Entropy **2005**, *7[1]* , 68-96

Entropy
ISSN 1099-4300
www.mdpi.org/entropy/

The meanings of entropy

Jean-Bernard Brissaud

Lab/UFR High Energy Physics, Physics Department, Faculty of Sciences, Rabat, Morocco.
email:jbbrissaud@softhome.net

Received: 19 November 2004 / Accepted: 14 February 2005 / Published: 14 February 2005

Abstract: Entropy is a basic physical quantity that led to various, and sometimes apparently conflicting interpretations. It has been successively assimilated to different concepts such as disorder and information. In this paper we're going to revisit these conceptions, and establish the three following results:
Entropy measures lack of information; it also measures information. These two conceptions are complementary.
Entropy measures freedom, and this allows a coherent interpretation of entropy formulas and of experimental facts.
To associate entropy and disorder implies defining order as absence of freedom. Disorder or agitation is shown to be more appropriately linked with temperature.

Keywords: Entropy; freedom; information; disorder.
MSC 2000 code: 94A17 (Measures of information, entropy)

"No one knows what entropy really is, so in a debate you will always have the advantage"

J. Von Neumann[1]

1 Introduction

Entropy, more than any other physical quantity, has led to various, and sometimes contradictory interpretations. Boltzmann assimilates it with disorder[2], Shannon with positive information[3], Brillouin with lack of information or ignorance[4], and other authors, although not numerous, to freedom[5]. Entropy is a fundamental quantity of modern physics[6][7], and appears in as diverse areas as biology, metaphysics or economy[8][9]. Hence great attention should be focused on the different interpretations of this concept. Entropy is the only physical quantity that always increases. It has such an importance that it can't stay dissociated from more familiar concepts.

In this paper,we will analyze the above interpretations and propose the following results:

(1) Entropy is appropriately associated with lack of information, uncertainty and indefiniteness. It is also appropriately associated with information. For an observer outside the studied physical system, entropy represents the lack of information about the state of the system. But for the system itself, entropy represents information, positively counted.

(2) Entropy measures freedom. This view provides a coherent interpretation of the various entropy formulas, and many experimental facts. A typical example is gas expansion: the freedom of position of the gas molecules increases with time.

(3) Entropy is inappropriately associated with disorder and even less with order. Nevertheless, disorder and agitation can be associated with temperature.

By "appropriately associated with a given concept", we mean an interpretation leading to correct predictions of the observed phenomena, and allowing a better understanding of the underlying equations. For instance, connecting entropy with lack of information is meaningful when studying the evolution of a gas in expansion; we have less and less information about the positions of the molecules. This description is also in agreement with Boltzmann entropy, since if there are more accessible microstates, there is less information about the specific microstate at a given time. Such a view is coherent with the main definitions of entropy, and agrees with the observed phenomena.

There are many definitions of entropy, and we will just consider the most famous ones. This paper is organized as follows:

In section 2, we'll adopt the quantum mechanical definition of entropy. We'll establish in this framework results (1) and (2), and use them to interpret the phenomenon of wave packet collapse, and the uncertainty principle. The fact that entropy represents the lack of information is today broadly accepted. Following Shannon's demonstration, we'll show that entropy can equally well represent the freedom of choice a quantum system or a message possesses, and that these two points of view are complementary, and not contradictory.

Section 3 is devoted to the study of statistical definitions of entropy discovered by Boltzmann and Jaynes, considered as generalizing Gibbs' work. We'll establish again the pertinence of results (1) and (2).

In section 4, we are going to study the link between entropy and disorder. Are freedom and disorder two sides of a same thing? Is order absence of freedom? The association of disorder with temperature, and the explanation of classical phenomena in terms of increasing freedom is the subject of the third part, devoted to classical thermodynamics. We'll establish that it is not entropy, but temperature, which measures disorder, and will invalidate the main arguments in favor of the analogy entropy/disorder. We'll then furnish a simple explanation of the third principle, based on the associations between entropy and freedom, temperature and disorder.

Last section is devoted to the conclusion.

2 Entropy in quantum mechanics

In this section we'll first review the classical theory of information and the meanings of the different quantities involved. We'll then consider the case of transmission of classical information with quantum states. The connection between entropy, information and freedom will be established. Lastly, we'll consider two quantum examples where assimilation of entropy to freedom is enlightening: the entropic uncertainty principle and the entropy of a black hole.

2.1 Classical theory of information

In this section, the notations and conventions used , and basic results of information theory are exposed.

Let $[p_i]_i$ be a probability distribution, which means that all p_i are positive or null, and $\sum_i p_i = 1$. The Shannon information of this distribution is defined as:

$$I[p_i]_i = \sum_i I(p_i) \qquad \text{with } I(x) = -x \log x$$

We have:

$$0 \leq I[p_i]_{i=1..n} \leq \log n$$

In this paper, log will always be base 2 logarithm, and information is measured in bits.

1 bit is the information given by the knowledge of one choice out of 2, or the freedom of making a choice out of 2. If we consider a person sending messages according to the distribution $[p_i]_i$, then $I[p_i]_i$ represents the average freedom she has in choosing which message to send next. If we consider a person receiving messages with the distribution $[p_i]_i$, then $I[p_i]_i$ represents the average information she gets in receiving a given message. $I[p_i]_i$ can be seen as the number of memory bits needed to represent this choice or this information.

$I[p_i]_i$ represents also the lack of information the receiver has before receiving the message. So we will have to be careful when we associate $I[p_i]_i$ with one of these concepts.

For the sender A, $I[p_i]_i$ represents the information she sends, or the freedom of choice she has. For the receiver B, $I[p_i]_i$ represents the lack of information she has before reception, or the information she has after reception.

However, in a transmission, the signal can be modified, and what is received is not what has been sent. We suppose now that a sender A emits messages according to the probability distribution $[p_i]_i$, and that a receiver B receives messages knowing the probability distributions $[q_j|p_i]_j$, which means that she knows the probability distribution of what she receives if she knows what A has sent.

We define the quantities:

$$a_{ij} = (q_j|p_i) \times p_i$$
$$q_j = \sum_i a_{ij}$$
$$(p_i|q_j) = a_{ij}/q_j$$
$$I(A) = I[p_i]_i$$
$$I(B) = I[q_j]_j$$
$$I(A \times B) = I[p_i \times q_j]_{i,j}$$
$$I(A, B) = I[a_{ij}]$$
$$I(A : B) = I(A \times B) - I(A, B)$$
$$I(A|B) =< I[p_i|q_j]_i >_j = \sum_j q_j \times I[p_i|q_j]_i$$
$$I(B|A) =< I[q_j|p_i]_j >_i = \sum_i p_i \times I[q_j|p_i]_j$$

The following relations always hold:

All these quantities are positive.

$$I(A : B) = I(A \times B) - I(A, B)$$
$$I(A : B) = I(B) - I(B|A)$$
$$I(A : B) = I(A) - I(A|B)$$

$I(A : B)$ represents the degree of correlation of A and B. It equals 0 when A and B are independent. Its maximum is $\min(I(A), I(B))$, it means that A and B are totally correlated.

$I(A|B)$ represents the average lack of information obtained by B after reception of a message. It equals zero when whatever is received determines what was sent. $I(A)$ represents the lack of information B has about A. So $I(A|B) - I(A)$ represents the average lack of information gained by B after reception of a message. And $I(A : B)$, its opposite, represents the average information gained by B after reception of a message.

Sometimes in the literature, $I(A : B)$ is defined as the opposite of our definition, and so is negative and decreasing with information gained by B.

Instead of people sending messages, we can consider the sender A as a physical system which can be in different states according to the probability distribution $[p_i]_i$, and the receiver B as a measuring apparatus which shows the different results of the experiment with probabilities $[q_j]_j$. The meanings of the different quantities are the same, the reception of a message being a measurement. This gives:

$I(A : B)$ represents the degree of correlation of A and B. It equals 0 when A and B are independent. Its maximum is $\min(I(A), I(B))$, it means that A and B are totally correlated. In this case, B can determine with certainty in which state A was before measurement.

$I(A|B)$ represents the average lack of information obtained by B after a measurement. It equals zero when whatever is measured, it determines the state of A. $I(A)$ represents the lack of information B has about A. So $I(A|B) - I(A)$, a negative quantity, represents the average lack of information gained by B after a measurement. And $I(A : B)$, its opposite, represents the average information gained by B after a measurement.

2.1.1 Entropy as positive information: Shannon and Brillouin

Shannon assimilates its definition of information $I = I[p_i]$ with thermodynamical entropy. In classical thermodynamics, entropy is usually calculated in $Joule/Kelvin$, and not in $bits$. However, this is only a matter of unity.

$$1\ bit = k \ln 2\ J/K$$
$$k = 1.4 \times 10^{-23} \text{ is Boltzmann constant.}$$

As we will see, the definitions of entropy in quantum mechanics and in statistical thermodynamics have exactly the same form as Shannon information. Shannon states that the function I, and hence entropy, measures information, and this is meaningful if we consider that the information is owned by the system. Brillouin thinks that the function I, and hence entropy, shows the lack of information, and he is equally right if we adopt the point of view of the observer. Brillouin was aware of his opposition with Shannon[12], but didn't try to find why Shannon's opinion could also be of interest.

The main point is the deep similarity between the communication model and the experiment model. For the sender A in the communication model, $I(A)$ is the information sent by A, while in the experiment model, $I(A)$ is the entropy of the system A. Shannon, having the point of view of the sender, who wants to compress the message, or add error-correcting data, sees entropy as positive information. For the receiver B in the communication model, $I(B)$ is the uncertainty about what is received, while in the experiment model, $I(B)$ is the entropy of measurement. Brillouin, having the point of view of the receiver of the message, sees entropy as lack of information.

Let's have a closer look at Brillouin's argument: during the communication along the channel, some information is lost, because of noise. Information, he says, is decreasing, while entropy increases.

The entropy of the message is effectively growing. Suppose the alphabet contains only two letters: 0 and 1. With noise, the 0s and the 1s become equiprobable, which gives a maximum entropy. But if the message which is sent has been optimally compressed, its entropy is maximum from the beginning, and the channel's noise can not make it grow. What is decreasing is our capacity to recover the original message.

This situation can be compared with the evolution of a thermodynamical system: with temperature, the system goes towards equilibrium, and our capacity to describe the initial state of the system from its current state decreases. The noise corresponds to the temperature. Higher the noise, more redundant the messages have to be (error-correcting codes, to preserve the initial information), lower is their entropy. Entropy flows better at low temperature. Information flows better in silence (the opposite of noise).

A confusion about the nature of entropy comes from the fact that a perfectly compressed message is of maximum entropy, containing a maximum of information, while a random sequence of 0s and 1s, also of maximum entropy, contains no information.

Shannon sees in a maximum entropy message a perfectly compressed message, while Brillouin sees it as a perfectly random message. These two points of view are correct, and can be compared with the points of view of the system and the observer of this system. For the system, a maximum entropy means the ability to transmit a maximum amount of information. For the observer, the maximum entropy of the system means a maximum ignorance about the result of a future measurement.

2.2 Entropy, information and freedom in QM

2.2.1 Entropy of a quantum system

In quantum mechanics (QM), the state of a system is represented by its wave function, which is a vector of length one in a Hilbert space. If we note $|\varphi>$ this vector, and $<\varphi|$ its conjugate, then we can define its density matrix $\varrho = |\varphi><\varphi|$[13]. This matrix is diagonal in an orthonormal base starting with $|\varphi>$. Its representation in this base is:

$$\varrho = \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}$$

The entropy of a quantum system is defined as the Von Neumann entropy:

$$S(\varrho) = -Tr(\varrho \log \varrho)$$

In general, the density matrix ϱ is Hermitian and as such can always be diagonalized, with real eigenvalues, and orthonormal eigenvectors as base.

$$\varrho = \begin{pmatrix} p_1 & & & \\ & p_2 & & \\ & & \ddots & \\ & & & p_n \end{pmatrix}$$

Its entropy is then

$$S(\varrho) = I[p_i]_i = \sum_i I(p_i) = -\sum_i p_i \log p_i$$

In the case of our quantum system represented by its wave function, we find that

$$S(\varrho) = I[1, 0, ..., 0] = 0.$$

A quantum system has zero entropy. We are interested by the entropy of an ensemble of states, each one arriving with a probability p_i, not by the entropy of a single state, which is null. For example, an entangled state $\frac{1}{\sqrt{2}}(|01> + |10>)$ has zero entropy. A photon which has just passed the slits of a two slit experiment is in a state $|\varphi_{up}> + |\varphi_{down}>$, which is a single state; this photon has zero entropy, besides the fact that it can be measured only in the up or down path. Even a quantum system with zero entropy is not deterministic when measured. But our knowledge of the state of the system is complete. We know its state vector.

A single state is called a pure state, while an ensemble is called a mixed state.

2.2.2 Notations and main results

We now consider an ensemble of quantum states $(|\varphi_i>)_i$, not necessarily orthogonal. A system A is in the quantum state $|\varphi_i>$ according to the probability distribution $[p_i]_i$, and is measured by an observer, or a measuring apparatus B. B measures the state $|\varphi_i>$ with an orthonormal base of vectors $(|j>)_j$. After measurement, the system is in one of the states $|j>$.

Before measurement, B knows the quantum states $(|\varphi_i>)_i$, and the base $(|j>)_j$ of measurement. The only thing he doesn't know is the probability distribution $[p_i]_i$. When B doesn't know the quantum states $(|\varphi_i>)_i$, this is the field of quantum information which we will not enter into.

We note $(q_j|p_i)$ the probability of the system being in state $|j>$ after measurement, knowing it was in the state $|\varphi_i>$ before measurement.

QM tells us that $(q_j|p_i) = |<j|\varphi_i>|^2$.

This can be stated in terms of the density matrix.

Let define:

$$\varrho_i = |\varphi_i><\varphi_i|$$
$$\varrho = \sum_i p_i \varrho_i$$

We have:

$$(q_j|p_i) = <j|\varrho_i|j>$$

We can now define a_{ij}, q_j, $(p_i|q_j)$, $I(A)$, $I(B)$, $I(A \times B)$, $I(A, B)$, $I(A|B)$ and $I(B|A)$ as before, and all the relations still hold:

All these quantities are positive.
$$I(A:B) = I(A \times B) - I(A, B)$$
$$I(A:B) = I(B) - I(B|A)$$
$$I(A:B) = I(A) - I(A|B)$$

We also have the following inequalities:

$$I(A:B) \leq S(\varrho) \leq I(B) [14][15][16]$$
$$S(\varrho) \leq I(A)$$

A more detailed description can be found in [17].

2.2.3 Interpretation

Every measurement can be predicted with the density matrix ϱ, but ϱ can be written in many ways as a sum of matrices. The only canonical representation of ϱ is its diagonal form. Since entropy is a state function, it is natural that it only depends on the density matrix, and not on the distribution $[p_i]_i$.

$S(\varrho) \leq I(A)$ We have an equality iff the states $|\varphi_i>$ are orthogonal: $<\varphi_i|\varphi_j>=\delta_{ij}$. When classical states are transmitted, there are all orthogonal, in the sense that they can be identified with certainty. But in QM, two non orthogonal states can't be identified with certainty. For a given probability distribution $[p_i]_i$, entropy is maximum when all the sent states are orthogonal.

$I(A:B) \leq S(\varrho)$ This inequality is known as the Kholevo (or Holevo) bound, and tells that the average information gain for B in each measurement is bounded by the entropy of the system A. This entropy is the maximum capacity of a channel transmitting information from A, the maximum information the sender A can expect to effectively send, or the maximum information the receiver B can expect to receive. We have equality iff each base vector $|i>$ of the measurement is equal to an original quantum state $|\varphi_i>$. As $<i|j>=\delta_{ij}$, it implies that the states $|\varphi_i>$ are orthogonal, and hence that $S(\varrho) = I(A)$.

$S(\varrho) \leq I(B)$ This inequality tells that the entropy of the system A is less than the entropy of any measurement made by B. The entropy of a measurement is defined as $I(B) = I[q_j]_j$, and, as we will see, can be interpreted as the manifested freedom of the system A. It is not entropy, in the sense of thermodynamical entropy, just an info-entropy. Here also, equality holds iff each base vector $|i>$ of the measurement is equal to an original quantum state $|\varphi_i>$. For instance, let us consider a pure state $|\varphi>$. Its entropy is zero. However, only one measurement with an orthogonal base containing $|\varphi>$ will tell if a system is in the state $|\varphi>$ with certainty. Other measurements will not. In other words, only one measurement has an entropy of zero, the others have a strictly positive entropy. Let's consider now an ensemble of 2 orthogonal states with equal probabilities in a 2-dimensional Hilbert space. The entropy of this ensemble is maximum: $I(A) = I[\frac{1}{2}; \frac{1}{2}] = 1$. So any measurement will give two results with equal probabilities.

$I(B)$ **as freedom of choice** $I(B)$ can be decomposed in two parts. $S(\varrho)$ is the freedom of choice manifested by the sender of the quantum states, as in the classical case. However,

$$\Delta I = I(B) - S(\varrho) \geq 0$$

is a freedom of choice deeply linked with the probabilistic nature of QM. The Bell inequalities prove mathematically, and the Aspect experiment proves practically that the probabilistic choice of a given result for a measurement is not due to our ignorance of a given mechanism which would determine this result[19][20]. We propose a simple demonstration of this in Annex 1[21]. One could argue that an ensemble is made of probabilities reflecting our ignorance of how the system was prepared. But the probabilities of a measurement's result are a very deep aspect of the world, not a manifestation of our ignorance. As this result is really chosen at the last moment by the measured system and is deeply non-deterministic, we call ΔI the manifested quantum freedom of the system A with this measurement.

$I(B)$, as the sum of these two freedoms, can be assimilated to the freedom of the system A with this measurement.

2.2.4 Entropy as information and freedom

The entropy of a system can be seen as information or freedom.

As information, entropy is the upper bound of the quantity of information a system can give, or an observer can get in a measurement.

As freedom, entropy is the lower bound of the freedom of choice a system manifests in a measurement.

We can summarize this with the following inequalities:

$$I(A:B) \leq S(\varrho) \leq I(B)$$
$$\text{information} \leq \text{entropy} \leq \text{freedom of choice}$$

These three quantities are measured in bits, 1 bit of freedom is the freedom of making freely (with equal probabilities) a choice out of two.

Probabilities and density matrices As we have said before, when the states sent are orthogonal, $S(\varrho)$ manifests our ignorance of how the system was prepared. However, we don't need the concept of ensemble to encounter density matrices. They appear naturally when we want to study one part of an entangled system spatially separated in several parts.

For instance, if the system is an entangled pair in a 2-dimensional Hilbert space, it is a pure state of zero entropy. Suppose its state is $\frac{1}{\sqrt{2}}(|00> + |11>)$. The state of, say the left part, is a mixed state $\frac{1}{2}(|0><0| + |1><1|) = \frac{1}{2}\mathbf{1}$, with an entropy of 1 bit. Note that we lose 'obvious' inequalities like $I(A) \leq I(A,B)$. Here, $I(A) = 1\ bit$, and $I(A,B) = 0\ bit$.

When part of an entangled system, mixed states are very different from our conception of an ensemble. The probabilities appearing are not due to our ignorance of how the system was prepared (we know that), but are of the same nature as the probabilities involved in a measurement, due to the quantum nature of the physical world. The freedom of choice $S(\varrho)$ is then of the same nature as $\Delta I = I(B) - S(\varrho)$, and $I(B)$, the entropy of the measurement, is really the manifestation of a pure freedom of the system+measuring device.

2.2.5 An example

Let us consider the case of a quantum ensemble A composed of two equiprobable non orthogonal states $|u>$ and $|v>$ of a two-dimensional Hilbert space, measured by B in an orthonormal base $(|i>, |j>)$:

$$|u>= \cos\theta|i> + \sin\theta|j>$$
$$|v>= \sin\theta|i> + \cos\theta|j>$$

This example is detailed in Annex 2.

Our main inequality

$$I(A:B) \leq S(\varrho) \leq I(B)$$
$$\text{information} \leq \text{entropy} \leq \text{freedom of choice}$$

reads:

$$1 - I[\sin^2\theta; \cos^2\theta] \leq I[\tfrac{1}{2} + \tfrac{\sin 2\theta}{2}; \tfrac{1}{2} - \tfrac{\sin 2\theta}{2}] \leq 1$$

For $\theta = 0$, we get $1 \leq 1 \leq 1$. $|u>=|i>$, $|v>=|j>$, each measurement gives one bit of information.

For $\theta = \frac{\pi}{4}$, we get $0 \leq 0 \leq 1$. $|u>=|v>$. A measurement gives no information, but the system still manifests freedom.

For $\theta = \frac{\pi}{6}$, we get $1 - I[\frac{1}{4}; \frac{3}{4}] \leq I[\frac{1}{2} - \frac{\sqrt{3}}{4}; \frac{1}{2} + \frac{\sqrt{3}}{4}] \leq 1$. A measurement gives less information than the system owns, and the system manifests more freedom than it owns.

2.3 Dangers and advantages of vulgarization

I emphasize the fact that in no case I suppose that a quantum system owns freedom, in some philosophical sense, that it 'thinks' and 'chooses' like humans do. Nor that it owns a given information in the semantic sense of the word, to which it would give meaning. The thesis defended here is just that, if entropy has to be designated by a more meaningful word, let us choose one that is as appropriate as possible. Thermodynamics is the physical science which links the microscopic and macroscopic worlds, and the meanings of entropy show a curious mirror effect according to the adopted point of view; what is information for one is lack of information for the other. A second mirror effect is between information and freedom. These two words design entropy from the internal point of view, and both are helpful for our understanding of entropy. Information denotes a static aspect of a system, the information it owns about its state, while freedom shows a more dynamic aspect of a system, the next state it can choose.

But why multiply the meanings of entropy? For some physicists, only the word entropy should be used, since others could throw people into confusion. However, the significance of entropy is such that the use of more meaningful words can't be avoided. A common interpretation of entropy today is disorder, and the choice of this word is actually a great source of confusion[22], as we will show. So we have to find a better word.

Lack of information or uncertainty is certainly a good choice. But information is equally good, and having both points of view gives a deeper understanding of what entropy is. So there is no reason to dismiss one or the other. And this 'informational metaphor' for entropy has proved to be so rich that it is difficult to eliminate it entirely.

Then why add another word, freedom, in addition to information? It has the same disadvantage, that is to be of broad significance, and can easily lead to misuses in other domains. However, we can give three reasons for this choice. First, entropy and quantum mechanics both use probabilities. The world is not deterministic, and quantum systems choose their next state when measured. It is really a free choice, not an illusion of choice due to the ignorance of some hidden determinism. The word information doesn't take into account this aspect of entropy. The word freedom does. It is natural to say that a system has more freedom if it can equally choose from more states, or if it can more equally choose from a given set of states. This freedom is a freedom of choice but, as we will see, manifests also as freedom of position, motion, ... A second reason to choose this word is the common unit it shares with information: the bit. A bit of freedom corresponds to a choice out of two possibilities. Measuring freedom of choice in bits is natural, and shows a deep connection between freedom and information. Lastly, as we will see in section 4, usual experiments in classical thermodynamics are better understood assimilating entropy with freedom. This word is useful for pedagogical reasons, and can be used from the beginning to advanced courses in thermodynamics.

2.4 Freedom and the uncertainty principle

2.4.1 Entropy of a probability density

We can define the info-entropy of a probability density $f(x)$:

$$I(f) = -\int f(x) \log f(x) dx \qquad \text{in bits}$$

For an outside observer measuring x, $I(f)$ is the ignorance or the uncertainty he has before the measurement, or the average information he gets with the measurement.

For a particle following the law $f(x)$, $I(f)$ is its freedom to choose a particular x value.

2.4.2 The entropic uncertainty principle

The uncertainty principle is a constraint on the variances of two conjugate quantities. $I(f)$ reflects more precisely than variance the fact that the density probability f is not localized. If f is localized on several peaks, its variance reflects the distance between the peaks, while its info-entropy depends only on the sharpness of the peaks. Hence an entropic uncertainty principle would reflect more accurately the fact that the probability density is not localized, even on several values.

An inequality, similar to the Heisenberg uncertainty principle, exists for entropy[23].

$$I(X) + I(P) \geq \log(\tfrac{eh}{2})$$

X is the probability distribution of positions and P that of momenta or, more generally, of two conjugate variables.

The entropic uncertainty principle is stronger than the Heisenberg uncertainty principle. The last can be derived from the first, using the fact that the Gaussian is the probability density with maximum info-entropy for a given standard deviation[24].

The info-entropy of a Gaussian G_σ, with standard deviation σ, is:

$$I(G_\sigma) = -\int G_\sigma(x)\log(G_\sigma(x))dx = \log(\sigma\sqrt{2\pi e})$$

Hence we have, for a probability density f with standard deviation σ:

$$I(f) \leq \log(\sigma\sqrt{2\pi e})$$
$$\sigma \geq \frac{2^{I(f)}}{\sqrt{2\pi e}}$$

Let X be the probability distribution of positions and P that of momenta, having respectively standard deviation σ and σ'. We have:

$$\sigma \times \sigma' \geq \frac{2^{I(X)}}{\sqrt{2\pi e}} \times \frac{2^{I(P)}}{\sqrt{2\pi e}} = \frac{2^{I(X)+I(P)}}{2\pi e}$$

Applying the entropic uncertainty principle, we get:

$$\sigma \times \sigma' \geq \frac{2^{\log(\frac{eh}{2})}}{2\pi e} = \frac{h}{4\pi} = \frac{\hbar}{2}$$

which is the uncertainty principle.

We can see the entropic uncertainty principle as a principle of guaranteed minimum freedom: a quantum system can't be entirely deterministic, can't have a freedom of zero. The entropic uncertainty principle imposes the minimum freedom manifested by two measurements linked with observables which do not commute. It says that it is impossible for a particle to have a completely deterministic behavior, whatever the measurement made. If this is true for a given measurement, this will be false for another.

2.5 Entropy of a black hole

The entropy of a black hole is proportional to the number of degrees of freedom it owns, itself proportional to the area of the event horizon[25]. One more time, entropy measures freedom.

Moreover, entropy manifests itself as a fundamental quantity, leading to one of the few formulas using so many fundamental physical constants, found by J. Bekenstein :

$$S = \frac{Ac^3}{4hG}/\ln 2 \qquad bits$$
$$c = 3 \times 10^8 \ m/s \qquad \text{is the speed of light}$$
$$h = 6.6 \times 10^{-34} \ Js \qquad \text{is Planck constant}$$
$$G = 6.7 \times 10^{-11} \ m^3kg^{-1}s^{-2} \qquad \text{is the constant of gravitation}$$

This black hole thermodynamics could be the bridge between quantum mechanics and general relativity, entropy and energy being the main concepts in this approach[6].

2.6 Summary: entropy seen as information and freedom

Having the point of view of the observer of the system, we can assimilate entropy to ignorance, or uncertainty about the result of a future measurement on the system, and this point of view has been largely developed. Instead, we'd like to insist on the system's point of view or, in other words, the point of view you would have if you were the studied quantum system. From the system's point of view, entropy measures a minimum freedom of choice: the system chooses (according to a probability set) what result it gives when measured. We have here the manifestation of a pure choice made by the system when measured. Doing so, it becomes deterministic regarding this measurement. The same measurement, just after the first one, always gives the same result.

The bit is a unit of information. It is also a unit of freedom, and this establishes interesting links between information and freedom. The freedom we are talking about here, which is a freedom of choice, is as big as the number of choices is high, and as they are equally accessible. If the quantum system owns a big freedom of choice, it also owns a big amount of information, and the choice it makes gives part of this information to an outside observer (it makes this choice when measured). As long as it hasn't made any choice, its behavior is unpredictable for the outside observer, its freedom is for her a lack of information, or uncertainty about its future behavior.

We will now show that this analogy between entropy and freedom/information can be maintained when we study systems with a great number of elements. This is the object of statistical thermodynamics.

3 Entropy in statistical thermodynamics

3.1 Boltzmann entropy

Statistical thermodynamics was born with Boltzmann, and his famous formula for entropy:

$$S = \log \Omega$$

We recall that entropy is measured in bits, and log is the base 2 logarithm.

Ω designates the number of possible microstates compatible with the macroscopic description of the system at equilibrium.

The bigger the entropy, the bigger the number of microstates, the more freedom the system owns regarding the microstate it is effectively in. Boltzmann formula is naturally interpreted as the system's freedom of choice.

We can even extend this point of view to each system's particle. Considering a system composed of P particles, with P_1 particles in state 1, ...,P_N particles in state N, letting $p_i = P_i/P$, and applying Stirling's formula (supposing $P_i >> 1$ for all i), by a classical calculus we get:

$$S = P \times (-\sum_{i=1}^{N} p_i \log p_i)$$

This leads us to define the average freedom of a single particle by the Shannon entropy:

$$s = -\sum_{i=1}^{N} p_i \log p_i$$

and to say that the system's entropy is the sum of the entropies of every particle which constitutes it.

So a thermodynamical system owns a given amount of information S, equals to the information an outside observer would have if he knew the microstate the system is in. Each particle can then be seen as having the average amount of information $i = s$. Each particle can also be seen as having the average freedom $f = s$, since the microstate of the system is the result of all the individual choices of its components. The freedom of the system is the sum of the freedom of each particle.

Once again, according to the point of view, entropy can be assimilated to freedom and information, or to lack of information.

For instance, a molecule of one liter of ideal monatomic gas like Helium 4 at normal pressure and temperature ($300°K$, 10^5 pascals) has an entropy of 17 bits[26]. Seen as information, this entropy means that all the parameters needed to encode the state of a molecule (position and speed) can be encoded with 17 bits. Seen as freedom, this entropy means that one molecule can choose its next state from $2^{17} = 131072$ possible states, or that it is free to make 17 independent binary choices to decide its next state.

3.2 MaxEnt: Entropy according to Jaynes

Jaynes, using Gibbs method, but interpreting it differently, holds the following reasoning: if an experiment always gives the same macroscopic result when starting with the same macroscopic state, this means that the microscopic state of the system contains no useful additional information. For logical reasons, we are led to define the system's microscopic state as one with maximum entropy satisfying the macroscopic knowledge we already have[27].

This principle of logic, applied in many fields, including non physical ones, is called MaxEnt (MAXimise ENTropy).

In the case where only the average value of energy is sufficient to describe the macroscopic state of the system, we have to find the probability law $p(E)$ which should satisfy:

$$\begin{cases} -\int p(E)\log(p(E))dE \text{ maximum} \\ \int p(E)dE = 1 \\ \int p(E)EdE = < E > \end{cases}$$

Using Lagrange multipliers method, we find the so-called canonical set distribution:

$p(E) = e^{-\beta E}/Z$ where Z is a normalization factor (the partition function) and β a parameter induced by the Lagrangian formalism. $\beta = 1/kT$, where T is the system's temperature and k the Boltzmann constant.

This definition of temperature is typical of modern thermodynamics, which defines temperature from entropy and energy. While in classical thermodynamics, temperature and energy are the basic concepts which are used to define entropy, the more recent approaches define temperature as a function of energy and entropy, considered as the fundamental properties of a system. More precisely, by:

$$\left(\frac{\partial S}{\partial U}\right)_V = \frac{1}{T} \qquad S \text{ in } J/K,\ U \text{ in Joules, } T \text{ in Kelvins}$$

or:

$$\left(\frac{\partial S}{\partial U}\right)_V = \frac{1}{kT\ln(2)} \qquad S \text{ in bits, } U \text{ in Joules, } T \text{ in Kelvins.}$$

Many relations found in classical thermodynamics involving entropy, energy and temperature can be found as consequences of the Lagrangian formalism[28].

Maximizing entropy (with the constraints) allows to describe equilibrium. But is it true that entropy is maximized at every moment, including far from equilibrium? MaxEnt-NESOM[29], which consists in maximizing quantum entropy at every moment , permits to recover the results of the close to equilibrium theories (Prigogyne's theorem of minimal entropy production, Onsager's reciprocity relations), and is experimentally verified in far from equilibrium situations. If this theory happens to be the correct description of a thermodynamical system, in equilibrium or not, this means that physical universe is ruled by a logical principle of maximization of the information/freedom of its elements.

4 Classical thermodynamics

4.1 Temperature, heat and disorder

Born in the XIX^{ith} century, classical thermodynamics was about the efficiency of heat engines, and the scientists of those days saw in entropy a source of limitations: entropy was lowering engines efficiency, forbid the perpetuum mobile, was the cause of things getting worn away, and led our universe to an inexorable thermal death. There has been confusion between entropy and disorder from the beginning, for the good reason that nobody knew what entropy was (and this point of view is still largely shared). While entropy was assimilated to disorder, temperature was a measure of molecular agitation, and heat was disordered energy. So the three fundamental quantities of thermodynamics - entropy, heat and temperature - were all linked with two closed concepts: disorder and agitation.

It is not possible to understand entropy without also understanding what temperature and heat mean, the three being tied by the famous equation:

$$dS = \frac{\delta Q}{kT\ln 2} \qquad bits$$

In this equality, dS is the entropy received by the system. Entropy is a state function, and dS an exact differential. δQ is just a little quantity of heat, not an exact differential, which is received in a reversible transformation. T is temperature and k Boltzmann constant. A reversible transformation is a transformation which can be drawn with a continuous curve in a (T,S) diagram (T function of S) or, equivalently for a gas, in a Carnot diagram (P function of V). If the system is not in equilibrium at some time during the transformation, it has no coordinates in such diagrams, the curve is not continuous and the formula doesn't hold.

4.1.1 Temperature is a measure of agitation, or disorder

To assimilate temperature with disorder, or agitation, is very meaningful. Low temperature systems are said to be very ordered. More fundamentally, temperature measures, for a gas, the part of the molecules motion that doesn't contribute to a possible global motion, and this is effectively a usual meaning of the words agitation, or disorder: a disordered motion is a motion made of many useless moves.

Since temperature is often seen in equations in the form $1/(kT)$, we designate this quantity by the word opposite to agitation, calm.

4.1.2 Heat and work

Internal energy U is the energy of the system at rest. It is not composed of heat and work. However, a small variation dU of this energy can be divided in small variations of heat and work. Heat is the part of the internal energy variation which contributes to entropy. One way of understanding this is to consider a quantum system with N energy levels (U_i) and probabilities (p_i). Then

$$S = -\sum_{i=1}^{n} p_i \log p_i$$
$$U = \sum_{i=1}^{n} p_i U_i$$

and

$$dU = \sum_{i=1}^{n} dp_i U_i + \sum_{i=1}^{n} p_i dU_i$$

We see that in the last sum only the first term contributes to a variation of entropy, and so represents the variation of heat, while the second one represents the variation of work.

As heat can not give work, it was called disordered energy, or useless energy. This denomination was confirmed by the fact that, for an ideal gas in absence of work, heat is tied to temperature by a linear relation (which coefficient is the calorific capacity): temperature measuring molecular agitation, heat became agitation for a given quantity of matter, disorder.

From the external point of view, work is the quantity of interest.

From the internal point of view, heat is the quantity of interest, since it can give freedom.

4.1.3 The equation $dS = \frac{\delta Q}{kT \ln 2}$

For a given amount of heat δQ, this equality says that entropy increases more at low temperature. Assimilating entropy with freedom, and temperature with agitation, it says that, for a given heat, freedom increases more in the calm.

4.2 Entropy doesn't measure disorder

Besides the fact that thousands of papers describe entropy as a measure of disorder, we can find more and more thermodynamics researchers and teachers stating explicitly that entropy is not disorder[22]. However, different reasonings lead to this misconception, and many great scientists still use it.

To clarify this point, we will detail three kinds of explanations which wrongly lead to this analogy: those which stand on an anthropomorphic vision of order, those based on examples for which temperature and entropy vary together (if temperature measures disorder, they don't prove anything), and those based on a definition of order as absence of freedom.

4.2.1 Anthropomorphic order

Justifications of analogy between entropy and disorder based on an anthropomorphic notion of order lack rigor. Seeing entropy as freedom helps to find counter-examples.

- decks of cards which are more and more 'disordered' during a shuffle are only so if we consider as 'ordered' a deck in which cards are in the same order as in a brand new deck. In fact, the more shuffled the deck is, the more freedom a card gets to have any position in the deck.

- the 'messy' student rooms fit in this category, the notion of 'well ordered room' being very subjective. Is a room where all is put in a corner 'well ordered'? In any case, its entropy is low.

- A cathedral, 'manifestly ordered', ends being a bunch of sand, 'manifestly disordered'. The trouble is that if the disorder of the cathedral is defined by its entropy, then there are many configurations of the sand more 'ordered' (a formless bloc, for instance). In fact, each grain of dust which detaches from the cathedral gains freedom. The bunch of sand itself is dynamic, with always grains flying off and others landing. Each grain of dust is in average more free in the bunch of dust than in the cathedral.

4.2.2 Confusion between entropy and temperature

Assimilation of entropy and disorder comes also from the fact that entropy and temperature often vary together.

However, we should notice that this is not a general law. It is false that entropy increases with temperature. Entropy doesn't vary with temperature (U is the internal energy):

$$\left(\frac{\partial S}{\partial T}\right)_U = 0$$

Entropy varies with energy:

$$\left(\frac{\partial S}{\partial U}\right)_V = \frac{1}{T}$$

However, energy is an increasing function of temperature, and even linear in the case of an ideal gas: $U = 3/2 \times kT$. This phenomenon makes think entropy is an increasing function of temperature. But this is not true at constant energy.

To be convinced that temperature, and not entropy, measures disorder, we have to look for situations where entropy and temperature don't vary together.

An example is the expansion of a gas: its entropy increases and its temperature decreases. When you use a vaporizer, the water spray acquires freedom and arrives cold on your face (besides being at ambient temperature in the container). This example shows that entropy measures freedom (here, freedom of position), and temperature disorder. The liquid is colder, the molecules are less agitated, their motions contribute more to the global motion of the spray.

A similar example, and maybe more important to the reader, is the expansion of the universe as a whole. Since the Big-Bang, entropy is always increasing and temperature decreasing. But the universe was extremely disordered at the beginning, and has become more and more ordered.

4.2.3 Order as lack of freedom

Here is a conception of order: if, in a population composed of N individuals, each one is free to choose from two colors of suits, the situation is more 'disordered' than if everybody wears the same color of suit[30]. For a physicist, it becomes: if N spin half particles, agitated by temperature, are equally distributed between their two possible states, the situation is more 'disordered' than if they all share the same state.

Another common example is the transition from solid to liquid state. It is clear that a molecule of water has more freedom of motion than a molecule of ice. When we say water is more disordered than ice, it is what we mean. To see the fact that the molecules of water can move everywhere in the liquid, and not in the solid, as disorder, is defining disorder as freedom of choice.

In these examples, the definition of order is exactly the antithesis of freedom, order being maximum when freedom is minimal, and reciprocally. Entropy being a measure of freedom, it is also a measure of this definition of disorder, and all examples confirm that.

The question is: what definition for disorder do we choose?

If we adopt as definition of disorder "what happens when there is freedom", then entropy is a measure of disorder and also of freedom, and order means absence of freedom.

If we adopt as definition of disorder "what doesn't serve the global motion", then it is temperature which measures disorder, and, for a given heat, freedom increases more in order.

As far as we know, there is no justification for the analogy of entropy and disorder, except to define order as the opposite of freedom (which we will not do). Our definition of disorder is "What doesn't contribute to the global motion".

We hope that a clear distinction between these two meanings of the word 'disorder' will clarify what it exactly means to assimilate entropy with disorder, will discourage authors to do so, and encourage them to see entropy as freedom, and temperature as agitation or disorder.

4.3 Study of a few classical experiments

4.3.1 Study of an ideal gas

An ideal gas can be determined with only two parameters, for instance energy and entropy (or volume, temperature, pressure), or defined by the equation: $PV = nkT$, where n is the number of particules. At equilibrium, each molecule of the gas owns a maximum freedom: it can equally be everywhere in the volume occupied by the gas. Each molecule also owns a maximum freedom of momentum, taking into account temperature (temperature is proportional to the variance of the momentum). So the probability law of the momentum is a Gaussian with variance $\propto T$, because with a given variance, the gaussian is the maximum entropy probability distribution.

Some authors call freedom of position configurational freedom, and freedom of momentum thermal freedom[5]. But as we can define an info-entropy for every observable (See section 2), we prefer to say explicitly the observable for which we measure freedom.

If we raise the temperature of an ideal gas at constant volume, its energy increases ($U = \frac{3}{2}kT$), and so increases its entropy. We find that:

$$dS \propto d\left(\ln(T)\right)$$

Its entropy of position has not changed (the molecules are still uniformly in all the available volume), but its entropy of momentum has increased (due to temperature)

If we raise the volume of a gas at constant temperature, its entropy also increases. But this time, its entropy of position increases (each molecule have more space), and its entropy of momentum does not change. We find that:

$$dS \propto d(\ln V).$$

4.3.2 Free energy

The second principle states that a phenomenon can occur spontaneously only if:

$$\Delta S \geq \tfrac{\Delta U}{T} \qquad (\text{free energy} = \Delta U - T\Delta S \leq 0) \qquad (1)$$

The phenomena of fusion, vaporization, ..., osmotic pressure, can be explained in terms of free energy. We can see every time that a phenomenon occurs if the system gets enough freedom, taking temperature into account[5].

For instance, in melting (solid$\rightarrow$liquid), molecules get freedom since they can go in all the liquid. They also get an energy ΔU, but if the temperature T is too low, ΔS - the gain in freedom - is less than $\frac{\Delta U}{T}$, and melting doesn't occur.

In the case of mixing liquids (and particularly solvents), molecules are as free as the concentration of their liquid is low. So, introducing a small quantity of solvent in a solution increases strongly the solvent's entropy, and softly the solution's one. This fact, combined with equation (1) above, allows to explain many experiments with solvents.

4.4 Degrees of freedom

Some molecules can rotate, and possess an energy and an entropy of rotation. Some can also vibrate, and so own an entropy of vibration. When making the entropic balance of a system, we have to consider entropy for every degree of freedom. Rotation, vibration, spin, ..., have to be taken into account. When we just calculate entropy differences, the degrees of freedom for which entropy is constant can be neglected.

Degrees of freedom are the classical version of the different tensors which compose the density matrix in quantum mechanics. Each degree of freedom corresponds to a measurement: position, spin, ... For each degree of freedom, there is an energy and an entropy.

Entropy depends upon the number of degrees of freedom of the studied system, which reinforces the idea that it characterizes the system's freedom.

4.5 The third principle

The Third Principle of thermodynamics, or Nernst's theorem, states that entropy is zero if temperature is zero[31].

When temperature goes towards zero, the system's particles are less and less agitated, the system is more and more ordered. At absolute zero, molecules are immobile, they never change state. So they don't have any freedom of choice (they can only have one state), and their entropy equals zero. The Third Principle simply says that if a system never changes state, it has no freedom.

However, this interpretation is only true in the classical case. In the quantum case, the uncertainty principle forbids a system to have zero freedom for all observables. This leads to the existence of vacuum fluctuations and zero point energy at zero temperature, which can be measured for instance using the Casimir effect[32][33].

5 Conclusion

Clarifying the meaning of entropy led us to distinguish two points of view: the external one, which is the one of the observer of the studied system, and the internal one, which is the one of the system itself.

The external point of view leads to largely admitted associations: entropy as lack of information, or indetermination about the microscopic state of the studied system.

The internal point of view, the one we should have if we were the studied system, leads to interpretations more rarely seen, and yet useful. Entropy is seen as a measure of information, or freedom of choice. These two analogies fit well together, and are tied by the duality of their common unit: the bit. A bit of information represents one possibility out of two, a bit of freedom represents one choice out of two. The entropy/information rehabilitates Shannon's memory, for whom entropy is positive information; the entropy/freedom takes into account the fundamental non-determinism of the measurement process in quantum mechanics. It leads to a natural interpretation of the different definitions of entropy, and of the usual experiments studied in thermodynamics.

Entropy is often assimilated to disorder, and this conception seems to us inappropriate. Instead, temperature is a good measure of disorder, since it measures molecular agitation, the part of the motion which doesn't contribute to a possible global motion. To assimilate entropy with disorder leads to another, unwise, definition of order, as absence of freedom, since entropy measures freedom.

The second principle states that $S_{initial} \leq S_{final}$, in which S stands for the total entropy of all the bodies involved in the process from equilibrium initial state to equilibrium final state. What is the domain of validity of this principle? Coren[34] establishes experimentally that information has always increased since the origin of universe. The author gives no thermodynamical justification to this, and finds that every major step of evolution (Big-Bang, the formation of planets, the birth of life, then homo sapiens, the discovery of writing, and computers) can easily be seen in terms of increasing amount of information.

We add that these different steps can also be seen in terms of increasing freedom: of action for the living beings, of speech or thought for human beings. Could the evolution, not only of the physical world, but also of at least some aspects of the living world and of humanity, be a manifestation of the second principle, seen as a principle of increasing freedom/information?

Acknowledgements

I would like to express my gratitude to Hassan Saidi, who helped me in clarifying this text and translating it. I would also thank the Descartes school for the means I had at my disposal.

Annex 1: probabilistic nature of the state of a quantum system

Quantum mechanics (QM) doesn't allow to predict the result of a measurement, but only the probability of the possible different results. So many physicists considered this theory as incomplete, in the sense that these probabilities are the manifestation of our ignorance of the underlying mechanisms which produces it. "God doesn't play dice" said Einstein; in a famous paper he signed with Podolski and Rosen[20], he describes a thought experiment (it was his speciality) to show that QM implies absurd results. The absurdity is due, according to Einstein, to the phenomenon of wave packet collapse which happens to a quantum system when measured: according to QM, it has to happen simultaneously everywhere in space. This bothered the discoverer of restrained relativity, who refutes the notion of simultaneity of two spatially separated events. The experiment they imagined relies on the possibility of emitting two photons going in opposite directions and described by a single non-factorizable wave function. A system made of two spatially separated subsystems can of course be described by QM, but the wave function which describes it can be written as a product of two wave functions, reflecting the fact that one can be measured without measuring the other (collapsing one of the wave function without collapsing the other). In the case of the two photons, the non-factorizability of the wave function means that measuring one collapses all the wave function, instantaneously, modifying the state, and therefore the result of a measurement of the other. As the two photons can be separated by light-years, this instantaneity implies a supra-luminal interaction, an heresy for Einstein.

In 1964, Bell proved that the results of QM can imply that it is impossible for a particle to be in such a state before a measurement that this state would determine (deterministically) the result of this measurement. It was theoretically possible to check the inexistence of 'hidden variables', with the following EPR-like experiment[19]:

5.1 Alan Aspect experiment (1981)

One photon goes left, the other goes right. Each one will be measured by one of the three observables A, B and C. In practice, the measured quantity is photon's polarization, which can only take two values for a given axis. A, B and C are the observables corresponding to three possibe directions of the measuring apparatus, oriented 120° from each other (their axis divide a circle in three equal parts). Let's call 0 and 1 the two possible results of the measurement of one of the two photons, with one of the three observables. For instance, in the first experiment, we choose observable A for the left photon and observable C for the right photon. We get two results, for instance 1 for the left photon, and 0 for the right one. This experiment is not deterministic. QM predicts the probability that, in this first experiment, the results are 1 on the left and 0 on the right.

If we repeat this experiment a great number of times, we verify that the results conform to the probability law predicted by QM.

This sequence of experiments is simple to make; but for it to prove the non existence of hidden variables, we should make certain that one photon can not tell the other which observable measured it before the other photon is being measured. The Alan Aspect experiment guarantees such a communication can't occur. The observable which measures the left photon is chosen at random, just when the photon is very close to the measuring apparatus, and it has no time to communicate (even at speed of light) this observable to the right photon before that one is measured, also by a random observable chosen at the last moment. This implies that the axis of the two measuring apparatus (the two observables) are chosen in a few nanoseconds, technology only lately available.

5.2 The experimental results

The Aspect experiment confirms QM's predictions, which are:

(a) For a given side (left or right) and a given observable, the probabilities of having 0 or 1 are equals. For instance, the probabilities that the right photon, measured with observable B, gives 0 or 1 are equal (and so equal 1/2).

(b) When the two photons are measured with the same observable (with the same axis), the results of the measurements are always different. If the left's measurement is 0, the right's one is 1, and reciprocally.

(c) When the two photons are measured with two different observables, the probability for the two results to be equal is 3/4.

5.3 The proof that there is no state preceding and determining the measurement

The trouble is that these probabilities are impossible if we suppose the existence of a state of the photon, prior to the measurement, which would determine every possible results for all observables. Here is the proof:

Let's suppose that when they split, or even later, but before being measured, the two photons are in a given 'state', which will determine the results of the different possible measurements. For instance, the left photon is in a state which determines that a measurement will give 0 with A, 1 with B and 0 with C. The fact (b) implies that the right photon has to be in such a state that a measurement will be 1 with A , 0 with B and 1 with C.

Generally speaking, the state of the left photon implies the state of the right one.

The number of different possible states for our two photons is then reduced to 8, written in the following table, one per line:

	$left$			$right$	
A	B	C	A	B	C
0	0	0	1	1	1
0	0	1	1	1	0
0	1	0	1	0	1
0	1	1	1	0	0
1	0	0	0	1	1
1	0	1	0	1	0
1	1	0	0	0	1
1	1	1	0	0	0

For instance, in the state $n°4$, if the left photon is measured with C, the obtained result will be 1, and if the right photon is measured with B, the obtained result will be 0.

Now, let's calculate the frequencies these 8 states should have to confirm the results (a) and (c). We call them $p(i)$, $i = 1$ to 8.

The result (a) implies that states $n°$ 1 and 8 should have equal probability, since 0 and 1 play the same role.

$$p(1) = p(8) = p/2.$$

If we suppose, which is fairly reasonable, that universe is isotropic (no privileged direction), then states n° 2 to 7 should also have equal probability, since they can be deduced from each other by permutation of A, B and C and of 0 and 1.

$$p(2) = p(3) = p(4) = p(5) = p(6) = p(7) = q/6.$$

The sum of probabilities makes 1:

$$p + q = 1.$$

Point (c) says that, in particular, when the left photon is measured with observable A and the right one with observable B, the probability of two equal results should be 3/4. Let's count the states verifying this property, the $left/A$ column in the table being equal to the $right/B$ one. We find states $n°$ 3, 4, 5 and 6. The probability for the two photons to be in one of these states is $4 \times q/6$. So we should have:

$$\frac{4q}{6} = \frac{3}{4}$$

$$q = \frac{9}{8} \qquad et \qquad p = 1 - q = -\frac{1}{8}$$

"It hurts, for probabilities"[21].

Here it is! We have proved the incredible, that the photons really choose the result they give, when measured, playing dice as Einstein would say; this choice can not have been made before the measurement, for every measurement. A quantum state is really its wave function. Probabilities in quantum mechanics are not a measure of our ignorance. They reflect a very deep property of our physical world, not a limitation of our knowledge.

But this makes result (b), for which a simple mechanism was easy to imagine (the photons choose opposite values when they split), even more incredible than what we found. If the left photon only chooses at the last moment to give for instance 1 when measured with observable A, how does the right one always correctly choose to give 0 if also measured with observable A?

To our knowledge, there is no fully satisfying explanation for this. Here are the facts. We really feel that the two photons have to communicate faster than light to be able to verify the results (a), (b) and (c).

However, Einstein relativity remains intact. Neither this experiment nor any other one of this kind allows the transmission of a single bit of information (of entropy :-) faster than light[35].

Annex 2

For those who are not used to the QM formalism, the best way to understand the notation is often to look at an example.

We consider the case of a quantum ensemble A composed of two equiprobable non orthogonal states $|u>$ and $|v>$ of a two-dimensional Hilbert space. It is measured by B in an orthonormal base $(|i>,|j>)$ such that the vectors $|u>$ and $|v>$ are in the first quadrant defined by $(|i>,|j>)$, $<i|u>=\cos\theta$ and $<v|j>=\cos\theta$.

$$|u>=\cos\theta|i>+\sin\theta|j>$$
$$|v>=\sin\theta|i>+\cos\theta|j>$$

We have:

$$<v|u>=\sin\theta\cos\theta+\cos\theta\sin\theta=\sin 2\theta=\cos(\tfrac{\pi}{2}-2\theta)$$

First we have to define the density matrix of system A:

$$\varrho=\tfrac{1}{2}(|u><u|+|v><v|)$$

Then, since $|u>$ and $|v>$ are not orthogonal, we have to put ϱ in diagonal form to calculate its entropy. Geometrically, we can guess that $|u>+|v>$ and $|u>-|v>$ are the eigenvectors.

$$\varrho(|u>+|v>)=\varrho(|u>)+\varrho(|v>)=\tfrac{1}{2}\left(|u>+\sin 2\theta|v>\right)+\tfrac{1}{2}\left(\sin 2\theta|u>+|v>\right)$$
$$\varrho(|u>+|v>)=\tfrac{1}{2}\left(1+\sin 2\theta\right)\left(|u>+|v>\right)$$

Similarly:

$$\varrho(|u>-|v>)=\tfrac{1}{2}\left(1-\sin 2\theta\right)\left(|u>-|v>\right)$$

The eigenvalues are $\frac{1}{2}+\alpha$ and $\frac{1}{2}-\alpha$ where $\alpha=\frac{\sin 2\theta}{2}$

$$S(\varrho)=I[\tfrac{1}{2}+\alpha;\tfrac{1}{2}-\alpha] \qquad \text{where } \alpha=\tfrac{\sin 2\theta}{2}$$

We now calculate the probability distributions:

$$p(|i>\,|\,|u>)=|<i|u>|^2=\cos^2\theta$$
$$p(|j>\,|\,|u>)=|<j|u>|^2=\sin^2\theta$$
$$p(|i>\,|\,|v>)=|<i|v>|^2=\sin^2\theta$$
$$p(|j>\,|\,|v>)=|<j|v>|^2=\cos^2\theta$$

$$p(|u>,|i>)=p(|i>\,|\,|u>)\times p(|u>)=\cos^2\theta\times\tfrac{1}{2}$$
$$p(|u>,|j>)=p(|j>\,|\,|u>)\times p(|u>)=\sin^2\theta\times\tfrac{1}{2}$$
$$p(|v>,|i>)=p(|i>\,|\,|v>)\times p(|v>)=\sin^2\theta\times\tfrac{1}{2}$$
$$p(|v>,|j>)=p(|j>\,|\,|v>)\times p(|v>)=\cos^2\theta\times\tfrac{1}{2}$$

$$p(|i>) = p(|u>,|i>) + p(|v>,|i>) = \frac{1}{2}$$
$$p(|j>) = p(|u>,|j>) + p(|v>,|j>) = \frac{1}{2}$$

$$p(|u> \mid |i>) = \frac{p(|u>,|i>)}{p(|i>)} = \cos^2\theta$$
$$p(|u> \mid |j>) = \frac{p(|u>,|j>)}{p(|j>)} = \sin^2\theta$$
$$p(|v> \mid |i>) = \frac{p(|v>,|i>)}{p(|i>)} = \sin^2\theta$$
$$p(|v> \mid |j>) = \frac{p(|v>,|j>)}{p(|j>)} = \cos^2\theta$$

We can now calculate the different info-entropies:

$$I(A) = I[p(|i>); p(|j>)] = I[\tfrac{1}{2}; \tfrac{1}{2}] = 1\ bit$$

We can check that $S(\varrho) \leq I(A)$, with equality when $\alpha = 0$, i.e. $\theta = 0[\frac{\pi}{2}]$

$$I(A|B) = p(|i>) \times I[p(|u> \mid |i>); p(|v> \mid |i>)] + p(|j>) \times I[p(|u> \mid |j>); p(|v> \mid |j>)]$$
$$I(A|B) = \tfrac{1}{2} I[\cos^2\theta; \sin^2\theta] + \tfrac{1}{2} I[\sin^2\theta; \cos^2\theta] = I[\sin^2\theta; \cos^2\theta]$$

$$I(B) = I[p(|u>); p(|v>)] = I[\tfrac{1}{2}; \tfrac{1}{2}] = 1\ bit$$

$$I(B|A) = p(|u>) \times I[p(|i> \mid |u>); p(|j> \mid |u>)] + p(|v>) \times I[p(|i> \mid |v>); p(|j> \mid |v>)]$$
$$I(B|A) = \tfrac{1}{2} I[\cos^2\theta; \sin^2\theta] + \tfrac{1}{2} I[\sin^2\theta; \cos^2\theta] = I[\sin^2\theta; \cos^2\theta]$$

$$I(A \times B) = I[p(|u>)p(|i>); p(|u>)p(|j>); p(|v>)p(|i>); p(|v>)p(|j>)]$$
$$I(A \times B) = I[\tfrac{1}{4}; \tfrac{1}{4}; \tfrac{1}{4}; \tfrac{1}{4}] = 4I(\tfrac{1}{4}) = 4 \times \tfrac{1}{2} = 2\ bits$$

$$I(A,B) = I[p(|u>,|i>); p(|u>,|j>); p(|v>,|i>); p(|v>,|j>)]$$
$$I(A,B) = I[\tfrac{\cos^2\theta}{2}; \tfrac{\sin^2\theta}{2}; \tfrac{\sin^2\theta}{2}; \tfrac{\cos^2\theta}{2}] = 2I[\tfrac{\sin^2\theta}{2}; \tfrac{\cos^2\theta}{2}]$$
$$I(A,B) = 2 \times (\tfrac{1}{2} I[\sin^2\theta; \cos^2\theta] + I(\tfrac{1}{2})) = 2 \times (\tfrac{1}{2} I[\sin^2\theta; \cos^2\theta] + \tfrac{1}{2})$$
$$I(A,B) = I[\sin^2\theta; \cos^2\theta] + 1$$

We now have three ways to calculate $I(A:B)$:
$I(A:B) = I(A \times B) - I(A,B) = 2 - (I[\sin^2\theta; \cos^2\theta] + 1) = 1 - I[\sin^2\theta; \cos^2\theta]$
$I(A:B) = I(A) - I(A|B) = 1 - I[\sin^2\theta; \cos^2\theta]$
$I(A:B) = I(B) - I(B|A) = 1 - I[\sin^2\theta; \cos^2\theta]$

Our main inequality $I(A:B) \leq S(\varrho) \leq I(B)$ reads:

$$1 - I[\sin^2\theta; \cos^2\theta] \leq I[\tfrac{1}{2} + \tfrac{\sin 2\theta}{2}; \tfrac{1}{2} - \tfrac{\sin 2\theta}{2}] \leq 1$$

For $\theta = 0$, we get $1 \leq 1 \leq 1$. $|u> = |i>$, $|v> = |j>$, each measurement gives one bit of information.

For $\theta = \frac{\pi}{4}$, we get $0 \leq 0 \leq 1$. $|u> = |v>$. A measurement gives no information, but the system still manifests freedom.

For $\theta = \frac{\pi}{6}$, we get $1 - I[\frac{1}{4}; \frac{3}{4}] \leq I[\frac{1}{2} - \frac{\sqrt{3}}{4}; \frac{1}{2} + \frac{\sqrt{3}}{4}] \leq 1$. A measurement gives less information than the system owns, and the system manifests more freedom than it owns.

Note: The function $I(x) = -x \log x$ is defined for all x real positive. It verifies $I(xy) = xI(y) + yI(x)$ and $I(x+y) \leq I(x) + I(y)$. As a consequence, we have:

1/ If $[x_a]_a$ is a distribution and k a real positive number:

$$I[kx_a]_a = kI[x_a]_a + I(k) \times \sum_a x_a$$

As a special case, useful to compare the different info-entropies, if $[p_a]_a$ is a probability distribution we have :

$$I[\tfrac{p_a}{k}]_a = \tfrac{1}{k} I[p_a]_a + I(\tfrac{1}{k})$$

2/ If $[x_a]_a$ and $[y_a]_a$ are distributions:

$$I[x_a + y_a]_a \leq I[x_a]_a + I[y_a]_a$$

References

[1] Shannon:"My greatest concern was what to call it. I thought of calling it 'information', but the word was overly used, so I decided to call it 'uncertainty'. When I discussed it with John Von Neumann, he had a better idea. Von Neumann told me, 'you should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage." as quoted in M. Tribus, E.C. McIrvine, Energy and information, Scientific American, 224 (September 1971), 178-184.

[2] Ludwig Boltzmann. http://www.entropylaw.com/entropydisorder.html

[3] Claude Shannon's "A mathematical theory of communication" Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, 1948.

[4] Léon Brillouin (1889-1969). Science and Information Theory. Editions Academic Press. 1956

[5] Gary L. Bertrand. Mass, Energy and Freedom: The coins of thermodynamics. http://web.umr.edu/~gbert/basic/entropy.htm

[6] P. E. Williams. Energy and Entropy as the Fundamentals of Theoretical Physics. Entropy 2002, 4. http://www.mdpi.net/entropy/htm/e4040128.htm

[7] T. L. Duncan. The Deep Physics Behind the Second Law: Information and Energy as Independent Forms of Bookkeeping. Entropy 2004, 6. http://www.mdpi.net/entropy/htm/e6010021.htm

[8] P. Garbaczewski. INFORMATION -Is There Anything Where Nobody Looks ? http://sympozjum-lfppi.if.pwr.wroc.pl/proceedings/KarpaczPG.pdf

[9] M.V. Mesquita, A.R. Vasconcellos, R. Luzzi, S. Mascarenhas. Systems Biology: an information-theoretic-based statistical approach. Brazilian Journal of Physics, vol. 34, n° 2A, June 2004. http://pcsbf1.sbfisica.org.br/bjp/Vol34/Num2a/v34_459.pdf

[10] R. Landauer. Irreversibility and heat generation in the computing process. IBM J. Res. Dev. 5, 183-191. 1961. http://www.research.ibm.com/journal/rd/053/ibmrd0503C.pdf

[11] C. H. Bennett. Demons, engines and the second law. Scientific American. 1987.

[12] Brillouin [1951a] p. 335 : " Shannon, however compares information with positive entropy, a procedure which seems difficult to justify since information is lost during the process of transmission, while entropy is increased." http://www.mpiwg-berlin.mpg.de/staff/segal/thesis/thesehtm/chap5/ch5b.htm

[13] Ole E. Barndor -Nielsen, Richard D. Gill. Quantum Statistics. 1996. http://www.math.uu.nl/people/gill/Preprints/chapter2.pdf

[14] Von Neumann Entropy. http://arxiv.org/pdf/math-ph/0102013

[15] Properties of the Von Neumann Entropy. http://red.csie.ntu.edu.tw/QC/peng/slides5.pdf

[16] Entropy of the Spin System. http://www.realtime.net/~drwolf/papers/dissertation/node88.html

[17] OJE Maroney. Information and Entropy in Quantum Theory. Ph.D. Thesis Birkbeck College University of London. http://www.bbk.ac.uk/tpru/OwenMaroney/thesis/thesis.pdf

[18] N. J. Cerf, C. Adami. Accessible information in quantum measurement. http://xxx.lanl.gov/pdf/quant-ph/9611032

[19] http://encyclopedia.thefreedictionary.com/EPR%20paradox

[20] Einstein, A.; Podolsky, B.; and Rosen, N. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" Phys. Rev. 47, 777-780, 1935.

[21] http://www.eleves.ens.fr:8080/home/madore/misc/best_of_GroTeXdieck/EPR. Mars 1999. In french.

[22] http://www.entropysite.com. A site dedicated to the teaching of entropy, based on the idea that entropy is not disorder, but a measure of the dispersal of energy.

[23] D.C. Brody, B. K. Meister. Discrete uncertainty relations. J. Phys. A:Math Gen 32. 1999. http://www.iop.org/EJ/S/UNREG/t7JlkvTukzd,50UZvzQ,aQ/article/0305-4470/32/26/310/a92610.pdf

[24] E.T. Jaynes. Probability Theory, The logic of science. Cambridge University Press. 2003. p. 207-210

[25] D. Domert. 2001. The holographic principle. www.teorfys.uu.se/courses/exjobb/HOLOGRAPHIC.pdf

[26] J. Machta. Entropy, information and computation. University of Massachussets, Amherst 1999.

[27] E.T. Jaynes. Probability Theory, The logic of science. Cambridge University Press. 2003. p.440-450

[28] E.T. Jaynes. Probability Theory, The logic of science. Cambridge University Press. 2003. p.358-365

[29] A Nonequilibrium Statistical Ensemble Formalism. Maxent-Nesom: Concepts, Construction, Application, Open Questions and Criticisms. Sept 1999. http://arxiv.org/PS_cache/cond-mat/pdf/9909/9909160.pdf.

[30] S. Galam. Petite présentation des systèmes désordonnés. In french. http://www.cenecc.ens.fr/EcoCog/Porquerolles2001/bases-galam.html

[31] http://www.psigate.ac.uk/newsite/reference/plambeck/chem2/p02042.htm

[32] T. Valone, Understanding Zero Point Energy. Integrity Research Institute, 1220 L Street NW, Suite 100, Washington. http://users.erols.com/iri/ZPEpaper.html

[33] The Energetic Vacuum: Implications For Energy Research. H.E. Puthoff, PhD. Institute for Advanced Studies, Austin. http://www.ldolphin.org/energetic.html

[34] R. L. Coren, Empirical Evidence for a Law of Information Growth. Entropy 2001, 3. http://www.mdpi.net/entropy/htm/e3040259.htm

[35] P. H. Eberhard, (1977) Nuovo Cimento, 38B, 75 et 46B, 392.

©2005 by MDPI (http://www.mdpi.org). Reproduction for noncommercial purposes permitted.