SELECTED ARTICLES

ENTROPY

ODED KAFRI

KSP BOOKS

Entropy

Selected Articles

Oded Kafri
Retired Professor, Israel

KSP Books
econsciences.com

Entropy
Selected Articles

Oded Kafri

KSP Books
econsciences.com

ISBN: 978-605-2132-51-7 (e-Book)
KSP Books 2018

Entropy: Selected Articles

Author(s): ***Oded Kafri***
Retired Professor, Israel

econsciences.com

Preface

Can physics describe economy? What are the physical driving force behind the evolution and our sociological and economic activity?

In this collection of papers I argue that entropy, under a certain condition, is information; and therefore information has a tendency to increase. Based on this, the statistical properties of information networks are calculated, and in contradistinction to the classic sparse system that yields the canonical bell-like distribution, information networks are found to have a universal long tail distribution. This distribution predicts, with great accuracy, the wealth distribution; the relative poverty; and wealth inequality in the economy and the popularity of sites on the internet. This long tail distribution, which is called Planck-Benford distribution, yields correctly Zipf law, Pareto law, and Benford law.

Since Planck- Benford distribution has no free parameter it is a universal distribution and it can be applied to the analysis of polls, bestseller lists, earthquakes, etc.

When the statistics of the information network in equilibrium is applied to the economy, the universality of the distribution means that the wealth distribution in equilibrium is independent of the wealth of the country. In addition, one can show that people tend to migrate from poor countries to rich countries, and money tends to flow from rich countries to poor countries. Therefore, one might conclude that the tools of mechanical statistics can treat sociological and economic system in similar ways that are done in physics.

Contents

Introduction

What is the physical driving force of the changes in the world? What is the driving force of evolution, and what is the driving force of economy? The only physical driving force that generates irreversible changes in the universe is the propensity of entropy to increase or, as it generally called, the second law of thermodynamics.

In my work, I study the effect of the second law of thermodynamics on wealth distribution among people in equilibrium (maximum entropy) and the dynamics of the social and economic network.

In many textbooks, there is the misconception that entropy is a disorder, and therefore many conclude that eventually, the second law will destroy the world and the economy; i.e. the book "Entropy: A New worldview" by Jeremy Rifkin and Ted Howard. However, in my papers, I argue that with contradistinction to quantum sparse systems in which entropy is a disorder, in dense classic system entropy is information, and hence when it increases it generates order. Papers 2, 3, 4 and 6 are discussions on the properties of information and how and under what conditions thermal heat becomes information.

Furthermore, the probability of states distribution in sparse systems is normal (bell-like) and in dense systems, it becomes the "long tail" distribution. This distribution, called “Planck-Benford distribution”, is universal and independent of the energy of the system and has no free parameter. Pareto law, Zipf law, Benford law and many other distributions that are found in many sociological statistics are shown to be private cases of it. In papers 9, 10 and 1, I derive the Planck-Benford distribution and show how it fits the above-mentioned famous laws.

In paper 8, I discuss the error that leads to the conception that power law cannot be obtained by Gibbs- Boltzmann statistics.

1.
Entropy Principle in Direct Derivation of Benford's Law

Bedford's law is an empirical uneven distribution of digits that is found in many random numerical data. Numerical data of natural sources that are expected to be random exhibit an uneven distribution of the first order digits that fits to the equation,

$$\rho(n) = \log_{10}(1+\frac{1}{n}) \text{, where } n = 1,2,3,4,5,6,7,8,9 \quad (1)$$

Namely, digit 1 appears as the first digit at probability $\rho(1)$, which isabout 6.5times higher than the probability $\rho(9)$ of digit 9 (Fig 1).

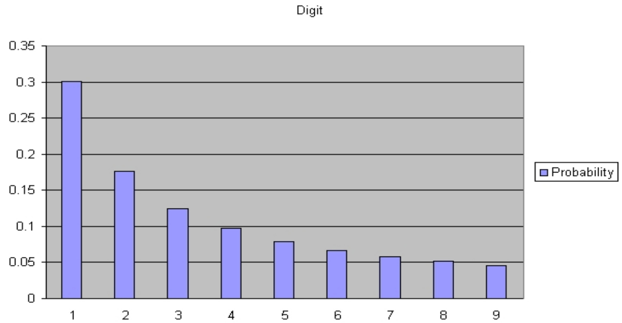


Figure 1. *Benford's law predicts a decreasing frequency of first digits, from 1 through 9.*

Eq. (1) was suggested by Newcomb in 1881 from observations of the physical tear and wear of books containing logarithmic tables (Newcomb, 1881). Benford further explored the phenomenain 1938, empirically checked it for a wide range of numerical data (Benford, 1938), and unsuccessfully attempted to present a formal proof. Since then Benford's law was found also in prime numbers (Cohen, & Talbot, 1984), physical constants, Fibonacci numbers and many more (Cohen, & Talbot, 1984; Zyga, 2007; Torres et. al., 2007).

Benford's law attracts a considerable attention (New York Times, 1998). Attempts for explanation are based on scale invariance (Cohen, 1976) and base invariance (Hill, 1995; Hill, 1996; Kossovsky, 2012) principles. However, there are no *a priori* well-defined probabilistic criteria when a data set should or should not obey the law (Zyga, 2007). Benford's distribution of digits is counterintuitive as one expects that a random numbers would result in uniformity of their digits distribution, namely, $\rho(n)=\frac{1}{9}$ as in the case of an unbiased lottery. This is the reason why Benford's law is used by income tax agencies of several nations and states for fraud detection of large companies and accounting businesses (Zyga, 2007; Nigrini, 1992; Nigrini, 1996). Usually, when a fraud is done, the digits are invoked in equal probabilities and the distribution of digits does not follow Eq. (1).

In this paper Benford's law is derived according to a standard probabilistic argumentation. It is assumed that, counter to common intuition (that digits are the logical units that comprise numbers) that the logical units are the 1's. For example, the digit 8 comprises of 8 units 1 etc. This model can be easily viewed as a model of balls and boxes, namely:

A) Digit n is equivalent to a "box" containing n none-interacting balls.

B) N sequence of such "boxes" is equivalent to a number or a numerical file.

C) All possible configurations of the boxes and balls, for a given number of balls, have equal probability.

The last assumption is the definition of equilibrium and randomness in statistical physics. In information theory it means that the file is in the Shannon limit (a compressed file).

A number is written as a combination of ordered digits assuming a given base B. When we have a number with N digits of base B, we can describe the number as a set of N boxes, each contains a number of balls n, when n can be any integer from 0 to $B-1$. We designate the total number of balls in a number as P.

An unbiased distribution of balls in boxes means an equal probability for any ball to be in any box. Hereafter, it is shown that this assumption is equivalent to assumption C and yields Benford law.

The "intuitive" distribution in which each box has an equal probability to have any digit n (n balls) does not means an equal probability for any single ball to be in any box, but an equal probability for any group of n balls (the digit n) to be in any box.

For example, for base B=4 there are four digits 0,1,2,3. The highest value of a 3 digits number in this base is |3|3|3|, which contains 9 balls. There is only one possible configuration to distribute the 9 balls in 3 boxes (because the limit of 3 balls per box). However, in the case of 3 balls in 3 boxes there are several possible configurations, namely: |3|0|0|, 0|3|0, 0|0|3, 2|1|0, 2|0|1, 1|2|0, 0|2|1, 1|0|2, 0|1|2, and 1|1|1. We see that digit 1 appears 9 times, digit 2 appears 6 times, and digit 3 appears 3 times. It is worth noting that the ratio of the digits 9:6:3, $\rho(1)=0.5$, $\rho(2)=0.33\bar{3}$, and $\rho(3)=0.16\bar{6}$, is independent of *N,* which is the reason why 0 is not included in Benford's law. As we see in the example above, each box has the same probability of having 1, 2 or 3 balls as the other boxes, however, the probability of a box of having 3 balls is smaller than the probability of a box to have 2 balls and the probability of a box of having one ball is the highest. The reason for this is that in order for a box to have several balls it has to score a ball several times, since the probability of a box to score *n* balls is smaller as *n* increases, lower value digits have higher probability. The formal calculation of the distribution of balls in the boxes was done by writing all possible ten configurations (in general $\frac{(N+P-1)!}{P!(N-1)!}$ configurations), and give each one of them an equal probability) and then counting the total number of each digit, regardless of its location.

The distribution of *P* balls in *N* boxes in equilibrium is a classic thermodynamic problem. Equilibrium is defined in statistical mechanics as a statistical ensemble in which all the possible configurations have an equal probability. The equilibrium distribution function $\rho(n)$ (the fraction of boxes having *n* balls) is calculated in a way that it yields maximum entropy, which means equal probability for all the configurations (microstates).

The standard way to find the distribution function in equilibrium is to maximize entropy S (S is proportional to the logarithm of the number of configurations Ω), under the constraint of a fixed number of balls *P.*

To do this, we apply the Stirling approximation; since $\Omega(N,P)=\frac{(N+P-1)!}{(N-1)!P!}$ we obtain,

$$S=\ln\Omega\cong N\{(1+n)\ln(1+n)-n\ln n\} \quad (2)$$

Where $n=\frac{P}{N}$. The number *n* is any integer (limited by *B*-1). If we designate the number of boxes having *n* balls by $\phi(n)$ then $P=\sum_{n=1}^{B-1}n\phi(n)$.

It should be noted that we count all the boxes in all the configurations excluding the empty boxes (the 0's) and the boxes that contain more balls then *B*-1.

To find $\phi(n)$ that maximizes S we will use the Lagrange multipliers method namely to define a function $f(n)=N\{(1+n)\ln(1+n)-n\ln n\}-\beta(P-\sum_{n=1}^{B-1}n\phi(n))$. The first term is Shannon entropy and the second term is the conservation of the balls. β is the Lagrange multiplier. To find $\phi(n)$ we substitute $\frac{\partial f(n)}{\partial n} = 0$, hence

$$\phi(n) = \frac{1}{\beta}\ln(\frac{n+1}{n}) \cdot \qquad (3)$$

We are interested in the normalized distribution, namely,

$$\rho(n) = \frac{\phi(n)}{\sum_{n=1}^{B-1}\phi(n)} \qquad (4)$$

Since $\sum_{n=1}^{B-1}\phi(n) = \frac{1}{\beta}\ln B$ it follows that

$$\rho(n) = \frac{\ln(1+\frac{1}{n})}{\ln B} = \log_B(1+\frac{1}{n}) \qquad (5)$$

namely, Benford's law.

In the normalization of $\phi(n)$ the quantity $\frac{N}{\beta}$ disappeared. That means that the distribution function is independent in the digit location in P and N and it is only a function of B. That is the reason why Benford law is so general.

Reconsidering the example of 3 balls in 3 boxes, we calculate from Eq,(5) that $\rho(1) = 0.5$; $\rho(2) \cong 0.29$ $\rho(3) \cong 0.21$. The total number of the none-zero digits (1, 2 and 3) is 18, and the distribution points to the ratio 9:5:4 as compared to the result of 9:6:3 that was obtained in the numerical example. The deviation from the theoretical calculation is explained by the fact that Sterling approximation yields a better fit as the number of digits grows.

Benford's law distribution was shown recently to be a special case of Planck distribution of photons at a given frequency (Kafri, 2016). It is intriguing that digits distribution of prime numbers also obeys the Planck statistics.

References

Benford, F. (1938). The law of anomalous numbers, *Proceedings American Mathematical Society,* 78(4), 551-572.

Cohen, D., & Talbot, K. (1984). Prime numbers and the first digit phenomenon, *Journal of Number Theory*, 18(3), 261-268. doi. 10.1016/0022-314X(84)90061-1

Cohen, D. (1976). An explanation of first digit phenomenon. *Journal of Combinatorial Theory, Series* A, 20(3), 367-370. doi. 10.1016/0097-3165(76)90032-7

Hill, T.P. (1995). Base-invariance implies Benford's law, *Proceedings of the American Mathematical Society,* 123(3), 887-895. doi. 10.1090/S0002-9939-1995-1233974-8

Hill, T.P. (1996). A statistical derivation of the significant-digit law, *Statistical Science*, 10(4), 354-363. doi. 10.1214/ss/1177009869

Kafri, O. (2016). Sociological inequality and the second law, *Journal of Economic and Social Thought*, 3(4), 476-480.

Kossovsky, A.E. (2012). Towards a better understanding of the leading digits phenomena. [Retrieved from].

Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers, *American Journal of Mathematics*, 4, 39-40. doi. 10.2307/2369148

New York Times, (1998, Aug. 4). Monday, November 24.

Nigrini, M. (1992). The detection of income evasion through an analysis of digital distributions, *Ph.D. thesis*, Dept. of Accounting, Univ. Cincinnati, Cincinnati OH. [Retrieved from].

Nigrini, M. (1996). A taxpayer compliance application of Benford's law, *Journal of American Taxation Association*, 18, 72-91.

Torres, J., Fernandez, S., Gamero, A., Sola, A. (2007). How do numbers begin? (The first digit law)." *European Journal of Physics*. 28(3), 17-25. doi. 10.1088/0143-0807/28/3/N04

Zyga, L. (2007). Numbers follow a surprising law of digits, and scientists can't explain why, [Retrieved from].

2.
The Second Law and Informatics

Introduction

Heat is the energy transferred from one body to another. The second law of thermodynamics gives us universal tools to determine the direction of the heat flow. A process is likely to happen if at its end the entropy increases. Similarly, energy distribution of particles will evolve to a distribution that maximizes the Boltzmann -*H* function (Huang, 1987) namely, an equilibrium state, where entropy and temperature are well defined.

Information technology (IT) is governed by energy flow. Processes like data transmission, registration and manipulation are all energy consuming. It is accepted that energy flow in computers and IT are subject to the same physical laws as in heat engines or chemical reactions. Nevertheless, no consistent thermodynamic theory for IT was proposed. Hereafter, a thermodynamic theory of communication is considered.

The discussion starts by drawing a thermodynamic analogy between a spontaneous heat flow from a hot body to a cold one and energy flow from a broadcasting antenna to receiving antennas. This analogy may look quiet natural. When a file is transmitted from a transmitter to the receivers, the transmitted file's energy, thermodynamically speaking, is heat. The Boltzmann entropy and the Shannon information have the same expression (Shannon, 1949), so we can think about information increase in broadcasting with analogy to entropy increase in heat flow. However, to complete the analogy, it is necessary to calculate a temperature to the broadcasting antenna and the receiving antenna.

To establish this analogy one has to calculate, for informatics systems, the thermodynamic quantities appearing in the second law, namely, entropy, heat, and temperature, and to define equilibrium. These informatics- thermodynamics quantities should comply with the Clausius inequality (Jaynes, 1988; Kestin, 1976) and to reproduce the Carnot efficiency.

The classical thermodynamics of heat transfer from a hot bath to a cold bath and the basic definitions of entropy, heat, temperature, equilibrium, and the Clausius inequality are a major part of the discussion of information thermodynamics. Therefore, in section 2 a brief review of these concepts is provided.

A file with a given bits distribution resembles a frozen two-level gas. The thermodynamics of a two-level gas in which the location of the exited atoms is constantly varying in time is well known. Deriving the thermodynamics of a file calls for comparison between the thermodynamics of a two-level gas and that of a binary file. In section 3, a calculation of the entropy, heat, temperature, and the definition of equilibrium for the transfer of a two-level gas from a hot bath to a cold bath according to statistical mechanics are provided and the compliance with the Clausius inequality is shown.

Following this review, in section 4, an analysis of the transfer of a binary file (a frozen two-level gas) from a broadcasting antenna (a hot bath) to receiving antennas (a cold bath) is provided. A temperature is calculated to the antenna that, together with the transmitted file information (entropy) and its energy (heat), is shown to be in accordance both with classic thermodynamics (i.e. the Clausius inequality) and information theory. The difference between randomness of a two-level gas, and randomness of a binary file and its effect on the thermodynamic quantities is discussed. It is concluded that the Shannon information is entropy.

Based on the results of section 4, in section 5 the second law of thermodynamics is defined for informatics. It is argued that reading/writing a file is equivalent to an isothermal compression/ expansion of an ideal gas and amplifying/attenuating a file is equivalent to an adiabatic compression/expansion of an ideal gas. An ideal amplifier cycle comprises of two adiabatic and two isotherms is shown to have the Carnot efficiency.

Finally in section 6, this theory is used to calculate a thermodynamic bound on the computing power of a physical device. This bound is found to be the Landauer's principle.

Classical thermodynamics of heat flow

In this section a short review of the quantities that will be used later for Informatics systems is provided (Jaynes, 1988). The second law of thermodynamics is a direct outcome of maximum amount of work ΔW that can be extracted from an amount of heat Q transferred from a hot bath at temperature T_H to a cold bath at temperature T_C (Jaynes, 1988). This amount of work can be calculated from the Carnot efficiency,

$$\eta \equiv W/Q\eta \leq 1 - T_C/T_H. \tag{1}$$

Namely, the maximum efficiency η of a Carnot machine depends only on the temperatures T_C and T_H. To obtain the maximum efficiency the machine has to work slowly and reach equilibrium at any time in a reversible way. Clausius (Kestin, 1976) defined the entropy, *S,* in equilibrium, such that it reproduces the Carnot efficiency, namely,

$$S \geq Q/T \tag{2}$$

If one dumps an amount of heat *Q*, to a thermal bath at temperature *T*, in a reversible way, thechange in the entropy of the bath is $S = Q/T$, and the system is in equilibrium. If one dumps the heat irreversibly the system is not in equilibrium and Q/T is smaller than ΔS as a result of efficiency lower then η. The entropy change S is equal to Q/T only in a reversible dumping in equilibrium. Therefore, if we assume that any system has a tendency to reach equilibrium, any system tends to increase Q/T. Taking a system out of equilibrium requires work, since the system will eventually reach equilibrium (namely, the energy of the work will be thermalized), therefore the entropy of a closed system tends to increase and cannot decrease. Temperature and entropy are defined in equilibrium and the temperature can be calculated as,

$$T = (Q/S)_{\text{equilibrium}} \tag{3}$$

Note that away from equilibrium entropy and temperature are not well defined (Jaynes, 1988).

Consider a simple example of the entropy increase in heat flow from a hot thermal bath to a cold one (see Fig 1).

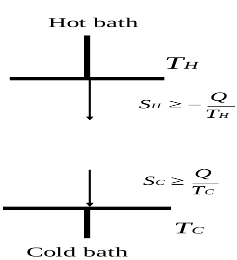


Figure 1: *The entropy increase in spontaneous energy flow from a hot thermal bath to a cold thermal bath.*

When we remove an amount of energy Q from the hot bath, the entropy reduction at the hot bath is Q/T_H .When we dump this energy to the cold bath, the entropy increases by Q/T_C .The total entropy increase is $S = Q/T_C - Q/T_H$. One can see that if the process is not in equilibrium $S > Q/T_C - Q/T_H$. In general

$S \geq Q/T_C - Q/T_H$ (4)

Hereafter, it is shown that inequality (4) is true both in statistical physics (sections 3) and in information theory (section 4).

Statistical Physics of a two-level gas

A binary file resembles a two-level gas. However, in a two-level gas particles exchange energy and in a binary file the energy distribution of the bits is fixed. Hereafter, a thermodynamic analysis of a two-level gas transmitted from a hot bath to a cold bath is reviewed. In section 4 the thermodynamic quantities that will be calculated for file transmission will be compared to those of a two-level gas and the origin of the differences is discussed.

Boltzmann has shown that the entropy of a system can be expressed as $S = -k\sum_{i=1}^{\Omega} p_i \ln p_i$ where i index the possible microscopic configuration of the system, p_i is the probability to be in the i^{th} configuration, Ω is the total number configurations and k is the Boltzmann constant. If all configurations are equally probable $p_i = 1/\Omega$, then $S = k\ln\Omega$, (Jaynes, 1988). This expression will be used to calculate the thermodynamic quantities appearing in the Clausius inequality for a system that resembles an informatics system. Consider a thermal bath at temperature T_H, which is in contact with a sequence of L states. n of the L states have energy ε and will be called "one".

$L - n$ of the states have no energy and will be called "zero". We analyze the thermodynamics of transferring this two-level gas from a hot bath at temperature T_H to a colder bath at temperature T_C with analogy to the heat flow analysis of section 2. To calculate the entropy we need to count the number of configurations of the two-level sequence, namely, the possible combinations of n, "one" particles in L states. As can be seen in Fig 2 this number is the n^{th} binomial coefficient. Namely, there are, $\Omega = L!/[\, n!\,(L-n)!]$ possible combinations.

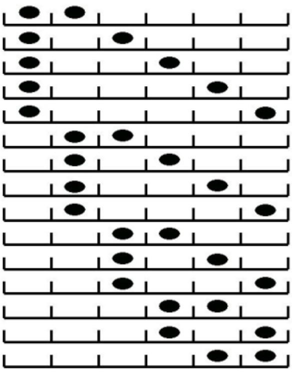


Figure 2: *A two-level gas with L=6 and n=2. In equilibrium all possible combinations have an equal probability. If some of the combinations have higher probability than others, the system is not in thermal equilibrium.*

The entropy of the system is $k\ln\Omega$, and the internal energy of the system is $U = n\varepsilon$. Since the gas is entirely removed from the thermal bath $Q=U$ and the temperature is given by Eq. (3). Using Stirlings formula we derive $\partial Q/\partial S$ to obtain T (Kompaneyets, 1961). The well-known result is,

$$\frac{n}{L-n} = e^{\frac{-\varepsilon}{kT}} \text{ or } T = \frac{\varepsilon}{k\ln\frac{L-n}{n}} \qquad (5)$$

Eq. (5) is the Maxwell Boltzmann distribution for a two-level gas (Gershenfeld, 2000). For a given ε, one parameter T represents all the knowledge on the two-level gas in equilibrium. This is a well-defined system with a well-defined entropy, temperature and energy. The equilibrium was invoked by giving an equal probability distribution to all the possible combinations Ω of the n particles in L states. If a system is not in equilibrium, there are certain combinations that are preferred and therefore the gas has a biased distribution. An unbiased distribution is the probability distribution, which describes the information we have about a system in the most honest way that allows us to make the best prediction about the property of a system. Jaynes (1957a; 1957b) has shown that unbiased distribution yields the Shannon information. In a biased distribution the actual combination span is smaller, and Ω of the gas is smaller. Boltzmann called the quantity $k\sum_{i=1}^{\Omega} p_i \ln p_i \geq -S$ calculated for a biased distribution the H function (Huang, 1987).

Hereafter, we calculate the entropy balance when a two-level gas is removed from a hot bath and is dumped into a cold bath for reversible and irreversible operation. It is shown that the process complies with the Clausius inequality

When the two-level gas is removed from the hot bath, the entropy is reduced by $S_H = n_H\varepsilon/T_H = kn_H \ln[(L-n_H) / n_H] \geq Q/ T_H$. When we dump it to the cold bath, we generate an entropy $S_C = n_H\varepsilon/T_C = kn_H \ln[(L-n_C) / n_C] \geq Q/ T_C$.

The total change in the entropy is,

$$S_c - S_H = \frac{kQ}{\varepsilon} \ln \frac{n_H(L-n_C)}{n_H(L-n_H)} = \frac{Q}{T_c} - \frac{Q}{T_H} \quad (6)$$

If T_C is lower than T_H, then $n_C > n_H$ and we see that Eq. (6) is positive with accordance with eq. (4), namely, the Clausius inequality (see Fig 3).

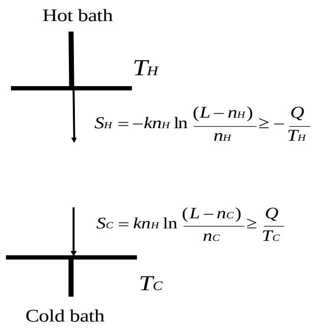


Figure 3: *The entropy increase, due to transmission of a two-level gas, from a hot bath to a cold bath is with accordance with the Clausius inequality.*

Thermodynamics of information

In the Shannon model a binary file is transferred from a transmitter to a receiver. A binary file can be viewed as a frozen two-level gas. A binary file is not in thermal equilibrium as only one possible combination of the bits is transmitted.

Shannon's first theorem deals with the maximum amount of information that can be coded in a given binary file of length L in a noiseless channel and in a noisy channel (Shannon second theorem). This amount of information is called the Shannon information and it was shown to have the same expression as the Boltzmann entropy, namely $I = -\sum_{i=1}^{L} p_i \log p_i$ (Shannon, 1949). Many papers were written on the connection between the Shannon information and the Boltzmann entropy (Jaynes, 1957; Brillouin, 1962; Rojdestvenski & Cottman, 2000; Kafri & Glatt, 1990). However in this paper a connection between Shannon information and the second law (i.e. the Clausius inequality) is discussed. The amount n of "one" bits, in a file of length L, has no unique relation to the amount of the Shannon information in the file. This is in contradistinction to a two-level gas in which the energy the temperature and the entropy are functions of n (see Fig. 3). For example, in a group of several files having the same n some may have a very small amount of Shannon information, i.e. when all the "one" bits are in the beginning of the file, and the rest of the file has zero bits or any other ordered combination (see Fig. 4) that can be regrouped effectively. Some other files may have a relatively high amount of Shannon information, if the distribution of the bits in the file is random as will be discussed later.

Figure 4: *Three possible binary files having the same energy. The higher two files have higher order and therefore contain little Shannon information. The lower file is random and is shown to be in equilibrium.*

The amount of the Shannon information in a file is a function of the randomness of the bits in it. The reason that Shannon obtained the same expression as Boltzmann is that, in a two-level gas in equilibrium we have no way to predict what combination of "one" particles will be at a certain time (see Fig.2), and in a random file we have no way to predict what bit will be at a certain time. The unpredictable sequence of bits is the useful information. When Alice is reading a binary file of length L she always obtains $L\ln 2$ useful nats (1bit = ln2 nat) even if the bits are ordered, because she lacks an *a-priori* knowledge of what bit will come next. However, Shannon first theorem is about Bob's ability to send a shorter file of length $L' \leq L$ from his knowledge of the sequence of the bits that he intends to send. The shortest file that can be recovered by Alice to reproduce the original file L is the amount of the Shannon information in the file. It is believed that the shortest file L' is a random file (namely there is a conjecture that a random file cannot be compressed).

Let's examine an antenna broadcasting a file composed of energy and includes information, which is received by several antennas. Consider the radiating antenna as a hot bath emitting energy and entropy. Similarly consider the receiving antennas as a cold bath that absorbs energy and entropy. It is argued that if we assume that information is entropy, the information balance obeys the Clausius inequality. To calculate the thermodynamic functions of an informatics system one need to calculate a temperature for the antenna. We assume that the antenna's temperature is identical to that of the file that it emits or absorbs. This is with analogy to the two-level gas transmission that was discussed in the previous section.

The calculations of the thermodynamic properties of a file and a two-level gas are different. In two-level gas Ω is the number of combinations of n particles in L states, in a file $\Omega = 2^L$. In a two-level gas there is a well-defined ratio between the entropy and the energy that enable to calculate T, for any n, because randomness means an equal-probability for any combination of n "one" particles in L states. For a file, randomness means an equal-probability for any bit; therefore, a random distribution means $n \approx L/2$. If n is not equal to $L/2$ there is no unique connection between the energy and the Shannon information and therefore a temperature cannot be calculated. For example, in Fig. 4 all the three files have the same amount of energy. The upper two have very little information as Bob can compress them significantly. The lower file is a random file and contains more information. For a random file the ratio between the energy and the information is unique as $n = L/2$ and $I = L\ln 2$, where I is the Shannon information. So by assigning energy ε to the "one" bit we obtain $Q = L\varepsilon/2$and $S = kI = kL\ln 2$. Using Eq. (3) we obtain for the temperature,

$$T=Q/S=\varepsilon/k2\ln 2. \tag{7}$$

In thermal systems, equilibrium is a state of randomness induced by collisions. Therefore, in analogy, it is assumed, that a random file is in equilibrium and has a well-defined temperature. This is in accordance with Clausius's result that temperature is defined only in equilibrium. The average energy per bit is $\varepsilon_n=\varepsilon/2\ln 2$, and therefore, Eq. (7) yields that for informatics $\varepsilon_n=kT$.

The derivation of the temperature in Eq. (7) is based on two major assumptions. The first one is that $S = kI$, namely, that the Shannon information is entropy as a consequence of the randomness of the bits in a file. The second assumption is that a random file is a state of equilibrium similarly tothermal systems in whichrandomness is a state of equilibrium. The obtained temperature $\varepsilon_n = kT$ is common in physics (i.e. harmonic oscillator). Nevertheless, it is necessary to show that these two assumptions encapsulated in the temperature of Eq. (7) comply with the Clausius inequality and the Carnot efficiency.

The broadcasting of a file to several antennas is equivalent to heat flow from a hot bath to a cold bath. Specifically a broadcasting antenna, which broadcasts a file having a high-energy bit, is a hot bath. A receiving antenna, which absorbs a lower energy bit file, is a cold bath. The entropy multiplies according to the number of the receivers. To calculate the entropy-information balance we consider an antenna broadcasting a binary file to N antennas. A possible realization of such system is a point-radiating antenna surrounded by a sphere, whose area is divided to N equal receivers. The hot antenna emits the broadcasted file at a temperature T_H. A receiver antenna receives the broadcasted file with a lower temperature $T_C=T_H/N$. Since $Q/T = kI$, we obtain from Eq. (4),

$$S \geq Q/T_C - Q/T_H = NkI - kI. \tag{8}$$

Eq. (8) shows that the file temperature, obtained in Eq. (7), yields correctly the increase in information in the broadcasting of a binary file to N receivers, which is $NI-I$. In fact the temperature is canceled out to give us simply the information balance from the thermodynamic quantities. In a "peer-to-peer" transmission, as in the Shannon model, no information increase is involved; therefore no thermodynamic considerations are necessary.

Now it is necessary to check if Eq. (8) behaves according to the Clausius inequality out of equilibrium. Namely, if broadcasting of a none-random file yields the inequality sign. When there are correlations between the bits, the amount of the Shannon information in the file is smaller. As a result, the same energy carries less Shannon information i.e. I is smaller than S/k. This shows that the second law of thermodynamics holds for informatics systems. Using Eq. (8) we can rewrite the Clausius inequality for informatics system as, $S \geq kI$.

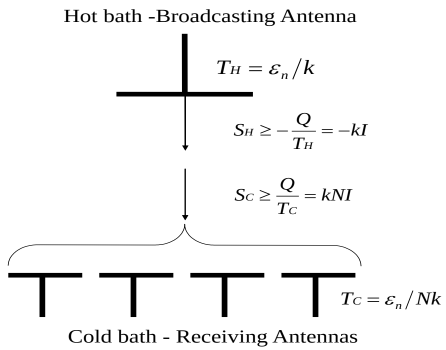


Figure 5. *The analogy between heat flow from a hot bath at temperature T_H to a cold bath at temperature T_C and an antenna broadcasting a file of bit energy ε to N antennas, each receiving the file with bit energy ε/N. In the thermal case the entropy increase is $S \geq Q/T_C - Q/T_H$. However, the same equation $S \geq Q/T_C - Q/T_H$ reproduces well the information balance when we use the temperature definition from Clausius inequality Q/S for a random binary file. The antennas deployment is drawn to emphasize the physics only.*

This implies that Information, like entropy, tends to increase. In a general case both informatics and thermal processes occurs simultaneously. In these cases a transformation of thermal entropy to informatics entropy and vice versa may occur. If we assume that entropy is an extensive quantity the Clausius inequality can be written as,

$$S \geq Q/T + kI \tag{9}$$

The first term on the RHS is the thermal entropy and represents the ensemble randomness due collisions. The second term is the informatics entropy and represents a quenched randomness of the nats in a sequence (a file). The amount of the Shannon information in a partially random file, with some correlation between bits, is equivalent to the Boltzmann $-H$ function, namely the "entropy" calculated out of equilibrium. Shannon, in his famous paper (Shannon, 1949), pointed out this analogy.

The second law for informatics-the Carnot cycle

In the previous section the energetic of a file broadcasted from one antenna to several antennas (a generalization of the Shannon theory) was studied. An analogy was drawn between information broadcasting from one antenna to several antennas to heat flow from a hot bath to a cold bath. What was shown is that:

1. The Shannon information content I of a file is equivalent to the Boltzmann -H function.
2. The transmitted file energy is equivalent to heat.
3. A random file is a state of equilibrium.
4. The temperature of the antenna is proportional to the average nat energy broadcasted from it or received by it.

These definitions comply with the Clausius inequality. We complete the analogy by demonstrating an informatics cycle, analogous to a Carnot machine.

The second law of thermodynamic is more renown in its verbal form, namely:

It is impossible to construct a cyclic machine whose net outcome is transferring heat from a cold bath to a hotter bath. Namely, work has to be invested, from outside of a system, to transfer heat from a cold bath to a hotter bath. This definition of the second law is a direct outcome of the Clausius inequality. If one transfers an amount of heat Q from a low temperature bath to a high temperature bath, S is negative with a violation of the second law. Machines with a negative entropy balance are called *perpetuum mobile* of the second kind.

The Clausius inequality was deduced from the efficiency of the Carnot machine. The Carnot machine comprises of a cylinder equipped with a piston filled with an ideal gas. The Carnot machine transfers energy from a hot bath to a cold bath and produces work. The piston is first in contact with a hot bath at a temperature T_H. At the first stage the piston expands slowly at a constant temperature (isothermal expansion). In this stage the piston removes energy and entropy from the hot bath into the gas. In the second stage the piston is isolated from the hot bath and expands until the gas is cooled to the temperature of the cold bath. During this expansion the piston produces work against an external pressure. Since the cylinder is isolated, no heat is exchanged with the gas, so that its entropy remains constant (this process is called an adiabatic expansion). The third step is an isothermal compression; the gas in the cylinder dumps heat and entropy to the cold bath. The cycle is completed by an additional adiabatic compression of the gas to the temperature of the hot bath by applying work. The total work and energy balance yields the Carnot efficiency.

It is now shown that it is possible to construct an informatics Carnot cycle consists of two isotherms and two adiabatic. Reading a file is an analog of an isothermal expansion. When a file is received at constant bit energy, the energy of the receiver increases but its temperature remains fixed, exactly as in an isothermal compression. When a file is amplified, its temperature is increased but its information content remains fixed. This is an adiabatic compression.

An analog of a Carnot cycle is found in a transmitter of a file over an optical fiber. The file is amplified periodically at a given distances due to the signal attenuation caused by energy loss in the fiber. Carnot was interested in extracting mechanical work from a temperature gradient to convey physical goods over the friction of the railroad. In an optical fiber transmission, one wants to invest electrical work to covey information over the intrinsic loses of light on its path.

The Carnot cycle for a file transmission over a fiber comprises of four steps identical to those of the original mechanical Carnot heat engine, see Fig.6.

1. A file is sent at high bit energy (high temperature) into the fiber. This is a writing process that is equivalent, as discussed above, to an isothermal expansion.
2. The file is transmitted through the fiber and during the transmission the file is attenuated. Its bit energy is reduced and the file is cooled. The information remains fixed and therefore this process is an adiabatic expansion. (In this example, the work is lost, outside the informatics system).
3. An amplifier reads the file at a low temperature. During this process the energy of the amplifier is increased but its temperature remains constant. This is an isothermal compression.
4. The file energy is amplified at fixed information content. This is an adiabatic compression. The file is ready for a new cycle.

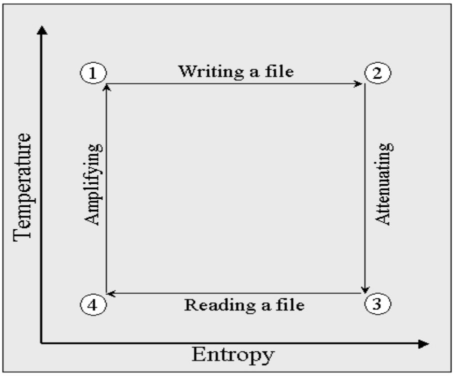


Figure 6. *The Carnot cycle of a file transmission along an optical fiber. Amplifiers are necessary to overcome the energy loss in the fiber. Each cycle of amplification is shown to be a Carnot cycle of two isothermals and two adiabatics.*

With analogy to the original Clausius formulation it is possible to define the second law of thermodynamic for informatics:

It is impossible to contract a cyclic amplifier whose net outcome is transferring heat from a low bit energy file to high bit energy file. Namely any amplifier requires outside work (i.e. a power source). If an amplifier receives a low temperature T_C file and increases its temperature to T_H, at the end there is a negative entropy balance, $S = Q/T_H - Q/T_C < 0$, because higher the energy of a bit means less bits per Q. Therefore, it requires adding to Q an extra energy in order to avoid negative entropy. To conserve the entropy one needs that

$Q_H/T_H = Q_C/T_C$. Namely, the information of the hot file is equal to that of the cold file. Designating $Q_C=Q$ and $Q_H= Q+W$, where W is the added work requires to avoid a negative entropy, we obtain that $W=Q\,(1 - T_C/T_H)$. Namely, the Carnot efficiency, of Eq. (1).

This formulation is applicable as well to optics. Every picture is comprised of combination of spatial modes (Kafri & Glatt, 1990) i.e. pixels. These spatial modes are independent light sources. If one detects an image, for a given time period, it is possible to assign energy to the pixels of the image and thus to calculate a temperature in addition to the Shannon information content. Therefore we can generalize the 2nd law for optics;

It is impossible to construct a passive imaging optical device that will produce an image with energy flux higher than that of the original image. The above phrases are by no means surprising or novel. However, it is shown that energy flow in computers and other informatics systems obeys the same physical laws as energy flow in steam engines and chemical reactions.

The Computing power of a physical device - The Landauer's Principle

Thermodynamic considerations can be used to calculate the maximum speed of a processor from the power P applied on it and its ambient temperature. In Turing model (Turing, 1936) erasing one bit and registering it again is an example of a logical operation. Therefore, the bits rate f of a file can be considered as the computing power of a physical device.

One can write the temperature of an emitter or a receiver as;

$$T = P/(k\,f\ln 2). \tag{10}$$

Every physical system is surrounded by a thermal bath that emits thermal noise at a temperature T_n. The higher the bit rate, the lower the temperature of the file as the bit energy is reduced. Since the temperature of the file must be kept above the temperature of the noise T_n, namely $T > T_n$, the frequency has an upper limit. From Eq. (10) we conclude that $f < P/(kT\ln 2)$ where T should be about 10 times higher than the noise temperature. Therefore, the upper bound on computing power of any device is,

$$f \leq P/(10\,kT_n\ln 2). \tag{11}$$

Namely the power applied on any computing device and its ambient temperature suffices to calculate a limit on its computing power. Von Neumann claimed that a computer operating at temperature T must dissipate at least $kT\ln 2$ energy per elementary act of information. In nature the ratio per nucleotide or amino acid is 20-100 $kT\ln 2$ (Bennet, 2003). The minimum energy dissipation per logical operation as suggested by Von Neumann is known as Landauer's principle (Bennett, & Landauer, 1985). It is seen that it obtains naturally from the second law of thermodynamics.

Summary

When a random binary file is removed from an emitter or absorbed by a receiver, its energy may be considered as heat and its Shannon information as entropy. The average nat energy of the file is kT, where T is shown to be the informatics temperature of the emitter or the receiver. If the binary file is not random, Shannon information is Boltzmann $-H$ function. This approach is shown to comply with the second law of thermodynamics, reproduces the Carnot efficiency and the Landauer's principle.

References

Bennet, C.H. (1982). Notes on Landauer's principle, reversible computation and Maxwell's demon, Int. J Theor. Phys. 21, 905. doi. 10.1016/S1355-2198(03)00039-X

Bennett, C.H. & Landauer, R. (1985). The fundamental physical limits of computation, *Scientific American*, 253(1), 48-56.

Brillouin, L. (1962). *Science and Information Theory*, Academic Press: New York.

Gershenfeld, N. (2000). *The Physics of Information Technology*, Cambridge University Press.

Huang, K. (1987). *Statistical Mechanics*, John Wiley: New York.

Jaynes, E.T. (1957a). Information theory and statistical mechanics I, *Physical Review*, 106(4), 620-630. doi. 10.1103/PhysRev.106.620

Jaynes, E.T. (1957b) Information theory and statistical mechanics II, *Physical Review*, 108(2), 171-190. doi. 10.1103/PhysRev.108.171

Jaynes, E.T. (1988). *The Evolution of Carnot's Principle*" in *Maximum-Entropy and Bayesian Methods in Science and Engineering,* (pp267), G.J. Erickson & C.R. Smith (Eds.), Kluwer, Dordrecht.

Kafri, O., & Glatt, I. (1990). *The Physics of Moiré Metrology*, Wiley & Sons Inc.

Kestin, J. (1976). *The second Law of Thermodynamics*, Dowden, Hutchinson & RossStroudsburg.

Kompaneyets, A.S. (1961). *Theoretical Physics*, Peace Publishers Moscow.

Rojdestvenski, I., & Cottman, M.C. (2000). *Mapping of Statistical Physics into Information Theory with Application to Biological Systems*, Journal of Theoretical Biology, 202(1), 43-54. doi. 10.1006/jtbi.1999.1042

Shannon, C.E. (1949). *A Mathematical Theory of Communication*, University of Illinois Press, Evanston, Ill.

Turing, A.M. (1937). On Computable Numbers with an Application to the Entscheidungsproblem, *Proceedings of the London Mathematical Society*, 42(1), 230-265. doi. 10.1112/plms/s2-42.1.230

3.
Informatics Carnot Machine in an Optical Fiber

Introduction

A sequence of light pulses transmitted through an optical fiber is widely used in communication (Palais, 1988). A random sequence of identical light pulses, representing "1" and vacancies, representing "0", is the physical entity of a transmitted binary file. The length L of the pulse's sequence (the number of the pulses and vacancies) and the randomness of the distribution of the pulses determine the amount of the Shannon information being transmitted (Shannon, 1949). In this communication a thermodynamic analysis of the transmission of a random sequence of a light pulses (a file) is considered. It shown that the Shannon information is entropy, and the amplification process done in an optical fiber is a Carnot cycle having the Carnot efficiency.

To calculate the entropy of the sequence of pulses, it is first necessary to calculate the entropy of a single, single-mode coherent pulse. It is assumed that the i^{th} pulse in a sequence has n_i photons, of energy $h\nu$, in a single mode. Since the photons are indistinguishable, the pulse is coherent. The temperature of the pulse will be assumed to be equal to that of a blackbody that emits n_i photons into a single-mode of frequency ν (Gershenfeld, 2000). Since a blackbody is in equilibrium with its radiation, a temperature can calculated. (Appropriate spatial and spectral filters may filter the other radiation modes). In this case,

$$n_i = \frac{1}{e^{h\nu / k_B T_i} - 1} \qquad (1)$$

The temperature of the coherent pulse obtained from eq. (1) to be

$$T_i = \frac{h\nu}{k_B \ln(1+\frac{1}{n_i})}, \quad (2)$$

The total energy of the pulse q_i is $n_i h\nu$. Therefore the entropy that that a single pulse carries away from the blackbody is $S_i = q_i / T_i$, or:

$$S_i = n_i k_B \ln(1+\frac{1}{n_i}) \quad (3)$$

Since $\lim_{n_i \to \infty} S_i = k_B$ it means that the entropy of a coherent pulse is identical to that of a classic harmonic oscillator (namely $q_i = k_B T_i$) and is not a function of its energy. Similarly, a mode without energy carries no entropy as $\lim_{n_i \to 0} S_i = 0$.

The total entropy of a sequence of pulses, some with a large number of photons, (having entropy k_B), and some with no photons (empty pulses having no entropy), is the Gibbs mixing entropy of the sequence, namely, $S = -k_B \sum_{j=1}^{\Omega} p_j \ln p_j$. Where Ω is the number of configurations of the pulses and p_j is the probability of the j^{th} configuration. It is seen that the Shannon information and entropy are equivalent in the classical limit (a large n_i) to the Gibbs entropy of mixing.

For example, to calculate the entropy of a random sequence of light pulses, of length L, it is necessary to consider the fact that each pulse has a probability of ½ to be "one" and probability of ½ to be "zero". Therefore, the mixing entropy term is $S_i = -k_B(\frac{1}{2}\ln\frac{1}{2} + \frac{1}{2}\ln\frac{1}{2}) = k_B \ln 2$. To find the total entropy of the sequence we can sum all the entropies of the pulses (because the entropy is extensive), namely, $S = \sum_{i=1}^{L} S_i = k_B L \ln 2$. The Shannon information I is defined as $I = -\sum_{j=1}^{\Omega} p_j \ln p_j$ (Shannon, 1949). If the probability of all the configurations is equal to $1/\Omega$ than $I = \log\Omega$. The number of the configurations of a binary file of a length L is 2^L, therefore the maximum amount of the information in the bits of a file having L pulses is $L\ln 2$ nats (1bit = ln2 nat). It is seen that thermodynamics and information theory yield the same result.

When the sequence of the bits is not random, the amount of information of the sequence is smaller. Therefore, in general, we obtain the Clausius inequality,

$$S \geq k_B I. \quad (4)$$

One can generalize this analysis and calculate the energy and the temperature of the whole sequence of pulses. This can be done easily for a random sequence. When n_i is large, the temperature T_i of a coherent pulse is $n_i h\nu/k_B = q_i/k_B$ where q_i is the energy of the pulse. Namely, with contradistinction to the entropy, T_i is a function of q_i. If we assume that all the energetic pulses have an equal energy q, the total energy of the sequence is $Q = \sum_{i=1}^{L} q_i = \frac{q}{2}L$. The entropy of the sequence is $S = k_B L \ln 2$ in nats or $k_B L$ in bits. The file temperature T is $Q/S = q/2k_B$. This means that the average bit's energy $q/2$ is equal to $k_B T$. This is the same relation as of a harmonic oscillator. It is worth noting that a random sequence of pulses is a non-coherent radiation, nevertheless it retains the thermodynamic properties of a harmonic oscillator.

Now it shown that this formalism complies with the second law of thermodynamics (Jaynes, 1988). Consider a long optical fiber in which a file comprising a sequence of light pulses, having a temperature T_H, with a pulse energy q_H, travels along the fiber. The pulse energy attenuated due to the loss in the fiber. Therefore, the energy and the temperature of the pulse reduced to T_C. Nevertheless, the amount of information (the entropy) remains intact. This process of cooling at constant entropy, thermodynamically speaking, is an adiabatic expansion. When the pulse energy reduces, the file requires amplification. To amplify the sequence of the pulses, the amplifier has to read the file first. The reading process is an energy transfer to the amplifier at constant bit energy. In this process, the amplifier increases its energy at a constant temperature. Thermodynamically speaking, this process is an isothermal compression. In the next stage the file is amplified back to T_H . This stage is an adiabatic compression in which we invest work to increase the energy of the pulses without increasing their information (entropy). Finally, at the last stage the amplifier writes (emits) the light pulses into the fiber. In this stage, the amplifier reduces its energy at a constant temperature T_H and the cycle starts again.

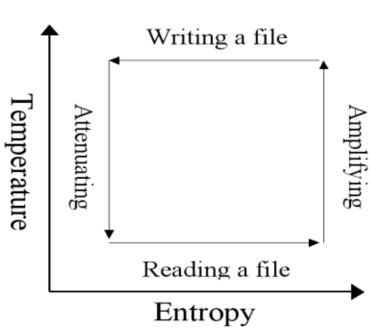


Figure 1. A *Carnot cycle for a file amplification. Amplifiers are necessary to overcome the energy loss along a fiber. Each cycle of amplification is a Carnot cycle of two isothermals and two adiabatic.*

Hereafter it shown that this cycle has an efficiency of the Carnot machine. Before entering the amplifier, the sequence of the light pulses has a relatively low temperature T_C, and its entropy is Q/T_C. After the amplification, it has a higher temperature T_H. If Q is unchanged, the entropy is smaller. Therefore, the entropy balance $\Delta S = Q/T_H - Q/T_C < 0$ is negative, which is a violation of the 2^{nd} law. The physical reason for the entropy reduction is that with a given amount of energy Q, one can write more low-energy bits than one can write high-energy bits. To conserve the entropy (a reversible operation), we have to keep $\Delta S = 0$. That means that we have to add more energy to Q. For a reversible operation $Q_H/T_H = Q_C/T_C$. Designating $Q_C = Q$ and $Q_H = Q+W$, we obtain,
$W = Q(1- T_C/T_H)$. In the irreversible case $Q_H/T_H > Q_C/T_C$, thus in general we obtain;

$$\eta \equiv \frac{W}{Q} \leq (1 - \frac{T_C}{T_H}) \qquad (5)$$

Eq. (5) is the Carnot efficiency.

Summary and discussion: The amount of entropy removed from a blackbody by a single radiation mode in the classical limit is k_B. If the radiation mode is empty, it does not remove entropy. The entropy removed from the blackbody is assumed to be equal to the entropy of the pulses. Therefore, it is argued that in the classical limit, an energetic pulse carries k_B entropy and a vacancy carries a zero entropy. When this assumption is applied to calculate the entropy of a sequence of pulses, the obtained entropy of the sequence is the Gibbs mixing entropy, which is identical to the logical Shannon information. The plausibility of this formalism demonstrated by presenting an informatics Carnot Cycle that yields the Carnot efficiency for an ideal amplifier cycle in an optical fiber.

Temperature and thermal equilibrium are concepts that used to describe random systems in equilibrium. In random systems, energy exchanged between particles by collisions. There is no energy exchange between photons. Nevertheless, the quenched randomness of the energetic bits and the zero bits behaves according to the present formalism as in equilibrium, namely, a state where it is possible to calculate a unique temperature.

It was shown previously that laser operation (Geusic, 1967; Levine & Kafri, 1974) and laser-cooling processes (Kafri & Levine, 1974) which involve a production or a usage of a coherent light, yield the Carnot efficiency, and therefore comply with the second law of thermodynamics. In these processes, the light was considered as work, as light radiation was assumed to be coherent. A coherent light beam has a single radiation mode (Kafri, & Glatt, 1990) and therefore it carries negligible amount of entropy. In the present study $L/2$ pluses, distributed randomly in L modes, carry entropy that is shown to be the Shannon information. The pulse sequence is not coherent, as it is random. The lower the coherence, the higher is the amount of information that can carried by the sequence. This communication suggests that the Shannon information can affect the efficiency of a Carnot machine.

Only when n_i is large the entropy is not a function of the energy and the temperature. In this limit, the entropy becomes a pure measure of the quenched randomness, exactly as the logical Shannon information. This is a vital condition in informatics, as the entropy should remain intact with the energy attenuation. When n_i is small, the entropy $S=S(Q)$ is smaller than that of the logical information. The entropy deficiency $k_B I - S(Q)$ is a loss of the logical information.

The Carnot efficiency of an amplifier can be tested experimentally. Calorimetric experiments of this kind require careful photon counting; nevertheless they are possible in the contemporary technology. This study suggests that the second law is applicable in the classical limit to informatics.

References

Gershenfeld, N. (2000). *The Physics of Information Technology,* Cambridge University Press.

Geusic, J.E., Schultz du Bois, E.O., & Scovil, H.E.D. (1967). Quantum Equivalent of the Carnot Cycle, Physical Review Letters, 156(2), 343. doi. 10.1103/PhysRev.156.343

Jaynes, E.T. (1988). The evolution of Carnot's principle, in *Maximum-Entropy and Bayesian Methods in Science and Engineering,* 1, G.J. Erickson & C.R. Smith (Eds.), (pp.267-281), Springer. doi. 10.1007/978-94-009-3049-0_15

Kafri, O., & Glatt I. (1990). *The Physics of Moiré Metrology,* J. Wiley & Sons Inc.

Kafri, O., &. Levine, R.D. (1974). Thermodynamics of adiabatic laser processes: Optical heaters and refrigerators, *Optics Communications*, 12(2), 118-122. doi. 10.1016/0030-4018(74)90371-X

Levine, R.D., & Kafri O., (1974). Thermodynamic analysis of chemical laser systems, *Chemical Physics Letters*, 27(2), 175-179. doi. 10.1016/0009-2614(74)90197-3

Palais, J.C. (1998). *Fiber Optics Communications,* McGraw-Hill Professional.

Shannon, C.E. (1949). *A Mathematical Theory of Communication*, University of Illinois Press, Evanston, Ill.

4. The Second Law as a Cause of the Evolution

Introduction

Until the late 17th century, the common hypothesis about the origin of life was Abiogenesis or a spontaneous generation of life from organic substances. For example, if we put an orange on a table, after a while we will probably find worms in it. The conclusion that the worms generated from the orange is only part of the truth. We know today that eggs of a fly called *Drosophila melanogaster* have to be present in the orange, in the first place in order for the worms to generate. Nevertheless, the "explanation" that the eggs are the reason for the generation of the worms in the orange does not diminish the appeal of the Abiogenesis hypothesis. For an observer in space looking at the earth over billions of years, the formation of life, buildings, roads, cars etc, cannot explained in a similar way as the formation of worms in the orange. For him, the explanation for life and even buildings, books etc. is a spontaneous generation. What are the "eggs" of these objects?

Contemporary science deals with evolution, namely how life evolved from the simple to the complex. Nevertheless, evolution theories do not deal and do not explain the reason for spontaneous generation of complex objects as life and artificial objects (Simeonov, 2010).

Sometimes life regarded as a thriving for order. It seems we are constantly fighting against the chaos invading our life, constantly looking for rules and laws, and if we do not find them, we invent them. However, there is no scientific definition of order. It seems odd that order not defined in science, while entropy, the entity that conceived as a measure of disorder, is so widely used. According to the second law of thermodynamics, any process that causes the entropy to increase is likely to happen. In other words, such processes are spontaneous. This is the reason for the bad reputation of the second law "which fights our tendency to order". The Internet is loaded with theological material claiming that life is a violation of the second law of thermodynamics.

In this paper, it argued exactly the opposite: The second law is responsible for life because life related to a spontaneous information increase, and information is a part of entropy. It argued that entropy comprises of two parts: the thermal entropy and the informatics entropy. While the thermal entropy and its connection to the 2nd law are well known, the connection between information and the 2nd law was only recently discussed (Kafri, & Kafri, 2013; Kafri, 2007a; 2007b).

Information conceived (erroneously) as order. It suggested that information contains null or even negative entropy (Brillouin, 1983). This "intuitive reasoning" is logically sound, as many understand information as the requirement that Alice will obtain the same value for the i^{th} bit each time Bob sends her an identical file. Since the bits location is fixed, the file is a frozen entity and thus contains zero entropy. Even more popular intuition is that information is negative entropy as was suggested by Brillouin (1962 and Woolhouse, 1967). The reasoning of this conclusion is that the file, before it received, is unknown, and thus contains entropy, which reduced with every bit that Alice reads. Therefore, information is a reduction of entropy.

Nevertheless, Shannon, in his first theorem, defined the information as the maximum amount of data that can transmitted in a noiseless channel. His expression, which is identical to the Gibbs entropy, represents the randomness of the distribution of the bits in a file (Shannon, 1949). The Shannon information tells us nothing about the actual content of a file and has no connection to it. When Alice receives a file of Λ bits, all we know is that there are maximum 2^{Λ} different possible (configurations) contents. In some of these 2^{Λ} files, the distribution of the bits correlated. In others, the distribution of the bits is random. When Alice receives a Λ bits random file, the amount of the Shannon information is $I = \Lambda$ bits. If the distribution of the bits is not random Bob can compress the file and send a shorter file of length I such that $I < \Lambda$ or in general $I \leq \Lambda$.

How many random distributions are there as compared with correlated distributions in a file? Jaynes has shown (Huang, 1987; Gupta, et al. 2005; Newman, 2006) that if we have a statistical ensemble, the most honest guess about its distribution is the Shannon information. The finding of Jaynes, in simple words, is that there are much more random distributions than there are correlated distributions. The work of Jaynes suggests that for a statistical ensemble the second law is a mere probabilistic effect. If we do not know anything about a statistical ensemble, our best bet is that it is random. If we reshuffle a distribution of something, it will become more random.

Jaynes work is applicable to information. If we add a noise to a file, it will probably increase the amount of the Shannon information in the file. Nevertheless, the subjective meaning of the message in the file may be lost.

While the similarity between the Gibbs entropy and the Shannon information is clear, there is a distinctive difference between them. Information is a logical quantity; the Shannon information is a mathematical entity. It is neither a function of the file energy nor a function of the temperature, as it is not made of a materialistic substance. A binary information file contains only "1"'s and "0"s. Entropy on the other hand, is a physical quantity and has a physical dimension. The physical meaning of the entropy derived from the second law. The basic outcome of the second law is that heat flows spontaneously from a hot bath to a cold bath; it does so, because in this process the entropy increases.

Another difference is that entropy is a dynamic quantity while information is a quenched quantity. Boltzmann obtained his approximation of the Clausius entropy for an ideal gas, which contains a large number of atoms exchanging energy constantly (Jaynes, 1988). The Boltzmann statistics contains inherently the canonic non-deterministic Maxwell-Boltzmann distribution. This distribution is a cornerstone of mechanical statistics as well as of quantum mechanics. Nevertheless, the canonic distribution is not applicable to information, which is a quenched quantity.

The paper in a nutshell

In other publications (Kafri, & Kafri, 2013; Kafri, 2007a; 2007b) it was shown that if we assign energy to the information bits, it is obtained, from the 2^{nd} law of thermodynamics, that the Shannon information is entropy, a random file is a state of equilibrium and the temperature is proportional to the bit's energy. In this paper, a toy model used to describe a generic file transmission from Bob to Alice using electromagnetic radiation. Bob is using a blackbody-based transmitter that enables him to control the temperature and the frequency of the radiation. In addition, Bob can modulate the radiation. It shown that when Bob transmits to Alice a low occupation number energy (the quantum limit), the thermodynamic functions of the energy transmission are the well-known canonic ones. However, when Bob increases the occupation number of the photons, a power-law distribution replaces the canonic distribution, information replaces the entropy, and the canonic statistical physics becomes a statistics of harmonic oscillators. In the high occupation limit, the obtained normalized thermodynamic functions are independent of any physical quantity and/or physical constant and therefore become purely logical functions. A qualitative discussion about whether nature prefers generation of information or generation of canonic entropy yields a conclusion that the two are equally welcome.

In section II, *a toy model,* in which Bob sends a collimated light beam to Alice, is described. Bob is using a blackbody radiation source that delivers energy per mode according to the Planck's statistics. Bob can change to his wish (without any physical limitations) the temperature of the source and select a frequency or frequencies of the radiation by a spectral filter. In addition, Bob has a shutter that enables him to modulate the radiation within the limitations of the laws of optics. It assumed that in equilibrium all the radiation modes that Alice receives have the same temperature.

1. In section III, *the quantum limit*, it assumed that the energy of the photons is much higher than the average energy. The Planck equation yields the familiar canonic distribution. The obtained entropy of the radiation is the Gibbs expression and the ratio between the number of the photons and the number of empty modes is the Maxwell-Boltzmann distribution.

2. In section IV, *the high occupation limit*, Bob is using his toy model to reduce the energy of the photon as compared with that of the average energy of the mode to the extent that the energy can added or removed smoothly to the mode. In this limit, the number of the photons is much larger than the number of the modes. The Bose-Einstein equation yields that each mode is a harmonic oscillator. It means that each mode's entropy is one Boltzmann constant, and the temperature is a linear function of the mode's energy.

3. In section IV-a, *modulation and information*, Bob is modulating the sequence of the harmonic oscillators to a binary file. The Shannon information is calculated. It shown that in a random file, when the number of the harmonic oscillators (energetic modes) is equal to the number of the vacancies (empty modes), the Shannon information is equal to the length of the file. In other cases, it shown that the amount of the Shannon information is smaller.

4. In section IV-b, *equilibrium and entropy*, the entropy of the file, which consists of harmonics oscillators and vacancies, is calculated. It shown that the Shannon information is the Gibbs mixing entropy. The Boltzmann H function is equivalent to the amount of information H in a correlated file.

In section IV-c, *logical quantities*, it shown that the normalized entropy in Bob's transmission is a function that does not contain any physical variable or constant. This is with contradistinction to a canonic entropy transmission. Therefore, the normalized entropy, in the high occupation limit, is a logical quantity.

In section IV-d, *logical equilibrium –the Benford's law*, Bob generates a set of modes in which each represents a digit. A possible way to construct such a set is to put in the mode that represents the digit N, N times more energy than the mode that represents the digit 1. To obtain equilibrium (equal temperature in the Planck distribution) Bob has to use either a different frequency for each digit-mode or alternatively a different density for each digit in a file. The obtained normalized distribution function of the digit-modes in equilibrium is a pure logical function, which is not a function of the initial temperature and/or frequency chosen by Bob. The result is identical to the famous Benford's law.

In section IV-e, *energy distribution, the power-law*, the log of the occupation number vs. the log of the photon energy divided by the average energy plotted for the Planck distribution. It is seen that in the low occupation number, a straight line of slop, –1, obtained. This behavior appears in many quantities in nature (a good example are the natural nets). With analogy to the momentum Gaussian distribution obtained from the canonic energy distribution if we consider the electric field of the radiation instead of the energy, a –2 slope obtained. Slopes around "-2" appear in many sociological statistics (Newman, 2006).

In section V, *mixed systems*, the question whether information can survive side by side with the canonic thermal entropy in equilibrium is discussed.

In section V-a, *Hooke's law harmonic oscillator*, a thermodynamic analysis of a mixed system consists of a single Hooke-law oscillator coupled to a heat bath at room temperature is presented. It is shown that the amplitude increase of a Hooke-law oscillator has the Carnot efficiency, similar to that of the amplification of a file (Kafri, 2007a; 2007b). A Hooke-law oscillator will relax its energy spontaneously to the heat bath as the relaxation increases the entropy.

In section V-b, *information vs. thermal entropy*, it shown that in the case of the blackbody emission, both the low occupation number photons as well as the high occupation number photons coexist in equilibrium. Therefore, it concluded that information and thermal entropy are equally welcome.

In section VI, *summary and discussion,* a table that shows the differences between the thermodynamic functions in the canonic distribution and in the high- occupation harmonic distribution presented. In view of these differences, the meaning of the logical quantities obtained in the thermodynamic theory of communication discussed. It concluded that, in equilibrium, inert quanta distributed in modes yield a power-law/Bedford-law distribution. It suggested that the informatics aspect of life is a tendency for reproduction and a compressed communication.

The Toy Model

In Fig. 1, the setup in which Bob sends Alice a flux of photons described. The analysis based on the classical Carnot Clausius thermodynamic. In the classical thermodynamic, the entropy is $S \geq Q/T$ where the equality sign stands for equilibrium, Q is the heat that Bob is sending or Alice is receiving, and T is the temperature of the transmitter or the receiver. This inequality is the Clausius inequality (Kestin, 1976) derived directly from the efficiency of the Carnot cycle (Jaynes, 1988).

Bob is sending a sequence of photons in a single longitudinal mode to Alice, as described in Fig.1. Bob has a blackbody at temperature T_H that emits a blackbody radiation. Bob attaches a pinhole filter (PH) of a diameter of λ^2 with a positive lens in order to obtain a collimated single longitudinal mode. After the pinhole, spatial filter Bob attaches a spectral filer and a polarizer (SF) that passes only the frequency ν, with a spectral width $\Delta\nu$. Here λ and ν are the wavelength and the frequency of the transmitted signal. The spectral width determines the number of the temporal modes. Bob can modulate the photons beam by using a mechanical or electro-optical shatter. Photons are bosons with a zero chemical potential. Therefore the number of photons n in a single mode obtained from the Planck distribution (Gershenfeld, 2000) is given by

$$n_i = \frac{1}{e^{h\nu/k_B T_H} - 1} \quad (1)$$

Where i index the temporal modes, $h\nu$ is the energy of the photon, T_H is the temperature of the Bob's source and k_B is the Boltzmann constant.

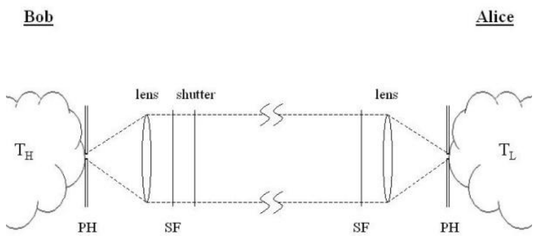


Figure 1. *A setup for a single transverse mode energy transmission from Bob to Alice*

Alice uses a detector to receive the message. In general, Alice uses a similar positive lens, a filter and a detector at the focal length of the lens. If the detector of Alice is at a temperature $T_L = T_H$, namely the temperature of the transmitter of Bob, the noise emitted by the detector will be as strong as the signal, and Alice will not be able to read the signals. Therefore, a prerequisite requirement for energy transmission from Bob to Alice is that $T_H > T_L$. In practice, Bob can heat his blackbody to a temperature limited by the physical properties of the blackbody's material. However, we assume that Bob does not have such limitations and he can produce a beam as hot as a laser beam to his wish. In addition, Bob can vary the frequency that he send. In practice, the wavelength of the radiation cannot exceed the diameter of the blackbody; nevertheless, we let Bob enjoy the benefit of a toy model. Hereafter, two limits discussed the quantum limit in which $h\nu \gg k_B T$ and the high occupation limit in which $h\nu \ll k_B T$.

The Quantum Limit

In the quantum limit, the energy of the photon $h\nu$ is much higher than the average energy, $k_B T$, of a mode emitted from a thermal bath (the blackbody). Therefore, classically it is impossible to emit a photon. However, when many modes are collecting their energies together they emit a single high-energy photon in an arbitrary (lucky) mode. In other words, when $n \ll 1$, it assumed that a group of $1/n$ modes will emit a single photon in an unknown mode of the group. Occupation numbers smaller than one exist in many systems in physics i.e. in ideal gas. In this case Eq(1) yields,

$$n_i = e^{-h\nu / k_B T_i} \quad , \tag{2}$$

Which is the canonic distribution. Consider a sequence of Λ temporal modes emitted from a radiation source, which is not in equilibrium. Non-equilibrium state means that any mode may have its own temperature. The number of photons in the sequence is $\sum_{i=1}^{\Lambda} n_i$. The average energy of a single mode is $q_i = n_i h\nu$. The temperature is calculated from Eq.(2) to be, $T_i = -h\nu / k_B \ln n_i$. The entropy of a single mode will be $S_i = q_i / T_i$, or $-k_B n_i \ln n_i$. Since the entropy is extensive, the total entropy of a sequence of Λ modes is,

$$S = \sum_{i=1}^{\Lambda} S_i = -k_B \sum_{i=1}^{\Lambda} n_i \ln n_i \cdot \tag{3}$$

When $n_i \ll 1$, $n_i = p_i$ and Eq.(3) is simply the Gibbs entropy. Assuming that all n_i are equal to n (which means an equilibrium state as all the temperatures T_i are equal to T), we obtain from Eqs. (3&2);

$$S = \sum_{i=1}^{\Lambda} S_i = \frac{\Lambda h\nu}{T} e^{-\frac{h\nu}{k_B T}} = \frac{\Lambda n h\nu}{T} = \frac{\Lambda q}{T}, \tag{4}$$

The entropy of the sequence of Λ temporal modes, in the quantum limit, is a function of the mode energy and its temperature. Any loss of a photon will change the entropy as well as any fluctuation of the source temperature.

The High Occupation Limit

When the source is hot and/or the frequency is low such that $h\nu \ll k_B T$, Eq(1) yields,

$$n_i h\nu = q_i = k_B T_i \quad \text{or} \quad S_i = k_B \tag{5}$$

This is the well-known relation of a harmonic oscillator. In this limit, the photon energy is negligible as compared with the average energy of the mode, and therefore energy can removed or added in a continuous way. The uncertainty of ½ $h\nu$ is also negligible. To some extent, it is a surprising result. Intuitively, we expect from one mode, which contains many photons, to have zero entropy. Nevertheless, one k_B is a very small amount of entropy, i.e. a laser mode, which sometimes contains as much as 10^{20} photons, has the same amount of entropy as one vibration mode of a single molecule. The Gibbs, Boltzmann or Von-Neumann entropies yield null entropy for a single mode radiation, as they are only an approximation of the entropy for a statistical ensemble (Jaynes, 1965). Finite entropy to a single mode is necessary, as entropy is an extensive quantity and the emission of entropy by a blackbody radiation is the sum of the single modes emission. Therefore, if a single mode would not carry entropy, a blackbody would not emit entropy as well.

When Bob is using his Blackbody (in this case, he will prefer a CW laser) and sending Alice Λ classical modes, the total entropy that is removed from his source is,

$$S = \sum_{i=1}^{\Lambda} S_i = \Lambda k_B, \qquad (6)$$

In the next two sections, it shown that entropy in the high occupation limit is not a simple sum of the entropies of the modes. A sequence of Λ oscillators is not always a state of equilibrium since we can add as many empty modes as we wish. Therefore, Eq.(6) is a lower bound of the entropy. The entropy defined only in equilibrium. The equation of the entropy can used away from equilibrium, however the obtained value (usually known as the Boltzmann -*H* function) is not unique and is always smaller than the entropy (Huang, 1987).

The entropy of a single oscillator in the high occupation limit is not a function of the energy or of the temperature. In fact, when a mode divided to *N* fractions, each fraction, when received, carries the same amount of entropy as the undivided mode. It was shown previously (Kafri, 2007a; 2007b) that this property is the basis of information transmission and is the cornerstone of IT.

Modulation and Information

A possible way for Bob to modulate his source is to use a shutter (Fig. 1). Every temporal mode has a duration of its coherence length, namely $\Delta t = c/\Delta \nu$. Where *c* is the speed of light. Therefore if the shutter is opened for a time interval Δt, an amount k_B of entropy is transmitted.

When Bob is transmitting a file, he possibly starts by sending a header to inform Alice about the file length Λ that he intends to send and some other information about the kind of compression or the language he uses. Usually Alice replies to confirm the acceptance of the header and her consent or refusal to receive the data and so on. However, although this handshaking process is vital to any communication, it is not discussed here. The discussion here assumes that Bob and Alice have pre-agreed language, compression, protocol and an open channel of communication.

If Bob sends *L* energetic bits in Λ modes where, $\Lambda > L$, there are several different messages that can be sent. The number of possible messages is the binomial coefficient $\Lambda!/(\Lambda-L)!\,L!$. The Shannon information is defined, in bits, as the shortest file *I* that has this number of messages. Therefore, $2^I = \Lambda!/(\Lambda-L)!\,L!$. Hereafter the information will be expressed in nats, namely, $e^I = \Lambda!/(\Lambda-L)!\,L!$. Stirling formula yields,

$I = \Lambda\ln\Lambda - L\ln L - (\Lambda - L)\ln(\Lambda - L).$ (7)

Under the assumption that all the combinations have equal probability, it is seen that if $\Lambda = 2L$ then $I = \Lambda\ln 2$ namely, a random file contains the maximum amount of information.

Equilibrium and Entropy

The basic definition of equilibrium obtained from Clausius inequality, namely $S \geq Q/T$. When the heat transmitted divided by the temperature is equal to the entropy, the system is in equilibrium. This implies that when Q/T is a maximum, the system is in equilibrium. If the system is not in equilibrium, the obtained temperature (that is always higher than T) is not unique and can yield different values for different histories of a system.

If we designate $p \equiv L/\Lambda$, then the RHS of Eq.(7) can be rewritten as,

$I = -\Lambda\{p\ln p + (1-p)\ln(1-p)\}$ (8)

Consider $p\Lambda$ oscillators, each carries k_B entropy, in a sequence of Λ modes. In the setup of Fig 1, each mode has the same frequency and temperature. That means that each mode is in equilibrium with the emitting Blackbody and with the other modes. However, there is mixing entropy of the energetic modes and the empty modes. The mixing entropy of the ensemble *a la* Gibbs is,

$H = k_B\Lambda\{p\ln p + (1-p)\ln(1-p)\}$ (9)

where H is the Boltzmann H function. The $-H$ function is the entropy calculated away from equilibrium such that $S \geq -H$. In equilibrium $p = ½$, H has a minimum and Eq.(8) yields that $S = k_B\Lambda\ln 2$.

Therefore it is possible to conclude that entropy and information are identical and a random sequence is a state of equilibrium. In the case that $p < ½$ one obtains,

$S \geq -H = k_B I$ (10)

Eq.(10) is the Clausius inequality for informatics.

It worth noting that Eq.(9) yields, in statistical physics, the canonic distribution (Kafri, 2007a; 2007b). Consider $p\Lambda$ energetic particles of energy $h\nu$ in Λ microstates. The energy of the sequence is $Q = p\Lambda\, h\nu$ .The temperature is defined *a la* Clausius as $T = \partial Q/\partial S$, Therefore,

$\partial S/\partial p = -\Lambda h\nu/T = \Lambda k_B\{\ln p - \ln(1-p)\}$ or $p/(1-p) = e^{-h\nu/k_BT}$ which is the canonic distribution of Eq.(2) for a two level system(see table 1) (Kafri, 2007a; 2007b).

What is the reason for such different results, in statistical physics and in informatics, obtained from the same Eq.(9)? The explanation is that in statistical physics, the canonic distribution prevails and the entropy of a mode is a function of the energy and the temperature, as is seen in Eq(4). Therefore, equilibrium (a maximum entropy) may obtained, for any value of $h\nu$ and T, for any $p \le ½$. In informatics the entropy of a mode is not a function of the energy or the temperature, therefore equilibrium exists only for a single value $p = ½$.

Logical Quantities

When Bob modulates the transmitted radiation of the setup of Fig.(1), he usually does not care about the coherence length of his radiation source. He transmits a sequence of pulses and vacancies of equal length. Therefore, each "1" bit will usually carry more entropy than one k_B. Practically it will carry $K = mk_B$ entropy units, were m is some integer (see Eq.(6)), therefore Eq.(10) can be rewritten as,

$$S \ge KI, \tag{11}$$

When Bob transmits a file, all he wants is for Alice to receive correctly one of the 2^{Λ} possible files in a Λ bits transmission. However, here we are interested for information of this particular transmission. A possible way to calculate the amount of information in the transmission is to use Eq.(9) to calculate $-H/S$ which is the normalized information. Which yield

$$-\frac{H}{S} = -\frac{p\ln p + (1-p)\ln(1-p)}{\ln 2} \le 1 \cdot \tag{12}$$

Eq.(12) is the logical Clausius inequality in informatics. It says that the maximum amount of information in a file that has a fraction p of the bits "1" or "0" is not a function of K. In fact, Eq.(12) is an inequality, free from any physical quantity.

Logical Equilibrium – The Benford's Law

Eq.(12) demonstrates that the fraction p of the "1" or "0" bits determines how far a file is from equilibrium. If p=1/2, it means that a file might be in equilibrium. Nevertheless, information transmission not done usually in bits. In our everyday life, we are using a much larger amount of symbols to communicate. An important set of symbols is the numerical digits. A possible way to form a set that represents the nine digits is to use nine kinds of modes. Each one contains 1,2,3,4,5,6,7,8,9 energy units respectively. What will be the relative distribution of these modes in equilibrium? If all the bosons have the same energy, each occupation number n yields a different temperature, which means a non-equilibrium state. Eq.(1) is rewritten as,

$$\Phi(n) = \Phi = \frac{h\nu}{k_B T} = \ln(1+\frac{1}{n}) \quad (13)$$

It seen that as *n* increases, the temperature increases. A possible way to obtain equilibrium, namely an equal temperature for all the digits, is to use a spectral filter, with nine different frequencies that can be obtained from Eq.(13). An alternative way to obtain an equilibrium state is to keep Φ constant and to distribute the nine symbols according to a density function $\rho(n)\Phi = \Phi(n)$. Such that,

$$\rho(n_i)\Phi = \ln(1+\frac{1}{n_i}) \quad (14)$$

The relative distribution of digits in equilibrium is,

$$\varphi(n_i) = \frac{\rho(n_i)\Phi}{\sum_{i=1}^{9}\rho(n_i)\Phi} = \frac{\rho(n_i)}{\sum_{i=1}^{9}\rho(n_i)} \quad (15)$$

From Eq.(15) it is seen that Φ disappears altogether, similarly to the way the normalized entropy is independent of the frequency and temperature and became information, in the high occupation limit. Since $\sum_{i=1}^{9}\rho(n_i) = \sum_{i=1}^{9}\ln(1+\frac{1}{n_i}) = \ln(10)$. Therefore,

$$\varphi(n_i) = \log_{10}(1+\frac{1}{n_i}) \quad (16)$$

Eq.(16) is the Benford's law (Hill, 1996; Benford, 1938; Hill, 1986) that was found empirically in many statistical ensembles of digits that originate from natural sources and are not produced by artificial randomizers.

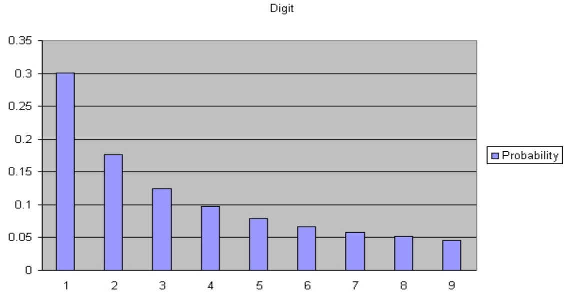


Figure2. *The Benford's law is the probabilities of the digits in many numerical data files.*

It is worth noting that with the cancellation of Φ we see that the normalized distribution function is independent, not only of the temperature that was chosen arbitrarily in Eq.(13), but also of the energy of the boson as well. Moreover, it is free of any physical variable and/or physical parameter.

Energy distribution – Power-law

For a canonic ensemble, the energy distribution obtained from the Planck statistics is, $n_i = e^{-h\nu_i/k_B T_i}$. The Planck distribution gives us the number of photons for any ratio $\Phi = h\nu/k_B T$. Φ may be viewed as the relative energy of a photon with respect to the average energy. Therefore, the occupation number n of all the modes having the same temperature T (in equilibrium) given by,

$$n_i = \frac{1}{e^{h\nu_i/k_B T} - 1} \tag{17}$$

The setup that describes Eq.(17) is the same one as in Fig.(1), but without the spectral filer, therefore all frequencies are transmitted. In high occupation number Eq.(17) can be approximate to, $\Phi_i = \ln(1+\frac{1}{n_i})$ or $\Phi_i \approx \ln n_i$. We expand $\ln n_i$ around 1 and obtain that $\ln\frac{1}{n_i} \approx \frac{1}{n_i}$ therefore $\frac{\partial \ln \Phi}{\partial \ln n} \approx -1$. In Fig.(2) a plot of $\ln n_i$ Vs. $\ln\Phi_i$ is shown, for in the classical regime a power-law like distribution is obtained, and moreover, the exponential truncation in the quantum regime appears. It is worth noting that the only assumption of this curve is equilibrium. Namely, all modes are at the same temperature (Gupta, et al. 2005).

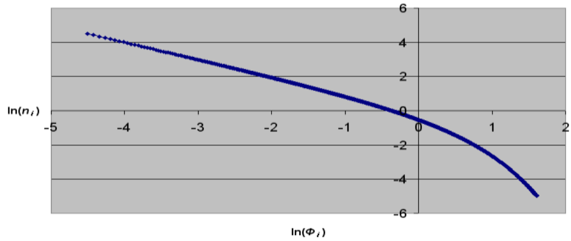


Figure 3. *A log-log plot of the occupation number versus the relative boson energy*

Since many phenomena that related to natural processes exhibit the power-law distribution (Newman, 2006) it attracts a considerable attention. To mention few: the frequency use of words, the number of hits of web sites, the copies of books sold, the population of cities etc. The slopes obtained from the empirical power-law distributions are around –1. This is similar to the Gaussian distribution of momentum that obtained from the exponential distribution of energy in the canonic limit. A connection of the Planck statistics to complex networks was discussed previously (Bianconi & Barabasi, 2001) and the similarity between the mapping of the Bose-Einstein gas and a network model was discussed. A possible explanation, based on the present theory, for the reason why so many phenomena exhibit a power-law distribution, found in section VI.

Combined Systems

In previous publications (Kafri, 2007a; 2007b) it was shown that for a classical ensemble the Shannon information is entropy, an amplifier is a Carnot cycle and broadcasting from one antenna to several antennas is a heat flow from a hot bath to a cold bath. In addition, an informatics perpetuum mobile of the second kind defined. Here it is shown that a classical ensemble has a power-law energy distribution while in the quantum limit when $\Phi >> 1$, the canonic distribution dominates and the regular Gibbs-Boltzmann thermal thermodynamics takes place. Therefore, the Shannon information and the thermal entropy are two faces of the same entropy.

Under what condition thermal entropy will generated and under what condition information will generated? When there are two possible ways to generate entropy in a system, the sum entropy will be the combination of the two that maximizes the total entropy of the system (Dawkins, 1976). Hereafter a simple example of such a combined system is considered.

Hooke's law Harmonic Oscillator

Consider a Hooke's law oscillator, at a room temperature T, having a spring constant κ and an amplitude A_L. The total energy of the oscillator is $E_L = \frac{1}{2}\kappa A_L^2$. The temperature of this oscillator from Eq(5) is $T_L = E_L / k_B$. To increase the amplitude of the oscillator to higher amplitude E_H, a work W should applied. The new amplitude will $E_H \leq E_L + W = \frac{1}{2}\kappa A_H^2 = k_B T_H$. The inequality stands for the situation in which the applied work is not with a resonance with the frequency of the oscillator and therefore part of the work wasted to heat. It is seen that,

$$\frac{W}{E_H} \leq 1 - \frac{T_L}{T_H} \,, \tag{18}$$

Namely, the efficiency of the amplification of the oscillator is the Carnot efficiency. Increasing the energy of the bits in a file was shown (Kafri, 2007a; 2007b) to be a classical Carnot cycle, which comprises of two isotherms and two adiabatic. Here it shown that single oscillator amplification is also a Carnot cycle. The Hooke oscillator has a weight of a finite mass that affects its frequency. The mass of the weight consists of a large number of particles; each particle has its own degrees of freedom. Each of these particles carries similar entropy to that of the whole Hooke's oscillator, which is a single oscillator. Therefore, the temperature of the Hooke oscillator is much higher than the thermal temperature of the weight, which is in the room temperature. The Hooke oscillator temperature is similar to that of antennas (Kafri, & Kafri, 2013) (for a typical cellular antenna was shown to be $\sim 10^{15}$K) or for a laser (for a 0.7μm laser with 10^{16} photons per mode, is $\sim 10^{20}$K) and is of the same order of amplitude, namely $\sim 10^{20}$K. These kinds of temperatures are impossible to obtain by heating up a blackbody by conventional means. Nevertheless, these temperatures can obtained by non-thermal resonance pumped sources.

Removing energy from the Hooke's oscillator does not change its entropy because it is a harmonic oscillator and therefore it has a constant entropy, k_B. However, dumping the oscillator's energy to a canonic ensemble increases the entropy according to Eq(4). Therefore, the Hooke oscillator will dump spontaneously its energy to its thermal bath. This example and similar phenomena are responsible for the common intuition that the information energy dumped spontaneously into a thermal energy. In fact, this is an example of heat flow from a hot harmonic oscillator to a cold thermal bath.

Information versus Thermal Entropy

Does nature prefers the informatics systems or the thermal canonic systems? This is an interesting question, as we know that our world consists of a mixture of the two. The common intuition, which based on the canonical thermal physics, suggests a pessimistic end to any closed system, namely, a canonic thermal equilibrium (the heat death that suggested by Kelvin). The common intuition suggests that informatics is a non-equilibrium phenomenon (Xiu-San, 2007). Since a file, is a sequence of harmonic oscillators, at the end, the information's energy will relax into a thermal equilibrium exactly as the Hooke's oscillator transfers its energy to its bath. However, this is not what the Planck's statistics suggests. As we see in Fig. 3, there are much more low energy bosons than high-energy bosons.

Eq.(13) suggests that for a given temperature T, when Φ is decreased, n is increased according to,

$$\Phi_i = \ln(1+\frac{1}{n_i}) \tag{19}$$

When Φ <1, it means that a boson has less energy than the average. When Φ >1 it means that a boson has more energy than the average. Eq(19) suggests that in equilibrium there are more poor energy classical bosons as compared with rich energy (lucky) canonic bosons. A blackbody radiation is a good example of a mixed system. The number of modes in the volume of a blackbody increases with the frequency cube; the wavelength of the light limited by the diameter of the blackbody. Therefore the occupation number decreases with the frequency according to Eq.(17). The result is the familiar Blackbody radiation spectrum curve that gives similar amount of energy to the poor photons and to the rich photons. Therefore, in blackbody emission, the number of the poor photons is much higher than that of the rich photons.

Summary and Discussion

Based on a toy model, it shown that the Planck distribution, in the quantum limit, yields the regular canonic thermodynamics. In the high occupation limit, the harmonic oscillator statistics replaces the canonic statistics. The harmonic oscillators' statistics differs in several aspects from the canonic statistics as is shown in table 1. An important feature of this statistics is that the normalized thermodynamic functions like entropy and particles distribution do not contain physical quantities. In the canonic entropy the exponential term does not canceled out in the normalization process. Therefore, the canonic entropy is a function of the temperature and the frequency. Any fluctuation of the energy and/or the temperature affects its magnitude. In the high occupation limit entropy, all the physical variables and parameters disappear and we obtain the Shannon information. Therefore, the entropy is not sensitive to any fluctuation in the occupation number, the source temperature and/or frequency. It is not even sensitive to the number of modes in a bit. This property of the entropy, in the high occupation limit, makes it appropriate to convey data.

Table 1. *The thermodynamic properties of the Bose-Einstein gas in equilibrium at temperature T for photons (with zero chemical potential) for the classical and the canonic distribution. p is the probability of the energetic modes and n is the occupation number.*

	High Occupation $n >> 1$	Canonic $n << 1$
Temperature	$T = \frac{nh\nu}{k_B}$	$T = -\frac{h\nu}{k_B \ln n}$
Equilibrium	$p = 1/2$	$\frac{p}{1-p} = e^{-\frac{h\nu}{k_B T}}$
Average mode entropy	$S = k_B \ln 2$	$S = \frac{h\nu}{T} e^{-\frac{h\nu}{k_B T}}$
Energy Distribution	Power-law	Exponential
Carnot cycle	Amplifie	Heat engine

The logical quantities, in the high occupation limit, are therefore applicable to many phenomena of our life. The Planck distribution of photons is a simple combinatory of states and particles without interactions. The only constraint encapsulate in it is the quantization.

Namely, it is possible to add or to remove energy from any mode in an integer amount of some undivided particle (a quant). As is seen in Eq.(15) when we keep the quant size fixed (a constant frequency) and we also assumed equilibrium (equal temperature for all the modes), than the normalized distribution of the photons is not a function of the energy, the frequency, the average energy or the temperature. The physics is faded away, and we remain with a statistical system of inert quanta. Systems like these are very common in life. Consider the distribution of the population of cities. Each city may considered as a mode. When we count the number of the peoples in a city, the peoples are, per definition, indistinguishable. Since the number of the peoples is quantized, therefore this system is identical to that of Eq.(15). Similarly, the number of books being sold in a certain period is a homological system to that of the population of cities. In this case, the number of the titles is the number of the modes and a single copy sold is a quant. The number of hits in the Internet is also a system of this kind as the number of the sites is the number of modes and a hit is a quant.

In the derivation of Benford's law Eq.(14) was used, namely, $\rho(n_i)\Phi = \ln(1+\frac{1}{n_i})$. This equation yields slope of "-1". In the normalization process Φ disappears. A slope "-2" is obtained if we substitute $\psi^2(n) = \rho(n)$, with a phenomenological analogy to the substitution of momentum instead of energy in canonic exponential distributions to obtain the Gaussian distribution.

The present model does not consider any interactions between the quantized particles. Nevertheless, interactions do exist. If we consider, for example, the distribution of the hits among the sites in the Internet, it is obvious that there are interactions between the visitors of the sites. The interactions might be advertisements by the sites and/or viral spread of the recommendations by the visitors. So what is the reason for a somewhat oversimplified model without interactions being so effective? A possible explanation is that the distribution of the hits is independent of the interactions; however, a specific rank of a certain site does depend on the interactions. Namely, the interactions are responsible only for the specific site location in the distribution. If that is true, removing a several popular sites will not change the normalized distribution. Other sites will take the place of the removed sites and the distribution will reach equilibrium again. Indeed, this is what seen in almost any economical system, namely, "there is no empty space". Unlike the derivation of Benford's law, the present model does not pretend to be a complete solution to the power-law distribution in social systems. Nevertheless, it argued that extensive equilibrium thermodynamics might predict the quantitative behavior of social systems.

Another notable property of the logical equilibrium is the quenched randomness. For the receiver, a random file is content. However, within the context of IT, a random file, which is a compressed file, is an ensemble of harmonic oscillators in equilibrium, as is seen in Eq(12). An outcome of this conclusion is that files should have a tendency to compress spontaneously. Aside from the natural spontaneous noise, we are obsessed with compression. In IT we compress files for economical reasons. However, an observer in space sees that most of the transmitted files on earth are compressed. This observer will rightly, conclude that files have a tendency to be compressed. Our tendency to compress seen also in art. We find ourselves impressed by an artist who can express a complex feeling with a few words, or by a painter who can represent a detailed picture with a few lines and colors. The artistic kind of compression known in IT as a lossy compression and is very popular in multimedia technology. The language is a most powerful compressor; sometimes, the amount of information in a short sentence is enormous considering the fact that it contains just a few bytes. A notable example is the mathematics, which enables to write relatively short formulas that describe complex logical processes. It is possible that our tendency for symbolism and mathematics is the natural tendency toward equilibrium.

The last issue and the most intriguing one is how the tendency of information to increase affects life. Conventional canonic thermodynamics explains how we decompose chemical compounds in order to produce mechanical work, and heat to enable our body to function properly. This paper suggests that we also want to increase information. The increase of information could done by reproduction and by broadcasting. It is clear that the present evolution theories are with full agreement with the present theory (Dawkins, 1976). The only modification required is that reproduction and evolution are spontaneous processes. It was shown previously (Kafri, 2007a; 2007b) that information is multiplied in broadcasting. Therefore, it is not surprising that we are obsessed with a desire to broadcast ourselves. When Bob broadcasts a file with I bits to N receivers, he will increase the information by NI. A receiver will increase the information by I bits. Therefore, it is better, thermodynamically, to broadcast than to receive.

It is an observable fact that information and life in their various forms increase with time; therefore, it is plausible that aside from the chemistry necessary for the existence, life means a reproduction and a compressed communication.

References

Bianconi, G., & Barabasi, A.L. (2001). Bose-Einstein condensation in complex networks, *Physical Review Letters*, 86(24), 5632. doi. 10.1103/PhysRevLett.86.5632

Benford, F. (1938). The law of anomalous numbers, *Proceedings of the American Mathematical Society*, 78(4), 551-572.

Brillouin, L. (1953). The negentropy principle of information, *Journal of Applied Physics*, 24, 1152. doi. 10.1063/1.1721463

Brillouin, L. (1962). *Science and Information Theory*, Academic Press NY.

Dawkins, R. (1976). *The Selfish Gene,* Oxford University Press.

Gershenfeld, N. (2000). *The Physics of Information Technology,* Cambridge University Press.

Gupta, H.M. Campanha, J.R., Pesce, R.A.G. (2005). Power-law distribution for the citation index of scientific publications and Scientists, Brazilizn Journal of Physics, 35(4), 981-986. doi. 10.1590/S0103-97332005000600012

Hill, T.P. (1996). A statistical derivation of the significant-digit law, Statistical Science, 10(4), 354-363.

Hill, T.P. (1986). The first digit phenomenon, American Scientist, 86(4), 358. doi. 10.1511/1998.4.358

Huang, K. (1987). *Statistical Mechanics,* John Wiley: New York.

Jaynes, E.T. (1957). Information theory and statistical mechanics I, *Physical Review Letters*, 106(4), 620. doi. 10.1103/PhysRev.106.620

Jaynes, E.T. (1957b). Information theory and statistical mechanics II, *Physical Review Letters,* 108(2), 171. doi. 10.1103/PhysRev.108.171

Jaynes, E.T. (1965). Gibbs vs. Boltzmann entropies, *American Journal of Physics*, 33, 391. doi. 10.1119/1.1971557

Jaynes, E.T. (1988). The evolution of Carnot's principle, in *Maximum-Entropy and Bayesian Methods in Science and Engineering,* 1, G.J. Erickson & C.R. Smith (Eds.), (pp.267-281), Springer. doi. 10.1007/978-94-009-3049-0_15

Kafri, O. (2007). The second law and informatics, *Unpublished Paper*, [Retrieved from].

Kafri, O. (2007b). Informatics Carnot cycle, *Unpublished Paper*, [Retrieved from].

Kafri, O., & Fishof, E. (2016). Economic inequality as a statistical outcome, *Journal of Economics Bibliography*, 3(4) 570-576.

Kafri, O., & Kafri, H. (2013). *Entropy - God's dice game*. CreateSpace, pp.154-157. [Retrieved from].

Kestin, J. (Ed.) (1976). *The Second Law of Thermodynamics,* Dowden, Hutchinson and RossStroudsburg.

Newman, M.E. (2006). Power-law, Pareto Distribution and Zipf's law, *Unpublished Paper*, [Retrieved from].

Shannon, C.E. (1949). *A Mathematical Theory of Communication*, University of Illinois Press: Evanston, Ill.

Simeonov, P.L. (2010). Integral biomathics: A post-Newtonian view into the logos of bios, *Progress in Biophysics and Molecular Biology*, 102(2-3), 85-121. doi. 10.1016/j.pbiomolbio.2010.01.005

Xiu-San, X. (2007). Spontaneous entropy decrease and its statistical physics, *Unpublished Paper*, [Retrieved from].

Woolhouse, H.W. (1967). Negentropy, information and the feeding of organisms, *Nature,* 213, 952. doi. 10.1038/213952a0

5.
Information Theoretic Approach to Social Networks

Introduction

The word network, like information, is overly used. For example, in Shannon's theory there are confusions originated from the fact that some conceive stored data as information, while Shannon information theory deals with a file transmission from a sender to a receiver. Similarly, we conceive a network as a static graphical diagram of links connecting nodes while actually a network characterized by a flow between nodes. For example, the electrical networks conceived as a static net of electrical cables connected together; transportation networks as a static net of roads and irrigation networks as a static net of pipes, etc. However, the flow of electricity, traffic or water is the essence of the networks. Many scientific papers ware published about various aspects of networks from Erdős & Rényi (1959) to Barabashi (1999; 2002; 2004). Many techniques applied in these researches, from graph theory of Erdős and Rényi to load distribution and statistics (Kafri, 2009). These diverse approaches apply to the many aspects of networks. However, here we discuss another kind of networks, which we call "social networks" which are in fact communication networks. In these networks, we overlook the physical wiring between the nodes and focus solely on the flow between them. An example to such nets is the data networks. Most of the people in the world are connected somehow by physical data networks. Eventually, everyone can communicate with almost everybody. However, the flow of the voice signals between the people is varying constantly in time and not distributed uniformly among them. These networks are similar to a two dimensional fluid.

Social Network: definition

In this paper we adopt a dynamic approach to nets. In our network there are N nodes which can communicate with all the other nodes with no wiring limitations. Moreover, each connection between node i and node j has a value R which is a measure of the flow intensity between the two nodes. For example, in social networks $R_{i,j}$ may be the number of communication channels used between node i and node j. In economical network R may represent the value of a transaction between two nodes.

Shannon, (1948) in his paper "a mathematical theory pf communication", describes in a quantitative way how Bob communicates with Alice via transmitting a "file" to her. The file is a sequence of bits where each bit can be either zero or one. When Alice receives the sequence of bits, she explores their value and extracts the content that Bob sent her. Shannon defined the entropy of the files as the logarithm of the possible different contents that the sequence may contain. Since N bits file has 2^N possible different contents, the entropy of the file is $N \ln 2$. Engineers are using, for their convenience, base 2 logarithm and therefore the Shannon entropy in this base is identical to the length of the file, N bits.

Basically, Shannon's theory deals with a one way communication between a sender (Bob) and a receiver (Alice) in which the sender send one or several bits to a receiver. Bits carry uncertainty which is expressed by the entropy. After reading and interpreting the file, the receiver can find its content

In this paper we describe a group of N senders. Each of these senders can send and also receive information from the other $N-1$ members of the group. We call the communication group of N senders/receivers a network. We also call each one of the N senders/receiver a node. In addition, we call a one way single communication channel connecting N_i to N_j a link. We designate $R_{i,j}$, as the number of links through which a sender i can send messages to a receiver j. Similarly, $R_{j,i}$ designates the number of links used from j to i. We assume that there is a total number of R links in the network and R can be any integer. The network can be described by a matrix:

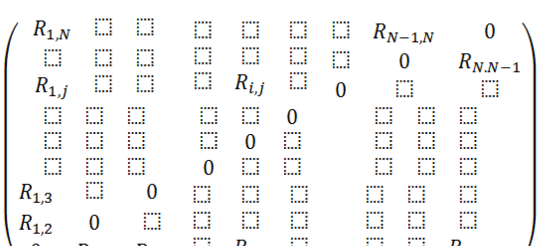


Figure 1. *Networks matrix.*

Where $R = \sum_{j=1}^{N} \sum_{i=1}^{N} R_{i,j}$.

The summation on a column i is the total links outgoing from node i to all other $N - 1$ nodes. Similarly, the summation on a row j is the total links entering node j from all other $N - 1$ nodes.

The network described above is different from our standard visualization of a net as a static diagram. $R_{i,j}$ can vary constantly like a two dimensional fluid matrix having two constraints;

A. $R_{i,i} = 0$

B. $R_{i,j} \leq R$.

Communication and economic networks have such a dynamic nature. In this aspect one may compare the network to a two dimensional fluid in which there are constant nodes and energetic links that are i.e. pulses (classical harmonic oscillator) or any other logical quantity such as money, etc. The number of links may represent the bandwidth of the communication channel or the amount of the money transferred.

We can imagine the networks as a two dimensional boson gas with $K = N^2 - N$ states and R particles. Therefore, we can calculate for it entropy, temperature, volume and pressure.

Large Networks Statistics

The number of microstates W of boson gas of R particles in K states is given by

$$W = \frac{(R+K-1)!}{(K-1)!R!} \tag{1}$$

Planck (1901) used this equation assuming that $(K + R) \gg 1$, and designatingthe "occupation number"$n \equiv \frac{R}{K}$, to obtain his famous result for the entropy;

$$S(R, K) = \ln W = K[(n + 1)\ln(n + 1) - n \ln n] \tag{2}$$

We define large network as a network in which$R \gg K$,In this network it is possible to remove energetic links from it with a negligible change in its statistical properties. The thermodynamic analogue to the large network is an infinite thermal bath.

The entropy of the large net is given by $S(R, K) = \ln W$ (Kafri, 2014). When one link added, the entropy is given by,

$$S(R + 1, K) = \ln \frac{(R+K)(R+K-1)!}{(R+1)R!(K-1)!} = \ln \frac{R+K}{R+1} + S(R, K) \tag{3}$$

In the case that$R$ is a large number than $R/(R + 1) \approx 1$and,

$$S(R+1,K) \approx S(R,K) + \ln\frac{n+1}{n} \quad (4)$$

Carnot Efficiency

Suppose we have two large networks H and L having occupation numbers n_H and n_L . We remove Q links from the L net and put them in the H net. If $n_H > n_L$ than the entropy of the Q, L net linksis higher than Q, H net link is, and the total entropy will be decreased. Therefore, we must add W links to the H net, in order to avoid entropy decrease such that,

$Q \ln[(n_L+1)/n_L] \leq (Q+W) \ln[(n_H+1)/n_H]$ or,

$$W \leq Q\{1 - \ln\frac{n_H(n_L+1)}{n_L(n_H+1)}\} \quad (5)$$

In the case that $n_{H,}$ and $n_L \gg 1$ then,

$$W \leq Q(1 - \frac{n_L}{n_H}) \quad (6)$$

Equation 6is Carnot inequality for networks.

Large Networks Temperature

The definition of temperature is related to the definition of entropy. In classical heat engine the Carnot efficiency is,

$$W \leq Q(1 - \frac{T_L}{T_H}) \quad (7)$$

Where W is the work, Q is the heat (energy removed or added) and T is the temperature. The occupation number $n$is related in the classical limit of blackbody radiation (photons) $n \gg 1$ to the temperature via,

$$nh\nu = k_B T \quad (8)$$

Here h is the Planck constant,ν is the oscillator frequency and k_B is the Boltzmann constant.

Therefore, if we substitute for a constant frequency, ν, in equation 6 we obtain equation 7.

We can calculate the temperature directly from,

$$T = \frac{Q}{S}$$

In equation 4 we obtained the entropy increase by adding one link $Q = 1$, namely,

$$\Delta S = \ln \frac{1+n}{n}.$$

Therefore,

$$T = \left(\ln \frac{1+n}{n}\right)^{-1} \quad (9)$$

This result can also be obtained from Planck equation (2),

$$T = \frac{\partial R}{\partial S}$$

$$\frac{\partial S}{\partial R} = \frac{1}{K}\frac{\partial}{\partial n} K[(n+1)\ln(n+1) - n \ln n] = \ln \frac{1+n}{n} = \frac{1}{T} \quad (10)$$

We see that the two ways yield the same result.

In the $\lim_{n\to\infty} \ln(1+\frac{1}{n}) = \frac{1}{n}$ and,

$$T = n \quad (11)$$

This result is consistent with equation 6.

Large networks Volume

The volume of the large net is the number of its states K. This is the major difference between a gas and a large net. In the linear world, and therefore in our intuition, the volume is an extensive quantity. However, in nets the number of nodes is the extensive quantity. Since $K_i = N_i(N_i - 1)$ and N is extensive, therefore K is not extensive, i.e. when we combine two nets 1 and 2, $N = N_1 + N_2$ and,

$$V = K = (N_1 + N_2)(N_1 + N_2 - 1) = K_1 + K_2 + 2N_1N_2$$

Or for large nets,

$$V \approx V_1 + V_2 + 2(V_1V_2)^{1/2} = (\sqrt{V_1} + \sqrt{V_2})^2 \quad (12)$$

Namely, the volume of the combined net is greater than the sum of their volumes. This is a counterintuitive result. When we combine two networks there is an expansion as a result of the increase of the number of states. Combining nets at constant number of links (adiabatic process) results in cooling and entropy increase. This is an explanation to a known phenomenon that networks tend to merge. It is well known that entropy increase in adiabatic process does not exist in ideal gas thermodynamics.

Large Networks Pressure

The gas law states that the pressure of the gas multiplied by its volume is a measure of the energy of the gas. In our case the particles are identical. Therefore, the energy of the net is the number of its links.

$$PV = R = VT \tag{13}$$

where P is the pressure and R is the number of the particles. With analogy we write

$$P = T \approx n \tag{14}$$

The pressure of a net is a measure of the tendency of two nets having different pressures to be combined together to equate their pressure and temperature to equilibrium, and thus to maximize the total entropy. Due to the nonextensivity of the volume, the combined pressure of two nets may be lower than the pressure of each one of them separately.

Large Networks Entropy

From equation 2 for large nets

$$S = \left[Vn \ln\left(\frac{1+n}{n}\right) + \ln(n+1)\right] \text{or,}$$

$$S = V \lim_{n\to\infty} \left[\ln\left(\frac{1+n}{n}\right)^n + \ln(n+1)\right] \approx [V + \ln(n+1)]\,) \approx V\ln(1+n) \tag{15}$$

Example

We take two large nets 1 and 2 with known pressure and volume. We combine them together. What will be the pressure and volume of the final net?

The solution for ideal gases is simple:

$$P_1V_1 + P_2V_2 = (R_1 + R_2)T$$

Or the temperature of the combined gases is,

$$T = \frac{P_1V_1 + P_2V_2}{(R_1 + R_2)k}$$

And the pressure of the final gas is,

$$P = \frac{P_1V_1 + P_2V_2}{V_1 + V_2}$$

For nets, the result is affected by the non-extensive nature of the nets volume. The temperature is the occupation number. Since R is extensive, therefore,

$$R = R_1 + R_2 = n_1 V_1 + n_2 V_2 \quad \text{and}$$

$$T \approx \mathrm{P} \approx n \approx \frac{P_1 V_1 + P_2 V_2}{\left(\sqrt{V_1} + \sqrt{V_2}\right)^2} \qquad (16)$$

Numerical example

Suppose we have two nets, each with 50 nodes; one has occupation number of 50 and the other of 100. The two nets are combined. What will be the value of the thermodynamic quantities in equilibrium of these two combined nets?

The net law is $\mathrm{P}V = R$

Where P is the pressure=temperature=occupation number, V is the number of states, R is the number of links.

For net 1: $V_1 = 50 \times 49 = 2{,}450$, $R_1 = 2450 \times 50 = 122{,}500$, $T_1 = \mathrm{P}_1 = 50$

Foe net 2: $V_2 = 50 \times 49 = 2{,}450$, $R_2 = 2450 \times 100 = 245{,}000$, $T_2 = \mathrm{P}_2 = 100$

In the combined net: $V = 100 \times 99 = 9{,}900$, $R = 367{,}500$, $T = \mathrm{P} = 37$

The entropy of net 1 is $S_1 = 2450[1 + \ln 51] = 12{,}034$, the entropy of net 2 is $S_2 = 2450[1 + \ln 101]$ =13,732, and the entropy of the combined net is $S = 9900[1 + \ln 38] = 45{,}648$. The entropy increase is then 19,882.

This result demonstrates the major difference between a net and an ideal gas. When we combine nets, the temperature and the pressure drop drastically as a result of the entropy increase originated from the states generation in the combined net. This exhibits the tendency of nets to combine.

Summary and Applications

Is there any value to thermodynamic analysis of networks? This question was probably asked about information theory 70 years ago. It was possible to send files from Bob to Alice without information theory. Actually Samuel Morse did it 100 years before Shannon's time. However, the quantitative work of Shannon enables to find limits on file's compression. Similarly, thermodynamic analysis of networks has already proved itself to be useful in showing that the distribution of links in the nodes in large networks is Zipfian (Kafri & Kafri, 2013). If we define the wealth of a node as the number of links that is has, we see that combining two nets does not increase the wealth but reduces the temperature. Reducing the temperature enables higher free links (free energy), and therefore higher data transfer on the same infrastructure. Equilibrium thermodynamics proved to be an important tool in engineering, chemistry and physics. Applying these tools to sociological networks dynamics may prove to be of some use. For example, defining temperature to a net may help in our understanding of data flow. Zipf distribution may help in finding the stable inequality of links (Kafri & Kafri, 2013).

In a previous paper (Kafri, 2014) a similar calculation was made for the entropy increase when a node is added to a net. The result obtained is similar to that of equation 15. Namely, each node generates about $2\ln(1+n)$ entropy. This result quantifies the entropic benefit of joining the crowd (high linkage nets or hot nets). In this paper we found that the entropy generation caused by adding a link to a net is $\ln(1+\frac{1}{n}) \approx \frac{1}{T}$. It means that with contradistinction to a node, a link will favor joining a network with lower linkage (colder net), which represents the tendency of links (energy) to flow from hot to cold. One should note that the entropy generation by adding link to a net is with accordance to Benford's law (Kafri & Kafri, 2013).

The concept of non-extensive volume can also describe an accelerated expansion without energy.

References

* This paper contains similar content to the paper by the same author (O.K.) "Money, Information and Heat in Networks Dynamics" published in *Mathematical Finance Letters* 6(4), 201-211. (2014) doi. 10.4236/apm.2016.64017 This paper was rewritten for the Information theory community.

Barabási, A.L., & Réka, A. (1999). Emergence of scaling in random networks, *Science*, 286(5439), 509-512. doi. 10.1126/science.286.5439.509

Barabási, A.L., & Frangos, J. (2002). *Linked: The New Science of Networks*. Perseus Books Group: New York.

Barabási, A.L., & Oltvai, Z. (2004). Network bioloy, *Nature Reviews Genetics,* 5, 101-113. doi. 10.1038/nrg1272

Erdős, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae,* 6, 290–297.

Kafri, O. (2009). The distributions in nature and entropy principle. [Retrieved from].

Kafri, O. (2014). Follow the Multitude - A Thermodynamic Approach. *Natural Science*, 6, 528-531. doi. 10.4236/ns.2014.67051

Kafri, O., & Kafri, H. (2013). *Entropy - God's dice game,* CreateSpace, pp. 208-210. [Retrieved from].

Planck, M. (1901). Über das Gesetz der Energieverteilungim Normalspectrum. *Annalen der Physik*, 309(3), 553-563. doi. 10.1002/andp.19013090310

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423. doi. 10.1002/j.1538-7305.1948.tb01338.x

6.
Information Theory and Thermodynamics

Introduction

Shannon information and Boltzmann entropy have the same mathematical expression. However, information conceived as an opposite of entropy. Brillouin in his book suggested that information is negative entropy (Brillouin, 1962). Many view information as a logical sequence of bits of some meaning as oppose to a thermal state, which is a state of randomness. The known scientific knowledge does not support this mystic idea. Shannon has shown that the higher the randomness of the bits in a file, the higher the amount of information in it (Shannon, 1949). The Landauer and Bennet school (Bennet, 2003) suggests that the randomness of the bits in a file related to Kolmogorov complexity (Li & Vitanyi, 1997). This claim may give an impression that the Shannon information is a meaningful subjective quantity. However, according to the Shannon theory a compressed file, containing meaningful information has similar amount of information as an identical file, with one flipped bit that cannot be decompressed and therefore, for us the receivers, it is just a noise.

In this paper, a thermodynamic theory of information is proposed. It is shown that Shannon information theory is a part of thermodynamics, and that information is the Boltzmann -*H* function. Therefore, information has a tendency to increase the same way as entropy.

The information increase observed in nature attributed to a specific mechanism rather than to a natural tendency. Here it is proposed that increase of information as increase of thermal entropy, is caused by the second law of thermodynamics.

To support this claim it is required to calculate, for informatics systems, the quantities in Clausius inequality (which is the formulation of the second law), namely, entropy, heat, and temperature and to define equilibrium.

In section I, the classical thermodynamics of heat transfer from a hot bath to a cold bath reviewed together with the basic definitions of entropy, heat, temperature, equilibrium, and the Clausius inequality.

In section II, a calculation of the entropy, heat, temperature, and the definition of equilibrium for the transfer of a one-dimensional two-level gas from a hot bath to a cold bath according to statistical mechanics provided and an analogy to Clausius inequality is shown.

In section III, an analysis of the transfer of a frozen one-dimensional two-level gas (a binary file) from a hot bath (a broadcasting antenna) to a cold bath (the receiving antennas) provided. A temperature is calculated to the antenna that, together with the transmitted file information (entropy) and its energy (heat), is shown to be in accordance with classic thermodynamics and the Clausius inequality. Therefore, It is concluded that in the absence of thermal equilibrium information is entropy

In section IV, these results used to calculate a thermodynamic bound on the computing power of a physical device and in section V, a thermodynamic bound on the maximum amount of information that an antenna can broadcast is calculated.

Classical thermodynamics of heat flow

Clausius deduced the second law of thermodynamics from Carnot's calculation of the maximum amount of work W that can be extracted from an amount of heat Q transferred from a hot bath at temperature T_H to a cold bath at temperature T_C (Kestin, 1976). Carnot used in his machine an ideal gas as a working fluid, and the gas law for his calculation. The Carnot efficiency is,

$$\eta \equiv W/Q\, \eta \leq 1 - T_C/T_H. \tag{1}$$

Namely, the maximum efficiency η of a Carnot machine depends only on the temperatures. To obtain the maximum efficiency the working gas should obey the gas law, therefore, the machine has to work slowly and reach equilibrium at any time. Clausius (Kestin, 1976) assumed that Carnot efficiency is always true no matter what mechanism or working fluid is used. That means that there is no dependence on the ideal gas law. Clausius concluded that if the Carnot efficiency is always true, there is a quantity, entropy S, that defined in equilibrium (the equality sign) and can calculated according to,

$$S \geq Q/T \tag{2}$$

When a system is not in equilibrium, Q/T is smaller than the entropy. This inequality reproduces the Carnot efficiency. However, it reveals more than one would expect. The entropy change S is equal to Q/T only in equilibrium. Out of equilibrium Q/T is smaller than entropy. Therefore, if we assume that any system has a tendency to reach equilibrium, any system tends to increase *Q/T*. Clausius assumed that taking a system out of equilibrium requires work, which will also eventually reach equilibrium (namely it will thermalized) and, therefore the entropy of a closed system tends to increase and cannot decrease. Temperature and entropy defined at equilibrium and the temperature can calculated as,

$$T = (Q/S)_{\text{equilibrium}} \quad (3)$$

This definition of temperature accepted to be always true.

Now I calculate a simple example of the entropy increase in heat flow from a hot thermal bath to a cold one (see Figure 1).

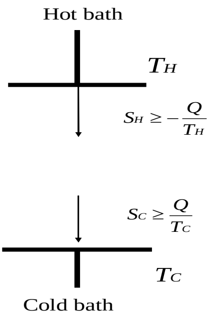


Figure 1. *The entropy increase in spontaneous energy flow from a hot thermal bath to a cold thermal bath.*

When we remove an amount of energy Q from the hot bath, the entropy reduction at the hot bath is Q/T_H . When we dump this energy to the cold bath, the entropy increases by Q/T_C . The total entropy increase is $S = Q/T_C - Q/T_H$. One can see that if the process is not in equilibrium S > $Q/T_C - Q/T_H$. In general,

$$S \geq Q/T_C - Q/T_H \quad (4)$$

In sections II and III I will give an analogy to this example for statistical physics and for information theory.

Statistical Physics of one-dimensional two-level gas

The entropy defined in statistical physics as $k\ln\Omega$, where Ω is the number of microstates (combimations) of a system and k is the Boltzmann constant (Landau & Lifshits, 1980). We will use this definition to calculate the thermodynamic quantities and the Clausius inequality for a system that resembles an informatics system.

We consider a thermal bath at temperature T_H, which is in contact with a sequence of L states. p of the L states have energy ε and that called "one". L-p of the states have no energy and are called "zero". We analyze the thermodynamics of transferring this two-level gas from a hot bath at temperature T_H to a colder bath T_C. The probability of the two-level sequence is simple, because the number of possible combinations of p, "one" particles in L states is the binomial coefficients (see Figure 2), namely, there are, $\Omega = L!/[\,p!\,(L-p)!]$ combinations.

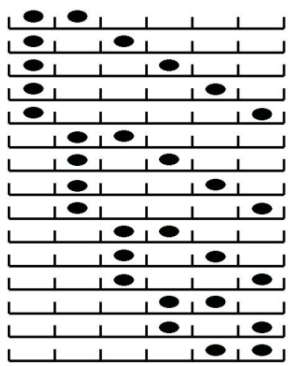


Figure 2. *One-dimension two-level gas with L=6 and p=2. In equilibrium all possible combinations have equal probability (the ergodic assumption). If some of the combinations have higher probability than others the system is not in thermal equilibrium.*

The entropy of the system is $k\ln\Omega$, and the energy of the system is $p\varepsilon$. The temperature is calculated from Eq. (3). Using Stirlings formula, we derive $\partial Q/\partial S$ to obtain T. The well-known result is,

$$p/(L-p) = \exp(-\varepsilon/kT) \text{ or } T = (\varepsilon/k)/\ln[(L-p)/p]. \qquad (5)$$

Namely, one parameter T represents all our knowledge on this one-dimensional two-level gas in equilibrium. This is a well-defined system with a well-defined entropy temperature and energy. The equilibrium was invoked by giving an equal probability to all the possible combinations of the p particles in L states (This assumption is called the ergodic assumption Plischke & Bergersen, (2006)). Eq. (5) is a famous result, but it should note that this derivation done by deriving Q (the heat) and not according to the internal energy of the gas as is done in most textbooks. The reason is that in this model, a two-level gas transferred from a hot bath to a cold bath, and therefore its energy is heat. If a system is not in equilibrium, there are certain combinations that are preferred (a biased distribution), and thus the actual combination span (phase space) is smaller. Therefore, the probability Ω of the gas not in equilibrium is smaller. Since the energy of the gas conserved, we obtain a higher effective "temperature". Boltzmann called the quantity $k\ln\Omega$ calculated for a biased distribution the $-H$ function (Huang, 1987).

When the two-level gas is removed from the hot bath, the entropy is reduced by $S_H = p_H\varepsilon/T_H = kp_H\ln[(L-p_H)/p_H]$. When we dump it to the cold bath, we generate entropy $S_C = p_H\varepsilon/T_C = kp_H\ln[(L-p_C)/p_C]$.

The total change in the entropy is,

$$S_C - S_H = (kQ/\varepsilon)\ln\{(p_H/p_C)[(L-p_C)/(L-p_H)]\} \geq Q/T_C - Q/T_H \quad (6)$$

It should be noted that part of the energy, $\varepsilon(p_H - p_C)$, was transferred from the gas to the cold bath. Obviously Eq. (6) is positive when T_C is lower than T_H , and we see that Eq. (6) is with accordance with Eq. (4), namely, the second law (see Fig 3). The inequality stands for the transmission of a one-dimensional two-level gas with a biased probability of the combinations of the gas. The probability in mechanical statistics considered sometimes as a time average on all possible combinations of p particles in L states. However, if we look at our one-dimensional two-level gas at a given short time, we will observe only one of the possible combinations.

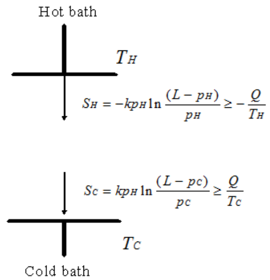


Figure 3. *The entropy increase, due to transmission of one-dimension two-level gas, from a hot bath to a cold bath.*

In fact, we will see a binary file. Adopting a slightly different point of view can solve this paradox, namely, instead of considering the system probability as a time average, we consider it as the probability of finding a given combination at a certain time. In the ergodic case, in equilibrium, any one of the possible combinations can pop up at a given time without any preference. In a biased system, not in equilibrium, certain combinations will have higher probabilities than others will. This approach does not affect the mathematical analysis; however, it will be very useful when we consider information.

Information theory of one-dimensional two-level sequence

The Shannon definition of information based on a model of a transmitter and a receiver. In his model, a binary file transferred from a transmitter to a receiver. A binary file is, in fact, a frozen one-dimensional two-level gas. The binary file is not in thermal equilibrium as it is highly biased to one possible combination of the bits. As oppose to two-level gas, where the energy of the bit ε fixed, in a binary file the bit energy may change continuously.

Shannon was interested in the maximum amount of information that can coded in a given binary file of length L. His famous result is that information has the same expression as entropy. However, no connection made between Shannon's entropy and thermodynamics. The amount p of "one" bits, in a file of length L, is not related to the amount of information in the file. This is in contradistinction to the two-level gas, in which the energy, the temperature and the entropy are functions of p, (see Figure 3). For example, several files having the same amount of "one" bits may have a small amount of information. For example if all the "one" bits are in the beginning of the file, and the rest of the file has zero bits or any other ordered combination (see Figure 4), and some other files may have a relatively high amount of information, if the distribution of the bits in the file is random.

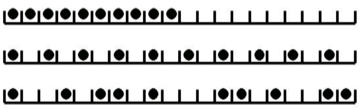


Figure 4. *Three possible binary files having the same energy. The higher two files have higher order and therefore contain little information and can be compressed effectively. The lower file is random and contains maximum information and therefore it cannot compressed and is in equilibrium.*

The amount of information in a file is a function of the randomness of the bits in it, and there is no unique connection between the number of bits and the amount of information. The reason that Shannon obtained the same expression as Boltzmann is that in two-level gas we have no way to predict what combination will be at a certain time, and in a random file we have no way to predict what bit will be at a certain time (the unpredictable sequence of bits is the useful information). Since information and entropy are probability calculations, the same expression obtained. Nevertheless, the calculations for the two-level gas and a file are different as will show hereafter. With analogy to the transmission of two-level gas, we start the thermodynamic analysis of a file transmission by considering a truly random sequence of L bits. In this case $p = L/2$, therefore the maximum information that a file of length L can contain according to Shannon is $I = L\ln 2$. In this case the ratio between the number of "one" bits (the energy) and the information (entropy) is unique. This is in contradistinction to a file with some correlations in which the number of "one" bits p does not determine the amount of information. So by assigning energy ε to the "one" bit we obtain $Q = L\varepsilon/2$ and $S = kL\ln 2$. Using Eq. (3) we obtain,

$$T = Q/S = (\varepsilon/k)/2\ln 2. \qquad (7)$$

Eq. (7) should be compared with Eq. (5) namely $T = (\varepsilon/k)/\ln[(L-p)/p]$. We can see that for a file, the temperature depends only on one variable, the bit energy. In a two-level gas, the temperature depends also on *p*. In two-level gas, lowering *p* reduces the temperature and increases the entropy. So according to the second law a two-level gas would tend to cool down. In a file, reducing the bit energy yields a similar result. Therefore, according to the second law, a file has a tendency to lower its bit energy. We complete the analogy by considering antenna broadcasting a binary file to *N* antennas. A possible deployment of such system is a point-radiating antenna surrounded by a sphere, whose area divided to *N* equal receivers. According to Turing model (1936), the hot antenna emits the broadcasted file. A receiver antenna receives the broadcasted file but with a lower bit energy. Therefore, it is equivalent to a cold bath. Using Eq. (4) we obtain,

$$S \geq Q/T_C - Q/T_H = NkI - kI. \qquad (8)$$

Eq. (8) shows that the file temperature obtained in Eq. (7) yields correctly the increase in information in the broadcasting of a binary file to *N* receivers (which is $NI - I$). In "peer-to-peer" transmission, as in Shannon model, no information increase was involved; therefore, no thermodynamic considerations are necessary. Out of equilibrium, there is a correlation between the bits, and the amount of information in the file is smaller. As a result, the same energy carries less information, therefore T is higher and I is smaller. Using Eq. (8) we can rewrite Clausius inequality for informatics system as, $S \geq kI$.

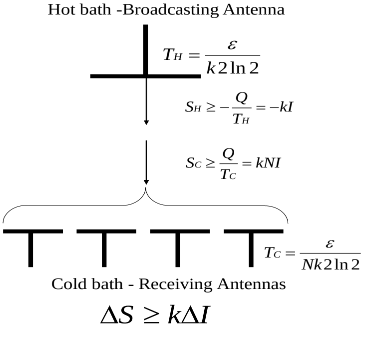


Figure 5. *The analogy between heat flow from a hot bath at temperature T_H to a cold bath at temperature T_C and an antenna broadcasting a file of bit energy ε to N antennas, each receiving the file with bit energy ε/N. In the thermal case the entropy increase is $\Delta S \geq \Delta Q/T_C - \Delta Q/T_H$. However, the same equation $\Delta S \geq \Delta Q/T_C - \Delta Q/T_H$ reproduces well the information balance when we use the temperature definition from Clausius inequality $\Delta Q/\Delta S$ for a compressed binary file. The antennas deployment is drawn to emphasize the physics only.*

This implies that Information, like entropy, tends to increase. In a general case in complex systems both informatics and thermal processes occurs simultaneously. In these cases a transformation of thermal entropy to informatics entropy and vice versa may occur. Thus, Clausius inequality can written as,

$$S \geq Q/T + kI \qquad (9)$$

It is worth noting that ideally compressed files rarely exist. In order to calculate the amount of information in a file, we have to find an ideal compressor (Huffman, 1977). Unfortunately, such a compressor does exist only asymptotically. The amount of information in an uncompressed file, with some correlation between bits, is equivalent to the Boltzmann –*H* function, namely the "entropy" of a system out of equilibrium with a biased distribution. Shannon, in his famous paper (Shannon, 1949), mentioned that information is the Boltzmann –*H* function, nevertheless it called by many entropy.

Example -The Computing power of a Physical device

The units used in communication are the power P and the frequency f in bits/sec of an emitter/transmitter and not the bit energy. Therefore, the temperature of emitter/transmitter can written as;

$$T = P/(k\, f \ln 2\,). \qquad (10)$$

It also assumed that any informatics system (i.e. computer) surrounded by a thermal bath that emits thermal noise at a temperature T_n. To calculate a bound on a computing power of a physical device Turing's model (1936) used. In Turing model erasing one bit and registering it again is an example of a logical operation. In our case the bits rate of the file is the maximum computing power. The higher the bit rate, the lower the temperature of the file as the bit energy reduced. Since the temperature of the file must be kept above the temperature of the noise T_n, the frequency has an upper limit. From Eq. (10) we conclude that $f \leq P/(k \ln 2\, T)$ where T should be about 10 times higher than the noise temperature. Therefore, the upper bound on computing power of any device is,

$$f \leq P/(10 k \ln 2\ T_n). \qquad (11)$$

The powers applied on any computing device, and its ambient temperature suffices to give a limit on its computing power. C.H. Bennett, in his review on the "Thermodynamics of computation" (Bennet, 2003), quotes from a Von Neumann talk that "a computer operating at temperature T must dissipate at least $k \ln 2\, T$ per elementary act of information". Later Bennett quotes that "in nature per nucleotide or amino acid the ratio is 20-100 $k \ln 2\, T$" with accordance to the present theory.

Example - the maximum information that an antenna can broadcast

Consider a point antenna broadcasting a compressed file. The bit energy at a receiving antenna having area A will be lower as the distance R between the transmitting antenna and the receiving antenna is higher. The temperature of the file at the receiver will be

$$T_r = T_t(A/4\pi R^2) = PA/[k \ln 2 f 4\pi R^2]. \tag{12}$$

The minimal size of both the transmitting and the receiving antenna is $\lambda = c/f$, where λ is the wavelength of the carrier radiation and c is the speed of light. If we assume that the bit energy at the receiver should be 10 times larger than kT (a conventional assumption for signal to noise requirements), we can calculate from Eq. (12) the maximum distance R that an antenna of power P and a frequency f can broadcast to a receiving antenna of area $A = \lambda^2$. As an example, consider a 50 W, 900 MHz radio transmitter. From Eq. (5) we find that the temperature of a broadcasted file via an antenna at the size of a wavelength is about $5\ 10^{15}$ K. We can cool the file from signal to noise considerations to about 3000 K (ten times the ambient temperature). We assume that the receiver has an antenna of area $A = \lambda^2$, and we obtain that the thermodynamic bound for the maximum distance $R \approx 100$ Km. This number may appear high, however, the receiving antenna is usually linear, and A is less than 1/100th of λ^2 and thus $R \approx 10$ Km.

Now we consider a large antenna of radius $R_t >> \lambda$. Because of the second law, it is not possible to detect a signal with a higher intensity (temperature) than that of the surface of the antenna. So we can replace T_r with T_t in Eq. (13), and calculate the entropy leaving the transmitting antenna. We can imagine the surface of the broadcasting antenna as a superposition of many small antennas of area λ^2 and substitute $A = \lambda^2$ in Eq. (13). The limit temperature of the broadcasted file can also calculated from Eq. (13). The maximum information transmission of an antenna over a time interval Δt is given by S/k and yields,

$$\Delta S \equiv P\Delta t/T_t = k \ln 2\ (f/\lambda^2)\ 4\pi R_t^2\ \Delta t \geq k\,\Delta I, \tag{13}$$

This is the Clausius inequality for a broadcasting antenna. This expression resembles the results of Bekenstein (1973) and Srednicki (1993) that a spherical emitter has entropy that is proportional to its area as a black hole or any imaginary sphere.

Summary

This paper deals with the energetic of a file broadcasted from one antenna to several antennas (a generalization of Shannon's theory). An analogy between information broadcasting from one antenna to several antennas to heat flow from a hot bath to a cold bath is drawn. We show that:

5. The Shannon information content *I* of a file is equivalent to the Boltzmann -*H* function.
6. The transmitted file energy is equivalent to heat.
7. A compressed file is a state of equilibrium.
8. The temperature of the antenna is proportional to the bit energy broadcasted from it or received by it.

Clausius inequality for an antenna and for informatics systems in general is calculated. In addition, a bound on the computing power of a physical device derived. The maximum information that can broadcasted from an antenna was calculated, and shown to be a function of its area.

References

Bekenstein, J.D. (1973). Black holes and entropy, Physical Review D, 7(8), 2333. doi. 10.1103/PhysRevD.7.2333

Bennet, C.H. (2003). 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*, 34(3), 501-510. doi. 10.1016/S1355-2198(03)00039-X

Brillouin, L. (1962). *Science and Information Theory*, Academic Press NY.

Huang, K. (1987). *Statistical Mechanics*, John Wiely: New York.

Huffman, D. (1952). A method for the construction minimum redundancy codes. *Proc. Institute of Radio Engineers*, 40, 1098-1101.

Kestin, J., (Ed). (1976). *The Second Law of Thermodynamics*, Dowden, Hutchinson and RossStroudsburg.

Landau, L.D. & Lifshits, E.M. (1980). *Statistical Physics*, 3rd Edition, Pergamon: New York.

Li, M. & Vitanyi, P.M.B. (1997). *An Introduction to Kolmogorov Complexity and its Applications*, Springer-Verlag, New York.

Plischke, M., & Bergersen B. (2006). *Equilibrium Statistical Physics*, 2nd Edition, World Scientific: London.

Shannon, C.E. (1949). *A Mathematical Theory of Communication*, University of Illinois Press, Evanston, Ill.

Srednicki, M. (1993). Entropy and arez, Physical Review Letters, 71(5), 666. doi. 10.1103/PhysRevLett.71.666

Turing, A.M. (1936). On computable numbers with an application to the entscheidungsproblem, Proc. London Math. Soc. 42(1), 230-265. doi. 10.1112/plms/s2-42.1.230

Ziv, J., & Lempel, A. (1977). A Universal algorithm for sequential data compression, *IEEE Transactions on Information Theory*, 23(3), 337-343.

7.
Economic Inequality as a Statistical Outcome

Introduction

Income distributive justice is a political subjective phrase related to an income distribution rather than to a scientific issue. Most people believe that income inequality should be as small as possible. Nevertheless, it is understood that a certain gap between the rich and the poor is necessary to stimulate competition between individuals. This competition is the invisible hand of any healthy economy. One may ask if there is an optimal inequality. This question is intriguing both from philosophical and practical points of view. Every society has a strong motivation to have a strong competitive economy on one hand and a social just on the other. These two factors are vital to the quality of life of the people. The governments regulate the net income distribution through taxation, and therefore it is of great importance to find if there is a theoretical criterion for an optimal wealth distribution. Moreover, history teaches us that a high income inequality may lead to political protests and even revolutions. In the words of philosopher Plutarch: "An imbalance between rich and poor is the oldest and most fatal ailment of all republics."

The income inequality research which probably started with Pareto golden rule at the end of the 19th century continues to these days (Ball, 2004). The contemporary physical approach to economy is based on statistical mechanics of ideal gas (Maxwell-Boltzmann), where as the distribution of incomeis compared to the distribution of energy-money among the particles-people (Dragulescu, & Petrova, 2000). However this approach that was applied by econophysicists (Ball, 2004) has not yield profound results. It was suggested previously (Kafri, 2014) that economy can be described more accurately as a network in which the money is a transient quantity exchanged between its nodes. In nature energy and transient energy, which is called heat, have different statistics. Energy has Maxwell Boltzmann statistics, and heat (i.e. photons, phonons and alike) obeys Planck statistics. In the network economy each node trades with the other nodes by transactions. Each transaction of money is represented by an integer number. The value of a number is the amount of money transferred and the sign of it is its direction. For example, if a transaction is +A it means that the node received A$, and if the transaction is –B its means that the node paid B$. In order to have an economy one need to add to this model a bank. The bank serves as the memory of the network in which all it transactions are registered. In addition of being a memory the bank is also aregulator. For example, the bank may decide that the balance of a given node, namely the sum of all itstransactions at any given time, cannot be negative. However, in order to have trade the bank should allow at least to some nodes to have a negative balance. In this case we say that the node receives a "credit" from the bank. When the bank issues the payment, it is registered as minus in the loaner-node's account. But, since the loaner pays with the loan to other nodes, and they deposit this money back in the bank, the total balance of the bank remains zero. We see that the bank is not really affected by crediting the nodes. In fact, the bank generated money from nothing by crediting the nodes, and therefore we may conclude that money is not subject to a conservation law.

At a first glance it seems that in this toy model there is no room for recessions, crisis, economic booms and alike. However, the total amount of money, which reflects the sum of all the transactions between the nodes, is not conserved, and therefore it may be changed due to psychological reasons like fear, optimism or even long period of prosperity that is expected to end.When the total amount of transactions reduces, there is an economic recession, and when it increases there isan economicgrowth.

The network economy model enables us to calculate the distribution of money between people exactly as it was done with the distribution of links among nodes (Kafri, 2014) and the distribution of energy among photons. This distribution, which is called Planck Benford's distribution (Kafri, 2016; Kafri, & Kafri, 2013), with accordance to the intuitive description of the network economy above, is also independent of the total amount of the money of the net or in the total amount of energy of the radiating object. That is to say; the ratio between the various income ranks is only a function of the number of the ranks. This is different from the normal distribution of energy between particles in ideal gas which varies with the total amount of energy of the gas.

The Planck-Benford distributionis basically a manipulation of Planck law (Planck, 1901) which describesthe equilibrium energy distribution in a finite number of radiation modes.The distribution of energy in the modes were calculated by maximizing the entropy (ME) of the radiating body (Kafri, 2016) namely,

$$\varepsilon(n) = \frac{\ln\left(1+\frac{1}{n}\right)}{\ln(N+1)} \quad (1)$$

Where N is the number of the modes, which are interpreted here as the chosen number of income ranks (which might be deciles, percentiles, tenth percentiles or any other positive integer), n is a serial number called here the rank number of the nodes where $n = 1,2, \ldots N$. Therefore, the people are the nodes in rank n and $\varepsilon(n)$ is their normalized wealth. If $N = 10$ then $\varepsilon(3)$ is the relative income of the third decile.

Eq. (1) was derived from Planck law (Planck, 1901) for photons, namely $n = 1/(\exp(\beta\varepsilon(n))-1)$; n is the occupation number which is the number of photons in a mode (mode is a radiation distinguishable state), β is a parameter related to temperature – which is determined by the total energy of the system, and ε is the energy-wealth of the photons. If we write Planck's equation differently, namely,

$\varepsilon(n) = \beta^{-1}\ln(1+\frac{1}{n})$, and considering that the normalization factor,

$\sum_{n=1}^{N}\varepsilon(n) = \beta^{-1}\ln(N+1)$, then $\varepsilon(n)/\sum_{n=1}^{N}\varepsilon(n)$ is the relative wealth distribution as expressed by Eq. (1). It is seen that smaller the rank number richer the people in it. Therefore when a number of people are divided randomly in N distinguishable groups, their wealth will decrease with the social rank n, according to Eq. (1).

Gini Index

Gini Index is the standard measure of income inequality for countries. It is a single number (ranges from 0 to 1) that is obtained from the relative net income distribution function $\varepsilon(n, N)$. If the income of x percent of the population is $\varepsilon(x)$, then one defines the Lagrange function as $L(x) = \int_0^x \varepsilon(x')dx'$.

Namely, $L(x)$ is the total income of all the population up to the fraction x. If the income is distributed equally, then $\varepsilon(x)$ is constant and $L(x) = x$.

Gini index is defined as $G = \int_0^1 [x - L(x)]\, dx$. If $\varepsilon(x)$ is constant then G is zero. Here we use a discrete version of the Gini index. We divide the population to 10 deciles according to the decreasing n, namely according to increasing income. We designate the fraction of the net income of the n decile by $\varepsilon(n)$ and the discrete Gini index is defined as

$$G = \sum_{i=1}^{10}\sum_{n=1}^{i}\left[\frac{n}{10} - \varepsilon(11-n)\right] \quad (2)$$

$L(11-n)$ is the discrete Lorentz curve, namely the fraction of the net income of all the deciles up to the $11-n$ decile, namely,

$L(i) = \sum_{n=1}^{i}\varepsilon(11-n)$ because ε is normalized $L(1) = 1$.

Now we calculate the Gini index for the Planck Benford's distribution of wealth in 10 ranks. Each rank represents a decile of the population having similar income.

In Fig. 1 we see the result of the substitution of Eq. (1) in Eq. (2)

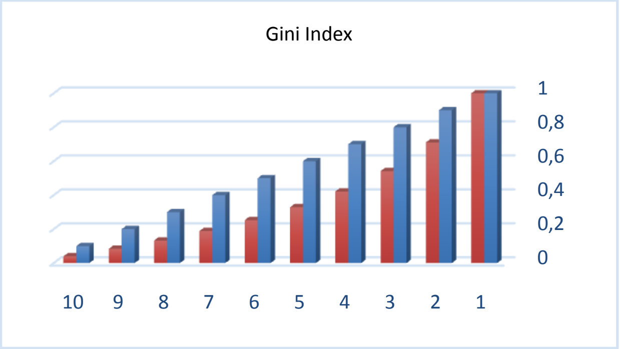


Figure 1. *The blue bars are the accumulated income of the deciles for the case where each decile has the same income. The orang bars are the accumulated income of the deciles with the distribution of Planck Benford. The sum of the differences between the blue bars and the orang bars is the Gini Index.*

This calculation yields G=0.327. It is quiet surprising that the average Gini index of the 35 countries of the OECD in 2012 is almost identical to that obtained here theoretically for network economy in equilibrium, namely G=0.32. Moreover, it is counterintuitive to think that in the free world the highly regulated income inequality will be similar to that of energy inequality among photons. The reason for the surprise is the influence of the governments on Gini index by taxation in order to increase equality and decrease Gini index. Most countries in the world also compensate poor people by supplementary income in addition to taxation. Yet the Gini index is almost identical.

The ratio between the incomes of the upper decile and the lowest decile

From Eq. 1 we calculated table 1 that present the relative wealth of the deciles. The richest decile $n = 1$ has 0.289 of the total wealth of the group, and the ratio between the highest income decile and the poorest, according to table 1, is 7.25. The average of the OECD for this ratio is 9.6, which is 32% higher than that of equilibrium countries. This point will be discussed later.

Table 1. *The relative income of deciles of ME society where the average of a decile is 0.1. The numbers calculated from Eq.(1). The left column is* n *and the right column is* $\varepsilon(n)$.

1.	0.289
2.	0.169
3.	0.120
4.	0.093
5.	0.076
6.	0.064
7.	0.056
8.	0.049
9.	0.044
10.	0.040

The Poverty

While Gini index and the ratio between deciles can be easily understood in terms of equilibrium society, poverty is harder to define. In the USA the poverty is defined as the inability to buy a certain amount of goods and services per unit time (i.e. a month). However, most countries define poverty as a relative quantity. In Europe a person is defined poor if his income is lower than 50% of the median income. The equilibrium network economy model cannot suggest the percentage of poor for the American absolute definition of poverty; however it can for the relative definition.

In Table 1 we see the equilibrium distribution of the wealth among the people according to their deciles. The median income which is given for a decile between the fifth and the sixth deciles is about 7 % of the total of the 10th deciles. Half of this amount is 3.5%. Therefore, according to this definition, in country in equilibrium about 9% are poor. Indeed in the OECD countries the average percentage of poor is about this number. One should remember that the calculation of poverty as done by the countries' institutions is not so simple as the calculation is done per capita while the income is calculated per family, therefore the number of children might change the numbers. Nevertheless, the equilibrium figures are with very good agreement with OECD economies (Murtin, & d'Ercole, 2015).

The wealth of the rich as compared to the average

Economists usually express the income of upper deciles, percentile and tenth percentile in terms of the average income. To calculate the average income in the ME distribution we have to find $\bar{n}$ in which the sum of all the incomes below it is equal to the sum of the incomes above it, namely,

$$\sum_{n=1}^{\bar{n}} \ln\left(1+\frac{1}{n}\right) = \sum_{n=\bar{n}}^{N} \ln\left(1+\frac{1}{n}\right) \qquad (3)$$

Which yields that; $2\ln(\bar{n}+1) = \ln(N+1)$ or

$$\bar{n} = \sqrt{N + 1} - 1 \quad (4)$$

Using Eq. (4) we can calculate the ratio R of the income of the richest and the average income.

$$R_N = \ln 2 / \ln(1 + 1/\bar{n}) \quad (5)$$

It worth noting that R is a function of N. The higher is N, the higher the gap between the rich and the average. From Eqs. 5 and 4 we calculated table 2 which is the ratio between the upper fractions to the average. The left column is N and the right is R.

Table 2. *The ratio between the upper fractions to the average.The left column is N and the right is the wealth of richest fraction as compared to the average. As N increases the ratio increases.*

10	2
100	7
1000	22
10000	69
100000	219
1000000	693

From Eq. (5) we can calculate the equilibrium net income of the richest. For example, if the average yearly income of a person is 30K$, we see from table 2 that for deciles in which $N = 10$, the ratio between the upper decile and the average is 2, therefore the upper decile will make 60K$. Similarly, the upper percentile will make $7 \times 30 =$ 210K$ and the upper tenth percentile annual income is $22 \times 30 =$ 660 K$.

CEO compensation

Eqs. (4) and (5) enable us to calculate the compensation of the CEO in terms of the average salary in his company and as a function of the number of employees N of the company. For example, Walmart has 2.2 million employees. In terms of an equilibrium company, its CEO should earn 1034 average salaries. Indeed, in reality Walmart's CEO makes a very similar number, namely 1028 average salaries (Link). In 23 companies of Fortune 100 the CEO compensation follows closely this formula. To mention few: Walmart, Macdonald's, Apple, Morgan Stanley, etc. Only 5 companies pay more than 2 times the equilibrium value, and in 5 companies the CEO makes less than 0.1 of this value. For example: W. Buffett salary is 0 in this scale. The average of the Fortune 100 companies is 0.87 as compared to 1 if all the companies would pay according to Planck Benford'slaw. Namely, on the average, for various reasons, there is a small tendency to pay a little less than the equilibrium salary for CEO's, the reasons probably are similar in their nature to that causing Mr. Buffett basically not to draw salary.

Pareto Law

Economists also calculate the distribution of wealth in term of the relative part of the total wealth held by the richest. The calculations in equilibrium society are done here for percentile using the equation $P = \ln 2/\ln 101$ for the first percentile, for the first ten percentiles $P = \ln 11/\ln 101,$ and for the first 20%, $P = \ln 21/\ln 101,$ of the wealth. This formula yields that the upper percentile has 15% of the total wealth, the upper decile has 52%, and the upper fifth has 66% of the wealth. In the OECD countries the average of the upper percentile has 18% of the wealth and the upper decile has 50% of the wealth (Murtin, & d'Ercole, 2015). 20% of the population in ME society has only about 66% of the wealth, a bit more justice for the poor than in the famous 80:20 Pareto law. According to this formalism 60% of the poorer population have 19.5% of the total wealth as compared to 13% in OECD.

Discussion

It is surprising that this oversimplified toy model yields such sound results. Yet, we have to point outthe limitations of this model. In this model there is only one bank and one country. In reality there are several banks and several countries trading between themselves. Moreover, the central bank takes no interest or any other fee for the loans. Yet it seems as if the single country inequality values are not affected by international trade or by the plurality of the banks or the charges of the bank. The second limitation is the differences between photons' energy and human wealth. With analogy to blackbody radiation in which all the photons in a given radiation modes have the same energy, the basic assumption of this model is that all the people in the same income rank earn exactly the same amount of money. The higher the ranknumber, the poorer the people (for large n the wealth ε is proportional to $1/n$ which is Zipf's law (Zipf, 1949; Gabarix, 1999). Therefore, when we divide the population to percentiles instead of deciles, we add more wealth scales of poor people that were not previously counted. For photons, the size of the blackbody determines the photons' minimum energy; similarly for people, the minimum money required to keep one alive determines the minimum wealth. This amount is lower than that of the formal definition of poverty in the OECD. This explains the differences of the ratio between the upper decile to the tenth decile, 9.6 in the OECD as compared to 7.25 of the present model, as some of the people that are poorer than the 10th decile of the model appears in the OECD statistics but not here. On the other hand, if we divide the people's wealth to percentiles instead of deciles, we count many poor people that are below the poverty that exists in the OECD. This limitation of "empty" percentiles of high rank number does not exist when calculating the CEO's compensation of companies in which the rank's number is low. The reason for it is that here we calculated the top salary in comparisonto the average salary which is substantially higher that the median salary. Generally, in the zone that $n \ll N$ the ratios of wealth will behave according to this model.

The same statistics was previously shown (Kafri, 2016) to be effective for voting. The distribution of the parliament seats among the 10 parties in Israel in the elections of 2015 is similar to that obtained by Eq.(1). In fact, the Gini index of inequality of seats among parties in the Israeli parliament, when is calculated according to Eq. (2) is 0.324. If so, one may ask whetherwe behave as a microcanonical ensemble after all. If we accept the assumption that the only physical law that causes irreversible changes in the universe is the second law, than the answer is that in equilibrium, maximum entropy distribution will be reached, and its probability should apply to economy which is a part of nature. As physicist Josiah Gibbs said (Kafri, & Kafri, 2013): "the whole is simpler than the sum of its parts".

Summary

In this note we calculate the thermodynamic equilibrium distribution of the wealth among people according to their income rank. We use a toy model economy of people randomly exchanging money between them selves. We make an analogy between this network economy and Planck's statistics in which the people/nodes are the photons, their energy is their wealth, and the social ranks are the radiation modes. We calculate for this distribution the indexes used by economists to describe the relative inequalities of income in countries and in companies. Namely, Gini Index, the ratio between highest income decile and the lowest income decile, relative poverty and the relative income compared to the average of the upper percentile and tenth percentile and the wealth held by the richest. We applied this formulation to calculate the executive compensation as a function of the number of employees and the average salary paid by the companies. The results fit well the inequities of wealth both in the OECD countries and in the Fortune 100 companies.

References

Ball, P. (2004). The physical modeling of human social systems - A Review, *ComPlexUs*, 1, 190-206. doi. 10.1159/000082449

Dragulescu, A. & Yakovenko, V.M. (2000). Statistical Mechanics of Money, *European Physical Journal*, 17, 723-729. doi. 10.1007/s100510070114

Gabarix, X. (1999). Zipf's law for cities: An Explanation, *The Quarterly Journal of Economics*. 114(3), 739-767. doi. 10.1162/003355399556133

Kafri, O., & Kafri, H. (2013) *Entropy - God's dice game*. CreateSpace, pp. 154-157.

Kafri, O. (2014). Money, information and heat in networks dynamics, *Mathematical Finance Letters* 4, 1-11.

Kafri, O. (2016). A novel approach to probability, *Advances in Pure Mathematics*, 6, 201-211. doi. 10.4236/apm.2016.64017

Ksenzhek, O., & Petrova S. (2008) Inequality and economic efficiency of society through the prism of thermodynamics, Hungarian Electronic Journal Economics, Manuscript No. ECO-080111-A.

Murtin, F., & d'Ercole, M.M. (2015). Household wealth inequality across OECD countries: New OECD evidence, *OECD Statistic Brief*, No.21.

Planck, M. (1901) On the law of distribution of energy in the normal spectrum, *Annalen der Physik*, 4, 553-562.

Zipf, G.K. (1949). *Human Behavior and the Principle of Least-Effort*, Addison-Wesley.

8.
A Comment on Nonextensive Statistical Mechanics

An Introduction to

There are two typical distributions observed in the macroscopic world. The first one is the bell-like function, which is a result of the canonic distribution, and the second one is the long tail, which is a result of a power law distribution. While some statistical quantities are bell-like (the human height etc.), many others, like the human wealth etc. have a long tail distribution. The long tail distribution is as common in nature as the bell-like distribution.

Apparently, many believe that the long tail distribution cannot be obtained from equilibrium thermodynamics. The reason for this misconception is the way the canonic distribution is derived in some textbooks (Back, 2009), namely to define a function as followed:

$$f(p_i) = \sum_{i=1}^{W} p_i \ln p_i + \alpha \sum_{i=1}^{W} p_i + \beta \sum_{i=1}^{W} p_i E_i \tag{1}$$

The derivation $\frac{\partial f(p_i)}{\partial p_i} = 0$ yields the canonic distribution. Here p is the probability, E is the energy, α and β are Lagrange multipliers and W is the number of microstates.

Eq. (1) looks exact, as the first term on the RHS represents the (-) Gibbs entropy, the second term is equivalent to the total number of particles, and the third term is the total amount of energy of the system. At a first glance, no approximations are made, and therefore, the only possible solution that maximizes the entropy for a given number of particles and a given amount of energy is the canonic distribution (Back, 2009). This implies that there is no way to obtain a power law distribution by maximizing Boltzmann-Gibbs entropy.

This is probably the reason for the enormous effort made to "generalize" the second law. The idea was to change the concept of entropy in a way that Eq. (1) will yield a power law distribution. This is the justification for Tsallis entropy (Tsallis, 1988), Renyi entropy (Lenzi, 2000), and more… The "entropy" of the highest impact is Tsallis entropy, which received, since it was suggested in 1988, according to Google scholar, more than 1250 citations. This warm welcome by the "community" is surprising as Tsallis entropy is nonextensive, which means a system in disequilibrium. The physical explanation for the nonextensivity is long-range interactions, which also implies disequilibrium.

Therefore, accepting nonextensive entropy means giving up the most important concept of thermodynamics, namely the tendency of any system to reach equilibrium. In other words, nonextensivity means giving up the second law of thermodynamics altogether!

Hereafter, it is shown that the assumption that canonic distribution is the only solution that maximizes Boltzmann-Gibbs entropy under the constraints of Eq. (1) is erroneous.

Eq. (1) should be written,

$$f(p_i, p_j) = \sum_{j=1}^{W} p_j \ln p_j + \alpha \sum_{i=1}^{N} p_i + \beta \sum_{i=1}^{N} p_i E_i \cdot \qquad (2)$$

Namely, Gibbs entropy should be summed over all possible different configurations W of the ensemble (the microstates). However, the summation over the energies $p_i E_i$ should be done over the states *N*, as the distribution that we are looking for is the distribution of energy among states and not microstates (all the microstates have equal energy!). Usually, W and N are different numbers. An ensemble of N states and *P* particles where $P < N$, and no more than one particle is allowed in a state, has a number of configurations,

$$W(P, N) = \frac{N!}{(N-P)!P!} \qquad (3)$$

Applying Stirling formula and using Boltzmann entropy $S = \ln W$, we obtain that

$$S \cong -N\{p \ln p + (1-p)\ln(1-p)\}, \text{ where } p = \frac{P}{N}.$$

Or in Gibbs formalism,

$$S=-\sum_{j=1}^{W}p_j\ln p_j\cong-\sum_{i=1}^{N}\{p_i\ln p_i+(1-p_i)\ln(1-p_i)\}\tag{4}$$

In the approximation $p<<1$, $(1-p)\ln(1-p)$ vanishes and the expression $-\sum_{i=1}^{N}p_i\ln p_i$ is entropy.

In this case Eq. (2) becomes, $f(p_i)\cong\sum_{i=1}^{N}p_i\ln p_i+\alpha\sum_{i=1}^{N}p_i+\beta\sum_{i=1}^{N}p_iE_i$ and yields the canonic distribution.

The conclusion up to this point is that the canonic distribution is not a law of nature, and it exists only at low occupation number systems.

Since, Eq. (1) is not always true, the legitimate way to look for other distributions is to calculate the number of microstates and their probabilities rather than changing the expression of the entropy.

Hereafter, it is shown that in fact, a power law distribution and its appropriate statistics exist in physics for over a century.

In the general case (neglecting degeneracy), we have to count all the configurations of P particles in N states for any value of n (here we replace the symbol p by n as we allow $\frac{P}{N}>1$). We follow the footsteps of Planck's seminal work from Planck (1901), namely,

$$W(P,N)=\frac{(N+P-1)!}{(N-1)!P!}\tag{5}$$

We apply again the Stirling formula as was done by Planck and obtain that $S\cong N\{(n+1)\ln(n+1)-n\ln n\}$, or in Gibbs formalism,

$$S=-\sum_{j=1}^{W}p_j\ln p_j\cong-\sum_{i=1}^{N}\{n_i\ln n_i-(n_i+1)\ln(n_i+1)\}\tag{6}$$

(Some may recall that this is Planck's derivation). If $n<<1$, we obtain again that the entropy is $-\sum_{i-1}^{N}p_i\ln p_i$, and therefore the canonic energy distribution is obtained as a private case. Since n is now interpreted as a number and not a probability we omit the second term in Eq. (2). By substituting the entropy of Eq. (6) in Eq. (2)

$$f(n_i)\cong\sum_{i=1}^{N}\{n_i\ln n_i-(n_i+1)\ln(n_i+1)\}+\beta\sum_{i=1}^{N}n_iE_i\tag{7}$$

we obtain the Planck equation namely,

$$n_i = \frac{1}{e^{\beta E_i} - 1} \quad (8)$$

Similarly, substituting the entropy calculated from the number of microstates of Eq. (3) in Eq. (7) yields the Fermi-Dirac distribution.

We designate $\Phi_i = \beta E_i$ and we plot $\ln n_i$ versus $\ln \Phi_i$ and we see that when $n > 1$, Planck equation yields a power law distribution with a slope -1.

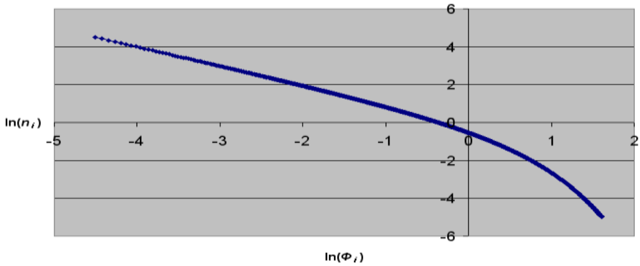


Figure1. *A log-log plot of the occupation number versus the relative energy.*

In Fig.1 it is seen that when the number of particles is higher than the number of states (high occupation numbers), a power low distribution is obtained, and at low occupation numbers the canonic distribution is obtained. In the classic Rayleigh-Jeans approximation the distribution of photons in a radiation mode is a long-tail distribution. In fact, the same statistics was used recently to derive Benford's law and the wealth distribution (Kafri, & Kafri, 2013; Kafri, 2009, 2016).

References

Back, C. (2009). Generalize information and entropy in physics, Contemporary Physics, 50(4), 495-510. doi. 10.1080/00107510902823517

Kafri, O., & Kafri, H. (2013). *Entropy - God's Dice Game*. CreateSpace, pp. 154-157. [Retrieved from].

Kafri, O. (2009). Entropy principle in direct derivation of Benford's law, Unpublished Paper, [Retrieved from].

Kafri, O. (2016). Sociological inequality and the second law, *Journal of Economic and Social Thought,* 3(4), xxx-xxx.

Lenzi, E.K., Mendes, R.S., & Silva, L.R. (2000). Statistical mechanics based on Renyi entropy, Phisica A, 280(3-4), 337-345. doi. 10.1016/S0378-4371(00)00007-8

Planck, M. (1901). On the law of distribution of energy in the normal spectrum, *Annalen der Physik*, 4, 553.

Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs entropy, Journal of Statistical Physics, 52(1), 479-487. doi. 10.1007/BF01016429

9.
Sociological Inequality and the Second Law

An Introduction to

It seems that nature dislike equality. In many cases distributions are uneven, a few have a lot and many have to be satisfied with little. This phenomenon was observed in many sociological systems and has many names. In economy it is called Pareto law (Bak, 1996; Newman, 2006), in Sociology it is called Zipf law (Troll, & Graben, 1998; Gunther, 1996) and in statistics it is called Benford law (Benford, 1938; Hill, 1996). These distributions differ from the canonic (exponential) distribution by a relatively moderate decay (a power-law decay) of the probabilities of the extremes that enables a finite chance to become very rich. Here it is shown that the power law distributions are a result of standard probabilistic arguments that are needed to solve the statistical problem of how to distribute P particles in N boxes. Intuitively one tends to conclude that P particle will be distributed evenly among N boxes, since the chance of any particle to be in any box is equal, namely, $\frac{1}{N}$. However, this is an incorrect conclusion, because the odds that each box will score the same amount of particle are very small. Usually there are some lucky boxes and many more unlucky ones. The distribution function of particles in boxes should maximize the entropy. This is because in nature, fairness does not mean an equal number of particles to all boxes N, but an equal probability to all the microstates (configurations) Ω. The equal probability of all the microstates is the second law of thermodynamics, which, exactly for this reason, causes heat to flow from a hot place to a cold place.

Calculating the distribution of P particles in N boxes with an equal chance to any configuration is not simple, as the number of the configurations $\Omega(P, N)$ is a function of both P and N namely,

$$\Omega(N,P) = \frac{(N+P-1)!}{(N-1)!P!} . \quad (1)$$

The derivation of the distribution function to Eq.(1) is not new. Planck published it in 1901 in his famous paper in which he deduced that the energy in the radiation mode is quantized (Planck, 1901; Kafri, & Kafri, 2013). Here the Planck's calculation is followed with the modifications needed to fit our, somewhat simpler, problem. Planck first expressed the entropy, namely $S = k_B \ln \Omega$ (k_B is the Boltzmann constant), as a function of the number of modes N and the number of light quanta P in a mode $n = \frac{P}{N}$. Using Stirling formula, he obtained that $S = k_B N\{(1+n)\ln(1+n) - n\ln n\}$. Then he used the Clausius inequality in equilibrium (Kestin, 1976) to calculate the temperature T, from the expression, $\delta S = \frac{\delta Q}{T} = N\frac{\delta q}{T}$, where Q is the energy of all the radiation modes and q is the energy of a single radiation mode. Therefore, the temperature is $T = N\frac{\partial q}{\partial S}$. Then, Planck made his assumption that $q = nh\nu$, namely $T = Nh\nu\frac{\partial n}{\partial S}$. Therefore, $\frac{\partial S}{\partial n} = k_B N \ln(\frac{n+1}{n}) = N\frac{h\nu}{T}$, this is the famous Planck equation, namely, the number of quanta in a radiation mode is, $n = \frac{1}{e^{\frac{h\nu}{k_B T}} - 1}$. The calculation of Planck is comprised of three steps. First he expressed the entropy S by the average number of quanta n in a box and the number of boxes (radiation modes) N. Next, he used the Clausius equality to calculate the temperature. The equality sign in Clausius inequality expresses the assumption of equilibrium in which all the configurations have the same probability. Then Planck added a new law that was verified by the data of the blackbody radiation that the energy of the quant is proportional to the frequency. This law is responsible for the observation that in the higher frequencies n is lower.

In our problem we do not have energies or frequencies. We just have particles and boxes. Therefore, we will write the dimensionless entropy, namely the Shannon information as a function of n and N, and obtain that $I = N\{(1+n)\ln(1+n) - n\ln n\}$. Parallel to Planck, we calculate the dimensionless temperature Θ according to $\Theta = \frac{\partial P}{\partial I} = N\phi(n)\frac{\partial n}{\partial I}$. Here we replace the total energy Q by P and q by $n\phi(n)$, where $\phi(n)$ is a distribution function that tells us the number of boxes having n particles. $\phi(n)$ is the analogue of Planck's $h\nu$. Changing the frequency enabled Planck to change the number of the particles in a mode at a constant temperature. Here we change the probability of a box with n particles at a constant temperature. The sociologic temperature $N\phi(n)\frac{\partial n}{\partial I} = \Theta$ is equal, in equilibrium, in all the boxes. Since, $\frac{\partial I}{\partial n} = N\ln(\frac{1+n}{n}) = \frac{N\phi(n)}{\Theta}$ one obtains that $\phi(n) = \Theta \ln\frac{n+1}{n}$.

	1	2	3	4	5	6	7	8	1 Average	2 Theoretical
A	55%	39%	47%	64%	46%	56%	65%	47%	52%	50%
B	32%	38%	31%	20%	37%	30%	19%	33%	30%	29%
C	13%	23%	22%	17%	17%	15%	16%	19%	18%	21%

This is the analogue of the Planck's equation, namely $n = \frac{1}{e^{\frac{\phi(n)}{\Theta}} - 1}$.

When P is large as in many statistical systems, we are interested in the normalized distribution. Since $\sum_{n=1}^{N}\phi(n) = \Theta\ln(N+1)$ we obtain that the normalized distribution function is,

$$\rho(n) = \frac{\ln(1+\frac{1}{n})}{\ln(N+1)} \qquad (2)$$

This is the main result of this paper. This result can be applied to any natural random distribution of inert particles in N boxes[1].

To check the validity of this distribution we start with Benford's law. Benford's law was found experimentally by Newcomb in the 19th century,

[1] The Plank derivation can be obtained using a more standard way namely, the Lagrange multipliers. In this method we write a function, $f(n) = \ln\Omega + \beta(P - \sum n\phi(n))$. The first term is the Shannon information and the second term is the conservation of particles. We substitute $\frac{\partial f(n)}{\partial n} = 0$ to find that, $\beta\phi(n) = \ln(\frac{n+1}{n})$. This is the maximum information solution that yields after normalization the Eq. (2) see Kafri (2009).

was extended later by Benford (1938) and explained on a statistical basis by Hill (1986, 1996). It says that in numerical data files, which were not generated by a randomizer, namely balance sheets, logarithmic tables, the stocks value etc, the distribution of the digits follows the equation $\rho(n) = \log(1+\frac{1}{n})$. For example, the frequency of the digit 1 is about 6.5 times higher than that of the digit 9. It is seen that if one substitute in Eq.(2) N=9 the Benford law is obtained. One can assume that the digit 1 is a box with n=1 particle and n=9 is a box with 9 particles. In fact, it is obvious that the equation valid for $n = C \times 1$, for the digit 1 and $n = C \times 9$ for the digit 9, where C is any number bigger than one.

Another way, intriguing even more, to check the informatics Planck distribution of Eq. (2) is to compare its results to polls statistics. In polls there are usually N choices and P voters that suppose to select their preferred choice. Usually each voter can select only one choice. A poll is not necessarily a statistical system. An example for a non-statistical poll is a poll with the three questions: 1. Do you prefer to be poor? 2. Do you prefer to be young, healthy and rich? 3. Do you prefer to be old and sick? In this poll one expects that most people will vote 2 (at least for themselves). However, it is clear that nobody will make the effort to make this poll, as its result is predictable. However, in the Internet there are many examples of multi- choice votes with unpredictable answers. Here we study three choices polls that were done on the Internet by the Globes Newspaper (2008) (an Israeli economical daily news) on variety of subjects between 10 Feb. 2008 and 10 Apr. 2008, for eight consecutive weeks on various issues. The results are presented in Fig 1.

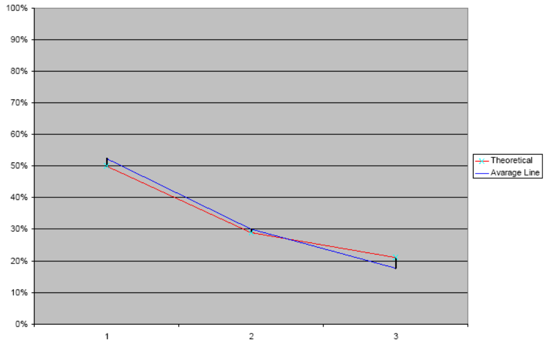


Figure 1. *The average distribution of votes of consecutive eight polls: Each poll has three choices selected by about 1500 voters. The blue line is the actual distribution. The red one is the theoretical calculation based on maximizing the Shannon information.*

It is seen that although the individual votes for the preferred choices A, B and C are quite different from the theoretical values, namely, 50%, 29% and 21% respectively. The average is with a good agreement with the experimental results. It is plausible that on the average, the polls reflect more uncertainty about the best choice than in an individual poll. Therefore, one expects that the average of the eight polls will be closer to equilibrium.

If we consider the number of particles in a box as an indicator of wealth, one may use Eq. (2) to calculate the theoretical particles wealth of boxes in equilibrium. For example, in a set of a million boxes the richest box will have a relative density of $\frac{\ln 2}{\ln 1000001} \cong 0.05$. Namely, 5% of the particles will be in one box. Similarly, the richest 10% will have $\frac{\ln 2}{\ln 11} \cong 0.29$. That means that 10% of the boxes will posses 29% of the particles. The richest half of the boxes will have about 63% of the wealth. The poorest 10% of the boxes will posses $\frac{\ln(1+\frac{1}{9})}{\ln 11} \cong 0.044$ of the particles, namely less than the richest single box. From the point of view of the boxes this is an unfair distribution. Nevertheless, from the point of view of the microstates (which are the configurations of boxes and particles) this is the just way to distribute the wealth.

It was shown previously that Planck formula yields a power law with slop 1 (Kafri, 2007). There are many publications that find power-law distributions with variety of slopes (Newman 2006). If we assume that the probability of the particles in a box is $\phi^{\alpha}(n)$, we can generalize this theory to a slop α power-law.

To conclude: the uneven distributions that are so common in life are partially an outcome of an unbiased distribution of configurations. This is the second law of thermodynamics as manifested by Boltzmann and Planck. Namely, the probability of all the microstates is equal. Not all the systems are in equilibrium, but systems in equilibrium are more stable. Thermal equilibrium is reached by the dynamics of the system. In blackbody, photons are emitted and absorbed constantly by the hot object, therefore one can expect to a thermal distribution. In economy the money exchanges hands all the time. The digits in numerical data are also changed by the number crunching operations. Nevertheless, the situation in polls is different. Voting in the Internet is a spontaneous non-interactive social activity; therefore, it is surprising that the solitary autonomic action of an individual yields a result of a statistical ensemble. A possible explanation is that our decision process mimics the behavior of a group, after all a human is a coalition of cells.

References

Bak, P. (1996). *How Nature Works: The Science of Self-Organized Criticality*, Springer-Verlag, New York.

Benford, F. (1938). The law of anomalous numbers, *Proceedings of the American Mathematical Society*, 78(4), 551-572.

Globes Newspaper, (2008). [Retrieved from]

Gunther, R., Levitin, L., Schapiro, B., & Wagner, P. (1996). Zipf's law and the effect of ranking on probability distributions, *International Journal of Theoretical Physics*, 35(2), 395-417. doi. 10.1007/BF02083823

Hill, T.P. (1986). The first digit phenomenon, American Scientist, 86(4), 358. doi. 10.1511/1998.4.358

Hill, T.P. (1996). A statistical derivation of the significant-digit law, Statistical Science, 10(4), 354-363.

Kafri, O., & Kafri, H. (2013). *Entropy - God's Dice Game*. CreateSpace, pp. 154-157. [Retrieved from].

Kafri, O. (2007). The second Law as a Cause of the Evolution, Unpublished Paper, [Retrieved from].

Kafri, O. (2009). Entropy principle in direct derivation of Benford's law, *Unpublished Paper*, [Retrieved from].

Kestin, J. (1976). *The Second Law of Thermodynamics,* Dowden, Hutchinson and RossStroudsburg.

Newman, M.E. (2006). Power-law, Pareto Distribution and Zipf's law, *Unpublished Paper*, [Retrieved from].

Planck, M. (1901). On the law of distribution of energy in the normal spectrum, *Annalen der Physik*, 4, 553-562.

Troll, G., & Graben, P.B. (1998). Zipf's law is not a consequence of the central limit theorem, *Physical Review*, 57(2), 1347-1355. doi. 10.1103/PhysRevE.57.1347

10.
The Distributions in Nature and Entropy Principle

Introduction

According There are two common distributions in life: The first one is the "bell-like" distribution, which is found in the distribution of IQ, human heights, human age at death etc. This "almost universal" distribution was introduced for the first time by Moivre in the 18th century and explored by Laplace and Gauss around 1800.

As opposed to the bell curve distribution, many quantities are distributed unevenly (Bak, 1996). For example, the probability to live in a big city is higher than the probability to live in a small village. Similarly, the probability to be poor is higher than the probability to be rich. Although intuitively it is logical for cities' population and wealth to have a bell curve distribution, it is not so. Their distributions are uneven and are characterized by a long tail to the right, in which few have a lot and many have quite a little. These distributions were observed by Pareto, Zipf, Newcomb and Benford about a century later and received their name accordingly: Zipf law (Zipf, 1949; Miller, & Newman, 1958), Pareto's rule (Pareto 1897; Jurgan, 1951), and Benford's law (Newcombs, 1881; Benford, 1938).

The first to discover it was Pareto. In 1896 he observed that the ownership of lands in Italy is distributed among the population in the ratio of around 20:80, namely, about 20% of the population own about 80% of the land. From his observations of other countries as well, he concluded that this ratio is general. Mussolini embraced the Italian Marquis Pareto because he believed that the Pareto's rule proves nature's preference of the fittest. Zipf - a Harvard professor of linguistics - found out that the ratio between the first most frequent word and the second one, in any text in many languages, is two. Similarly, the ratio between the second most frequent word and the fourth one is also two, etc. He claimed that the shortest and most "efficient" words appear more frequently (Zipf, 1949).

Zipf believed in the evolutionary philosophy, i.e. the most "useful" and "efficient" words are the winners, in the spirit of "the survival of the fittest". On the other hand, many people and political movements believe that Pareto's rule is unfair and the wealth should be shared more equally, namely, as in the bell curve distribution. The discovery of Newcomb about the uneven frequency of digits in logarithmic table in 1881 (Newcomb, 1881), (the higher the value of a digit, the lower its frequency) raises some doubts as for the real reason for the uneven distributions. Later, in 1938, Benford confirmed Newcomb's uneven distribution of digits in a wide range of numerical data (Benford, 1938). He attempted, unsuccessfully, to present a formal proof to Newcomb's equation, see Eq. (12). Since then, this distribution was found also in prime numbers (Cohen, 1984), physical constants, Fibonacci numbers and many more (Kossovsky, 2012).

In this paper it is argued that the "bell-like" distribution and the long tail distribution are the boundaries of the same probability distribution. This probability function is obtained by a fair and unbiased random distribution of particles in boxes.

We consider a set of N boxes scoring P particles; it is assumed that all the boxes have an equal probability to score a particle, namely, the probability of a box to score a particle is $= 1/N$. Therefore, the probability to score n particles is $q_n = (\frac{1}{N})^n$. It is clear that $q_n < q$. This is the basic reason why the rich are fewer than the poor. In the case of $P \ll N$, where a multiple score is negligible, the "bell-like" distribution is obtained; and in the case of $P \gg N$, a long tail distribution is obtained.

How P particles are distributed in N boxes?

The answer to it is not new: the particles are distributed in a way that maximizes the entropy (Planck, 1901).

According to Boltzmann, entropy is proportional to the maximum possible number of the different configurations (microstates) of a set. Namely,

$$S = \ln \Omega \quad (1)$$

(we take here the Boltzmann constant $k_B \equiv 1$). A microstate is one possible distinguishable configuration of a set of boxes and particles. Boltzmann entropy is obtained from the Gibbs-Shannon entropy by assuming that all the microstates have an equal probability. The Gibbs-Shannon entropy is given by:

$$S = -\sum_{j=1}^{\Omega} p_j \ln p_j \quad (2)$$

where p_j is the probability of the microstate j and Ω is the number of microstates to be maximized. If all the microstates have an equal probability, namely, $p_j = 1/\Omega$, Boltzmann entropy $\ln\Omega$ is obtained.

Therefore, the distribution of particles that maximizes Boltzmann entropy means an equal probability to any configuration as well as an equal probability to any particle to be in any box.

The number of microstates (different configurations) of P particles in N states is given by the Planck expression (Planck, 1901) namely,

$$\Omega(P,N) = \frac{(N+P-1)!}{p!(N-1)!} \quad (3)$$

To visualize the problem we start with a numerical example; namely, calculating the distribution of 3 particles in 3 boxes that maximizes entropy. According to Eq. (3) the number of microstates $\Omega(3,3) = 10$ as follows:

|300| 030| 003| 210| 201| 120| 021| 102| 012| and |111|.

We see that although each box has an equal chance to score 1, 2, or 3 particles, the boxes with 1 particle appear 9 times, those with 2 particles appear 6 times, and those with 3 particles appear 3 times. The relative frequency of the boxes with one particle in a set of three boxes is therefore $f(1)$=0.5; with two particles $f(2)$=0.333and with three particles $f(3)$=0.166.

To calculate the relative frequencies $f(n)$, we designate $n = P/N$, where n is the number of particles in a box, and apply the Stirling's formula

$\ln N! \cong N \ln N - N$. We obtain (Planck, 1901) from Eqs.(1) and (3) that,

$$S \cong N\{(1+n)\ln(1+n) - n\ln n\} \cong \sum_{n=1}^{N}\{(1+n)\ln(1+n) - n\ln n\} \quad (4)$$

Now we write the Lagrange equation,

$$F(n) \cong \sum_{n=1}^{N}\{(1+n)\ln(1+n) - n\ln n\} - \beta\{P - \sum_{n=1}^{N} n\phi(n)\} \quad (5)$$

The first term on the RHS is the entropy and the second term is theconstraint of the number of particles. Namely, $P = \sum_{n=1}^{N} n\phi(n)$ is the number of particles, $\phi(n)$ is the number of boxes that scored n particles and β is a

Lagrange multiplier. $\phi(n)$ can be interpreted as the probability of a box to have n particles. The normalized (n) , $f(n)$ is the relative frequency of the boxes that scored n particles. From $\frac{\partial(F(n)}{\partial n} = 0$ one obtains,

$$\phi(n) = \beta^{-1} \ln(1 + \frac{1}{n}) \tag{6}$$

Eq. (6) is the analogue of Planck equation (Kafri, 2007, 2009, 2016), namely,

$$n = 1/[e^{\beta\phi(n)} - 1] \tag{7}$$

Hereafter, we examine three cases:

In the first case we assume that n>> 1. Here one can expect to find a large number of particles (limited by P) in any of the boxes. For example, if we conduct a popularity poll between the N words among P authors, and there are many more authors than words, then the maximum entropy distribution of the votes between the words is shown to be the Zipf law.

In the second case we consider the intermediate zone where n is in the range of the number of the boxes. This case fits well to the distribution of ranks, namely, Pareto's rule and Benford's law.

In the third case we consider n<< 1, where the number of particles is negligible as compared to the number of boxes. This case fits well to the probability of guessing correctly the IQ of a person in a single guess based only on the knowledge of the average. This case yields the "bell-like" distribution.

Zipf law

Consider the case where $P \gg N$ where $\gg 1$. In this case $\beta\phi \ll 1$, therefore from Eq. (7) $\phi(n)$ can be approximated to,

$$n\phi(n) = \frac{1}{\beta} \tag{8}$$

Eq.(8) is the Zipf law. Namely, the ratio in the frequencies between n=1 (the most frequent word) and n=2 (the second most frequently word) is 2 which is identical to the ratio between $n = 2$ and $n = 4$ etc. This ratio is not a function of β as $\frac{\phi(1)}{\phi(2)} = \frac{\phi(2)}{\phi(4)} = \cdots \frac{\phi(n)}{\phi(2n)} = \cong 2$.

Pareto's rule

to calculate the relative frequency of Eq.(6), namely, $f(n)$ we have to divide $\phi(n)$ by the sum over all the M occupied boxes $M \leq N$, namely,

$$\sum_{i=1}^{M} \phi(n) = \frac{1}{\beta}(\ln\frac{2}{1} + \ln\frac{3}{2} + \cdots .. + \ln\frac{M+1}{M}) = \frac{1}{\beta}\ln\frac{M+1}{M} \tag{9}$$

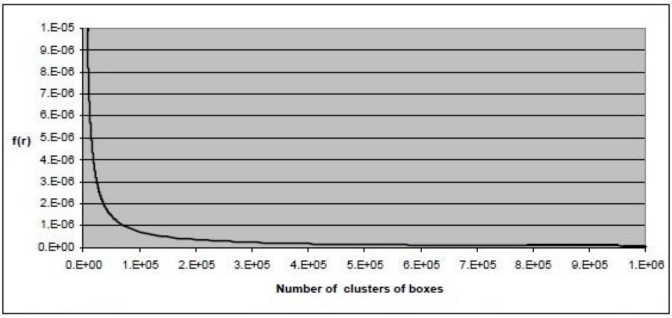
Therefore,

$$f(n) = \frac{\ln(1+\frac{1}{n})}{\ln(M+1)} \tag{10}$$

Like in the Zipf law, for integer *n*'s, the relative frequency is not a function of β .We define a rank $r \equiv nN/P$where $r = 1, 2, 3, \ldots. R$. By defining the ranks we combined the boxes into clusters of boxes such that each cluster will contain $r = 1, 2, 3, \ldots. R$ groups of P/N particles. Therefore $r = 10$ means 10 times more particles than *r*= 1. We can repeat the calculation of the frequency again but instead of using *n*, we will use *r*, and obtain;

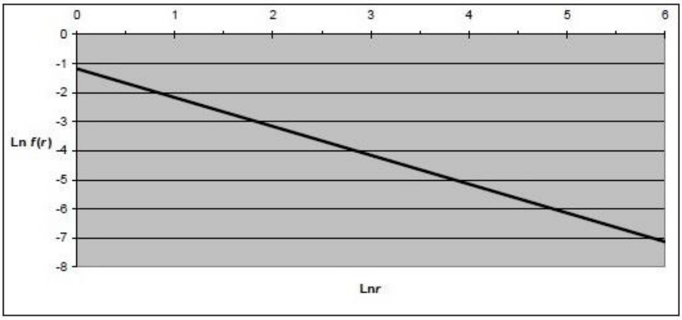
$$f(r) = \frac{\ln(1+\frac{1}{r})}{\ln(R+1)} \tag{11}$$

In Graph.1 The relative frequencies$f(r)$ for a set of R=10^6 clusters and *r*= 1,2,3,...., *R*according to Eq.(11) is plotted. A long tail distribution is demonstrated.



Graph 1. *A million clusters and their probabilities. The rank increases as its probability decreases.*

Eq.(11) "behaves" as a power law, this is so because a plot of the logarithm of the cluster *r* versus the logarithm of its probability yields a straight line as demonstrated for a million ranks.



Graph 3. *Log-Log plot of frequency versus the rank for R=million is a straight line.*

The Pareto's 20:80 rule of thumb was proved to be correct not only in wealth distribution but in many other phenomena as well. For example, it is believed that 20 percent of customers yields 80 percent of the revenue; 20 percent of the drivers cause 80 percent of the accidents; etc (Jurgan, 1951). In order to find the ratio obtained from Eq. (11) we divide the boxes into 10 ranks. Each rank contains 1, 2, 3,….,9, 10 equal groups of particles. We construct the table below from $f(r) = \frac{\ln(1+\frac{1}{r})}{\ln(11)}$

Table 1. *The relative frequencies of 10 ranks*

r	10	9	8	7	6	5	4	3	2	1
f(*r*)%	4	4.4	4.9	5.6	6.6	7.6	9.3	12	16.9	28.9

The total number of groups is $\sum_{r=1}^{10} r = 55$. However, the richest five ranks contain $\sum_{r=6}^{10} r = 40$ groups. Their total frequencies are $\sum_{r=6}^{10} f(r) = 25.5\%$, which means that about 73% of the packages are in the hands of about 25% of the boxes. This is a typical behavior of the Pareto's rule but with a small deviation from the empirical rule of thumb of 20:80, namely, a 25:75 rule.

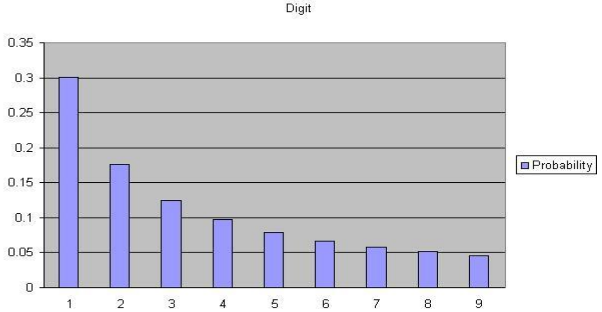
Benford's Law

Another application of Eq. (11) is Benford's law. Newcomb suggested Benford's law in 1881 from observations of the physical tear and wear of books containing logarithmic tables (Newcomb, 1881). Benford further explored the phenomenon in 1938, and empirically checked it for a wide range of numerical data. The main application of Benford's distribution is based on its existence in numerous random numerical files like financial data, street addresses, etc. Since one intuitively expects to obtain an even distribution of digits, as would be in the case of an unbiased lottery, some income tax authorities are looking at balance sheets for digit distributions in order to detect fraud detection. If the balance sheets don't fit to Benford's law, a further inspection is done (Nigrini, 1996).

In the derivation of Benford's law we assume that a digit is a box with *n* particles. This assumption is logical as a digit, unlike a word, has an absolute meaning as compared to other digits, exactly as the meaning of the number of particles in a box. There is a constraint though: the number of particles in a digit cannot exceed 9. The digit zero does not appear in Benford's law distribution of the first order. In Eq. (11) *r* may have any number. In digits, per definition, $r \leq 9$, therefore, it is legitimate to calculate the equilibrium distribution of the occupied boxes and to add as many empty boxes without affecting the distribution. In this case *R* is 9 and Eq. (11) yields the relative frequency,

$$f(r) = \frac{\ln\left(1+\frac{1}{r}\right)}{\ln(10)} = \log(1 + \frac{1}{r}) \qquad (12)$$

This is the Benford's law.



Graph 3. *Benford's law predicts a decreasing frequency of first digits, from 1through 9.*

"Bell-like" distribution

Zipf law, Pareto's rule and Benford's lawoccurs where the number of particles is larger than the number of boxes. Hereafter, the case where $P << N$ is considered.

In this case $n << 1$, we neglect the boxes that scored several particles, because, practically there are no such boxes. We want to find the probability distribution of *N* boxes to score one particle. In this limit, $e^{\beta\phi} \gg 1$ and Eq. (7) can be approximated to,

$$n_i = e^{-\beta\phi_i} \qquad (13)$$

Here n_i is the fraction of a particle in a box and the frequency $\phi_i = \phi\ (n_i)$ is the probability to find this fraction. The total number of particles *P* is given by the same expression that we used in the Lagrange equation (5) namely,

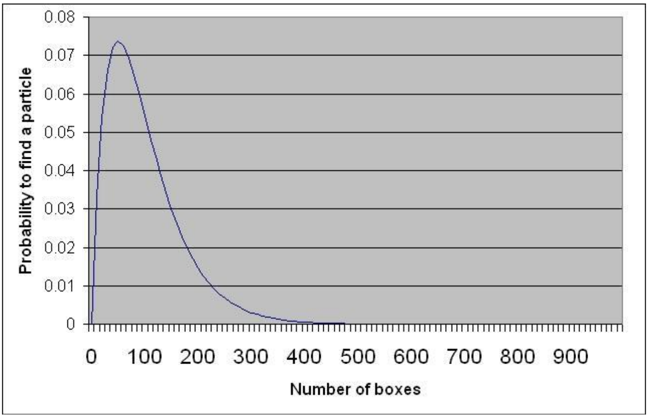
$$P = N\phi_i n_i = N\phi_i e^{-\beta\phi_i} \qquad (14)$$

in the limit $\beta \rightarrow 0$ one obtains that all the frequencies ϕ_i of the boxes are equal, namely $\phi_i = P/N$. This is an even distribution. The even distribution isthe intuitive distribution that one expects to find in a distribution of particles in boxes. This distribution causes us to believe that uneven distributions are counterintuitive.

In the case where β is finite

$$P = \sum_{i=1}^{N} \phi_i e^{-\beta\phi_i} = \sum_{i=1}^{N} P(\phi_i, \beta) \quad (15)$$

$\frac{P(\phi_i,\beta)}{P}$ is the relative probability to find a particle in a box. From Eq. (15) it is seen that $P(\phi_i, \beta)$ has two components, the first is the frequency ϕ_i of the fraction n_j of the particle and the second is the fraction of particles. As opposed to the case where $P>>N$, the frequency $\phi(n)$ itself is not the probability to find n particles but the probability to find a fraction of a particle. To find the probability of a single particle we have to multiply the frequency by the fraction of the particle namely $\phi_i n_i$. When the frequency increases the associate fraction of particles decreases exponentially with the frequency. The larger the β, the steeper is the decay. Since $P(\phi, \beta)$ is a linearly increasing function of ϕ_i multiplied by an exponentially decay function of ϕ_i, the distribution of particles in a box has a definite maximum.

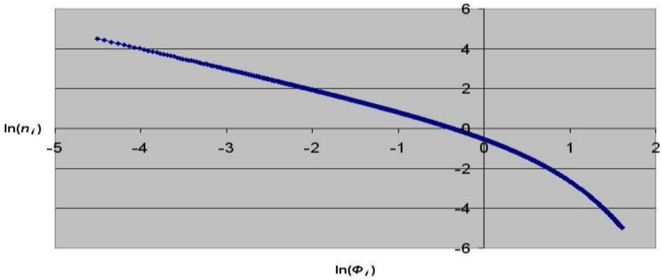


Graph 4. *The number of boxes and their probability to find a single particle for N=1000 and β= 1/50*

The maximum probability is obtained from $\frac{\partial P}{\partial \phi} = e^{-\beta\phi} - \beta\phi e^{-\beta\phi} = 0$ and is given by $\phi_{max} = 1/\beta$. In Graph (4) we see that the obtained curves is typical to the distributions of velocity of molecules, human age at death etc.

Discussion

The long tail distribution attracts a considerable attention because it is so ubiquitous [15]. Sometimes it is called a power law distribution and scale-free distribution. This is because a Log-Log presentation of the distribution yields a straight line as seen in Fig.2. When a power law fits are done, different slopes obtained for different statistics. For example, in Zipf law the ratio between the frequency of the 1[st] and the frequency of the 2[nd] is 2; in Pareto's rule and in Benford's law this ratio is about 1.7. Namely, in different regimes of P/N different "slopes" are obtained as is seen in Graph 5. Another notable point is that the normalized frequencies $f(n)f(n)$ for$P \gg N$ are not a function of β .This is with contradistinction to the case $P<<N$ in which the distribution is a function of β.



Graph 5. *A plot of* ln ϕ *versus* ln *n for high values of n a "power law" decay is obtained, however for low values of n an exponential decay is obtained.*

The Lagrange multiplier β has a meaning. In ther modynamic the temperature is related to it via $T \propto \frac{1}{\beta}$.We see that in the case of Zipf law the frequency multiplied by the number of particles is proportional to the temperature. In the case of $n \ll 1$ the temperature is proportional to the frequency in which the probability to find a particle is the highest. This is the main difference between the long tail distribution and the "bell-like" distribution. In the long tail the temperature means the average wealth of a box. In the bell curve the temperature means the maximum probability.

Summary

The distribution of P non-interacting particles in N boxes iscalculated for a fair system. Since there is no preference to any configuration of particles and boxes, the entropy principle can be applied. It is shown that when the number of the particles is negligible as compared to the number of boxes, the "bell-like" distribution (which prefers the average) is obtained. However, when the number of particles is higher than the number of boxes, a long tail distribution is obtained. The obtained long tail distribution yields correctly Zipf law, Pareto's rule and Benford's law.

The Pareto's rule usually is conceived as an evolutionary law. Namely, the 20% of the drivers that cause 80% of the accidents are the bad drivers.

Maybe the personality of these drivers is the reason for their excessive involvement in car accidents. Similarly, there might be good reasons for the fact that few people get rich and the majority remains poor. These kinds of questions cannot be answered by this kind of analysis. However, one should bear in mind that particles without personality, interactions or statistical bias are also distributed in the same way.

References

Bak, P. (1996). *How Nature Works: The Science of Self-Organized Criticality*, Springer-Verlag, New York.

Benford, F. (1938). The law of anomalous numbers, *Proceedings of the American Mathematical Society*, 78(4), 551-572.

Cohen, D., & Talbot K. (1984). Prime numbers and the first digit phenomenon, *Journal of Number Theory*, 18(3), 261-268. doi. 10.1016/0022-314X(84)90061-1

Jurgan, J.M. (1951). *Quality Control Handbook*, McGraw-Hill Book Company.

Kafri, O. (2007). The second law as a cause for the evolution, *Unpublished Paper*, [Retrieved from].

Kafri, O. (2009). Entropy principle in direct derivation of Benford's law, *Unpublished Paper*, [Retrieved from].

Kafri, O. (2016). Sociological inequality and the second law, *Journal of Economic and Social Thought,* 3(4), xxx-xxx.

Kossovsky, A.E. (2012). Towards a better understanding of the leading digits phenomena, *Unpublished Paper*, [Retrieved from].

Miller, G.A. & Newman E.B. (1958). Tests of a statistical explanation of the rank-frequency relation for words in written English, *American Journal of Psychology*, 71(1), 209-218. doi. 10.2307/1419208

Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers, *American Journal of Mathematics*, 4, 39-40.

Newman, M.E. (2006). Power-law, Pareto Distribution and Zipf's law, *Unpublished Paper*, [Retrieved from].

Nigrini, M. (1996). A taxpayer compliance application of Benford's law, Journal of the American Taxation Association, 18(1), 72-91.

Pareto, V. (1897). *Coursd' Economie Politique*. Droz, Geneva Switzerland, 1896, Rouge, Lausanne et Paris.

Planck, M. (1901). On the law of distribution of energy in the normal spectrum, *Annalen der Physik*, 4, 553-562.

Zipf, G.K. (1949). *Human Behavior and the Principle of Least-Effort*, Addison-Wesley.

Conclusion

In this collection of papers, it shown that the "order"which is generated around us by evolution and the results of our doing can be explained by the propensity of entropy to increase; namely, by the second law of thermodynamics. Entropy is conceived by many as a disorder, however, that is true only for spares systems. In a dense system, entropy is information thatis characterized by a long tail distribution.

In order to apply the second law to sociology and economy, we have to define a sociological net. With analogy to the internet, in which, in principle, every site can receive and broadcast information to any other site we can describe economic network as a group of bank accounts that each one of them canreceive or pay money to any other account.

The distribution of links between the sites is similar to that of money in the bank accounts. This long tail distribution,whichis obtained by maximizing the entropy of the net, is called Planck-Benford distribution. It is also shown that Planck-Benford distribution can predict polls distribution; Gini inequality Index in the OECD countries; the percentage of the relative poverty; the salaries of the CEO's relative to the average salaries and the number of employees.

Moreover, the Planck-Benford income distribution, being an equilibrium distribution (Max Entropy),can provide a standard tool for estimatingthe stability of the economy of a given country namely closer the income distribution of a country to Planck Benford distribution closer the economy to equilibrium.

ISBN: 978-605-2132-51-7 (e-Book)
KSP Books 2018

Oded Kafri

Dr. Kafri was a senior scientist and a group leader at the Nuclear Research Center-Negev in Israel. He has published about 150 scientific papers, among them two pioneering papers on "Visual Cryptography" and on "Moire deflectometry" and holds numerous patents. Dr. Kafri has received several international awards, among them the CeBit Highlights award in 1994. Dr. Kafri founded three high tech companies and wrote three books: The Physics of Moiré Metrology, Entropy – God's Dice Game, and Money: Physics and Distributive Justice" which is a description of the economy as physical phenomenon.

KSP BOOKS

e-ISBN
978-605-2132-51-7