entropy

Article

Entropy, Shannon's Measure of Information and Boltzmann's H-Theorem

Arieh Ben-Naim

Department of Physical Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel; arieh@fh.huji.ac.il

Academic Editors: Geert Verdoolaege and Kevin H. Knuth
Received: 23 November 2016; Accepted: 21 January 2017; Published: 24 January 2017

Abstract: We start with a clear distinction between Shannon's Measure of Information (SMI) and the Thermodynamic Entropy. The first is defined on any probability distribution; and therefore it is a very general concept. On the other hand Entropy is defined on a very special set of distributions. Next we show that the Shannon Measure of Information (SMI) provides a solid and quantitative basis for the interpretation of the thermodynamic entropy. The entropy measures the uncertainty in the distribution of the locations and momenta of all the particles; as well as two corrections due to the uncertainty principle and the indistinguishability of the particles. Finally we show that the H-function as defined by Boltzmann is an SMI but not entropy. Therefore; much of what has been written on the H-theorem is irrelevant to entropy and the Second Law of Thermodynamics.

Keywords: entropy; Shannon's measure of information; Second Law of Thermodynamics; H-theorem

---

1. Introduction

The purpose of this article is to revisit an old problem, the relationship between entropy and Shannon's measure of information. An even older problem is the question about the *subjectivity* of entropy which arose from the association of entropy with the general concept of information. Finally, we discuss the H-theorem; its meaning, its criticism, and its relationship with the Second Law of Thermodynamics.

The paper is organized in four parts. In Section 2, we present a brief introduction to the concept of SMI. In Section 3, we derive the thermodynamic entropy as a special case of SMI. In Section 4, we revisit the Boltzmann H-theorem. In light of the SMI-based interpretation of entropy; it will become clear that the function $-H(t)$ is identical with the SMI of the velocity distribution. The entropy is obtained from $-H(t)$ *after* taking the limit of $t \to \infty$, i.e., the value of $-H(t)$ at equilibrium.

Because of its central importance we state our conclusion here:

> It is absolutely necessary to distinguish between SMI and the entropy. Failing to make such a distinction has led to too many misinterpretations of entropy and the Second Law, as well as assigning properties of SMI to entropy, and in particular misunderstanding the H-theorem, discussed in Section 4.

In 1948, Shannon sought and found a remarkable *measure* of *information*, of *uncertainty* [1,2] and *unlikelihood*. It was not a measure of *any* information, not *any* uncertainty about *any* proposition, and not the unlikelihood about the occurrence of *any* event. However, because the quantity he found has the same mathematical *form* as the entropy in statistical mechanics, he called his measure, as allegedly suggested by von Neumann: "entropy". This proved to be a grievous mistake which had caused a great confusion in both information theory and thermodynamics.

The SMI is defined for any probability distribution. The entropy is defined on a tiny subset of all the possible distributions. Calling SMI entropy leads to many awkward statements such as: The *value* of the maximal entropy at equilibrium is the *entropy* of the system.

The correct statement concerning the entropy of an isolated system is as follows: An isolated system at equilibrium is characterized by a fixed energy E, volume V and number of particles N (assuming a one-component system). For such a system, the entropy is *determined* by variables E, V, N. In this system the entropy is fixed. It is not a function of time, it does not change with time, and it does not tend to a maximum.

Similarly, one can define the entropy for any other well defined thermodynamic system at equilibrium [3,4]. This is exactly what is meant by the statement that entropy is a *state function*.

For any isolated system not at equilibrium one can define the SMI on the probability distributions of locations and velocities of all the particles. This SMI changes with time [2]. At equilibrium, it attends a maximal value. The maximal value of the SMI, attained at equilibrium is related to the entropy of the system [1,2]. In this article, whenever we talk about SMI we use the logarithm to the base 2, but in thermodynamics we use, for convenience the natural logarithm $\log_e x$. To convert to SMI we need to multiply by $\log_2 e$, i.e., $\log_2 x = \log_2 e \, \log_e x$.

Parts of this article have been published before in [3–5]. Specifically, the derivation of the entropy function of an ideal gas based on the SMI, was published by the author in 2008 [3]. The discussion of the Boltzmann H-Theorem in terms of SMI is new. We do not discuss relations with the huge field of thermodynamics of irreversible processes. This whole field is based on the assumption of *local equilibrium*, which, in the author's opinion was never fully justified. Therefore, in this article we use the concept of entropy only for *macroscopic equilibrium systems*, while the SMI may be used for any system.

2. A Brief Introduction to the Concept of SMI

In this section, we present a very simple definition of the SMI. We then discuss its various interpretations. For any random variable X (or an experiment, or a game, see below), characterized by a probability distribution: $p_1, p_2, \ldots, p_n$, we define the SMI as:

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{1}$$

If X is an experiment having n outcomes, then p_i is the probability associated with the occurrence of the outcome i.

We now discuss briefly the three interpretations of SMI. The first is an average of the *uncertainty* about the outcome of an experiment; the second, a measure of the unlikelihood; and the third, a measure of information. It is ironic that the "informational" interpretation of SMI is the least straightforward one, as a result it is also the one most commonly misused. Note that the SMI has the form of an average quantity. However, this is a very special average. It is an average of the quantity $-\log p_i$ using the probability distribution $p_1, \ldots, p_n$.

2.1. The Uncertainty Meaning of SMI

The interpretation of H as an *average uncertainty* is very popular. This interpretation is derived directly from the meaning of the probability distribution [2,5,6].

Suppose that we have an experiment yielding n possible outcomes with probability distribution $p_1, \ldots, p_n$. If, say, $p_i = 1$, then we are *certain* that the outcome i occurred or will occur. For any other value of p_i, we are *less certain* about the occurrence of the event i. *Less certainty* can be translated to *more uncertainty*. Therefore, the larger the value of $-\log p_i$, the larger the extent of uncertainty about the occurrence of the event i. Multiplying $-\log p_i$ by p_i, and summing over all i, we get an *average uncertainty* about *all* the possible outcomes of the experiment [6].

We should add here that when $p_i = 0$, we are *certain* that the event *i will not* occur. It would be awkward to say in this case that the *uncertainty* in the occurrence of *i* is zero. Fortunately, this awkwardness does not affect the value of H. Once we form the product $p_i \log p_i$, we get zero when either $p_i = 1$, or when $p_i = 0$.

Yaglom and Yaglom [7] suggest referring to $-p_i \log p_i$ as the uncertainty in the event i. In this view, the SMI (referred to as "entropy" by Yaglom and Yaglom) is a *sum* over all the uncertainties in the outcomes of the experiment.

This interpretation is invalid for the following reason. As we noted above, it is plausible to interpret $-\log p_i$ as a measure of the extent of uncertainty with respect to the occurrence of the outcome i. Since $-\log p_i$ is a monotonically decreasing function p_i, Figure 1a, larger p_i, or smaller $-\log p_i$ means smaller uncertainty (or larger certainty). In this view, the SMI is an *average* uncertainty over *all* possible outcomes of the experiment.

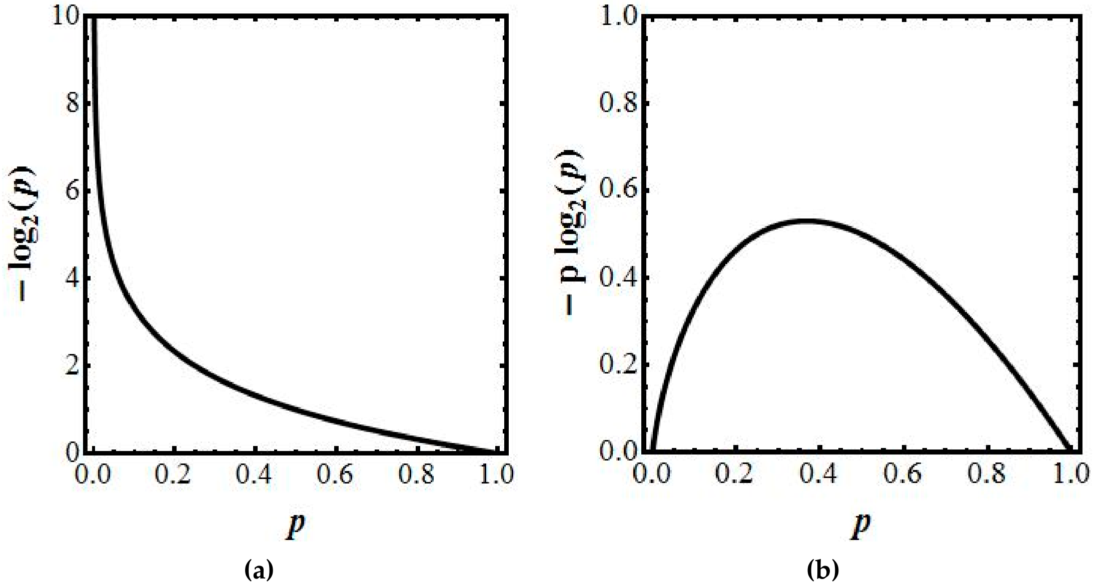

Figure 1. The functions (**a**) $-$Log (p) and (**b**) $-p$ Log (p).

The quantity $-p_i \log p_i$ on the other hand, is not a monotonic function of p_i, Figure 1b. Therefore, one *cannot* use this quantity to measure the extent of uncertainty with respect to the occurrence of the outcome i.

2.2. The Unlikelihood Interpretation

A slightly different but still useful interpretation of H is in terms of *likelihood* or *expectedness*. These two are also derived from the meaning of probability. When p_i is small, the event i is unlikely to occur, or its occurrence is less expected. When p_i approaches one, we can say that the occurrence of i is more likely. Since $\log p_i$ is a monotonically increasing function of p_i, we can say that the larger the value of $\log p_i$, the larger the likelihood or the larger the expectedness for the event. Since $0 \le p_i \le 1$, we have $-\infty \le \log p_i \le 0$. The quantity $-\log p_i$ is thus a measure of the *unlikelihood* or the *unexpectedness* of the event i. Therefore, the quantity $H = -\sum p_i \log p_i$ is a measure of the *average unlikelihood*, or *unexpectedness*, of the entire set of the outcomes of the experiment.

2.3. The Meaning of the SMI as a Measure of Information

As we have seen, both the uncertainty and the unlikelihood interpretation of H are derived from the meaning of the probabilities p_i. The interpretation of H as a measure of information is a little trickier and less straightforward. It is also more interesting since it conveys a different kind of *information* on the Shannon measure of *information*. As we already emphasized, the SMI is not *information* [8]. It is also not a measure of *any* piece of information, but of a very particular kind of information. The confusion of SMI with information is almost the rule, not the exception, by both scientists and non-scientists.

Some authors assign to the quantity $-\log\ p_i$ the meaning of information (or self-information) associated with the event i.

The idea is that if an event is rare, i.e., p_i is small and hence $-\log\ p_i$ is large, then one gets "more information" when one knows that the event has occurred. Consider the probabilities of the outcomes of a die as shown in Figure 2a. We see that outcome "1" is less probable than outcome "2". We may say that we are less uncertain about the outcome "2" than about "1". We may also say that outcome "1" is less likely to occur than outcome "2". However, when we are informed that outcome "1" or "2" occurred, we cannot claim that we have received more or less information. When we know that an event i has occurred, we have got the *information* on the occurrence of i. One might be surprised to learn that a rare event has occurred, but the *size* of the *information* one gets when the event i occurs is not dependent on the probability of that event.

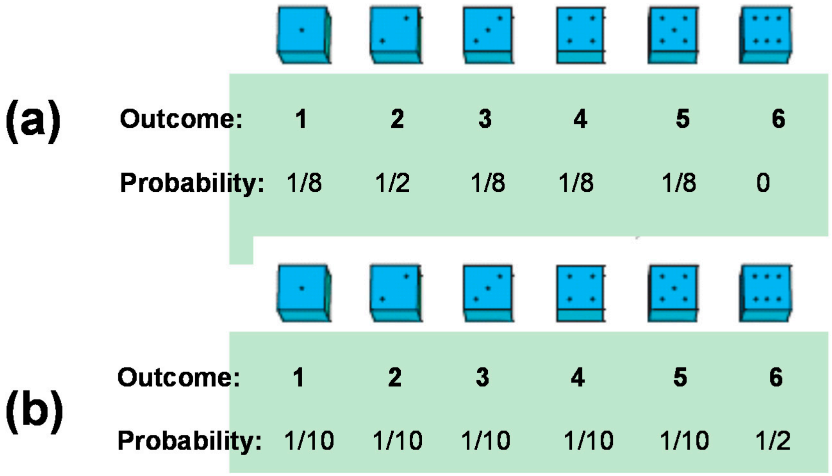

Figure 2. Two possible distributions of an unfair die.

Both p_i and $\log\ p_i$ are measures of the uncertainty about the occurrence of an event. They do not measure *information* about the events. Therefore, we do not recommend referring to $-\log\ p_i$ as "information" (or self-information) associated with the event i. Hence, H should not be interpreted as *average information* associated with the experiment. Instead, we assign "informational" meaning directly to the quantity H, rather than to the individual events.

It is sometimes said that removing the *uncertainty* is tantamount to obtaining *information*. This is true for the entire experiment, i.e., to the entire probability distribution, not to individual events.

Suppose that we have an unfair die with probabilities $p_1 = \frac{1}{10}, p_2 = \frac{1}{10}$, $p_3 = \frac{1}{10}, p_4 = \frac{1}{10}, p_5 = \frac{1}{10}$ and $p_6 = \frac{1}{2}$, Figure 2b. Clearly, the uncertainty we have regarding the outcome $i = 6$ is less than the uncertainty we have regarding any outcome $i \neq 6$. When we carry out the experiment and find the result, say $i = 3$, we removed the uncertainty we had about the outcome before carrying out the experiment. However, it would be wrong to argue that the *amount* of information we got is larger or smaller than if another outcome had occurred. Note also that we talk here about the *amount* of information, not the information itself. If the outcome is $i = 3$, the information we got is: The outcome is "3". If the outcome is $i = 6$, the information is: The outcome is "6". These are different information, but one cannot claim that one is larger or smaller than the other.

We emphasize again that the interpretation of H as *average uncertainty* or *average unlikelihood* is derived from the meaning of each term $-\log p_i$. The interpretation of H as a measure of information is not associated with the meaning of each probability p_i, but with the *entire distribution* $p_1, \ldots, p_n$.

We now describe in a qualitative way the meaning of H as a *measure* of *information associated with the entire experiment*.

Consider any experiment or a game having n outcomes with probabilities $p_1, \ldots, p_n$. For concreteness, suppose we throw a dart at a board, Figure 3. The board is divided into n regions, of areas $A_1, \ldots, A_n$. We know that the dart hit one of these regions. We assume that the probability of hitting the ith region is $p_i = A_i/A$, where A is the total area of the board.

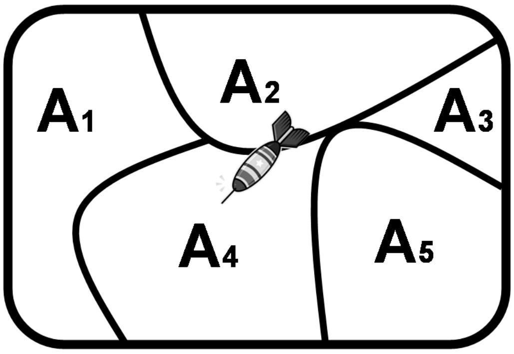

Figure 3. A board divided into five unequal regions.

Now the experiment is carried out, and you have to find out where the dart hit the board. You know that the dart hit the board, and you know the probability distribution $p_1, \ldots, p_n$. Your task is to find out in which region the dart is by asking binary questions, i.e., questions answerable by Yes, or No.

Clearly, since you do not *know* where the dart is, you *lack information* on the location of the dart. To acquire this information you ask questions. We are interested in the *amount* of *information* contained in this experiment. One way of measuring this amount of information is by the *number* of *questions* you need to ask in order to obtain the required information.

As everyone who has played the 20-question (20Q) game knows, the number of questions you need to ask depends on the *strategy* for asking questions. Here we shall not discuss what constitutes a strategy for asking questions [9]. Here we are only interested in a measure of the "amount of information" contained in this experiment. It turns out that the quantity H, to which we referred to as Shannon's measure of information (SMI), provides us with a measure of this information in terms of the minimum number of questions one needs to ask in order to find the location of the dart, given the *probability distribution* of the various outcomes [2,8].

For a general experiment with n possible outcomes, having probabilities $p_1, \ldots, p_n$, the quantity H is a measure of how "difficult" it is to find out which outcome has occurred, given that an experiment was carried out. It is easy to see that for experiments having the same total number of outcomes n, but with different probability distributions, the amount of information (measured in terms of the number of questions) is different. In other words, knowing the probability distribution gives us a "hint" or some partial information on the outcomes. This is the reason why we refer to H as a measure of the amount of information *contained* in, or *associated* with a given probability distribution. We emphasize again that the SMI is a measure of information associated with the *entire* distribution, not with the individual probabilities.

3. Derivation of the Entropy Function for an Ideal Gas

In this section we derive the entropy function for an ideal gas. We start with SMI which is definable to any probability distribution [9,10]. We apply the SMI to two molecular distributions; the locational and the momentum distribution. Next, we calculate the distribution which maximizes the SMI. We refer to this distribution as the *equilibrium* distribution. Finally, we apply two corrections to the SMI, one due to Heisenberg uncertainty principle, the second due to the indistinguishability of the particles. The resulting SMI is, up to a multiplicative constant equal to the entropy of the gas, as calculated by Sackur and Tetrode based on Boltzmann definition of entropy [11,12].

In previous publication [2,13], we discussed several advantages to the SMI-based definition of entropy. For our purpose in this article the most important aspect of this definition is the following:

The entropy is *defined* as the maximum value of the SMI. As such, it is not a function of time. We shall discuss the implication of this conclusion for the Boltzmann H-theorem in Section 4.

3.1. The Locational SMI of a Particle in a 1D Box of Length L

Suppose we have a particle confined to a one-dimensional (1D) "box" of length L. Since there are infinite points in which the particle can be within the interval (0, L). The corresponding locational SMI must be infinity. However we can defined, as Shannon did, the following quantity by analogy with the discrete case:

$$H(X) = -\int f(x) \log f(x) dx \tag{2}$$

This quantity might either converge or diverge, but in any case, in practice we shall use only differences of this quantity. It is easy to calculate the density which maximizes the locational SMI, $H(X)$ in (2) which is [1,2]:

$$f_{eq}(x) = \frac{1}{L} \tag{3}$$

The use of the subscript *eq* (for equilibrium) will be cleared later, and the corresponding SMI calculated by (2) is:

$$H(\text{locations in } 1D) = \log L \tag{4}$$

We acknowledge that the location X of the particle cannot be determined with absolute accuracy, i.e., there exists a small interval, h_x within which we do not care where the particle is. Therefore, we must correct Equation (4) by subtracting $\log h_x$. Thus, we write instead of (4):

$$H(X) = \log L - \log h_x \tag{5}$$

We recognize that in (5) we effectively defined $H(X)$ for a finite number of intervals $n = L/h$. Note that when $h_x \to 0$, $H(X)$ diverges to infinity. Here, we do not take the mathematical limit, but we stop at h_x small enough but not zero. Note also that in writing (5) we do not have to specify the units of length, as long as we use the same units for L and h_x.

3.2. The Velocity SMI of a Particle in 1D "Box" of Length L

Next, we calculate the probability distribution that maximizes the continuous SMI, subject to two conditions:

$$\int_{-\infty}^{\infty} f(x) dx = 1 \tag{6}$$

$$\int_{-\infty}^{\infty} x^2 f(x) dx = \sigma^2 = constant \tag{7}$$

The result is the Normal distribution [1,2]:

$$f_{eq}(x) = \frac{\exp\left[-x^2/\sigma^2\right]}{\sqrt{2\pi\sigma^2}} \tag{8}$$

The subscript eq. for equilibrium will be clear later. Applying this result to a classical particle having average kinetic energy $\frac{m<v_x^2>}{2}$, and identifying the standard deviation σ^2 with the temperature of the system:

$$\sigma^2 = \frac{k_B T}{m} \tag{9}$$

We get the equilibrium velocity distribution of one particle in 1D system:

$$f_{eq}(v_x) = \sqrt{\frac{m}{2mk_B T}}\, \exp\left[\frac{-mv_x^2}{2k_B T}\right] \tag{10}$$

where k_B is the Boltzmann constant, m is the mass of the particle, and T the absolute temperature. The value of the continuous SMI for this probability density is:

$$H_{max}(\text{velocity in } 1D) = \frac{1}{2}\log(2\pi e k_B T/m) \tag{11}$$

Similarly, we can write the momentum distribution in 1D, by transforming from $v_x \to p_x = mv_x$, to get:

$$f_{eq}(p_x) = \frac{1}{\sqrt{2\pi m k_B T}}\, \exp\left[\frac{-p_x^2}{2mk_B T}\right] \tag{12}$$

and the corresponding maximal SMI:

$$H_{max}(\text{momentum in } 1\ D) = \frac{1}{2}\log(2\pi e m k_B T) \tag{13}$$

As we have noted in connection with the locational SMI, the quantities (11) and (13) were calculated using the definition of the *continuous* SMI. Again, recognizing the fact that there is a limit to the accuracy within which we can determine the velocity, or the momentum of the particle, we correct the expression in (13) by subtracting $\log h_p$ where h_p is a small, but infinite interval:

$$H_{max}(momentum\ in\ 1D) = \frac{1}{2}\log(2\pi e m k_B T) - \log h_p \tag{14}$$

Note again that if we choose the units of h_p (of momentum as: *mass length/time*) the same as of $\sqrt{mk_B T}$, then the whole expression under the logarithm will be a pure number.

3.3. Combining the SMI for the Location and Momentum of One Particle in 1D System

In the previous two sections, we derived the expressions for the locational and the momentum SMI of one particle in 1D system. We now combine the two results. Assuming that the location and the momentum (or velocity) of the particles are independent events we write

$$H_{max}(location\ and\ momentum) = H_{max}(location) + H_{max}(momentum) = \log\left[\frac{L\sqrt{2\pi e m k_B T}}{h_x h_p}\right] \tag{15}$$

Recall that h_x and h_p were chosen to eliminate the divergence of the SMI for a continuous random variables; location and momentum.

In (15) we assume that the location and the momentum of the particle are independent. However, quantum mechanics imposes restriction on the accuracy in determining both the location x and the corresponding momentum p_x. In Equations (5) and (14) h_x and h_p were introduced because we did not care to determine the location and the momentum with an accuracy greater that h_x and h_p, respectively. Now, we must acknowledge that nature imposes upon us a limit on the accuracy with which we can determine both the location and the corresponding momentum. Thus, in Equation (15),

h_x and h_p cannot both be arbitrarily small, but their product must be of the order of Planck constant $h = 6.626 \times 10^{-34}\ J\ s$. Thus we set:

$$h_x h_p \approx h \tag{16}$$

And instead of (15), we write:

$$H_{max}(location\ and\ momentum) = \log\left[\frac{L\sqrt{2\pi e m k_B T}}{h}\right] \tag{17}$$

3.4. The SMI of a Particle in a Box of Volume V

We consider again one simple particle in a box of volume V. We assume that the location of the particle along the three axes x, y and z are independent. Therefore, we can write the SMI of the location of the particle in a cube of edges L, and volume V as:

$$H(location\ in\ 3D) = 3H_{max}(location\ in\ 1D) \tag{18}$$

Similarly, for the momentum of the particle we assume that the momentum (or the velocity) along the three axes x, y and z are independent. Hence, we write:

$$H_{max}(momentum\ in\ 3D) = 3H_{max}(momentum\ in\ 1D) \tag{19}$$

We combine the SMI of the locations and momenta of one particle in a box of volume V, taking into account the uncertainty principle. The result is

$$H_{max}(location\ and\ momentum\ in\ 3D) = 3\log\left[\frac{L\sqrt{2\pi e m k_B T}}{h}\right] \tag{20}$$

3.5. The SMI of Locations and Momenta of N Independent Particles in a Box of Volume V

The next step is to proceed from one particle in a box to N independent particles in a box of volume V. Giving the location (x, y, z), and the momentum (p_x, p_y, p_z) of one particle within the box, we say that we know the *microstate* of the particle. If there are N particles in the box, and if their microstates are independent, we can write the SMI of N such particles simply as N times the SMI of one particle, i.e.:

$$\text{SMI}(\text{of } N \text{ independent particles}) = N \times \text{SMI}(\text{one particle}) \tag{21}$$

This Equation would have been correct when the microstates of all the particles where independent. In reality, there are always correlations between the microstates of all the particles; one is due to *intermolecular interactions* between the particles, the second is due to the *indistinguishability* between the particles. We shall discuss these two sources of correlation separately.

(i) Correlation Due to Indistinguishability

Recall that the microstate of a single particle includes the location and the momentum of that particle. Let us focus on the location of one particle in a box of volume V. We have written the locational SMI as:

$$H_{max}(location) = \log V \tag{22}$$

Recall that this result was obtained for the continuous locational SMI. This result does not take into account the divergence of the limiting procedure. In order to explain the source of the correlation due to indistinguishability, suppose that we divide the volume V into a very large number of small cells each of the volume V/M. We are not interested in the exact location of each particle, but only in which cell each particle is. The total number of cells is M, and we assume that the total number of particles is $N \ll M$. If each cell can contain at most one particle, then there are M possibilities to put the first particle in one of the cells, and there are $M-1$ possibilities to put the second particle in the remaining empty cells. Altogether, we have $M(M-1)$ possible microstates, or configurations for two particles. The probability that one particle is found in cell i, and the second in a different cell j is:

$$\Pr(i,j) = \frac{1}{M(M-1)} \tag{23}$$

The probability that a particle is found in cell i is:

$$\Pr(j) = \Pr(i) = \frac{1}{M} \tag{24}$$

Therefore, we see that even in this simple example, there is correlation between the events "one particle in i" and one particle in j":

$$g(i,j) = \frac{\Pr(i,j)}{\Pr(i)\Pr(j)} = \frac{M^2}{M(M-1)} = \frac{1}{1-\frac{1}{M}} \tag{25}$$

Clearly, this correlation can be made as small as we wish, by taking $M \gg 1$ (or in general, $M \gg N$). There is another correlation which we cannot eliminate and is due to the indistinguishability of the particles.

Note that in counting the total number of configurations we have implicitly assumed that the two particles are labeled, say red and blue. In this case we count the two configurations in Figure 4a, as *different* configurations: "blue particle in cell i, and red particle in cell j", and "blue particle in cell j, and red particle in cell i".

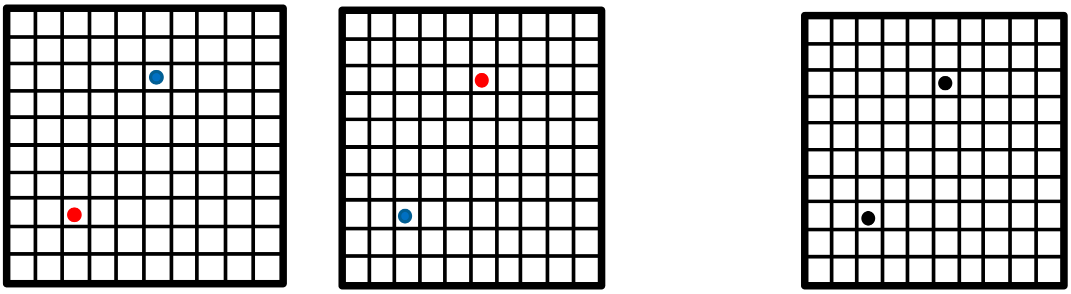

Figure 4. Two different configurations are reduced to one when the particles are indistinguishable.

Atoms and molecules are indistinguishable by nature; we cannot label them. Therefore, the two microstates (or configurations) in Figure 4b are indistinguishable. This means that the total number of configurations is not $M(M-1)$, but:

$$number\ of\ configurations = \frac{M(M-1)}{2} \to \frac{M^2}{2},\ \text{for large } M \tag{26}$$

For very large M we have a correlation between the events "particle in i" and "particle in j":

$$g(i,j) = \frac{\Pr(i,j)}{\Pr(i)\Pr(j)} = \frac{M^2}{M^2/2} = 2 \tag{27}$$

For N particles distributed in M cells, we have a correlation function (For $M \gg N$):

$$g(i_1, i_2, \ldots, i_n) = \frac{M^N}{M^N/N!} = N! \tag{28}$$

This means that for N indistinguishable particles we must divide the number of configurations M^N by $N!$. Thus in general by removing the "labels" on the particles the number of configurations is *reduced* by *N!*. For two particles the two configurations shown in Figure 4a reduce to one shown in Figure 4b.

Now that we know that there are correlations between the events "one particle in i_1", "one particle in i_2" … "one particle in i_n", we can define the *mutual information* corresponding to this correlation. We write this as:

$$I(1;2;\ldots;N) = \ln N! \tag{29}$$

The SMI for N particles will be:

$$H(N\ particles) = \sum_{i=1}^{N} H(one\ particle) - \ln N! \tag{30}$$

For the definition of the mutual information, see [2].

Using the SMI for the location and momentum of one particle in (20) we can write the final result for the SMI of N indistinguishable (but non-interacting) particles as:

$$H(N\ \text{indistinguishable}) = N \log\ V\left(\frac{2\pi mek_BT}{h^2}\right)^{3/2} - \log N! \tag{31}$$

Using the Stirling approximation for $\log N!$ in the form (note again that we use the natural logarithm):

$$\log N! \approx N \log\ N - N \tag{32}$$

We have the final result for the SMI of N indistinguishable particles in a box of volume V, and temperature T:

$$H(1,2,\ldots N) = N \log\left[\frac{V}{N}\left(\frac{2\pi mk_BT}{h^2}\right)^{3/2}\right] + \frac{5}{2}N \tag{33}$$

By multiplying the SMI of N particles in a box of volume V at temperature T, by the factor $(k_B\ \log_e 2)$, one gets the *entropy*, the *thermodynamic entropy* of an ideal gas of simple particles. This equation was derived by Sackur and by Tetrode in 1912, by using the Boltzmann definition of entropy.

One can convert this expression into the entropy function $S(E,V,N)$, by using the relationship between the total energy of the system, and the total kinetic energy of all the particles:

$$E = N\frac{m\langle v\rangle^2}{2} = \frac{3}{2}Nk_BT \tag{34}$$

The explicit entropy function of an ideal gas is:

$$S(E,V,N) = Nk_B \ln\left[\frac{V}{N}\left(\frac{E}{N}\right)^{3/2}\right] + \frac{3}{2}k_BN\left[\frac{5}{3} + \ln\left(\frac{4\pi m}{3h^2}\right)\right] \tag{35}$$

(ii) Correlation Due to Intermolecular Interactions

In Equation (35) we got the entropy of a system of non-interacting simple particles (ideal gas). In any real system of particles, there are some interactions between the particles. One of the simplest interaction energy potential function is shown in Figure 5. Without getting into any details on the function $U(r)$ shown in the Figure, it is clear that there are two regions of distances $0 \le r \lesssim \sigma$ and $0 \le r \lesssim \infty$, where the slope of the function $U(r)$ is negative and positive, respectively. Negative slope correspond to repulsive forces between the pair of the particles when they are at a distance smaller than σ. This is the reason why σ is sometimes referred to as the *effective diameter* of the particles. For larger distances, $r \gtrsim \sigma$ we observe attractive forces between the particles.

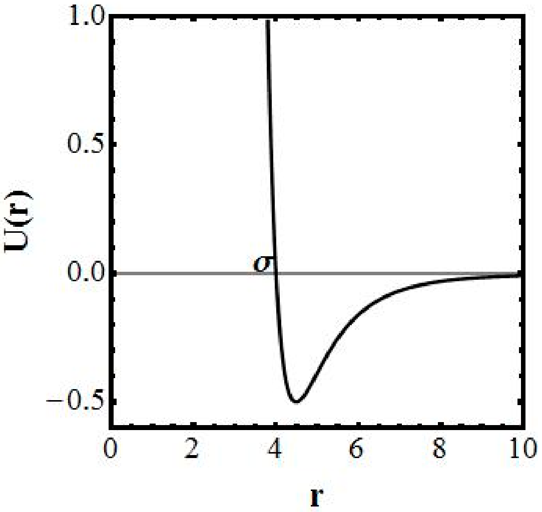

Figure 5. The general form of the pair-potential between two particles.

Intuitively, it is clear that interactions between the particles induce *correlations* between the locational probabilities of the two particles. For hard-spheres particles there is infinitely strong repulsive force between two particles when they approach to a distance of $r \leq \sigma$. Thus, if we know the location $\boldsymbol{R_1}$ of one particle, we can be sure that a second particle, at $\boldsymbol{R_2}$ is not in a sphere of diameter σ around the point $\boldsymbol{R_1}$. This *repulsive* interaction may be said to introduce *negative correlation* between the locations of the two particles.

On the other hand, two argon atoms *attract* each other at distances $r \lesssim 4A$. Therefore, if we know the location of one particle say, at $\boldsymbol{R_1}$, the probability of observing a second particle at $\boldsymbol{R}_2$ is *larger* than the probability of finding the particle at $\boldsymbol{R_2}$ in the absence of a particle at $\boldsymbol{R_1}$. In this case we get *positive correlation* between the locations of the two particles.

We can conclude that in both cases (attraction and repulsion) there are correlations between the particles. These correlations can be cast in the form of *mutual information* which reduces the SMI of a system of N simple particles in an ideal gas. The mathematical details of these correlations are discussed in Ben-Naim [3].

Here, we show only the form of the mutual information (MI) in very low density. At this limit, we can assume that there are only *pair correlations*, and neglect all higher order correlations. The MI due to these correlations is:

$$I(due\ to\ correlations\ in\ pairs) = \frac{N(N-1)}{2} \int p(\boldsymbol{R_1}, \boldsymbol{R_2}) \log g(\boldsymbol{R_1}, \boldsymbol{R_2}) d\boldsymbol{R_1} d\boldsymbol{R_2} \tag{36}$$

where $\boldsymbol{g}(R_1, R_2)$ is defined by:

$$g(\boldsymbol{R_1}, \boldsymbol{R_2}) = \frac{p(\boldsymbol{R_1}, \boldsymbol{R_2})}{p(\boldsymbol{R_1})p(\boldsymbol{R_2})} \tag{37}$$

Note again that log g can be either positive or negative, but the average in (36) must be positive.

3.6. Conclusions

We summarize the main steps leading from the SMI to the entropy. We started with the SMI associated with the *locations* and *momenta* of the particles. We calculated the distribution of the locations and momenta that *maximizes* the SMI. We referred to this distribution as the *equilibrium distribution*. Let us denote this distribution of the locations and momenta of all the particles by $f_{eq}(\boldsymbol{R}, \boldsymbol{p})$.

Next, we use the equilibrium distribution to calculate the SMI of a system of N particles in a volume V, and at temperature T. This SMI is, up to a multiplicative constant $(k_B \ln 2)$ identical with the *entropy* of an ideal gas at equilibrium. This is the reason we referred to the distribution which maximizes the SMI as the *equilibrium distribution*.

It should be noted that in the derivation of the entropy, we used the SMI twice; first, to calculate the distribution that maximize the SMI, then evaluating the maximum SMI corresponding to this distribution. The distinction between the concepts of SMI and entropy is essential. Referring to SMI (as many do) as entropy, inevitably leads to such an awkward statement: the maximal value of the entropy (meaning the SMI) is the entropy (meaning the thermodynamic entropy). The correct statement is that the SMI associated with locations and momenta is defined for any system; small or large, at equilibrium or far from equilibrium. This SMI, not the entropy, evolves into a maximum value when the system reaches equilibrium. At this state, the SMI becomes proportional to the entropy of the system.

Since the entropy is a special case of a SMI, it follows that whatever interpretation one accepts for the SMI, it will be automatically applied to the concept of entropy. The most important conclusion is that entropy is not a function of time. Entropy does not change with time, and entropy does not have a tendency to increase.

We said that the SMI may be defined for a system with any number of particles including the case $N = 1$. This is true for the SMI. When we talk about the entropy of a system we require that the system be very large. The reason is that only for such systems the entropy-formulation of the Second Law of thermodynamic is valid. This topic is discussed in the next section.

3.7. The Entropy Formulation of the Second Law

In the previous section, we derived and interpreted the concept of entropy. Knowing what entropy is leaves the question of "why entropy always increases," unanswered.

This question is considered to be one of the most challenging one. This property of the entropy is also responsible for the mystery surrounding the concept of the entropy. In this section, we discuss very briefly the origin of the increase in entropy in one specific process. The correct answer to the question of "why entropy always increases" removes much of the mystery associated with entropy.

In this section, we "derive" the correct *answer* to the correct *questions*; when and why entropy of a system increases?

Consider the following process. We have a system characterized by E, V, N. (This means N particles, in a volume V having total energy E). We assume that all the energy of the system is due to the kinetic energy of the particles. We neglect any interactions between the particles, and if the particles have any internal energies (say, vibrational, rotational, electronic, nuclear, etc.), these will not change in the process. We now remove a partition between the two compartments, as in Figure 6, and observe what happens. Experience tells us that once we remove the partition, the gas will expand to occupy the entire system of volume $2V$. Furthermore, if both the initial and the final states are equilibrium states, then we can apply the entropy function to calculate the change in the entropy in this process, i.e.:

$$\Delta S(V \to 2V) = N k_B \ln \frac{2V}{V} = N k_B \ln 2 \tag{38}$$

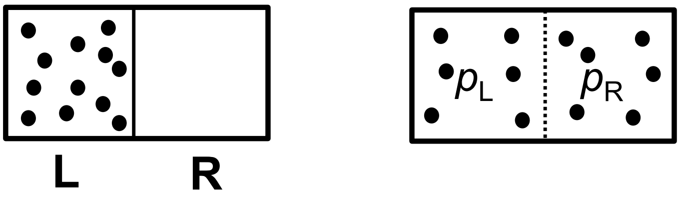

Figure 6. Expansion of an ideal gas from V to $2V$.

Note carefully that this entropy change corresponds to the *difference* in the entropy of the system at two equilibrium states; the initial and the final states, Figure 6.

The informational interpretation of this quantity can be obtained by dividing ΔS by the constant factor $k_B \ln 2$ and we get:

$$\Delta H(V \rightarrow 2V) = \frac{\Delta S}{k_B \ln 2} = N \tag{39}$$

This means that the SMI of the system increased by N bits. The reason is simple. Initially, we know that all N particles are in a volume V, and after removal of the partition we lost one bit per particle. We need to ask one question to find out where a particle is: in the right (R), or the left (L) compartment.

Now that we understand the meaning of this entropy change we turn to study the *cause* for this entropy change. Specifically, we ask:

Why does the entropy of this process increase? Before we answer this question we will try to answer the more fundamental question:

Why did this process occur at all?

We shall see that an answer to the second question leads to an answer to the first question.

Clearly, if the partition separating the two compartments is not removed nothing will happen; the gas will not expand and the entropy of the system will not change. We can tentatively conclude that having a system characterized by (E, V, N) the entropy is *fixed* and will not change with time.

Let us examine what will happen when we remove the partition separating the two compartments in Figure 6.

Instead of removing the entire partition, we open a small window between the two compartments. This will allow us to follow the process in small steps. If the window is small enough, we can expect only one particle at the time to pass through it.

Starting with all the N particles on the left compartment, we open the window and observe what will happen.

Clearly, the first particle which crosses the window will be from the left (L) to the right (R) compartment. This is clear simply because there are no particles in the R compartment.

After some time, some particles will move from L to R. Denote the number of particles in R by n and the number in L by $N - n$. The pair of numbers $(N - n, n)$ may be referred to as a *distribution* of *particles* in the two compartments. Dividing by N, we get a pair of numbers $(p_L, p_R) = (1 - p, p)$ where $p = \frac{n}{N}$. Clearly, this pair of numbers is a *probability distribution* $(p_L, p_R \geq 0, p_L + p_L = 1)$. We can refer to it as the temporary probability distribution, or simply the *state distribution* (More precisely, this is the locational state of the particles. Since we have an ideal gas, the energy, the temperature, and the velocity or momentum distribution of the particles will not change in this process).

For each *state distribution*, $(1 - p, p)$ we can define the corresponding SMI by:

$$H(p) = -p \log\ p - (1 - p) \log p \tag{40}$$

Note that p changes with time, as a result also $H(p)$ will change with time. If we follow the change of the SMI we will observe a nearly monotonic increasing function of time. For actual simulations, see Ben-Naim (2008, 2010) [3,9]. The larger N, the more monotonic the curve will be and once n reaches the value: $N/2$, the value of the SMI will stay there "forever." For any N, there will be fluctuations, both on the way up to the maximum, as well as after reaching the maximum. However, for very large N these fluctuations will be unnoticeable. After some time we reach an equilibrium state. The equilibrium states is reached when the locational distribution is such that it maximizes the SMI, namely:

$$p_{eq} = \frac{N}{2} \tag{41}$$

and the corresponding SMI is:

$$H_{max} = N\left[-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}\right] = N \tag{42}$$

Note again that here we are concerned with the locational distribution with respect to being either in L, or in R. The momentum distribution does not change in this process.

Once we reached the equilibrium state, we can ask: What is the *probability* of finding the system, such that there are $N - n$ in L, and n in the R? Since the probability of finding a specific particle in either L or R is 1/2, the probability of finding the probability distribution (*N–n, n),* is:

$$\Pr(N-n,n) = \frac{N!}{n!(N-n)!}\left(\frac{1}{2}\right)^{N-n}\left(\frac{1}{2}\right) = \left(\begin{array}{c} N \\ n \end{array}\right)\left(\frac{1}{2}\right)^{N^n} \tag{43}$$

It is easy to show that this probability function has a maximum at $n = \frac{N}{2}$.

Clearly, if we sum over all n, and use the Binomial theorem, we get, as expected:

$$\sum_{n=0}^{N}\Pr(N-n,n) = \sum_{n=0}^{N}\left(\begin{array}{c} N \\ n \end{array}\right)\left(\frac{1}{2}\right)^{N} = \left(\frac{1}{2}\right)^{N}2^{N} = 1 \tag{44}$$

We now use the Stirling approximation:

$$\ln N! \approx N\ln N - N \tag{45}$$

To rewrite (43) as:

$$\ln\Pr(1-p,p) \approx -N\ln 2 - N[(1-p)\ln(1-p) + p\ln p] \tag{46}$$

or equivalently, after dividing by $\ln 2$, we get:

$$\Pr(1-p,p) \approx \left(\frac{1}{2}\right)^{N}2^{NH(p)} \tag{47}$$

If we use instead of the approximation (45) the following approximation:

$$\ln N! \approx N\ln N - N + \frac{1}{2}\ln(2\pi N) \tag{48}$$

We get instead of (47) the approximation:

$$\Pr(1-p,p) \approx \left(\frac{1}{2}\right)^{N}\frac{2^{NH(p)}}{\sqrt{2\pi Np(1-p)}} \tag{49}$$

Note that in general the probability Pr of finding the distribution $(1-p,p)$ is related to the SMI of that distribution. We now compare the probability of finding the state distribution $\left(\frac{1}{2},\frac{1}{2}\right)$ with the probability of finding the state distribution (1, 0). From (49) we have:

$$\Pr\left(\frac{1}{2},\frac{1}{2}\right) = \sqrt{\frac{2}{\pi N}} \tag{50}$$

For the state (1,0) we can use the exact expression (43):

$$\Pr(1,0) = \left(\frac{1}{2}\right)^{N} \tag{51}$$

The ratio of these two probabilities is:

$$\frac{\Pr\left(\frac{1}{2},\frac{1}{2}\right)}{\Pr(1,0)}=\sqrt{\frac{2}{\pi N}}2^N \tag{52}$$

Note carefully that $\Pr\left(\frac{1}{2},\frac{1}{2}\right)$ *decreases* with N. However, the ratio of the two probabilities in (52) *increases* with N.

The corresponding difference in the SMI is:

$$H\left(\frac{1}{2},\frac{1}{2}\right)-H(1,0)=N-0=N \tag{53}$$

What about the entropy change? To calculate the entropy difference in this process, let us denote by $S_i=S(E,V,N)$ the entropy of the initial state. The entropy at the final state $S_f=S(E,2V,N)$ may be obtained by multiplying (53) by $k_B\ln 2$, and add it to S_i:

$$S_f=S_i+(k_B\ln 2)N \tag{54}$$

The change in entropy is therefore:

$$\Delta S=S_f-S_i=Nk_B\ln 2 \tag{55}$$

which agrees with (38). It should be emphasized that the ratio of probabilities (52) and the difference in the entropies in (55) are computed for *different* states of the same system. In (55), S_f and S_i are the entropies of the system at the *final* and *initial equilibrium* states, respectively. These two equilibrium states are $S(E,2V,N)$ and $S(E,V,N)$, respectively. In particular, $S(E,V,N)$ is the entropy of the system *before* removing the partition.

On the other hand, the ratio of the probability in (52) is calculated at equilibrium *after* removing the partition.

We can now answer the question posed in the beginning of this section. After the removal of the partition, the gas will expand and attend a new equilibrium state. The *reason* for the change from the initial to the final state is probabilistic. The probability of the final state $\left(\frac{1}{2},\frac{1}{2}\right)$ is overwhelmingly larger than the probability of the initial state (1,0) *immediately after the removal of the partition*. As a result of the monotonic relationship between the probability $\Pr(1-p,p)$, and the SMI, whenever the probability increases, the SMI increases too. At the state for which the SMI is maximum, we can calculate the change in entropy which is larger by $Nk_B\ln 2$ relative to the entropy of the initial equilibrium state S_i, i.e., *before* the removal of the partition. We can say that the process of expansion occurs *because* of the overwhelmingly larger probability of the final equilibrium state. The increase in the entropy of the system is a *result* of the expansion process, not the *cause* of the process.

3.8. Caveat

Quite often, one might find in textbooks the Boltzmann definition of entropy in terms of the number of states:

$$S=k_B\ln W \tag{56}$$

W, in this equation is often referred to as probability. Of course, W cannot be a probability, which by definition is a number between zero and one. More careful writers will tell you that the ratio of the number of states is the ratio of the probabilities, i.e., for the final and the initial states, one writes:

$$\frac{W_f}{W_i}=\frac{\Pr(f)}{\Pr(i)} \tag{57}$$

This is true but one must be careful to note that while W_i is the number of states of the system *before* the removal of the partition, the corresponding probability $\Pr(i)$ pertains to the same system *after* the removal of the partition.

Very often you might find the erroneous statement of the second law based on Equation (56) as follows: the number of states of the system tends to increase, therefore the entropy tends to increase too. This statement is not true; both W and S in (56) are defined for an equilibrium state, and both do not have a tendency to increase with time!

4. Boltzmann's H-Theorem

Before we discuss Boltzmann's H-theorem, we summarize here the most important conclusion regarding the SMI.

In Section 3, we saw that the entropy is obtained from the SMI in four steps. We also saw that the entropy of a thermodynamic system is related to the *maximum* value of the SMI defined on the distribution of locations and velocities of all particles in the system:

$$S = KMaxSMI(locations\ and\ velocities) \tag{58}$$

where K is a constant $(K = k_B \ln 2)$.

We know that every system tends to an equilibrium state at very long time, therefore we identify the Max SMI as the time limit of the SMI, i.e.:

$$S = K \lim_{t \to \infty} SMI(locations\ and\ velocities) \tag{59}$$

The derivation of the entropy from the SMI is a very remarkable result. But what is more important is that this derivation reveals at the same time the relationship between entropy and SMI on one hand, and the fundamental difference between the two concepts, on the other hand.

Besides the fact that the SMI is a far more general concept than entropy, we found that even when the two concepts apply to the distribution of locations and velocities, they are different. The SMI can evolve with time and reaches a limiting value (for large systems) at $t \to \infty$.

The entropy is proportional to the maximum value of the SMI obtained at equilibrium. As such entropy is not, and cannot be a function of time. Thus, the "well-known" mystery about the "entropy always increase with time," disappears. With this removal of the mystery, we also arrive at the resolution of the "paradoxes" associated with the Boltzmann H-theorem.

In 1877 Boltzmann defined a function $H(t)$ [14–16]:

$$H(t) = \int f(v,t) \log[f(v,t)] dv \tag{60}$$

and proved a remarkable theorem known as Boltzmann's H-theorem. Boltzmann made the following assumptions:

1. Ignoring the molecular structure of the walls (ideal. perfect smooth walls).
2. Spatial homogenous system or uniform locational distribution.
3. Assuming binary collisions, conserving momentum and kinetic energy.
4. No correlations between location and velocity (assumption of molecular chaos).

Then, Boltzmann proved that:

$$\frac{dH(t)}{dt} \leq 0 \tag{61}$$

and at equilibrium, i.e., $t \to \infty$:

$$\frac{dH(t)}{dt} = 0 \tag{62}$$

Boltzmann believed that the behavior of the function $-H(t)$ is the same as that of the entropy, i.e., the entropy always increases with time, and at equilibrium, it reaches a maximum. At this time, the entropy does not change with time. This theorem drew a great amount of criticism, the most well-known are:

I. The "Reversal Paradox" States:

"The H-theorem singles out a preferred direction of time. This is inconsistent with the time reversal invariance of the equations of motion". This is not a paradox because the statement that $H(t)$ always changes in one direction is false.

II. The "Recurrence Paradox", Based on Poincare's Theorem States:

After sufficiently long time, an isolated system with fixed E, V, N, will return to arbitrary small neighborhood of almost any given initial state.

If we assume that $dH/dT < 0$ at all t, then obviously H cannot be periodic function of time.

Both paradoxes have been with us ever since. Furthermore, most popular science books identify the Second Law, or the behavior of the entropy with the so-called *arrow* of *time*. Some even go to the extremes of *identifying* entropy with time [8,17,18].

Both paradoxes seem to arise from the conflict between the *reversibility* of the equations of motion on one hand, and the apparent *irreversibility* of the Second Law, namely that the H-function decreases monotonically with time. Boltzmann rejected the criticism by claiming that H does not always decrease with time, but only with high probability. The irreversibility of the Second Law is not absolute, but also highly improbable. The answer to the recurrence paradox follows from the same argument. Indeed, the system can return to the initial state. However, the recurrence time is so large that this is never observed, not in our lifetime, not even in the life time of the universe.

Notwithstanding Boltzmann's correct answers to his critics, Boltzmann and his critics made an enduring mistake in the H-theorem, a lingering mistake that has hounded us ever since. This is the very identification of the $-H(t)$ with the behavior of the entropy. This error has been propagated in the literatures until today.

It is clear, from the very definition of the function $H(t)$, that $-H(t)$ is a SMI. And if one identifies the SMI with entropy, then we go back to Boltzmann's identification of the function $-H(t)$ with entropy.

Fortunately, thanks to the recent derivation of the *entropy function*, i.e., the function $S(E, V, N)$, or the Sackur-Tetrode equation for the entropy based on the SMI, it becomes crystal clear that the SMI is not entropy! The entropy is obtained from the SMI when you apply it to the distribution of locations and momenta, then take the limit $t \to \infty$, and only in this limit we get entropy function which has no traces of time dependence.

Translating our findings in Section 3 to the H-theorem, we can conclude that $-H(t)$ is SMI based on the velocity distribution. Clearly, one cannot identify $-H(t)$ with entropy. To obtain the entropy one must first define the $-H(t)$ function based on the distribution of both the locations and momentum, i.e.:

$$-H(t) = -\int f(\boldsymbol{R}, \boldsymbol{p}, t) \log f(\boldsymbol{R}, \boldsymbol{p}, t) d\boldsymbol{R} d\boldsymbol{p} \tag{63}$$

This is a proper SMI. This may be defined for a system at equilibrium, or very far from equilibrium. To obtain the entropy one must take the limit $t \to \infty$, i.e., the limit $-H(t)$ at equilibrium, i.e.:

$$\lim_{t \to \infty} [-H(t)] = Max\ SMI\ (at\ equilibrium) \tag{64}$$

At this limit we obtain the entropy (up to a multiplicative constant), which is clearly not a function of time.

Thus, once it is understood that the function $-H(t)$ is an SMI and not entropy, it becomes clear that the criticism of Boltzmann's H-Theorem were addressed to the evolution of the SMI and not to the entropy. At the same time, Boltzmann was right in defending his H-theorem when viewed as a theorem on the evolution of SMI, but he was wrong in his interpretation of the quantity $-H(t)$ as entropy.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; The University of Illinois Press: Chicago, IL, USA, 1949.
2. Ben-Naim, A. *Information Theory*; World Scientific: Singapore, 2017.
3. Ben-Naim, A. *A Farewell to Entropy: Statistical Thermodynamics Based on Information*; World Scientific: Singapore, 2008.
4. Ben-Naim, A.; Casadei, D. *Modern Thermodynamics*; World Scientific: Singapore, 2016.
5. Ben-Naim, A. *Entropy and the Second Law. Interpretation and Misss-Interpretationsss*; World Scientific: Singapore, 2012.
6. Ben-Naim, A. *Discover Probability. How to Use It, How to Avoid Misusing It, and How It Affects Every Aspect of Your Life*; World Scientific: Singapore, 2015.
7. Yaglom, A.M.; Yaglom, I.M. *Probability and Information*; Jain, V.K., Reidel, D., Eds.; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1983.
8. Ben-Naim, A. *Information, Entropy, Life and the Universe. What We Know and What We Do Not Know*; World Scientific: Singapore, 2015.
9. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [CrossRef]
10. Jaynes, E.T. Information theory and statistical mechanics, Part II. *Phys. Rev.* **1957**, *108*, 171–189. [CrossRef]
11. Ben-Naim, A. The entropy of mixing and assimilation: An information-theoretical perspective. *Am. J. Phys.* **2006**, *74*, 1126–1135. [CrossRef]
12. Ben-Naim, A. An Informational-Theoretical Formulation of the Second Law of Thermodynamics. *J. Chem. Educ.* **2009**, *86*, 99–105. [CrossRef]
13. Ben-Naim, A. *Entropy, the Truth the Whole Truth and nothing but the Truth*; World Scientific: Singapore, 2016.
14. Boltzmann, L. *Lectures on Gas Theory*; Dover Publications: New York, NY, USA, 2012.
15. Brush, S.G. *The Kind of Motion We Call Heat. A History of the Kinetic Theory of Gases in the 19th Century, Book 2: Statistical Physics and Irreversible Processes*; North-Holland Publishing Company: Amsterdam, The Netherlands, 1976.
16. Brush, S.G. *Statistical Physics and the Atomic Theory of Matter, from Boyle and Newton to Landau and Onsager*; Princeton University Press: Princeton, NJ, USA, 1983.
17. Ben-Naim, A. *Discover Entropy and the Second Law of Thermodynamics. A Playful Way of Discovering a Law of Nature*; World Scientific: Singapore, 2010.
18. Ben-Naim, A. Entropy: Order or Information. *J. Chem. Educ.* **2011**, *88*, 594–596. [CrossRef]