second paradox comes from statistical mechanics. The Boltzmann entropy of the disjoined state is $S_D = 2S = 2N \ln V$, whereas it is $S_J = 2N \ln 2V$ for the joined state. Leading to the difference $\Delta S = 2N \ln 2$. This, whatever the gas species, identical or not. If statistical mechanics solves the paradox of discontinuity, it raises another: Why do the Boltzmann and Clausius entropies differ when they should be the same?

2.1.2. Usual solutions

Concerning the first paradox, the consensus is that the discontinuity is not problematic. In fact, the dissimilarity of two species of atoms is discontinuous, thus that of entropy is not a problem. In term of the classification of Quine [35], the discontinuity is treated in the literature as a veridical paradox (the two premises are correct but not inconsistent).

As for the second paradox, it is most of the time treated as a falsidical paradox (at least one premise is wrong). As phenomenology is the final arbiter, the calculation of the cardinality of the phase spaces must be reconsidered (corrected).

Justifications for this correction are mainly of two kinds. Denote $\mathcal{V}_D = \mathcal{V}^2$ and $\mathcal{V}_J$ the cardinality of the phase spaces for systems D and J, respectively, in the case where all the particles are different and clearly identified with a label, such as a serial number. Also, denote $\mathcal{W}_D = \mathcal{W}^2$ and $\mathcal{W}_J$ the corresponding cardinality in the case where all particles are identical.

The first approach to justify a correction is based on the notion of indistinguishability of particles that comes from quantum mechanics [36]. For N indistinguishable particles, their $N!$ permutations give the same microstate which must therefore be counted only once leading to $\mathcal{W} = \mathcal{V}/N!$. This is known as the correct Boltzmann counting, leading to:

$$\mathcal{W}_D = \frac{\mathcal{V}_D}{N!N!}, \quad \mathcal{W}_J = \frac{\mathcal{V}_J}{(2N)!} \tag{18}$$

The second approach remains within the framework of classical mechanics where particles (even identical) are always distinguishable. In the sense that they always have distinct trajectories, allowing them to be (in principle) traceable and thus identified at any time. When partitioning the system into two compartments, particles can be combined in $(2N!)/N!N!$ different manners into the two separate compartments [37–39]. It follows that the number of possible results for the disjoined state is increased by this multiplicative factor, leading to:

$$\mathcal{W}_D = \mathcal{V}_D \frac{(2N)!}{N!N!}, \quad \mathcal{W}_J = \mathcal{V}_J \tag{19}$$

It follows that with the two approaches, the entropy of mixing two identical gases is the same:

$$\Delta S = \ln\left(\frac{\mathcal{W}_J}{\mathcal{W}_D}\right) = \ln\left(\frac{\mathcal{V}_J}{\mathcal{V}_D}\frac{N!N!}{(2N)!}\right) \tag{20}$$

but for different reasons [40]. In both cases, the second Gibbs paradox is claimed to be solved, because by using an approximation of the Stirling formula one gets $\ln(N!N!/(2N)!) = 2N \ln(N) -$

$2N \ln(2N) = -2N \ln 2$. So that the excess of entropy (Eq.17) obtained with the Boltzmann equation is corrected.

It is important to outline that Eq.19 for $\mathcal{W}_D$ counts the number of all possible disjoined microstates. That is to say, $\mathcal{W}_D$ is the cardinality of the phase space viewed as an ensemble of different possibilities including the different possible combinations in the repartition of particles. But once the partition is in place, a given disjoined microstate thus obtained will never by itself have a dynamic trajectory allowing it to reach another repartition (the repartition is frozen). In other words, the dynamics of a disjoined state cannot allows all the possibilities accounted for by Eq.19 to be explored. The disjoined state is no longer ergodic (this is noticed in [41]). It follows that Eq.19 is implicitly valid if the corresponding entropy is an uncertainty about the actual state of the disjoined system. Probabilities are prior-probabilities and not frequencies. In a classical mechanics framework, the above solution automatically places us implicitly in a "subjectivist" rather than in a "frequentist" position. That is to say, there is no solution of the 2nd Gibbs paradox in the framework of classical statistical mechanics and frequentist (ergodic) inference. The solution is necessarily quantum or based on a probabilistic inference. It is up to the reader to decide which one is more consistent and natural. This often goes unnoticed.

The "subjectivist" position to solve the 2nd Gibbs paradox is explicit in some papers, which are nevertheless largely in a minority (see e.g. [42–45]). But even in these latter papers, the paradox of discontinuity is either eluded of treated as veridical.

2.1.3. Yet Another Solution

The aim of this section is to show how from the information point of view, the dissimilarity of two gaseous contents is not bivalued but gradual and the Shannon's entropy too, that is, as close as possible to a continuous function with atomistic matter. So that the paradox of discontinuity is actually falsidical. In doing so, the 2nd paradox is also solved by considering it to be veridical.

Let us start by observing that thermodynamics considers cycles performed repeatedly and reproducibly. Therefore, if the two gases are identical, the representation of a joining/disjoining cycle does not care about:

1.   the exact number of particles in each compartment up to the standard deviation $\sqrt{N}$ of the binomial distribution;
2.   the traceability of particles (this information is only relevant if the two gases initially differ).

The correct image for the joining-disjoining cycle is given by Figure 2 (instead of Figure 1). It is to this very cycle that thermodynamics refers in Eq.16. The calculation of ΔS in term of probabilities has thus to be done by accounting for these two useless pieces of information. Accounting for the latter is very common and leads to the correct Boltzmann counting ($-\ln N!$ term), accounting for the former requires in addition the use of the exact Stirling formula $\ln(N!) = N \ln(N/e) + \ln(\sqrt{2\pi N}) + o(1)$ (rather than the usual approximation that consists in the first term only). It can be found in [46]. Here, a different derivation is proposed that avoids the term $-\ln(N!)$ and allows us in doing so to solve the paradox of discontinuity.
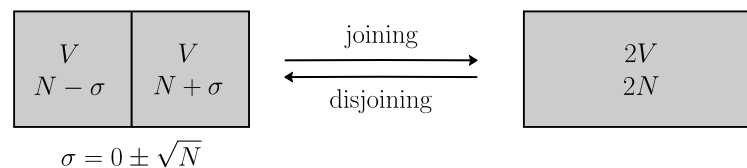

Figure 2. Correct image of a joining-disjoining thermodynamic cycle supposed to be repeatable (to be compared to Figure 1): putting back the partition leaves each compartment with the same number of particles up to the standard deviation $\sqrt{N}$.

Consider a first kind of cycle (see Figure 3 top) that consists in just moving the partition between the two compartments, denoted A and B. Let t_0 be the time just before the first cycle starts. After the first cycle is ended, all pieces of information about the exact contents of A and B at time t_0 is lost. At the end of each cycle in the disjoined state, the number of particles per compartment is always $N \pm \sqrt{N}$. Also after t_0, any information about the traceability of particles is lost. So that the uncertainty concerning these two features, exact number of particles and traceability, is unchanged by further cycles, and so the Shannon entropy. Therefore by considering these further cycles, the Shannon entropy of mixing is zero, just like Clausius entropy:

$$\Delta S_1 = 0 \tag{21}$$

Imagine that we know for certain that initially the two compartments had exactly the same number of particles $N \pm 0$ and that we want to retrieve this information when restoring the disjoined state (see Figure 3 middle). The procedure for the gas partition can be the following:

1. put all particles in a separate box;
2. partition the empty volume $2V$ into A and B;
3. take iteratively one pair (a, b) of particles, put one particle (either a or b) in compartment A and the other in B.
4. after N iterations, the two compartments have exactly the same number of particles.

There are four possibilities to arrange a and b in two boxes: $\{.|ab, a|b, b|a, ab|.\}$, and only $a|b$ or $b|a$ are convenient. Therefore each iteration divides the number of possibilities by 2 and gives us 1 bit of information. The N iterations of the overall procedure and Eq.12 provide the corresponding entropy of mixing:

$$\Delta S_2 = N \ln 2 \tag{22}$$

Note that the procedure can be stopped at any iteration, if we are satisfied by the uncertainty on N would lead a random repartition of the rest of particles. So that depending on our wish, the entropy of mixing can take any value from 0 to $N \ln 2$ by step of $\ln 2$.

Imagine that we know for certain that at t_0, compartment A was filled with isotope a and compartment B with isotope b (see Figure 3 bottom). So that, we are not satisfied by the previous procedure and want to restore exactly the original state. In other words, we want to preserve the traceability. To achieve this, among the two possibilities $\{a|b, b|a\}$ in the previous procedure, we must choose $a|b$. Here again, at each iteration the number of possibilities is divided by 2 and gives us 1 additional bit of information. So that at the end, compared to the previous state the entropy has decreased by an additional amount $N \ln 2$. Finally, if we consider traceability as crucial the Shannon entropy of mixing is:

$$\Delta S_3 = 2N \ln 2 \tag{23}$$

Here again, the procedure can be stopped at any iteration according to which degree of impurities is acceptable.

Depending on our knowledge about the original state or depending on what we consider as being important about it, the mixing-unmixing cycle differs and the Shannon entropy of mixing too. The latter can take gradually any value from 0 to $2N \ln 2$ by step of $\ln 2$. The two Gibbs paradoxes are solved.
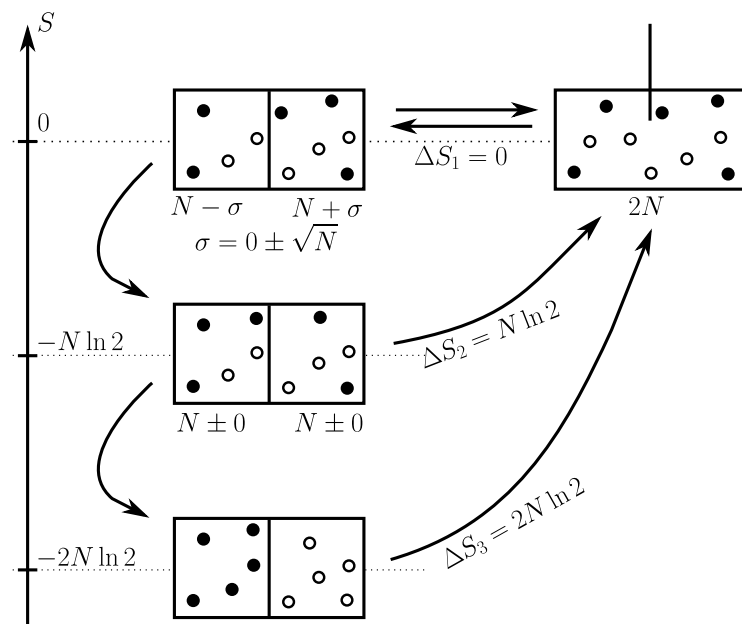
Figure 3. Mixing-unmixing cycle of a gas made of two species. The cycle depends on the information we had and do not want to lose. Top: information about exact number of particles per compartment and traceability is lost at the end of the first cycle, therefore further cycles leave it unchanged ($\Delta S_1 = 0$). Middle: retrieving exactly the same number of particles in each compartment has a minimal energy cost of $N \ln 2$ ($\Delta S_2 = N \ln 2$). Bottom: preserving particles traceability has an additional minimal energy cost also equal to $N \ln 2$ ($\Delta S_3 = 2N \ln 2$)

2.2. Paradoxes Related to H-Function

Boltzmann was the first [4] to write a quantity defined at any time t, named H-function, that takes the form of an entropy. For that, he considers the phase-space of one single particle ($\Gamma_1 \subset \mathbb{R}^6$) and the probability $p(\mathbf{x})$ to find a particles at $\mathbf{x} \in \Gamma_1$ at time t (with a time scale supposed to be discretized like the phase space). The H-function is defined as:

$$H(t) = \sum_{\mathbf{x} \in \Gamma_1} p(\mathbf{x}, t) \ln(1/p(\mathbf{x}, t)) \tag{24}$$

which is nothing other that the Shannon's entropy of the particle density p.

Starting from the idea of the Maxwell's kinetics theory of gases that the motion of colliding rigid particles is governed by the equations of classical mechanics, but also that their large number allows a statistical treatment, Boltzmann has obtained an integro-differential equation (named Boltzmann transport equation) for the time variation of the density p of a dilute gas (for a reference book see [47]). Boltzmann's equation allowed him to prove that for a bounded phase-space (an isolated system):

$$\frac{\mathrm{d}H}{\mathrm{d}t} \geq 0 \tag{25}$$

with the equality corresponding to the equilibrium state defined as stationary. Eq. 25 is known as the H-theorem. *"Its proof is clever and beautiful, although not perfectly rigorous."* (Villani [48]). In fact, the proof in the general case is still in progress. But for physicists, the H-theorem is quite natural and can be viewed as another expression of the 2nd law of thermodynamics [49] (concerning another distribution than that of microstates) plus a definition of equilibrium that warrants its stability. The 2nd law states that the entropy of an isolated system cannot spontaneously decrease. So that, even if classical thermodynamic (that of Clausius) says nothing about what exactly happens during a

spontaneous process, but only deals with the entropy before and after (the time variable is not present in classical thermodynamic equations), it is legitimate to say that entropy increases with time during a spontaneous process. For instance, consider a gas which is initially confined in a small box inside a larger room. Opening the box allows the gas to expand freely over an increasing volume ($\mathrm{d}H / \mathrm{d}t > 0$) until it uniformly occupies the entire room ($\mathrm{d}H / \mathrm{d}t = 0$) in a stationary state. In doing so, its entropy continuously increases.

The Boltzmann transport equation and H-theorem constitute the first attempt to demonstrate the macroscopic 2nd law of thermodynamics from what happens at the microscopic scale. Against this attempt, two objections have been raised: the reversibility paradox and the recurrence paradox.

2.2.1. Loschmidt's Reversibility Paradox

This paradox was originally stated [50] under the form of a thought experiment. Consider the free expansion of a gas enclosed initially at time t_0 in a box placed in a larger room. Once the system is at equilibrium, after a certain time τ, imagine that the direction of the velocity of each particle is reversed, without changing its magnitude. The operation does not change the macroscopic properties of the gas, such as temperature or volume. So that neither heat nor work are provided to the system. But, once this has been done, the gas particles goes backward through the same sequence of collisions than the previous one. So that at time 2τ its original microstate is restored. The gas is returned inside the box in contradiction with the 2nd law of thermodynamics and with the observation that this never happens.

The problem can be viewed in two different manners. First, to decide whether or not the process violates the 2nd law, we must wonder how the operation of reversing the velocities of particles is possible and whether it can be done without energy expenditure (certainly not if the operation is physically performed with a device that obeys to the 2nd law). This problem, thus posed in term of an operating "demon", will be discussed in section 2.3.2.

The other viewpoint is that this paradox basically raises the question of how from time-symmetrical equations of motion (those of mechanics) it is possible to obtain time-asymmetrical results. The consensual answer [49,51,52] is that, within the ingredients that permit to write the Boltzmann transport equation, the time-asymmetry is already presents under the form of the "hypothesis of molecular chaos": the velocities of two particles before their collision are fully uncorrelated but of course fully correlated and determined by mechanics after the collision. Fundamentally, the Boltzmann transport equation (and thus the H-theorem) is obtained by moving the time asymmetry of the 2nd law of thermodynamics from the macroscopic to the microscopic scale. The 2nd law is phenomenological and comes from an inductive reasoning which basically is a generalization of observations. By moving at the microscopic scale, it becomes a postulate or a hypothesis allowing to build a deductive reasoning. This looks like a circular reasoning, but actually it is a progress for a theory in terms of economy of though and potential unification of different areas of physics (for instance unification of thermodynamics and fluids mechanics).

However, for the purpose of this paper two points are worth emphasizing. The first is that, in the spirit of the mechanical approach, the independence of probabilities for the velocities of pre-colliding particles results from the impossibility to reach a sufficient accuracy about the initial conditions, i.e. it results from an incomplete knowledge (in this mechanistic conception, real stochasticity does not exist). The underlying conception of probabilities is therefore much closer to that of subjectivists than to that of objectivists (depiste a frequentist ambition).

The second point is that, whatever its origin, that is to say either incomplete knowledge (usual meaning of chaos) or a true stochastic process, the molecular chaos results over time in a loss of correlation between microstates along the trajectory in the phase space. The system is no longer deterministic in the sense that the chain of causality, representing the trajectory of the system in the phase space, is broken. The volume-preserving dynamics (Eq.14) and the ergodic hypothesis can still be postulated but they have lost their principal physical justification and the corresponding inductive reasoning has lost its strength and is much more hypothetical.

2.2.2. Poincaré-Zermelo Recurrence Paradox

Here, comes into play the Poincaré recurrence theorem. Consider a system, with a bounded phase space Γ, that continuously undergoes over time a transformation *F* that preserves the volume of any subset of Γ. Hamiltonian systems obey this condition according to the Liouville's theorem, but here the condition is more general and can apply not only to deterministic but also to stochastic (purely random) systems like the Ehrenfest urn model [53]. Then, Poincaré shows that the system will recurrently pass to any point of Γ already visited.

Going back to the example above of the free expansion of a gas from a small box into a larger room, yes the gas expands, but it is also expected to return in the box on its own, without any demon. Although it would take a long time, it is not impossible, not just once but recurrently.

Hence the paradox stated by Zermelo (for an historical perspective see [54]): How can such a recurrence be consistent with a continuously increasing H-function? How, also, can it be consistent with a stable state of equilibrium?

Different arguments have been put forward to resolve the Poincaré-Zermelo paradox. For instance, no system has a strictly bounded phase space, even the universe is expanding. Or, in the thermodynamical limit of an infinite number of particles, the time of recurrence is also infinite. The argument initiated by Boltzmann himself is that for concrete thermodynamic systems with a very large number of particles, the calculated average time of recurrence is greater than the age of the universe. Practically, the gas never returns to the box on its own. So that everything is a question of time-window: *"The range of validity of Boltzmann's equation ... is limited in time by phenomena such as the Poincaré recurrence"* (Villani [55]), but this limitation is never reached.

All these arguments are valid for resolving the paradox, they all amount to saying that in practice there is no recurrence. But it remains a problem. If there is no recurrence, how to conceive probabilities as frequencies? Or in the reverse manner, if there is recurrence, how to reconcile it with the 2nd law (H-theorem) and with a stable equilibrium? The probabilist inference offered by information theory avoids this inconsistency.

2.3. Demons

Demons observe thermodynamic systems, acquire information about them, and use it to act on them. In doing so they can possibly produce energy. Where does this energy come from? In fact, energy is an abstraction only defined by a conservation principle [56]. So that, if something is missing in an energy balance, it means that we have discovered a new form of energy. Demons, in their own, demonstrate the link between information and energy. The same idea can be expressed in another manner: *"In so far as the Demon is a thermodynamic system already governed by the Second Law, no further supposition about information and entropy is needed to save the Second Law."* Earman & Norton [57]. In other words, the very definition of a principle is that everything conforms to it, by definition a principle is inviolable.

This is an application of pure logic with which I completely agree. However, given the great expenditure of gray matter devoted to this question, such an answer cannot suffice. *"How does it happen that there are people who do not understand mathematics? If the science invokes only the rules of logic, those accepted by all well-formed minds, if its evidence is founded on principles that are common to all men, and that none but a madman would attempt to deny, how does it happen that there are so many people who are entirely impervious to it?"* (Poincaré [58] p.46). In fact, demons raise paradoxes that exist, and continue to exist even after they have been "resolved", by the mere fact that they have been stated. So that we cannot shrug them off only by pure logic, we need more. Thus in what follows, we do not use the shortcut given in the preamble and rather examine whether what we know from the encoding problem is sufficient for understanding how demons can operate in accordance with the 2nd law. No further supposition about information and entropy is needed to save the 2nd law [57], but we suppose that we already have information theory at our disposal to solve the inconsistencies raised in the previous section. So that here, we just want to check its consistency with demons.

With the encoding problem, the quantity of information needed to represent a system, or equivalently the uncertainty about its state, is identified with its entropy, thus linked to energy. In particular, acquiring information, i.e. reducing uncertainty, requires an energy expenditure (Eq.11-12). This acquired information is just like potential energy, it is stored and could be used in return. Increasing the potential energy requires providing work, but to use it in return, something else is needed: know-how. Otherwise the potential-information-energy is simply dissipated at the end of a cycle, that is when the information is outdated. In short, it is not the acquisition of information which directly has an effect on the system, it is the acquisition plus the action which depends on it. A misunderstanding of this point is at the origin of ill-founded criticisms of information theory (e.g. [59]).

Demons are supposed to know how, but the realization requires a physical implementation, not only of the action itself but also of all the information processing chain. By physical implementation of information processing, I mean a black box including everything necessary for measurement, storage, transmission, eventual erasure, etc, that necessarily falls under the 2nd law of thermodynamics. It is this physical implementation that is responsible for the minimum energy expenditure for demons to operate.

From the 2nd law of thermodynamics expressed in terms of information (Eq.11-12), the acquisition of 1 bit of information has a minimum energy cost equal to $T \ln 2$. Therefore, to check if demons work in accordance with the 2nd law, it is enough to check if the quantity of information necessary for their action is consistent with the energy that can be obtained in return.

2.3.1. Maxwell's Demon

The family of thermodynamical demons [60] was born with the temperature-demon of Maxwell [61]. Imagine a gas in an insulating container separated in two compartments A and B by a thermally insulating wall in which there is a small hole. A demon *"can see the individual particles, opens and c1oses this hole, so as to allow only the swifter particles to pass from A to B, and only the slower ones to pass from B to A. He will thus, without expenditure of work, raise the temperature of B and lower that of A in contradiction to the second law of thermodynamics."* ([61] p.308). The temperature difference between the two compartments can eventually be used for running a thermodynamic cycle and producing work.

A simplified version is the pressure-demon: particles whatever their speed can only pass from A to B. This results in a pressure difference between the two compartments, which can be used for producing mechanical work. Alternatively, in this simplified version the demon can be replaced by a concrete device, either by a one-way valve as proposed by Smoluchowski [60], or by a ratchet-pawl mechanism [62], or by an electric diode and the gas particles by electrons [18,63], then if the two compartments communicate by an additional channel, the device is expected to rectify thermal fluctuations and produce a net current of particles, which can deliver useful energy. For the last two concrete devices it has been experimentally shown that they can work, provided that the rectifier (ratchet-pawl or diode) is cooled at a lower temperature than the rest of the system [64,65] in return for the entropy decrease. Demons work in the same way once physically implemented.

The quantitative verification of the correspondence between, on the one hand, the information necessary for the demon and, on the other hand, the energy that can be obtained in return is simplified with the device proposed by Szilard [66]. It is composed of a single particle in a volume $2V$ (see Figure 4). The demon does not care about the velocity of the particle but only the compartment it is in. This information is encoded with only 1 bit which costs at least $T \ln 2$ (Eq.12). In the opposite compartment, the demon installs a piston which encloses the particle in a volume V. The system can return to the original state by an isothermal expansion that provides to the surroundings a work equal to at best $T \ln 2$. The overall cycle is consistent with thermodynamics.
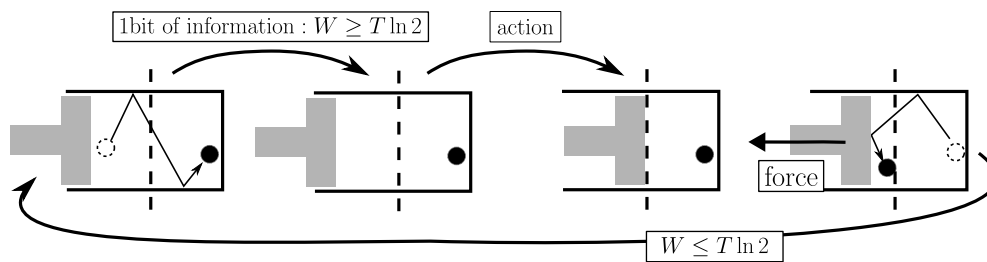
Figure 4. Szilard demon detects when the particle is in the suitable compartment and then installs a piston, allowing the device to subsequently produce work.

2.3.2. Loschmidt's Demon

Let us return to the Loschmidt's paradox of reversibility (§2.2.1), this time with an operating demon actually able to reverse the velocities of particles. We wonder whether the quantity of information (in terms of the minimum number of bit to encode it) necessary for the demon to operate is in agreement with the mechanical work that the gas could produce with a subsequent isothermal expansion.

Let V_1 and V_2 denote respectively the volume of the initial box and that of the room in which the gas was expanded and λ such as $V_2/V_1 = 2^\lambda$. So that the mechanical work per particle provided by an expansion is equal to at best $\lambda T \ln 2$, and for N particles:

$$W \leq N\lambda T \ln 2 \qquad (26)$$

For the demon to reverse the velocity of one particle in volume V_2, he must intercept the trajectory of the particle with an elastic wall (a mirror) having the correct direction (perpendicular to the trajectory), the correct orientation ($+$ or $-$), and the correct position (that of the particle) [67]. All the necessary information resides in the recording of the corresponding microstate of the particle. From V_1 to V_2, the number of bit required for encoding the velocity of one particle remains unchanged, but encoding its position requires λ extra bits (the cardinality of the phase space of one particle increases by a factor V_2/V_1). For N particles, $N\lambda$ extra bits are needed. From Eq.12 their acquisition costs at least $N\lambda T \ln 2$ in agreement with the work that can be obtained in return and with the 2nd law of thermodynamics.

2.3.3. Landauer's "Principle"

Equation 12, obtained in the sole framework of Shannon's information theory (plus the 2nd law of thermodynamics), resembles strongly to another one known as Landauer's "principle" [68–71] (that also uses the 2nd law but is free of Shannon's information theory and of the encoding problem). Clarification is therefore necessary to avoid confusion.

Landauer considers the physical implementation of a logical bit under the form of a one-to-one mapping between the two logical states (0 and 1) and two thermodynamical states materialized for instance by a particle in a bistable potential. In this framework, it was shown that the irreversible logical operation ERASE (or RESET TO 0) of the bit can be split into two steps:

1. put the bit into an undetermined state by flattening the potential;
2. set the bit to 0 by applying a bias and then raising back the energy barrier.

The point is that during the first step the probability distribution of the particle undergoes a leakage comparable to the irreversible adiabatic free expansion of a gas (by a factor 2). So that neither heat nor work are exchanged with the surroundings. Whereas the second step can be quasistatic. Suppose that the initial state was 0, the second step closes a thermodynamic cycle. So that to be in agreement with

the 2nd law, it must have an energy cost at least equal to $T \ln 2$. It follows that the total energy cost of the operation ERASE (of the cycle) is such as:

$$W_{\text{erase 1 bit}} \geq T \ln 2 \qquad (27)$$

This result is known as the Landauer's "principle" despite the fact that it cannot be a general principle (hence the quotes) but only applies to this particular physical implementation. Actually, to avoid any leakage from one potential hole to its neighbor, it is enough to design a physical implementation based on a two-to-one mapping between logic and thermodynamic states [72]. With only one potential hole there is no leak.

Equation 12 has general validity and concerns the acquisition of a data bit, whatever the way in which it was carried out and including all the steps necessary for this acquisition. Equation 27 concerns the erasure of a data bit with a particular physical implementation consisting of a bistable potential and results from the thermodynamics of this particular case. The difference between "acquisition" and "erasure" should be clear enough to avoid confusion. But, we can conceive certain particular data acquisition procedures (in particular that envisaged by Landauer and Bennet [69,71] by which they propose to replace the thermostatistic demons) which require erasing the bit before writing a new value there. In this case equations 12 and 27 lead to the same result. Hence a possible confusion.

The generic black box of the demons based on Eq.12, that dissipates $T \ln 2$ per bit, includes everything necessary for measurement, storage, transmission, eventual erasure etc. Different physical implementations correspond to different places where dissipation could occur. There is absolutely no clue allowing to suggest that this place is universal. Brillouin analyzed the physical limits for an observation through numerous examples of measurement procedure that could be implemented [18]. He showed that in all cases the energy expenditure to reach a given accuracy and the corresponding decrease of entropy that this information would allow are consistent with the 2nd law of thermodynamics.

But currently, the most popular physical implementation is that of Landauer. The functioning of the Landauer's black box is such as the measurement is free and only its recording under the form of bit-values causes energy dissipation. For one bit of data, this functioning is as follows (see Figure 5):

1. The bit is materialized with a bistable potential. Erasing the bit dissipates at least $T \ln 2$ (Eq.27).
2. The recording procedure requires the bit to be erased before new data is recorded.

This functioning is a doubly special case: a special case of bit implementation and a special case of recording procedure. Based on the 2nd law, it is obviously in full agreement with Eq.12. So that, if it is claimed that a solution using Landauer's "principle" is found for the paradoxes introduced by demons; then, the same solution is valid using the sole framework of Shannon's information theory and Eq. 12. But this time in a more direct way with general validity.

Landauer's "principle" is presented as the key point for definitively resolving the paradoxes caused by demons, therefore to definitively link information to energy [67,73–79]. In addition to the previous objection that it is not a general principle, let me focus on the second point of the functioning of Landauer's black box for demons.

When I was a teenager I had a boombox to record my favorite music. With this device it was possible to fully erase an already used cassette to start with an almost blank tape (a standard state). The idea behind this was that if you leave a blank between two pieces of music, you will not hear the old music when listening to the new. But erasing the cassette was not mandatory. The cassette could be directly overwritten, for example if a long concert was recorded. In this case, the silence between two pieces (as silences within a given piece) is not a blank (an absence of message) but a message in itself. In other words, it is possible to record and process data without having to erase anything (the fact that the data are digitized or not, does not change anything). The injunction to avoid overwriting (and thus for the need of erasing), which we find in recent literature (e.g. [67]) is unfounded.
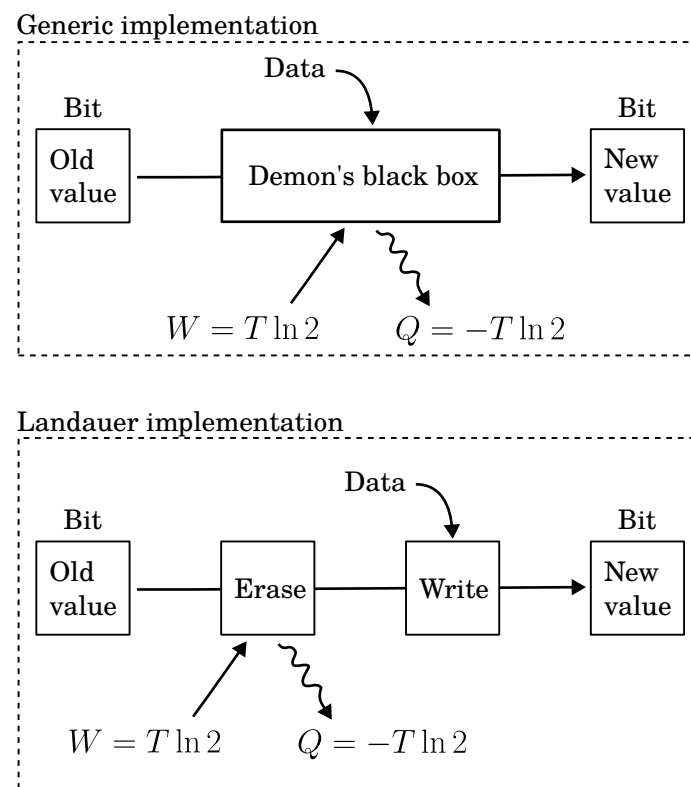
Generic implementation

$$W = T \ln 2 \qquad Q = -T \ln 2$$

Landauer implementation

$$W = T \ln 2 \qquad Q = -T \ln 2$$

Figure 5. Generic implementation (Eq.12) versus Landauer implementation (Eq.27) of a demon.

3. About Subjectivity in Physics

The approach of information theory to statistical mechanics avoids all inconsistencies of the alternative frequentist position. But, we are forced to notice an opposition to this idea. This opposition goes beyond science and is epistemological. The subjectivity introduced by information theory is the sticking point. Among many quotes, I note these which are particularly clear in this regard:

"The Jaynes approach [that of maximum entropy principle] is associated with a philosophical position in which statistical mechanics is regarded as a form of statistical inference rather than as a description of objective physical reality." (Penrose [2])

"A number of scientists and information theorists have maintained that entropy is a subjective concept and is a measure of human ignorance. Such a view, if it is valid, would create some profound philosophical problems and would tend to undermine the objectivity of the scientific enterprise." (Denbigh & Denbigh [17])

"This [the maximum entropy principle] is an approach which is mathematically faultless, however, you must be prepared to accept the anthropomorphic nature of entropy." (Lavis [80])

The aim of this section is to show that the type of subjectivity introduced into physics by information theory is in reality not new at all. It is in line with an ancient development of a general conception of what science is, which I propose to clarify.

3.1. Representationalism

The first subjective character introduced by information theory lies in the problem of coding a representation of the system which allows its behavior to be reproduced.

This representation depends on our knowledge, in the sense that it depends on the state of the art of the devices used to measure and probe the parameters needed for this representation. It also depends on the parameters that we consider relevant for this representation. Consider for example the unavoidable impurities in any chemical substance. A correct representation of the system must take them into account. But below a given threshold which depends on our tools of measurement,

impurities are no longer detectable and cannot be part of our representation of the system. But we can suppose them to be still present in an objective being of the system. Impurities are present in the real system, but not in our representation of it. Additionally, impurities can be isotopes. What about the representation of the system before their discovery?[43]. Also, in the thermodynamics of motors, pumps etc., most of the time we do not care about isotopes (actually we do not care about atoms either, before their discovery thermodynamic engines already worked very well), so that their presence or not is a useless information for the representation of the system (that does not need to be encoded). In this sense, information is subjective. It depends on the state of knowledge of the observer or on which level of details we (collectively) consider as relevant to describe a system. But it does not depend on the personality of the observer, it is not a personal element [15].

The subjectivity of entropy was already acknowledged by Maxwell (*"The idea of dissipation of energy depends on the extent of our knowledge"* [81]) and Gibbs (*"It is to states of systems thus incompletely defined that the problems of thermodynamics relate."* [34]). This is believed to contrast with other physical quantities considered as objective properties. Actually, information theory does nothing other than an explicitly introduction of representationalism (also named indirect-realism) into these problems.

The basic idea of indirect-realism is that our only access to reality is that provided by our senses (in a broad sense that includes all laboratory instruments). Following Einstein, *"all knowledge about reality begins with experience and terminates in it"* [24]. If "experience" is understood as a conscious event that passes through our senses, it follows that the concern of science is not the reality but the representation our senses give us of it. Fifty years before Shannon, Mach wrote: *"The law always contains less than the fact itself, because it does not reproduce the fact as a whole but only in that aspect of it which is important for us, the rest being either intentionally or from necessity omitted."* [82]. He was not talking about entropy or thermodynamics, he was talking about the laws of physics in general.

Information theory formalizes this idea that would otherwise remain unclear and implicit. Entropy itself is a very objective property well defined mathematically. But it is an objective property of a subjective representation of the reality. In this, according to indirect-realism entropy does not differ from other physical quantities. The above argument should be able to answer the question asked by some: *"Thermodynamic entropy is not different, in regard to its status of objectivity, from physical properties in general. How came it then that so many scientists have held, and still hold, the opposite opinion?"* (Denbigh & Denbigh [17] p.18)

Regarding all physical quantities other than entropy, from a purely scientific point of view the difference between direct and indirect-realism is just a question of vocabulary. For direct realists, science is directly about the real world. For indirect realists, we do not have access to the real world but only to a representation of it, so that only this representation is the subject matter of science. This difference can be ignored for a daily practice of science. Rename "representation of reality" to "reality" and both realisms are talking about the same thing. In both frameworks, theories concern the same object and their experimental examinations are equally achieved from our interactions with this object. The difference is only epistemological. Poincaré wrote: *"Does the harmony [the laws of nature] the human intelligence thinks it discovers in nature exist outside of this intelligence? No, beyond doubt, a reality completely independent of the mind which conceives it, sees or feels it, is an impossibility. A world as exterior as that, even if it existed, would for us be forever inaccessible. But what we call objective reality is, in the last analysis, what is common to many thinking beings, and could be common to all; this common part can only be the harmony expressed by mathematical laws. It is this harmony then which is the sole objective reality, the only truth we can attain."* ([83] p.15).

When it comes to entropy as seen by information theory, what becomes troubling is that the difference between the direct and indirect-realism views can no longer be ignored, even in science. This is the only point on which entropy is so special compared to all other physical quantities. Shannon entropy quantifies the complexity of the representation itself (in this case quantity of information is a particular case of complexity [13]). Doing so, it makes the notion of representation crucial and explicit.

3.2. Induction

Science is linked to knowledge, understood as a set of statements recognized as true. According to logical empiricism (or logical positivism), we have two possible sources of knowledge, each linked to a type of reasoning to assert that a statement is true: the first is purely logical (deductive reasoning) and the second empirical (inductive reasoning). Here, we will not enter into the debate on the justification of deduction, that is to say on the origin of the elementary rules of natural logic (which can possibly be empirical), we will focus only on induction, but we will need deduction for comparison.

Induction is unavoidable and omnipresent in natural sciences: generalization, interpolation, regression analysis, analogy etc., all are inductions based on known experimental facts. *"Without generalisation, prediction is impossible"* (Poincaré [84]). In fact, induction is the reasoning that allows us from our current knowledge to predict new observations or answer new questions. At the basis of phenomenological laws, but also theoretical hypothesis, postulates or principles, there is always inductive reasoning, at least implicit.

However, if the truth of a deductive statement can be proved and verified (provided the premises and the logical rules are right), the verification of an inductive statement can never be definitively achieved because this would imply a infinite non-countable set of experimental facts. The truth of a deduction is certain, that of an induction is at best probable. *"By generalization, every fact observed enables us to predict a large number of others; only, we ought not to forget that the first alone is certain, and that all the others are merely probable"* (Poincaré [84]). If known experimental facts make it possible to base an inductive reasoning, new or upcoming ones can only either confirm or refute it, but never definitely prove it. Inductions are by essence provisional and likely to be updated or replaced by better ones as progress is made.

If an inductive statement can never be verified (proven to be true), how can we make the difference between a well-founded scientific claim and another ill-founded and irrational? How can we make a hierarchy between different reasonings? What is the best? This is known as the problem of induction (for a recent book on this topic see [85]).

A first piece of answer was provided by the falsificationism of Popper [86]. Since verifiability cannot be required for induction, Popper instead suggests replacing it with falsifiability. A valid inductive reasoning must be falsifiable (or refutable): it must be able to be confronted with the experiments. This is the first condition, if it is met, an induction remains "true" until proven otherwise. The requirement of falsifiability of an induction entails another, that of not being tacit or hidden, but explicit. Otherwise, we make them without any chance of attempting to refute them [84]. But this is still not enough to establish a hierarchy of inductive reasoning.

Confirmation or refutation of inductive reasoning passes though experiments. At first glance, the refutation seems clear-cut, while the confirmation seems gradual (incremental) as more and more facts consistent with an induction reinforce it. However, both are conditioned on the validity of experimental results, themselves conditioned on confidence intervals (errors bars). This automatically introduces a link between the notion of "degree of confirmation" or "degree of belief" and that of "probability of truth" of an induction [87]. Hence the claim: all inductive reasoning in science falls under the same universal pattern as that of probabilistic inferences. The best is the most probable according to our present knowledge, that is to say the one having the highest prior probability of be true. This is the essence of Bayesianism [88] (named after Bayes and his theorem about the probability of an event conditioned on prior knowledge) and its derivatives in spirit, among which the maximum entropy inference can be classified.

Not everyone agrees with the existence of such a universal pattern for induction. For example, Norton [89] introduces a material theory of induction which professes that the logic of induction is determined by facts specific to each case and which cannot always be expressed in terms of probability. To which it has been opposed [90] that as soon as the confirmation procedure (and then the updating of the induction) involves data and measurements, probabilities come into play.

Whether maximum entropy inference is a starting point to produce a first prior probability distribution necessary for Bayesian updating, or whether it is a generalization of Bayesian inference, unless it is the other way around[91], is beyond the scope of this paper. The main point is the universal aspect of all inductive reasoning, that of being ultimately probabilistic, that of involving prior knowledge and prior probabilities, that of being subjective (in the sense given to this word in this paper).

Although scientists are aware of the problem of induction and adopt probabilistic inductive reasoning for their daily practice (personally, I don't know any scientist who would prefer to work on the option she believes has the lowest probability of success), this practice is not necessarily conscious and the problem of induction is often (temporarily) forgotten or denied. Below are some quotes from the recent literature of interest here: *"Experimental verification of Landauer's principle linking information and thermodynamics."* (Bérut et al.[73]). *"Information and thermodynamics: experimental verification of Landauer's Erasure principle."* (Bérut et al.[74]). *"We experimentally demonstrate a quantum version of the Landauer principle."* (Yan et al.[76]). *"Landauer's principle has been recently verified experimentally"* (Binder [67]), etc. I cannot imagine that these authors ignore or disagree with the impossibility to verify an induction. Instead, I interpret these quotes as language facilities that are not innocent but reveal a reluctance to inductions. Physicists prefer deductions, proofs and definitive verifications, all things expected of hard-sciences.

The problem with information theory is that here again, as for representationalism, everything is explicit: we cannot feign ignorance of our complete dependence on subjective probabilities in natural sciences.

Conclusion

The subjectivity of information theory, as it has been defined in this paper, that is to say something which is not personal but simply refers to the role played by our knowledge, allows us to resolve the inconsistencies present in thermostatistics from the start. At the same time, it is this subjectivity that worries some for epistemological rather than scientific reasons.

The role played by subjectivity should not be so surprising, at least in this area. Thermodynamics from the beginning refers to anthropocentric concepts and vocabulary such as energy grades, useful energy, energy cost, work, dissipation... In addition, thermodynamics is a science of the macroscopic scale. This term itself is anthropocentric, since macroscopic only designates our human scale. Indeed, in practice, the role that a certain subjectivity can play is admitted in science. But we are so steeped in positivism and with the ambition to be objective, that when subjectivity becomes too explicit it becomes annoying. In fact, science is a human construct. The "Laws of Nature" do not come from Nature, they come from us. The mere fact that these laws are provisional and subject to being continually replaced (updated) by better ones as science and our knowledge progress proves this.

Finally, there is another source of reluctance towards information theory that can be perceived in light of certain recent publications. It was not mentioned in this article but probably deserves special attention. It is also linked to the ambition of objectivity, but not in the same way as the refusal of representationalism and Bayesianism was. It is due to a particular meaning given to the word "physical", understood as "materialized", as opposed to virtual or non-tangible. There is nothing more "objective" than matter. After Landauer, ("Information is physical" [69]), many authors interpreted his principle as the missing element they were waiting for to materialize information. Probably the most recent development of this idea can be found in the "mass-energy-information principle" [92,93]. In short: information is energy, energy is mass ($E = mc^2$), therefore information would have a mass. The authors propose measuring the mass of a hard drive before and after erasing 1 TB of data. In my opinion, this demonstrates a reluctance to introduce into physics a concept (that of information) which would remain virtual. It is likely not superfluous to recall that the same problem has already been encountered concerning potential energy. What is the mass of potential energy? [94–96]. If there is one, it is likely not localized.

Yes, information is energy. But no more than potential energy, that is to say a concept introduced in physics in order to satisfy a principle of conservation. Exactly like information was with the historical demons.

References

1. Gibbs, J. *Elementary principles in statistical mechanics*; Charles Scribner's sons, 1902.
2. Penrose, O. Foundations of statistical mechanics. *Reports on Progress in Physics* **1979**, *42*, 1937–2006. https://doi.org/10.1088/0034-4885/42/12/002.
3. Clausius, R. *The mechanical theory of heat*; Macmillan & Co, London, UK, 1879.
4. Boltzmann, L. *Lectures on gas theory*; Dover ed., New York, NY, USA, 1964.
5. Planck, M. *The theory of heat radiation*; P. Blakiston's son, 1914.
6. Shannon, C.E. A mathematical theory of communication. *The Bell System Technical Journal* **1948**, *27*, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.
7. Ben-Naim, A.; Casadei, D. *Modern thermodynamics*; World Scientific, 2016. https://doi.org/10.1142/10300.
8. Sekerka, R. *Thermal Physics*; Elsevier, 2015. https://doi.org/10.1016/c2014-0-03233-9.
9. Swendsen, R.H. *An Introduction to Statistical Mechanics and Thermodynamics*; Oxford University Press, 2019. https://doi.org/10.1093/oso/9780198853237.001.0001.
10. Olafsen, J. *Sturge's Statistical and Thermal Physics*; CRC Press, 2019. https://doi.org/10.1201/9781315156958.
11. Luscombe, J.H. *Statistical Mechanics*; CRC Press, 2021. https://doi.org/10.1201/9781003139669.
12. Pathria, R.; Beale, P. *Statistical Mechanics*; Elsevier, 2022. https://doi.org/10.1016/c2017-0-01713-5.
13. Grunwald, P.; Vitanyi, P. Shannon Information and Kolmogorov Complexity **2004**. https://doi.org/10.48550/ARXIV.CS/0410002.
14. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630. https://doi.org/10.1103/PhysRev.106.620.
15. Jaynes, E.T. Prior Probabilities. *IEEE Transactions on Systems Science and Cybernetics* **1968**, *4*, 227–241. https://doi.org/10.1109/TSSC.1968.300117.
16. Shore, J.; Johnson, R. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory* **1980**, *26*, 26–37. https://doi.org/10.1109/tit.1980.1056144.
17. Denbigh, K.; Denbigh, J. *Entropy in Relation to Incomplete Knowledge*; Cambridge University Press, 1985. https://doi.org/10.1119/1.14692.
18. Brillouin, L. *Science and Information Theory*; Dover Publications: Mineola, N.Y., 1956.
19. Fano, R. The transmission of information. Technical Report 65, Massachusetts Institute of Technology, Research Laboratory of Electronics, 1949.
20. Claude E. Shannon, an oral history conducted in 1982 by Robert Price. IEEE History Center, Piscataway, NJ, USA, 1982.
21. Lairez, D. A short derivation of Boltzmann distribution and Gibbs entropy formula from the fundamental postulate. *arXiv* **2022**, [2211.02455v3].
22. Mach, E. *The Science of Mechanics*; The Open Court Publishing Company, 1919.
23. Duhem, P. *The Aim and Structure of Physical Theory*; Princeton University Press, 2021. https://doi.org/10.2307/j.ctv1nj34vm.
24. Einstein, A. On the method of theoretical physics. *Philosophy of Science* **1934**, *1*, 163–169. https://doi.org/10.1086/286316.
25. Jaynes, E.T. How Does the Brain Do Plausible Reasoning? In *Maximum-Entropy and Bayesian Methods in Science and Engineering*; Springer Netherlands, 1988; pp. 1–24. https://doi.org/10.1007/978-94-009-3049-0_1.
26. Keynes, J. *A treatise on probability*; Macmillian, 1921.
27. Jaynes, E.T. The well-posed problem. *Foundations of Physics* **1973**, *3*, 477–492. https://doi.org/10.1007/bf00709116.
28. Balian, R. *From microphysics to macrophysics*; Springer Berlin Heidelberg, 1991. https://doi.org/10.1007/978-3-540-45475-5.

29. Uffink, J. Can the maximum entropy principle be explained as a consistency requirement? *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **1995**, *26*, 223–261. https://doi.org/10.1016/1355-2198(95)00015-1.

30. Norton, J.D. Causation as Folk Science. *Philosophers' Imprint* **2003**, *3*, 1–22.

31. Entropy, Special Issue 'Gibbs Paradox and its resolutions', 2009.

32. Entropy, Special Issue 'Gibbs Paradox 2018', 2018.

33. Planck, M. *Treatise of thermodynamics*; Longmans, Green and Co., 1903.

34. Gibbs, J.W. *On the equilibrium of heterogeneous substances : first [-second] part*; Connecticut academy of arts and sciences, 1874. https://doi.org/10.5479/sil.421748.39088007099781.

35. Quine, W.V. *The ways of paradox, and other essays*; Harvard University Press: Cambridge, Massachusetts, 1976.

36. Huang, K. *Statistical Mechanics*, 2nd ed.; J. Wiley & sons, 1991.

37. Cheng, C.H. Thermodynamics of the System of Distinguishable Particles. *Entropy* **2009**, *11*, 326–333. https://doi.org/10.3390/e11030326.

38. Versteegh, M.A.M.; Dieks, D. The Gibbs paradox and the distinguishability of identical particles. *American Journal of Physics* **2011**, *79*, 741–746. https://doi.org/10.1119/1.3584179.

39. Frenkel, D. Why colloidal systems can be described by statistical mechanics: some not very original comments on the Gibbs paradox. *Molecular Physics* **2014**, *112*, 2325–2329. https://doi.org/10.1080/00268976.2014.904051.

40. Dieks, D. The Logic of Identity: Distinguishability and Indistinguishability in Classical and Quantum Physics. *Foundations of Physics* **2014**, *44*, 1302–1316. https://doi.org/10.1007/s10701-014-9814-0.

41. Peters, H. Demonstration and resolution of the Gibbs paradox of the first kind. *European Journal of Physics* **2013**, *35*, 015023. https://doi.org/10.1088/0143-0807/35/1/015023.

42. van Kampen, N. The Gibbs Paradox. In *Essays in Theoretical Physics*; Elsevier, 1984; pp. 303–312. https://doi.org/10.1016/b978-0-08-026523-0.50020-5.

43. Jaynes, E.T. The Gibbs Paradox. In *Maximum Entropy and Bayesian Methods*; Springer Netherlands, 1992; pp. 1–21. https://doi.org/10.1007/978-94-017-2219-3_1.

44. Tseng, C.Y. Yet another resolution of the Gibbs paradox: an information theory approach. In Proceedings of the AIP Conference Proceedings. AIP, 2002. https://doi.org/10.1063/1.1477057.

45. Dieks, D. Is there a unique physical entropy? micro versus macro. In *New Challenges to Philosophy of Science*; Springer Netherlands, 2013; pp. 23–34. https://doi.org/10.1007/978-94-007-5845-2_3.

46. Lairez, D. Plea for the use of the exact Stirling formula in statistical mechanics **2022**. [2211.11447v4].

47. Cercignani, C. *The Boltzmann Equation and Its Applications*; Springer New York, 1988. https://doi.org/10.1007/978-1-4612-1039-9.

48. Villani, C. H-Theorem and beyond: Boltzmann's entropy in today's mathematics. In *Boltzmann's Legacy*; EMS Press, 2008; pp. 129–143. https://doi.org/10.4171/057-1/9.

49. Weaver, C. In Praise of Clausius Entropy: Reassessing the Foundations of Boltzmannian Statistical Mechanics. *Foundations of Physics* **2021**, *51*. https://doi.org/10.1007/s10701-021-00437-w.

50. Darrigol, O. Boltzmann's reply to the Loschmidt paradox: a commented translation. *The European Physical Journal H* **2021**, *46*. https://doi.org/10.1140/epjh/s13129-021-00029-2.

51. Uffink, J. Compendium of the foundations of classical statistical physics. In *Philosophy of Physics*; Butterfield, J.; Earman, J., Eds.; Handbook of the Philosophy of Science, North-Holland: Amsterdam, 2007; pp. 923–1074. https://doi.org/https://doi.org/10.1016/B978-044451560-5/50012-9.

52. Weaver, C. Poincaré, Poincaré recurrence and the H-theorem: A continued reassessment of Boltzmannian statistical mechanics. *International Journal of Modern Physics B* **2022**, *36*. https://doi.org/10.1142/s0217979222300055.

53. Karlin, S.; McGregor, J. Ehrenfest urn models. *Journal of Applied Probability* **1965**, *2*, 352–376. https://doi.org/10.2307/3212199.

54. Brown, H.R.; Myrvold, W.; Uffink, J. Boltzmann's H-theorem, its discontents, and the birth of statistical mechanics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **2009**, *40*, 174–191. https://doi.org/10.1016/j.shpsb.2009.03.003.

55. Villani, C. (Ir)reversibility and Entropy. In *Time*; Springer Basel, 2012; pp. 19–79. https://doi.org/10.1007/978-3-0348-0359-5_2.

56. Feynman, R.P.; Leighton, R.B.; Sands, M. *The Feynman lectures on physics*; Addison-Wesley, Reading, MA, 1966; chapter 4.

57. Earman, J.; Norton, J.D. EXORCIST XIV: The Wrath of Maxwell's Demon. Part II. From Szilard to Landauer and Beyond. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **1999**, *30*, 1–40. https://doi.org/10.1016/s1355-2198(98)00026-4.

58. Poincaré, H. *Science and method*; T. Nelson & sons, London, 1914.

59. Callender, C. A Collision Between Dynamics and Thermodynamics. *Entropy* **2004**, *6*, 11–20. https://doi.org/10.3390/e6010011.

60. Rex, A. Maxwell's demon - a historical review. *Entropy* **2017**, *19*, 240. https://doi.org/10.3390/e19060240.

61. Maxwell, J.C. *Theory of heat*, 3d ed. ed.; Longmans, Green and Co.: London, 1872.

62. Feynman, R.P.; Leighton, R.B.; Sands, M. *The Feynman lectures on physics*; Addison-Wesley, Reading, MA, 1966; chapter 46, pp. 1–9.

63. Brillouin, L. Can the rectifier become a thermodynamical demon? *Physical Review* **1950**, *78*, 627–628. https://doi.org/10.1103/physrev.78.627.2.

64. Bang, J.; Pan, R.; Hoang, T.M.; Ahn, J.; Jarzynski, C.; Quan, H.T.; Li, T. Experimental realization of Feynman's ratchet. *New Journal of Physics* **2018**, *20*, 103032. https://doi.org/10.1088/1367-2630/aae71f.

65. Gunn, J.B.; Staples, J.L. Spontaneous reverse current due to the Brillouin emf in a diode. *Appl. Phys. Lett.* **1969**, *14*, 54–56. https://doi.org/10.1063/1.1652709.

66. Szilard, L. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Behavioral Science* **1964**, *9*, 301–310. https://doi.org/10.1002/bs.3830090402.

67. Binder, P. The Reversibility Paradox: Role of the Velocity Reversal Step. *International Journal of Theoretical Physics* **2023**, *62*. https://doi.org/10.1007/s10773-023-05458-x.

68. Landauer, R. Irreversibility and Heat Generation in the Computing Process. *IBM Journal of Research and Development* **1961**, *5*, 183–191. https://doi.org/10.1147/rd.53.0183.

69. Landauer, R. Information is Physical. *Physics Today* **1991**, *44*, 23–29. https://doi.org/10.1063/1.881299.

70. Bennett, C.H. The thermodynamics of computation—a review. *International Journal of Theoretical Physics* **1982**, *21*, 905–940. https://doi.org/10.1007/bf02084158.

71. Bennett, C.H. Notes on Landauer's principle, reversible computation, and Maxwell's Demon. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **2003**, *34*, 501–510. https://doi.org/10.1016/s1355-2198(03)00039-x.

72. Lairez, D. Thermodynamical versus Logical Irreversibility: A Concrete Objection to Landauer's Principle. *Entropy* **2023**, *25*, 1155. https://doi.org/10.3390/e25081155.

73. Bérut, A.; Arakelyan, A.; Petrosyan, A.; Ciliberto, S.; Dillenschneider, R.; Lutz, E. Experimental verification of Landauer's principle linking information and thermodynamics. *Nature* **2012**, *483*, 187–189. https://doi.org/10.1038/nature10872.

74. Bérut, A.; Petrosyan, A.; Ciliberto, S. Information and thermodynamics: experimental verification of Landauer's Erasure principle. *Journal of Statistical Mechanics: Theory and Experiment* **2015**, *2015*, P06015. https://doi.org/10.1088/1742-5468/2015/06/p06015.

75. Ciliberto, S.; Lutz, E. The physics of information: from Maxwell to Landauer. In *Energy Limits in Computation*; Springer International Publishing, 2018; pp. 155–175. https://doi.org/10.1007/978-3-319-93458-7_5.

76. Yan, L.; Xiong, T.; Rehan, K.; Zhou, F.; Liang, D.; Chen, L.; Zhang, J.; Yang, W.; Ma, Z.; Feng, M. Single-Atom Demonstration of the Quantum Landauer Principle. *Physical Review Letters* **2018**, *120*. https://doi.org/10.1103/physrevlett.120.210601.

77. Proesmans, K.; Ehrich, J.; Bechhoefer, J. Finite-Time Landauer Principle. *Physical Review Letters* **2020**, *125*. https://doi.org/10.1103/physrevlett.125.100602.

78. Giorgini, L.T.; Eichhorn, R.; Das, M.; Moon, W.; Wettlaufer, J.S. Thermodynamic cost of erasing information in finite time. *Physical Review Research* **2023**, *5*. https://doi.org/10.1103/physrevresearch.5.023084.

79. Oriols, X.; Nikolić, H. Three types of Landauer's erasure principle: a microscopic view. *The European Physical Journal Plus* **2023**, *138*. https://doi.org/10.1140/epjp/s13360-023-03855-z.

80. Lavis, D.A., Frontiers in Fundamental Physics; Universities Press, India,, 2007; Vol. vol 3, chapter Equilibrium and (Ir)reversibility in Classical Statistical Mechanics.

81. Maxwell, J.C. Diffusion. *Encyclopedia Britannica, reproduced in Scientific papers* **1878**, *2*, 625–646. https://doi.org/10.1017/CBO9780511710377.064.

82. Mach, E. *Popular scientific lectures*; The Open Court Publishing Co., 1898.

83. Poincaré, H. *The value of science*; The science Press, 1907.

84. Poincaré, H. *Science and hypothesis*; The Walter Scott Publishing Co., 1905.

85. Huber, F. *A logical introduction to probability and induction*; Oxford University Press, 2019.

86. Popper, K. *The Logic of Scientific Discovery*; Routledge, 2005. https://doi.org/10.4324/9780203994627.

87. Sprenger, J. Confirmation and Induction. In *The Oxford Handbook of Philosophy of Science*; Humphreys, P., Ed.; Oxford University Press, 2016; chapter 9, pp. 185–209. https://doi.org/10.1093/oxfordhb/9780199368815.013.10.

88. Sprenger, J.; Hartmann, S. *Bayesian Philosophy of Science: Variations on a Theme by the Reverend Thomas Bayes*; Oxford University Press, 2019.

89. Norton, J.D. *The Material Theory of Induction*; University of Calgary Press, 2021. https://doi.org/10.2307/j.ctv25wxcb5.

90. Dawid, R. Turning Norton's Dome Against Material Induction. *Foundations of Physics* **2015**, *45*, 1101–1109. https://doi.org/10.1007/s10701-015-9943-0.

91. Jaynes, E.T. The Relation of Bayesian and Maximum Entropy Methods. In *Maximum-Entropy and Bayesian Methods in Science and Engineering*; Springer Netherlands, 1988; pp. 25–29. https://doi.org/10.1007/978-94-009-3049-0_2.

92. Vopson, M.M. The mass-energy-information equivalence principle. *AIP Advances* **2019**, *9*, 095206. https://doi.org/10.1063/1.5123794.

93. Vopson, M.M. Experimental protocol for testing the mass-energy-information equivalence principle. *AIP Advances* **2022**, *12*, 035311. https://doi.org/10.1063/5.0087175.

94. Brillouin, L. The actual mass of potential energy, a correction to classical relativity. *Proceedings of the National Academy of Sciences* **1965**, *53*, 475–482. https://doi.org/10.1073/pnas.53.3.475.

95. Brillouin, L. The actual mass of potential energy II. *Proceedings of the National Academy of Sciences* **1965**, *53*, 1280–1284. https://doi.org/10.1073/pnas.53.6.1280.

96. Hecht, E. Relativity, potential energy, and mass. *European Journal of Physics* **2016**, *37*, 065804. https://doi.org/10.1088/0143-0807/37/6/065804.