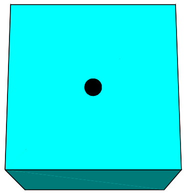

Chapter 1

Introduction

As stated in the preface, the aim of this book is to show that statistical thermodynamics will benefit from the replacement of the concept of entropy by "information" (or any of its equivalent terms; see also Section 1.3 and Chapter 3). This will not only simplify the interpretation of what entropy is, but will also bring an end to the mystery that has befogged the concept of entropy for over a hundred years. In the first section (1.1) of this chapter, we shall briefly survey the historical milestones in the understanding of the key concept of thermodynamics: temperature and entropy. I shall try to convince the reader that had the molecular theory of gases (the so-called kinetic theory of heat) preceded the definition of the absolute temperature, the entropy would have been defined as a dimensionless quantity which, in turn, would have eased the acceptance of the informational interpretation of the entropy. In the following two sections (1.2 and 1.3), we shall discuss the pros and cons of the two main groups of interpretations of the entropy — one based on the idea of order-disorder (or any related concept), and the second on the concept of information or missing information (or any related concept).

1.1 A Brief History of Temperature and Entropy

Temperature and entropy are the two fundamental quantities that make the entire field of thermodynamics deviate from other branches of physics. Both of these are statistical quantities, crucially dependent on the existence of atoms and their properties. Temperature, like pressure, is an intensive quantity; both can be felt by our senses. Entropy, like volume, is an extensive quantity depending on the size of the system. Unlike pressure and volume, temperature and entropy (as well as the Second Law), would not have existed had matter not been made up of an immense number of atoms and molecules. It is true that both temperature and entropy were *defined* and *measured* without any reference to the atomic constituency of matter. However, the *understanding* of these quantities and in fact, their very existence, *is* dependent on the atomic constituency of matter.

Perhaps, the first and the simplest quantity to be explained by the dynamics of the particles is the pressure of a gas. The pressure is defined as the force exerted on a unit area. This definition is valid not only without reference to the atomic constituency of matter, but also if matter were not atomistic at all, i.e., if matter were continuous. Not so for the temperature. Although temperature can be sensed and measured without reference to atoms, its very existence and certainly its explanation is intimately dependent on the atomistic constituency of matter. During the 19^{th} century, it was widely believed that heat was a kind of substance that flows from a hot to a cold body. The association of the notion of temperature with motions of particles was one of the most important achievements of scientists of the late 19^{th} century.

Robert Boyle (1660) found that at a given temperature of the gas, the product of the volume and pressure is constant. This is now known as the Boyle–Marriote law:

$$PV = constant. \tag{1.1.1}$$

Pressure could easily be explained by the particles incessantly bombarding the walls of the container. The pressure $\boldsymbol{P}$ is defined as force $\boldsymbol{F}$ per unit area, i.e.,

$$\boldsymbol{F} = \boldsymbol{P}A. \tag{1.1.2}$$

The force exerted by a moving particle of mass m and velocity v_x colliding with a wall perpendicular to the x-axis is equal to

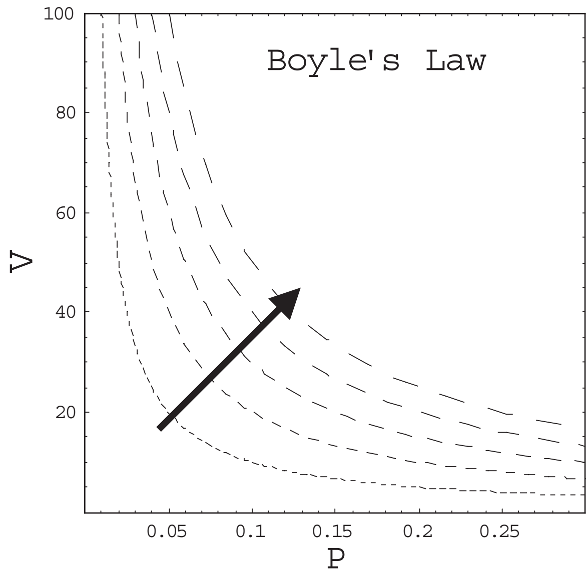

Figure 1.1. Volume as a function of pressure for a gas at different temperatures. Temperature increases in the direction of the arrow.

the change in the momentum per unit of time. An elementary argument[1] leads to the expression for the pressure in terms of the average square velocity of the particles[2]

$$P = \frac{\rho m \langle v^2 \rangle}{3}, \tag{1.1.3}$$

where ρ is the number density of the gas (i.e., the number of particles per unit volume), P is the (scalar) pressure, and

$$\langle v^2 \rangle = \langle v_x^2 \rangle + \langle v_y^2 \rangle + \langle v_z^2 \rangle = 3\langle v_x^2 \rangle. \tag{1.1.4}$$

This explanation was first given by Daniel Bernoulli (1732), long before the mechanical interpretation of the temperature.

Although the sense of hot and cold has been experienced since time immemorial, its measurement started only in the 17$^{\text{th}}$ and 18$^{\text{th}}$ centuries. Various thermometers were designed by Ole Christensen

[1]See, for example, Cooper (1968).

[2]We shall use both the notations $\langle x \rangle$ and $\bar{x}$ for the average of the quantity x.

Rømer (1702), Daniel Gabriel Fahrenheit (1714), Anders Celsius (1742), and many others.

Following the invention of the thermometer, it became possible to make precise measurements of the temperature. In the early 19$^{\text{th}}$ century, Jacques Charles and Joseph Louis Gay-Lussac discovered the experimental law that at constant pressure, the volume is linear in the temperature t (in °C), i.e.,

$$V = C(t + 273), \tag{1.1.5}$$

where C is a constant, proportional to the amount of gas.

From the graphs in Figure 1.2, it is clear that if we extrapolate to lower temperature all the curves converge to a single point. This led to the realization that there exists a minimal temperature, or an absolute zero temperature.

Combining the two empirical laws, one can write the equation of state of the ideal gas as:

$$PV = C \times T = nRT, \tag{1.1.6}$$

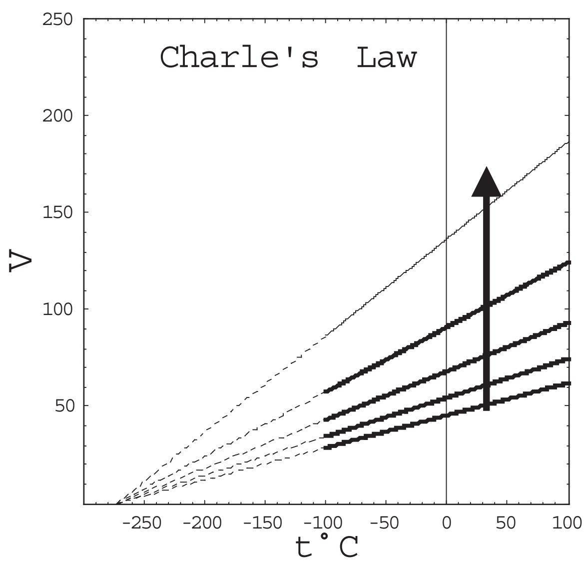

Figure 1.2. Volume as a function of temperature at different pressures. The pressure decreases in the direction of the arrow.

where T is the absolute temperature, n the number of moles, and the "constant," R, for one mole of gas is

$$R = 8.31 \times 10^7 \ \ \text{erg/mol K}. \tag{1.1.7}$$

Having a statistical mechanical theory of pressure, i.e., an interpretation of the pressure in terms of the average kinetic energy of the particles, on the one hand, and the equation of state of the gas on the other, one could derive the expression for the temperature in terms of the average kinetic energy of the particles. Thus, from (1.1.3) and (1.1.6), one can derive the equation

$$k_B T = \frac{2}{3} N \frac{m\langle v^2 \rangle}{2}, \tag{1.1.8}$$

where $k_B = 1.380 \times 10^{-23}$ J/K is now referred to as the Boltzmann constant, and N is the number of particles of the gas. In 1811, Amedeo Avogadro published his hypothesis that the average number of molecules in a given volume of *any* gas is the same (at a fixed temperature and pressure). With this finding, one can relate the Boltzmann constant k_B to the gas constant R by the relation

$$R = N_{AV} k_B \tag{1.1.9}$$

where $N_{AV} = 6.023 \times 10^{23}$ (particles/mol) is the Avogadro number.

Sadi Carnot published a seminal work on the ideal heat engine in 1824.[3] Carnot's prosaic style of writing was reformulated in mathematical terms by Clapeyron (1834). The latter laid the basis for the formulation of the Second Law by Clausius and by Kelvin.

William Thomson, later known as Lord Kelvin, established the existence of the absolute temperature (1854) and showed that by extrapolating from Charles' and Gay-Lussac's law (see Figure 1.2), there exists an absolute zero temperature at about −273°C (today the Kelvin scale is defined in such a way that its value at the triple point of water is, by definition, 273.16 K).

During that period, the First Law of Thermodynamics was established by many workers, notably by Julius Robert Mayer, William Thomsom, James Prescott Joule, and John James Waterston.

[3]Sadi Carnot (1824).

The Second Law was first formulated by Clausius in 1849. A second formulation by William Thomson (Lord Kelvin) followed in 1850. In 1854, Clausius studied the quantity dQ/T in connection with heat engines, where dQ is a small amount of heat transferred and T, the absolute temperature. However, the concept of entropy was introduced only in 1863.

James Clerk Maxwell published the dynamical theory of gases in 1860. This was extended by Ludwig Boltzmann in 1867. These works established both the existence and the form of the equilibrium velocity distribution of the molecules in the gas, known today as the Maxwell–Boltzmann distribution. This distribution established the connection between the average kinetic energy of the molecules with the absolute temperature.

During the late 19$^{\text{th}}$ and the early 20$^{\text{th}}$ centuries, the molecular interpretation of the entropy was developed by Ludwig Boltzmann and by Josiah Willard Gibbs. At that time, the molecular theory of matter was far from being universally accepted. A turning point in the acceptance of the atomistic theory of matter was the publication of the theory of Brownian motion by Albert Einstein (1905) followed by the experimental corroboration by Jean Perrin (1910).

As we have seen in this brief history of temperature and entropy, the concept of temperature, the various thermometers and the various scales came *before* the establishment of the Maxwell–Boltzmann (MB) distribution of velocities. The *acceptance* of the Maxwell–Boltzmann distribution as well as Boltzmann's expression for the entropy came later. It is clear from this sequence of events that the units of temperature were determined at least two decades before the recognition and acceptance of the molecular interpretation of the temperature.

When Clausius formulated the Second Law in terms of dQ/T, it was only natural to define the entropy in units of energy divided by temperature. From that, it followed that Boltzmann's entropy had to be defined with the constant k_B (later known as Boltzmann's constant) having the same units as Clausius' entropy, i.e., energy divided by temperature or Joule/Kelvin, or J/K.

Events could have unfolded differently had the identification of temperature with the average velocity of the atoms come earlier, in which case it would have been more natural to define temperature in units of energy,[4] i.e., one would have arrived at the identity

$$\frac{m\langle v^2\rangle}{2} = \frac{3}{2}T \tag{1.1.10}$$

(instead of $3k_B\, T/2$ on the right-hand side). Having this identification of the temperature would not have any effect on the formal expression of either the efficiency of Carnot's engine,

$$\eta = \frac{T_2 - T_1}{T_1}, \tag{1.1.11}$$

or on the definition of Clausius' entropy,

$$dS = \frac{dQ}{T}. \tag{1.1.12}$$

The only difference would have been that the entropy S would now be a dimensionless quantity. It would still be a state function, and it would still make no reference to the molecular constituency of matter. Clausius could have called this quantity entropy or whatever. It would not have changed anything in the formulation of the Second Law, in either the Clausius or in the Kelvin versions. The molecular interpretation of the entropy would have to wait however, until the close of the 19$^{\text{th}}$ century, when the foundations of statistical mechanics were laid down by Maxwell, Boltzmann and Gibbs.

What would have changed is the formal relationship between the entropy S and the number of states W. To relate the two, there will be no need to introduce a constant bearing the units of energy divided by the temperature. Boltzmann would have made the correspondence between the Clausius' thermodynamic entropy and the statistical entropy simply by choosing a dimensionless constant, or better, leaving the base of the logarithm in the relation $S = \log\, W$ unspecified (it will be shown in Chapter 3 that the choice of

[4] As actually is the practice in many branches of physics and in statistical mechanics. It should be noted that the relation between the temperature and the average kinetic energy of the particles is true for classical ideal gases. It is not true in general. For instance, ideal Fermi gas at T=0 K, has non-zero energy, see also Leff (1999) and Hill (1960).

base 2 has some advantages in connection with the interpretation of entropy in terms of missing information; see Section 3.5).

The removal of the units from Clausius' entropy would not have any effect on the interpretation of the entropy within the classical (or the non-atomistic) formulation of the Second Law. It would have a tremendous effect on the interpretation of Boltzmann's entropy, in terms of probability, and later in terms of the missing information.

The units of Clausius' entropy were, for a long time, the stumbling blocks that hindered the acceptance of Boltzmann's entropy. In one, we have a quantity that is *defined* by the ratio of energy and temperature; in the other, we have a quantity representing the number of states of the system. The two quantities seem to be totally unrelated. Had entropy been *defined* as a dimensionless quantity, the identification between the two entropies and eventually the interpretation of entropy as a measure of information would have become much easier.

The redefinition of the temperature in units of energy will also render the (erroneous) interpretation of the *entropy* as either "heat loss" or "unavailable energy" as superfluous. Consider for simplicity that we have an ideal mono-atomic gas so that all the internal energy consists of only the kinetic energy. We write the Helmholtz energy in the traditional forms as[5]

$$A = E - TS = \frac{3N}{2}k_{\mathrm{B}}T - TS. \tag{1.1.13}$$

In this form, and under certain conditions, TS may be referred to as "heat loss" or "unavailable energy." In the conventional definition of the temperature, it is the *entropy* that bears the "energy" part of the units. Therefore, it is almost natural to ascribe the *energy*, associated with TS, to the entropy S. This leads to the common interpretation of the entropy as either "heat loss" or "unavailable energy" as it features in many dictionaries. If, on the other hand, one defines T in units of energy and S as a dimensionless quantity, this erroneous assignment to S can be avoided. In the example of an ideal gas, we would have written, instead of (1.1.13),

[5]See, for example, Denbigh (1966).

the equivalent relation

$$A = \frac{3}{2}NT - TS, \tag{1.1.14}$$

where T itself is related to the average kinetic energy of the particles by

$$\frac{3}{2}NT = N\frac{m\langle v^2\rangle}{2} \tag{1.1.15}$$

or equivalently

$$T = \frac{m\langle v^2\rangle}{3}. \tag{1.1.16}$$

Hence

$$\begin{aligned} A &= N\frac{m\langle v^2\rangle}{2} - N\frac{m\langle v^2\rangle}{3}S/N \\ &= N\frac{m\langle v^2\rangle}{2}\left[1 - \frac{2S}{3N}\right]. \end{aligned} \tag{1.1.17}$$

In this form, the squared brackets on the right-hand side of (1.1.17) includes the entropy, or rather the missing information (MI). Since S is an extensive quantity, S/N is the MI per particle in this system. The Helmholtz energy A is viewed as the *total* kinetic energy, which when multiplied by the factor in the squared brackets in Equation 1.1.17, is converted to the "free energy."

1.2 The Association of Entropy with Disorder

During over a hundred years of the history of entropy, there have been many attempts to interpret and understand entropy. We shall discuss the two main groups of such interpretations of entropy.

The earliest, and nowadays, the most common interpretation of the entropy is in terms of disorder, or any of the related concepts such as "disorganization," "mixed-upness," "spread of energy," "randomness," "chaos" and the like.

Perhaps, the earliest association of changes in entropy in spontaneous processes with increase of disorder is already implied in Boltzmann's writings. It is commonly believed that Bridgman (1953) was the first to spell out the association of entropy with disorder. Leff (1996, 2007) has recently advocated in favor of the idea of "spread of energy".

The association of entropy with disorder is at best a vague qualitative, and highly subjective one. It rests on observations that in some simple spontaneous processes, when viewed on a molecular level may be conceived as a disordering that takes place in the system. This is indeed true for many processes, but not for all.[6]

Consider the following processes.

Expansion of an ideal gas

Figure 1.3 shows schematically three stages in a process of expanding of an ideal gas.

On the left-hand side, we have N atoms in volume V. In the middle, some of the N atoms have moved to occupy a larger volume $2V$, and on the right-hand side, the atoms are spread evenly in the entire volume $2V$. Take a look. Can you tell which of the three systems is the more ordered? Well, one can argue that the system on the left, where the N atoms are gathered in one half of the volume, is more ordered than N atoms spread in the entire volume. That is plausible when we associate entropy with missing information (see below), but as for order, I personally do not see either of the systems in the figures to be more ordered, or disordered than the other.[7]

The mixing of two different gases

Consider the two systems (Figure 1.4). In the left system, we have N_A blue, and N_B red particles. In the right, we have all the particles

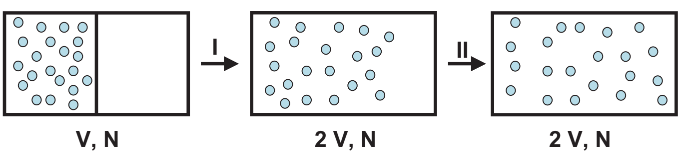

Figure 1.3. Three stages in the process of an expanding ideal gas.

[6]It is interesting to note that Landsberg (1978) not only contended that disorder is an ill-defined concept, but actually made the assertion that "it is reasonable to expect 'disorder' to be an intensive variable."

[7]This is certainly true if you ask people, who have never heard of the concept of entropy, to rate the degree of order in the three systems in Figure 1.3, as the author did.

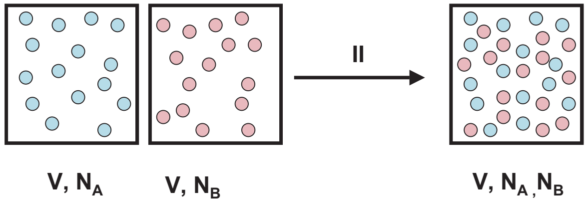

Figure 1.4. Mixing of two different gases.

mixed up in the *same* volume V. Now, which is more ordered? In my view, the left side is more ordered — all the blues and all the reds are separated in different boxes. On the right-hand side, they are mixed up. "Mixed-up" is certainly a disordered state, colloquially speaking. In fact, even Gibbs himself used the word "mix-upness" to describe entropy (see Sections 6.4–6.7). Yet, one can prove that the two systems have *equal* entropy. The association of mixing, with increase in disorder, and hence increase in entropy, is therefore only an illusion. The trouble with the concept of order and disorder is that they are not well-defined quantities — "order" as much as "structure" and "beauty" are in the eyes of the beholder!

"Mixing" of the same gas

Consider the following two processes. In the first (Figure 1.5a), we start with one atom of type A in each compartment of volume V,

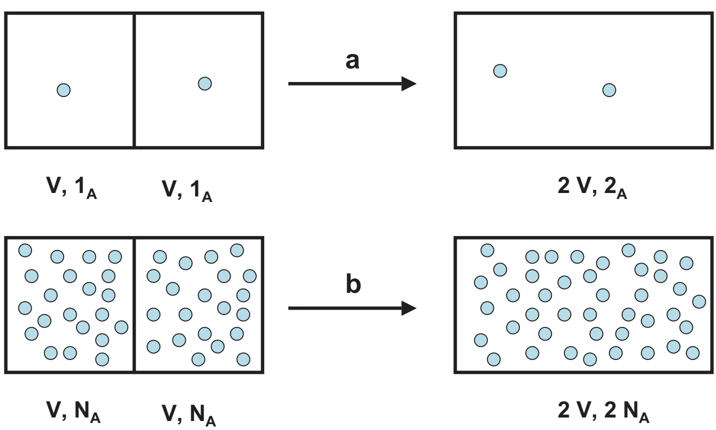

Figure 1.5. Assimilation, or "mixing" of two gases of the same kind.

separated by a partition. In the second, we have N atoms of the same kind in the two compartments. We now remove the partition. Something certainly happens. Can we claim that disorder increases or decreases? Clearly, in Figure 1.5a, the system is initially more "ordered," each particle is in a different box, and in the final stage, the two particles are mixed up. The same holds true for the case of N particles in each box, Figure 1.5b. Therefore, using the concept of "disorder," we should interpret the changes that take place in this process as *increase* in disorder. Yet in terms of thermodynamics, the entropy, in this process, should not be noticeably changed. As we shall see in Chapters 4 and 6, both processes involve an *increase* in MI; it is only in the macroscopic limit of very large N that the change in the MI in the process in Figure 1.5b becomes negligibly small and hence unmeasurable.

Extensivity of disorder?

Consider the two systems in Figure 1.6. We have two boxes of equal volume V and the same number of particles N in each box. Each of the boxes looks disordered. But can we claim that the "disorder" of the *combined* system (i.e., the two boxes in Figure 1.6) is twice the disorder of a single box? It is difficult to make a convincing argument for such an assertion since disorder is not well defined. But even as a qualitative concept, there is no reason to claim that the amount of disorder in the combined system is the sum of the disorder of each system. In other words, it is difficult to argue that disorder should have an additive property, as the entropy and the MI have.

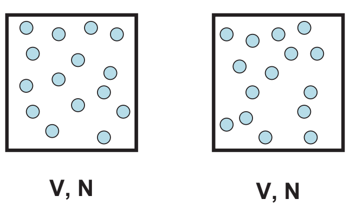

Figure 1.6. Is the extent of the disorder of the combined systems the sum of the disorder of each system?

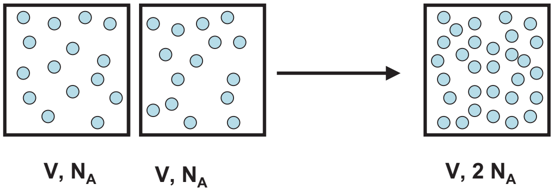

Figure 1.7. A pure assimilation process.

The assimilation process

Consider the process depicted in Figure 1.7. We start with two boxes of the same volume V and the same number of N particles in each. We bring all the particles into one box. In Section 6.3, we shall refer to this process as a pure assimilation process. Here, we are only interested in the question: Is the final state less or more disordered than the initial state? Again, one can argue either way. For instance, one can argue that having N particles in each different box is a more ordered state than having all the $2N$ particles mixed up in one box. However, it can be shown that, in fact, the entropy as well as the MI *decreases* in this process. Therefore, if we associate decrease in entropy with decrease in disorder, we should conclude that the final state is more ordered.

Heat transfer from a hot to a cold gas

As a final example, we discuss here an experiment involving change of temperatures. This experiment is important for several reasons. First, it is one of the classical processes in which entropy increases: in fact, this was the process for which the Second Law was first formulated. Second, it is a good example that demonstrates how difficult it is to argue that disorder is associated with entropy. The process is shown in Figure 1.8. Initially, we have two isolated

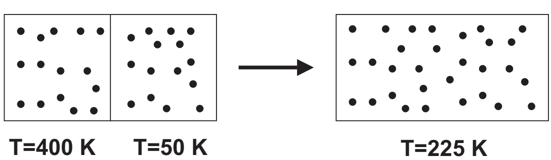

Figure 1.8. Heat transfer from a hot to a cold gas.

systems, each having the same volume, the same number of particles, say Argon, but with two different temperatures $T_1 = 50\,\mathrm{K}$ and $T_2 = 400\,\mathrm{K}$. We bring them into contact. Experimentally, we observe that the temperature of the hot gas will get lower, and the temperature of the cold gas will get higher. At equilibrium, we shall have a uniform temperature of $T = 225\,\mathrm{K}$ throughout the system.

Clearly, heat or thermal energy is transferred from the hot to the cold gas. But can we understand the changes that occur in the system in terms of disorder?

We know that temperature is associated with the distribution of molecular velocities. In Figure 1.9, we illustrate the distribution of velocities for the two gases in the initial state (the left-hand side of Figure 1.9). The distribution is sharper for the lower temperature gas, and is more dispersed for the higher temperature gas. At thermal equilibrium, the distribution is somewhat intermediate between the two extremes, and is shown as a dashed curve in Figure 1.9.

The heat transfer that we observed experimentally is interpreted on a molecular level as the change in the distribution of molecular velocities. Some of the kinetic energies of the hotter gas is transferred to the colder gas so that a new, intermediary distribution is attained at equilibrium.

Now look at the two curves on the right-hand side of Figure 1.9, where we plotted the velocity distribution of the entire system before and after the thermal contact. Can you tell which

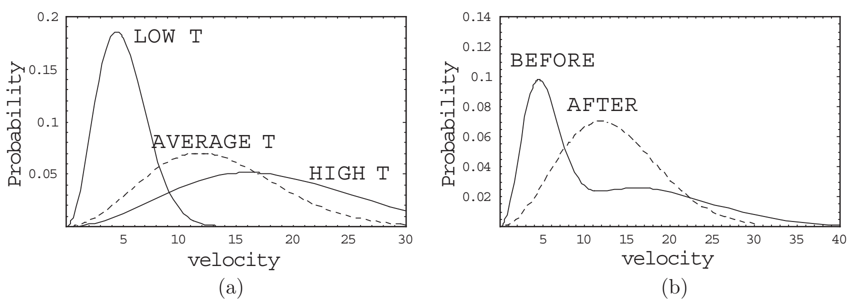

Figure 1.9. Velocity distributions for a gas at different temperatures. (a) The distributions corresponding to the left-hand side of Fig. 1.8. (b) The overall distribution of velocities, before and after the process of heat transfer.

distribution is more ordered or *disordered*? Can you tell in which distribution the *spread* of kinetic energy is more even, or over a larger range of velocities? To me, the final distribution (the dashed curve) looks more ordered, and the distribution looks less spread out over the range of velocities. Clearly, this is a highly subjective view. For this reason and some others discussed in Chapters 3 and 6, I believe that neither "disorder," nor "spread of energy" are adequate descriptions of entropy. On the other hand, information or MI is adequate. As we shall see in Chapter 6, the increase in the entropy in this process can be interpreted as an increase in the MI. It will be shown that the final distribution of velocities is that with the minimum Shannon's *information* or maximum MI. Although this result cannot be *seen* by looking directly at the system, nor by looking at the velocity distribution curves, it can be *proven* mathematically.

Order and disorder are vague and highly subjective concepts. It is true that in many cases, increase in entropy can be qualitatively correlated with increase in disorder, but that kind of correlation is not always valid. Furthermore, the commonly encountered statement that "nature's way is to go from order to disorder," is to say the same as "nature's way is to go from low to high entropy." It does not *explain* why disorder should increase in a spontaneous process. My objection to the association of entropy with disorder is mainly due to the fact that order and disorder are not well-defined and fuzzy concepts. They are very subjective, sometimes ambiguous, and at times totally misleading.

Ever since Shannon introduced the definition of the measure of information, it was found very useful in interpreting entropy (more on this in Section 1.3 and Chapter 3). In my opinion, the concept of missing information not only contributes to our understanding of what is the *thing* that changes and which is called entropy, but it also brings us closer to the last and final step in understanding entropy's behavior (more on this in Section 6.12). This view however is not universal.

On this matter, referring to the informational approach to statistical mechanics, Callen (1984), on page 384 writes:

> "*There is a school of thermodynamics who view thermodynamics as a subjective science of prediction.*"

In a paragraph preceding the discussion of entropy as disorder, Callen explains the origin of this subjectivity and writes:

> "*The concept of probability has two distinct interpretations in common usage. 'Objective probability' refers to a* ***frequency****, or a* ***fractional occurrence****; the assertion that 'the probability of newborn infants being male is slightly less than one half' is a statement about census data. 'Subjective probability' is a measure of* ***expectation based on less than optimum information****. The (subjective) probability of a* ***particular yet unborn*** *child being male,* ***as assessed by a physician****, depends upon that physician's knowledge of the parents' family histories, upon accumulating data on maternal hormone levels, upon the increasing clarity of ultrasound images, and finally upon an educated, but still subjective, guess.*"

I have quoted Callen's paragraph above to show that his argument in favoring "disorder" is essentially fallacious. I believe Callen has mis-applied a probabilistic argument to deem information as "subjective" and to advocate in favor of "disorder," which in his view is "objective" (more on this in Chapter 2).

I am not aware of any precise *definition* of order and disorder that can be used to validate the interpretation of the entropy in terms of the extent of disorder. There is one exception however. Callen (1985) writes on page 380:

> "*In fact, the conceptual framework of 'information theory' erected by Claude Shannon, in the late 1940, provides a basis for interpretation of the entropy in terms of Shannon's measure of* ***disorder****.*"

And further, on page 381, Callen concludes:

> "*For closed systems the entropy corresponds to Shannon's quantitative measure of the maximum possible disorder in the distribution of the system over its permissible microstates.*"

I have taught thermodynamics for many years and used Callen's book as a textbook. It is an excellent book. However, with all due respect to Callen and to his book, I must say that Callen misleads the reader with these statements. Shannon never defined nor referred to "disorder" in his writings. In my opinion, Callen is fudging with the definition of disorder in the quoted statement and in the rest of that chapter. What for? To "*legitimize*" the usage of *disorder* in interpreting entropy. That clearly does not jibe with Shannon's writings. What Callen refers to as Shannon's definition of *disorder* is, in fact, Shannon's definition of *information*. As we shall see in Chapter 3, the measure of "information" as defined by Shannon also retains some of the flavor of the meaning of information as we use in everyday life. This is not the case for disorder. Of course, one can *define* disorder as Callen does, precisely by using Shannon's definition of *information*. Unfortunately, this definition of "disorder" does not have, in general, the *meaning* of disorder as we use the word in our daily lives, as was demonstrated in the examples above.[8]

As we have seen above, even *mixing*, under certain conditions cannot be associated with an increase in disorder nor with the increase in entropy. In fact, we shall show in Chapter 6 that if mixing can be associated with increase in entropy, then also demixing should be associated with increase in entropy. However, mixing by itself has nothing to do with the increase in entropy.

Callen defined "*Shannon's disorder*" by (for more details, see Chapters 2 and 3)

$$\text{disorder} = -\sum p_i \log p_i. \tag{1.2.1}$$

This is clearly a distortion of Shannon's definition of information. Shannon never defined nor discussed disorder in connection with his work on information theory. I would make this statement even stronger by claiming that neither Shannon, nor anyone else *could have* arrived at such a definition of "disorder."

As we shall see in Chapter 3, Shannon did not start by defining information (or *choice* or *uncertainty*, certainly not *disorder*). What Shannon did was to start with the general concept of "information"

[8]McGlashan (1979) has also expressed reservations regarding the association of entropy with disorder. He excluded two cases however: "When, if ever, has the (number of states) anything to do with any of these words like mixed-upness." He then continues: "It does, but only for two very special cases. These are mixtures of perfect gases, and crystals at temperatures close to zero."

or rather the quantity of information transmitted along communication lines. Shannon then asked himself: if a quantitative measure of information exists, what properties must it have? After pronouncing several plausible properties of such a measure, he found that such a quantity must have the form $-\sum p_i \ln p_i$ (see Chapter 3). Had he set his aim to construct a measure of *disorder*, it is hard to believe that he, or anyone else, would have reached the same definition for *disorder*. In fact, it is not clear at all what requirement should a measure of disorder fulfill. It is not clear that the property of additivity required from information would also be required for disorder[9]; it might also be meaningless to require the "consistency property" (or the "independence on grouping" — see Chapter 3) from a quantity such as disorder.

Thus, even if Shannon would have pondered the plausible requirements for "disorder" to make it a quantitative measure, it is unlikely that he would have arrived at the definition (1.2.1), as claimed by Callen. Moreover, as we shall see in Chapter 3, many of the *derived* properties of information, such as conditional information, mutual information and the like would not and could not be assigned to disorder.

It is unfortunate and perhaps even ironic that Callen dismisses "information" as subjective, while at the same time embracing Shannon's definition of information, but renaming it as disorder. By doing that, he actually replaces a well-defined, quantitative and objective quantity by a more subjective concept of disorder. Had Callen not used *Shannon's definition of information*, the concept of disorder would have remained an undefined, qualitative and highly subjective quantity.

In my view, it does not make any difference if you refer to *information* or to *disorder*, as subjective or objective. What matters is that order and disorder are not well-defined scientific concepts. On the other hand, information is a well-defined scientific quantity, as

[9]See examples above. Note also that Callen in his 1987 edition, had made a list of requirements similar to those made by Shannon that "disorder" should subscribe to. In my opinion these requirements, although valid for information, are not even plausible for disorder. Similar reference to "Shannon's measure of disorder" may be found in Lee (2002).

much as a point or a line are *scientific* in geometry, or mass or charge of a particle are *scientific* in physics.

In concluding, we can say that increase in disorder (or any of the equivalent terms) can sometimes, but not always, be associated qualitatively with increase in entropy. On the other hand, "information" or rather MI can *always* be associated with entropy, and therefore it is superior to disorder.

1.3 The Association of Entropy with Missing Information

As in the case of order and disorder, the involvement of the notion of information is already implied in Boltzmann's expression for the entropy. It is also implicit in the writings of Maxwell. Maxwell was probably the first to use probabilistic ideas in the theory of heat. As we shall see in Chapter 3, probability is only a short step from information.

More explicitly, although not stated as such, the idea of Maxwell's demon, a being that can defy the Second Law, contains elements of the notion of information.

The demon was introduced by James Clerk Maxwell in the 1871 book, "Theory of Heat"[10]:

> "*Starting with a uniform temperature, let us suppose that such a vessel is divided into two portions or by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole so as to allow only the swifter molecules to pass from A to B, and only the slower ones pass from B to A. He will thus, without expenditure of work raise the temperature of B and lower that of A in contradiction to the second law of thermodynamics.*"[11]

---

[10]From Leff and Rex (1990), *Maxwell's Demon, Entropy, Information, Computing*, Adam-Hilger, Bristol, UK.

[11]It should be noted that in this statement Maxwell recognized the relation between the temperature and the average kinetic energies of the particles. Note also that intuitively, Maxwell felt that if such a Demon could have existed, he could use his *information* to achieve a process that violates the Second Law of Thermodynamics.

William Thomson (1874) referred to this imaginary being as "Maxwell's intelligent demon." Although Maxwell himself did not discuss the connection between the Second Law and information, this is implicitly implied by his description of what the demon is capable of doing, i.e., using information to lower the entropy.

Szilard (1929) in analyzing the implications of Maxwell's demon to the Second Law, referred to the "intervention of an intelligent being who uses information to defy the Second Law." Many articles have been written on Maxwell's demon and entropy.[12] However, the association of Maxwell's demon with information and with entropy was quite loose. The association of entropy with information was firmed up and quantified only after Shannon introduced the quantitative measure of information in 1948.

Previous to 1948, perhaps the most eloquent and explicit identification of entropy with information (albeit not the quantitative Shannon's measure) is the statement of G. N. Lewis in 1930:

> "*Gain in entropy always means loss of information, and nothing more. It is a subjective concept, but we can express it in a least subjective form, as follows. If, on a page, we read the description of a physico-chemical system, together with certain data which help to specify the system, the entropy of the system is determined by these specifications. If any of the essential data are erased, the entropy becomes greater; if any essential data are added, the entropy becomes less. Nothing further is needed to show that the irreversible process neither implies one-way time, nor has any other temporal implications. Time is not one of the variables of pure thermodynamics.*"

This is an almost prophetic statement made eighteen years before information theory was born. Lewis' statement left no doubt that he considered entropy as *conceptually* identical with information. Note however that Lewis claims that entropy is a subjective concept. This is probably a result of his usage of the *general* concept of information and not the specific measure of information as defined by Shannon.

[12]See Leff and Rex (1990).

The first to capitalize on Shannon's concept of information was Brillouin. In his book "Science and Information" (1956), Brillouin expressed the view that Clausius' entropy and Shannon's entropy are identical. Brillouin (1962) also suggested referring to information as neg-entropy. In my view, this amounts to replacing a simple, familiar and informative term with a vague and essentially misleading term. Instead, I would have suggested replacing entropy with either neg-information, missing information, or uncertainty.

Ilya Prigogine (1997) in his recent book "End of Certainty" quotes Murray Gell–Mann (1994) saying:

> "*Entropy and information are very closely related. In fact, entropy can be regarded as a measure of ignorance. When it is known only that a system is in a given macrostate, the entropy of the macrostate measures the degree of ignorance the microstate is in by counting the number of bits of additional information needed to specify it, with all the microstates treated as equally probable.*"[13]

I fully agree with the content of this quotation by Gell–Mann, yet Ilya Progogine, commenting on this very paragraph writes:

> "*We believe that these arguments are untenable. They imply that it is our own ignorance, our coarse graining, that leads to the second law.*"

Untenable? Why?

The reason for these two diametrically contradictory views by two great scientists has its sources in the confusion of the *general* concept of information, with the specific measure of information as defined by Shannon. I shall discuss this issue in greater detail below and in Chapter 3.

In my opinion, Gell–Mann is not only right in his statement, he is also careful to say "entropy *can* be regarded as a measure of ignorance ... Entropy ... measures the degree of ignorance." He does not say *"our own ignorance,"* as misinterpreted by Prigogine.

Indeed information, as we shall see in Chapter 3, is a measure that *is there* in the system. Within information theory, the term

[13] Gell–Mann (1994).

"information" is not a subjective quantity. Gell–Mann uses the term "ignorance" as a synonym to "lack of information." As such, ignorance is also an objective quantity that belongs to the system and is not the same as "*our own ignorance*," which might or might not be an objective quantity.

The misinterpretation of the information-theoretical entropy as a subjective information is quite common. Here is a paragraph from Atkins' preface from the book "The Second Law."[14]

> "*I have deliberately omitted reference to the relation between information theory and entropy. There is the danger, it seems to me, of giving the impression that entropy requires the existence of some cognizant entity capable of possessing 'information' or of being to some degree 'ignorant.' It is then only a small step to the presumption that entropy is all in the mind, and hence is an aspect of the observer. I have no time for this kind of muddleheadedness and intend to keep such metaphysical accretions at bay. For this reason I omit any discussion of the analogies between information theory and thermodynamics.*"

Atkins' comment and his rejection of the informational interpretation of entropy on the grounds that this "relation" might lead to the "presumption that entropy is all in the mind" is ironic. Instead, he uses the terms "disorder" and "disorganized," etc., which in my view are concepts that are far more "in the mind."

The fact is that there is not only an *analogy* between entropy and information, but an *identity* between the thermodynamic entropy and Shannon's measure of information.

The reason for the confusion is that the term information itself has numerous interpretations. We can identify at least three "levels" in which we can interpret the term "information." At the most general level, information is any knowledge that we can obtain by our senses. It is an abstract concept which may or may not be subjective. The information on "the weather conditions in New York state" might have different significance, meaning or value to different persons. This information is *not* the subject of interest of information theory. When Shannon sought a quantity to measure

[14]Atkins (1984).

information transmitted across communication lines, he was not interested in the *content* or the *meaning* of information, but in a quantity that measures the *amount* of information that is being transmitted.

Leaving aside the *content* or the meaning of information and focusing only on the *size* or the amount of information, we are still far from Shannon's measure of information. One can speak of different amounts of information, or information on different amounts of something. Consider the following two messages:

A: Each of the ten houses in the street costs one million dollars.
B: The first house in this street costs one million dollars, the second house costs one million dollars,... and the tenth house costs one million dollars.

Clearly, the two messages A and B carry the same information. The *size* of A is however much smaller than the size of B. Consider the next message:

C: Each of the houses in this town costs one million dollars.

Clearly, C contains more information than B. The message C tells us the price of more houses than the message B, yet it is much shorter than B (shorter in some sense that will be discussed in Chapter 3).
Consider the next message:

D: Each of the houses in this *country* costs one billion dollars.

This message conveys information on *more houses* and on *more money*, yet it is roughly of the same size as the messages C or A.

Information theory is neither concerned with the content of the message, nor with the amount of information that the message conveys. The only subject of interest is the *size* of the message itself. The message can carry small or large amounts of information, it can convey important or superfluous information; it may have different meanings or values to different persons or it can even be meaningless; it can be exact and reliable information or approximate and dubious information. All that matters is some measure of the size of the message. In Chapter 3, we shall make the last statement more quantitative. We shall also see that the size of the message can be expressed as the number of binary questions one needs to ask in order to retrieve the message. We shall be interested in the information *on* the message and not on the information carried by the message. We shall also see that entropy is nothing but the amount of missing information (MI).

Thus, neither the *entropy*, nor the Shannon measure of MI, are subjective quantities. In fact, no one has claimed that either of these is subjective. The subjectivity of information enters only when we apply the concept of information in its broader sense.

Jaynes pioneered the application of information theory to statistical mechanics. In this approach, the fundamental probabilities of statistical mechanics (see Chapter 5) are obtained by using the principle of maximum entropy. This principle states that the equilibrium distributions are obtained by maximizing the entropy (or the MI) with respect to all other distributions. The same principle can be applied to derive the most non-commital, or the least biased, distribution that is consistent with all the given information. This is a very general principle that has a far more general applicability. See also Chapters 4–6.

In his first paper on this subject, Jaynes wrote (1957)[15]:

> "*Information theory provides a constructive criterion for setting up probability distributions on the basis of partial knowledge and leads to a type of statistical inference which is called the maximum-entropy estimate.*"

And he added:

> "*Henceforth, we will consider the terms 'entropy' and 'uncertainty' as synonyms." "The thermodynamic entropy is identical with information theory — entropy of the probability distribution except for the presence of Boltzmann's constant.*"
>
> "*. . . we accept the von-Neumann–Shannon expression for entropy, very literally as a measure of the amount of uncertainty represented by the probability distribution; thus entropy becomes the primitive concept. . . more fundamental than energy.*"

[15]It should be noted that Jaynes' approach was criticized by several authors, e.g., Friedman and Shimony (1971) and Diaz and Shimony (1981).

Regarding the question of the subjectivity of "information," Jaynes writes:

> "*In Euclidean geometry the coordinates of a point are 'subjective' in the sense that they depend on the orientation of the observer's coordinate system; while the distance between two points is 'objective' in the sense that it is independent of the observer's orientation.*"

In 1983 Jaynes writes[16]:

> "*The function H is called entropy, or better, information entropy of the distribution* $\{p_{\mathrm{i}}\}$. *This is an unfortunate terminology which now seems impossible to correct. We must warn at the outset that the major occupational disease of this field is a persistent failure to distinguish between information entropy, which is a property of any probability distribution, and experimental entropy of thermodynamics, which is instead a property of a thermodynamics state as defined, for example, by such observed quantities as pressure volume, temperature, magnetization of some physical system. They should never have been called by the same name; the experimental entropy makes no reference to any probability distribution, and the information entropy makes no reference to thermodynamics. Many textbooks and research papers are fatally flawed by the author's failure to distinguish between these entirely different things, and in consequence proving nonsense theorems.*"
>
> "*The mere fact that the same mathematical expression* $-\sum p_i \log p_i$ *occurs both in statistical mechanics and information theory does not in itself establish any connection between these fields. This can be done only by finding new viewpoints from which thermodynamic entropy and information theory entropy appear as the same concept.*"

And later, Jaynes writes:

> "*It perhaps takes a moment of thought to see that the mere fact that a mathematical expression like*
>
> $$-\sum p_i \log p_i$$
>
> *shows up in two different fields, and that the same inequalities are used in two different fields does not in itself establish any*

[16] Jaynes (1983).

> *connection at all between the fields. Because, after all* e^x, $\cos\theta$, $J_0(z)$ *are expressions that show up in every part of physics and engineering. Every place they show up... nobody interprets this as showing that there is some deep profound connection between, say, bridge building and meson theory.*"

I generally agree with the aforementioned statements by Jaynes. Indeed, the fact that $\cos\theta$ appears in two different fields, say in the propagation of an electromagnetic wave, and in the swinging of a pendulum, does not imply a deep and profound connection between the two fields. However, in both fields, the appearance of $\cos\theta$ indicates that the *phenomena* are *periodic*.

Likewise, the appearance of $-\sum p_i \log p_i$ in two different fields does not indicate that there exists a deep profound connection between the two fields. This is true not only between communication theory and thermodynamics. The measure of information $-\sum p_i \log p_i$ in linguistics makes no reference to the distribution of coins in boxes, or electrons in energy levels, and the measure of information $-\sum p_i \log p_i$ in thermodynamics makes no reference to the frequencies of the alphabet letters in a specific language; the two fields, or even the two subfields (say, in two different languages) are *indeed* different. The information is about different *things*, but all are measures of information nonetheless!

Thus, although it is clear that the types of information discussed in thermodynamics and in communication theory are different, they are all measures of information. Moreover, even in the same field, say in analyzing the information in two languages in linguistics, or even the locational and the momenta information of a thermodynamics system, the *types* of information are different. But in all cases that the expression $-\sum p_i \log p_i$ appears, it conveys the meaning of a measure of information. The information is *about* different things in each case, but conceptually they are measures of information in all cases.

To conclude, the concept of information is very general: it can be applied to many different fields, and in many subfields. The same is true of periodic phenomena which are very general and can occur in many fields. However, the significance of the concept of information and the significance of the periodic phenomena are the same in whatever fields they happen to appear. Specifically, the entropy, or better the thermodynamic entropy, or even better the thermodynamic missing information, is one particular example of the general notion of information.

There is of course a conceptual difficulty in identifying the Boltzmann entropy with the Clausius entropy. However, it has been shown in numerous examples that *changes* in one concept are equivalent to changes in the other, provided that the constant k is chosen as the Boltzmann constant, with the dimensions of energy over absolute temperature (K). Once one accepts the identification of the Clausius entropy with the Boltzmann entropy, then the interpretation of the entropy as uncertainty, information or MI is inevitable.

It is now about 100 years since Gibbs and Boltzmann developed statistical mechanics *based* on probability. It is over fifty years since Shannon laid down the foundation of information theory *based on* probability. Jaynes used the measure of information as defined by Shannon to develop statistical mechanics from the principle of maximum entropy.

I have already mentioned von Neumann's "*disservice*" to science in suggesting the term entropy.[17] My reason for agreeing with this statement is different to the ones expressed by Denbigh. I would simply say that I shall go back to Clausius' choice of the term, and suggest that he should have not used the term entropy in the first place. This term was coined at the time of the pre-atomistic view of thermodynamics. However, once the foundations of statistical mechanics were firmly established on probability and once information theory was established, von Neumann should have suggested to the scientific community *replacing* entropy by information (or by uncertainty or unlikelihood; see Chapter 3). In my view, the suggestion of the term entropy, is not only a disservice, but a corruption of a well-defined, well-interpreted, intuitively appealing and extremely useful and general concept of information with a term that means *nothing*.[18]

[17]See preface, page xviii
[18]See preface, page xvi

In an article entitled "How subjective is entropy?", Denbigh discussed the extent of subjectivity of information and the objectivity of entropy.[19]

Indeed there is a valid question regarding the two types of probabilities, subjective and objective. However, once we restrict ourselves to discussing only scientific probabilities, we regard these as objective quantities in the sense that everyone given the same information, i.e., the same conditions, will necessarily reach the same conclusion regarding the probabilities. See more on this in Chapter 2.

Information is defined in terms of probability distribution. Once we agree to deal only with scientific and objective probabilities, the corresponding measure of information becomes objective too.

The best example one can use to define Shannon's measure of information is the game of hiding a coin in M boxes (see Chapter 3). I hide a coin in one of the M boxes and you have to ask binary questions to be able to locate the coin. The missing information is: "Where is the coin?" Clearly, since I know where the coin is while you do not know where it is, I have *more information* on the location of the coin than you have. This information is clearly subjective — your ignorance is larger than mine. However, that kind of information is not the subject matter of information theory. To define the missing information, we have to formulate the problem as follows: "Given that a coin is hidden in one of the M boxes, how many questions does one need to ask to be able to find out where the coin is?" In this formulation, the MI is *built-in* in the problem. It is as objective as the *given* number of boxes, and it is indifferent to the person who hid the coin in the box.

Denbigh and Denbigh (1985), who thoroughly discussed the question of the subjectivity of information, asked:

> "*Whose information or uncertainty is being referred to? Where does the entropy reside?*"

[19]Denbigh, K. (1981), "How subjective is entropy," *Chemistry in Britain* **17**, 168–185.

And their conclusion was:

> "*The problem about whose information is measured by H remains obscure.*"

In my view, the question of "whose information is being referred to" is not relevant to the quantity H, as defined by Shannon. Regarding the second question: Wherever the *number of boxes* (or the number of states of a thermodynamic system) "resides," there resides also the quantity H. Again, this question of the residency of H is irrelevant to the meaning of H.

Before ending this section on entropy and information, I should mention a nagging problem that has hindered the acceptance of the interpretation of entropy as information. We recall that entropy was *defined* as a quantity of heat divided by temperature. As such, it bears the units of energy divided by the absolute temperature (K) (i.e., Joules over K or J/K, K being the unit of the absolute temperature on the Kelvin scale). These are two tangible, measurable and well-defined concepts. How come "information," which is a dimensionless quantity, a number that has nothing to do with either energy or temperature, could be associated with entropy, a quantity that has been *defined* in terms of energy and temperature? I believe that this is a very valid point of concern which deserves some further pondering. In fact, even Shannon himself recognized that his measure of information becomes identical with entropy only when it is multiplied by the constant k (now known as the Boltzmann constant), which has the units of energy divided by temperature. This, in itself, does not help much in identifying the two apparently very different concepts. I believe there is a deeper reason for the difficulty of identifying entropy with information.

First, note that in the process depicted in Figure 1.8, the change in entropy does involve some quantity of heat (energy) transferred and temperature. But this is only one example of a spontaneous process. Consider the expansion of an ideal gas in Figure 1.3 or the mixing of two ideal gases in Figure 6.6. In both cases, entropy increases. However, in both cases, there is no change in energy, no heat transfer, and no dependence on temperature. If we carry out these two processes for an ideal gas as in an isolated system, then the entropy change will be fixed, independent of the temperature of the system, and obviously no heat is transferred from one body to another. These examples are only indicative that changes in entropy do not *necessarily* involve units of energy and temperature.

Second, the units of entropy (J/K) are not only unnecessary for entropy, but they *should not* be used to express entropy at all. The involvement of energy and temperature in the original definition of the entropy is a historical accident, a relic of the pre-atomistic era of thermodynamics.

Recall that temperature was defined earlier than entropy and earlier than the kinetic theory of heat. Kelvin introduced the absolute scale of temperature in 1854. Maxwell published the molecular distribution of velocities in 1859. This has led to the *identification* of temperature with the mean kinetic energy of the atoms or molecules in the gas. Once the identification of the temperature as a measure of the average kinetic energy of the atoms had been confirmed and accepted,[20] there was no reason to keep the old units of K. One should redefine a new absolute temperature, denote it tentatively as $\overline{T}$, to replace kT. The new temperature $\overline{T}$ would have the units of energy and there should be no need for the Boltzmann constant. The equation for the entropy would be simply $S = \ln W$,[21] and entropy would be rendered dimensionless!

Had the kinetic theory of gases preceded Carnot, Clausius and Kelvin, Clausius would have defined changes of entropy as energy divided by temperature. But now this ratio would have been dimensionless. This will not only simplify Boltzmann's formula for the entropy, but will also facilitate the *identification* of the *thermodynamic* entropy with Shannon's measure of information.

Thus, without entering into the controversy about the question of the subjectivity or objectivity of information, whatever it is, I believe that the entropy is *identical*, both conceptually *and* formally, with Shannon's measure of information. This identification is rendered possible by redefining temperature in units of

[20]see footnote 4 on page 7.

[21]see cover design.

energy.[22] This will automatically expunge the Boltzmann constant k_B from the vocabulary of physics. It will simplify the Boltzmann formula for the entropy, and it will remove the stumbling block that has hindered the acceptance of entropy as information for over a hundred years. It is also time to change not only the units of entropy to make it dimensionless,[23] but the term "entropy" altogether. Entropy, as it is now recognized does not convey "transformation," nor "change," nor "turn." It does convey some measure of *information*. Why not replace the term that means "nothing" as Cooper noted, and does not even convey the meaning it was meant to convey when selected by Clausius? Why not replace it with a simple, familiar, meaningful, and precisely defined term?" This will not only remove much of the mystery associated with the unfamiliar word entropy, but will also ease the acceptance of John Wheeler's view of regarding "*the physical world as made of information, with energy and matter as incidentals*."[24]

Finally, it should be said that even when we identify entropy with information, there is one very important difference between the thermodynamic information (entropy) and Shannon's information, which is used in communications or in any other branch of science. There is a huge difference of the order of magnitudes between the two. The concept of information is, however, the same in whatever field it is being used. We shall discuss in Chapter 4 the passage from a system with a small number of degrees of freedom to thermodynamic systems having a huge number of degrees of freedom.

[22] As is effectively done in many fields of physics.

[23] Note that the entropy would still be an extensive quantity, i.e., the entropy of the system would be proportional to the size of the system.

[24] Quoted by Jacob Bekenstein (2003).