

Preface

Nowadays, it is difficult to justify the writing of a new textbook on statistical thermodynamics. A quick glance at the bibliography at the end of this book shows an abundance of such textbooks. There are also a few books on statistical thermodynamics that use information theory such as those by Jaynes, Katz, and Tribus. These books use the principle of maximum entropy to “guess” the “best” or the least biased probability distributions of statistical mechanics.

The main purpose of this book is to go one step forward, not only to use the principle of maximum entropy in predicting probability distributions, but to replace altogether the concept of entropy with the more suitable concept of information, or better yet, the missing information (MI).

I believe that the time is ripe to acknowledge that the term *entropy*, as originally coined by Clausius, is an unfortunate choice. Moreover, it is also a misleading term both in its meaning in ancient and in contemporary Greek.¹ On this matter, I cannot do any better

¹In the Merriam-Webster Collegiate Dictionary (2003), “Entropy is defined as: “change,” literary “turn,” a measure of the “unavailable energy” in a closed thermodynamic system. . . a measure of the system’s degree of order. . .”

From Merriam-Webster online:

1. A measure of the unavailable energy in a closed thermodynamic system.
2. A measure of the system’s disorder.
3. The degradation of matter and energy in the universe to an ultimate state of inert uniformity. $\epsilon\nu = \textit{in}$, $\tau\rho\sigma\eta = \textit{trope} = \textit{transformation}$ $\epsilon\nu\tau\rho\sigma\iota\alpha = \textit{entropy} = \textit{transformation inwards}$ in modern usage, *entropy*: turn into, or turn to be or evolves into, the way something will turn out, will change; could be *evolves*.

than Leon Cooper (1968). Cooper cites the original passage from Clausius: in choosing the word “Entropy,” Clausius wrote:

*“I prefer going to the ancient languages for the names of important scientific quantities, so that they mean the same thing in all living tongues. I propose, accordingly, to call S the **entropy** of a body, after the Greek word “**transformation**.” I have designedly coined the word entropy to be similar to **energy**, for these two quantities are so analogous in their physical significance, that an analogy of denominations seems to be helpful.”*

Right after quoting Clausius’ explanation on his reasons for the choice of the word “Entropy,” Cooper commented:

*“By doing this, rather than extracting a name from the body of the current language (say: **lost heat**), he succeeded in coining a word that meant the same thing to everybody: **nothing**.”*

I fully agree with Cooper’s comment; however, I have two additional comments, and contrary to Cooper, I venture into taking the inevitable conclusion:

First, I agree that “entropy means the same thing to everybody: *nothing*.” But more than that, entropy is also a misleading term. The two quantities “energy” and “entropy” are not analogous in their physical significance; hence, there is no reason for using analogous denominations.

Second, I do not believe that Cooper’s apparently casual suggestion that “lost heat” might be a better choice, is a good idea, as much as the more common “unavailable energy” interpretation attached to “entropy” in most dictionaries.

As we shall discuss in Chapter 1, both the “heat loss” and “unavailable energy” may be applied under certain conditions to $T\Delta S$ but not to entropy. The reason it is applied to S rather than to $T\Delta S$, is that S , as presently defined, contains the units of energy and temperature. This is unfortunate. If entropy had been recognized from the outset as a measure of information, or of uncertainty, then it would be dimensionless, and the burden of carrying the units of energy would be transferred to the temperature T .

Finally, I believe that the time has come to reach the inevitable conclusion that “entropy” is a misnomer and should be replaced by either *missing information* or *uncertainty*. These are more appropriate terms for what is now referred to as “entropy.”

Unfortunately, there is a vigorous ongoing debate on the very *interpretation* of entropy as information, let alone the replacement of entropy by information. This aspect will be discussed at length in Chapter 1. In Chapter 1, I shall also discuss the more common interpretation of entropy as disorder, mixed-upness, disorganization, chaos and the like. In my opinion all these terms are also inappropriate interpretations of entropy.

Today, the concept of entropy is used in many fields far from, and unrelated to, thermodynamics. Even in a superficial survey of the applications of the term entropy in various fields, from communications to economics, sociology to psychology, linguistics, arts and many more, one immediately realizes that the *concept* that is used is *information*, and not entropy as defined by Clausius and as is used in thermodynamics.

I can understand the continual application of the term “entropy” by practitioners in thermodynamics and statistical mechanics. It is a tradition that is hard to change.

I fail to understand why so many authors use the term entropy where in fact what they really mean is *information or uncertainty* (that includes Shannon himself who, as the story goes, renamed his measure for information as entropy; see below). To me, the usage of the term entropy is a corruption of the meaningful concept of information. Entropy does not mean what it meant in ancient Greek, does not mean what it presently means in modern Greek, and does not mean what it was supposed to mean when Clausius made this unfortunate choice. To use a concept that means “nothing,” to replace a simple, familiar and meaningful concept such as information, by entropy, is at best, a perversive practice. The origin of this practice is found in Tribus² story:

“What’s in a name? In the case of Shannon’s measure the naming was not accidental. In 1961 one of us (Tribus) asked Shannon

²Tribus (1971).

what he had thought about when he had finally confirmed his famous measure. Shannon replied: “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. In the second place, and more important, no one knows what entropy really is, so in a debate you will always have the advantage.’”

On von Neumann’s suggestion, Denbigh (1981) comments:

“*In my view von Neumann did science a disservice!*” adding “there are, of course, good mathematical reasons why information theory and statistical mechanics both require functions having the same formal structure. They have a common origin in probability theory, and they also need to satisfy certain common requirements such as additivity. Yet, this formal similarity does not imply that the functions necessarily signify or represent the same concepts. The term ‘entropy’ had already been given a well-established physical meaning in thermodynamics, and it remains to be seen under what conditions, if any, thermodynamic entropy and information are mutually inconvertible.”

I agree with the first sentence by Denbigh and indeed: *In my view von Neumann did science a disservice.*

My reason for embracing Denbigh’s statement is that *information* (or choice, or uncertainty) is a simple, familiar, meaningful and well-defined concept. Renaming it *entropy* merely corrupts the term information. I shall delve into this more in Chapter 1.

The term “information” is indeed a far more general, and “overly used,” concept than the quantity defined by Shannon. This is probably the reason why Shannon sought the advice of von Neumann for an appropriate term for the quantity he defined to measure information (or choice, or uncertainty). It would have been helpful if an appropriate term could have been found that had retained the qualitative properties of information on the one hand, yet was restricted to usage for that specific quantity as defined by Shannon, on the other. Unfortunately, the choice of “entropy” certainly does not fulfill this requirement, and in my opinion is an

inadequate term. The choice made in this book is “missing information,” or MI, or better yet, the amount of MI. This term captures the meaning of the concept sought by Shannon as a measure of the missing information, and therefore it is an appropriate term to replace the traditional term entropy. Perhaps, in the future, a term that combines the concepts of “measure” and of “information” (or of “measure” and of “uncertainty”) should be coined. After all, Shannon himself did not discuss the *information* itself, but some measure of the size of a message that may or may not carry information. Perhaps a term like “**enformetry**,” which has a part from **entropy**, a part the root word of **information**, and **metry** that indicates a measure of the *size* of the message, would serve better than entropy. In Section 3.3, we shall see that the most appropriate interpretation of “entropy” is the number of binary questions that one needs to ask to acquire the missing information.

It seems to me that if the absolute temperature had been defined *after* the development of the kinetic theory of gases, and recognized as a measure of the average kinetic energy of the particles, it would have been bestowed with the units of energy.³ These units are more “natural” and more appropriate units for the temperature. Having temperature in units of energy would automatically render the Boltzmann constant superfluous and the Boltzmann entropy a dimensionless quantity. Such a dimensionless quantity would still be a state function. In this case, it would be easier to accept the interpretation of entropy as information. Furthermore, it would ease the acceptance of information as a cornerstone of the fabric of reality among matter and energy. The replacement of “entropy” by “information,” in itself, would not provide an explanation of the Second Law of Thermodynamics. However, being an intangible quantity, it would be easier to accept the fact that information, or rather the MI increases in one direction only; it would also remove much of the mystery associated with entropy.

The aim of this book is to develop statistical thermodynamics from the cornerstone concept of information, as defined by

³This is true for ideal gases. See also Section 1.1. Note that “*units of energy*”, is not the same as the *energy* of the system.

Shannon, not only as a means of guessing the best (or the least biased) probability distributions, but to replace entropy altogether.

To achieve that, I have allotted more space to present the elements of probability and information theory than to the foundations of statistical mechanics.

In Chapter 1, I will present the pros and cons of the usage of information in statistical mechanics. I will also discuss briefly the usage of disorder and related concepts in connection with entropy. Chapter 2 presents the elements of probability theory, enough to convince the reader of its usefulness and its beauty, enough to understand what information, as defined by Shannon, is. Probability theory is a rich, fascinating and extremely useful branch of mathematics. I have chosen some examples as exercises (most worked out in detail). Some are simple and straightforward, while some are more difficult, or “brain-teasers.” All are useful for understanding the rest of the book. In Chapter 3, I will present the concept of information, the definition, the meanings and the application of this concept in statistical mechanics. We shall not touch on the application of information in communication theory, but only on those aspects that are relevant and potentially useful to statistical thermodynamics.

Information as defined by Shannon is a real objective quantity as much as mass and energy in physics, a point or a circle in geometry. In fact, I believe that the measure of information as defined by Shannon is more “objective” than a “mass” in physics, or a “point” in geometry. Different persons would measure different masses of the same piece of matter; it is hard to claim an *exact* value of its mass.⁴ Similarly, a point in geometry is conceived as an idealization, or as a mathematical limit, but not a real physical object. However, a coin hidden in one of eight identical boxes *defines* a Shannon measure of the MI, which is $\log_2 8 = 3$. It is a real, objective and exact quantity. It would be futile to argue that this measure is “subjective,” or “all in the mind,” as claimed by some authors (see also Chapter 1). Of course, information, as uncertainty, is always *about* something. That “something” could be chosen in different

⁴Denbigh and Denbigh (1985).

ways even for the same system. In information theory, that “something” is irrelevant to the theory. What matters is only the *size* of the message carrying the information. (See also Chapter 3.)

If one accepts the probabilistic interpretation of the entropy, and agrees on the meaning of Shannon’s information, then the interpretation of the *thermodynamic entropy* as *thermodynamic information* becomes inevitable.

It is sometimes argued that information is not uncertainty. Indeed it is not. The information one has about an event is different from the uncertainty about that event. However, both of these concepts are valid and plausible interpretations of the quantity $-\sum p_i \log p_i$, which features in information theory. This, in itself, does not make the two concepts identical, but it allows us to use the two interchangeably whenever we refer to the quantity $-\sum p_i \log p_i$. Both of these concepts subscribe to the same requirements originally put forward by Shannon when constructing the measure of information, or uncertainty. It should be noted however that order and disorder do not subscribe to these requirements, and therefore should not be used to describe the quantity $-\sum p_i \log p_i$ (for more details, see Section 1.2).

Chapter 4 is a transition chapter. Instead of plunging directly into statistical mechanics, I decided to gradually transform the concept of the amount of missing information, denoted by H , as defined by Shannon, and as is used in many fields, into the more specific usage of information in thermodynamics. The transition from the general concept of MI to the specific thermodynamic MI is carried out along three (not necessarily independent) routes.

- (i) From a small number of states (or events, or configuration) to a very large number of states.
- (ii) From one type of information to two types.
- (iii) From discrete to continuous information.

This transitional chapter will culminate in the re-derivation of the well known Sackur–Tetrode equation for the entropy. In contrast to the traditional derivation, the new one is based on information–theoretical arguments, and therefore deserves to be renamed the

equivalent of the Sackur–Tetrode equation for the thermodynamic amount of missing information of an ideal gas.

In Chapter 5, we present the fundamental structure of statistical mechanics. This is a standard subject. Therefore, we shall refrain from any details that can be found in any textbook on statistical mechanics. We shall be brief and almost sketchy. The general framework of the structure of statistical thermodynamics, and some standard applications for ideal gases, will be mentioned briefly only if they shed new light on the entropy being replaced by MI.

Chapter 6 can be viewed either as a collection of simple applications of statistical thermodynamics to a few simple processes, or as exercises for practicing the calculation of informational changes in these elementary processes. Some of the processes are discussed in most textbooks, such as expansion, mixing or heat transfer. In this chapter, however, we shall emphasize the informational changes accompanying these processes.

By doing that, we shall stumble upon two important findings; the first is the illusion regarding the irreversibility of mixing ideal gases. This has to do with the common and almost universally accepted conclusion that mixing of ideal gases is an inherently irreversible process, and the interpretation of the quantity $-\sum x_i \ln x_i$ as “entropy of mixing.” It will be shown that mixing, in an ideal gas system can be reversible or irreversible, as can be demixing. The association of the increase of entropy with mixing or with increase of disorder is only an illusion.

The second is the illusion associated with the sense of loss of information in the assimilation process. This is a deeper illusion which has its roots in Maxwell’s and Gibbs’ writings regarding the apparent loss of information due to the assimilation process. We shall discuss this in Chapter 6. We shall show that both Maxwell’s statement as well as Gibbs’ have originated from the intuitive feeling that in an assimilation process, there is a sense of loss of information due to the loss of *identity* of the particles. This illusion is a result of “thinking classically” about identical particles; it can be resolved by properly interpreting the role of indistinguishability of the particles. Both of these illusions have misled many authors, including Gibbs, to reach the wrong conclusions.

Chapter 6 ends with a formulation of the Second Law of Thermodynamics in terms of probabilities and missing information. In the over-a-hundred-years of the history of the Second Law, people were puzzled by the apparent conflict between the reversibility of the equations of motion, and the irreversibility associated with the Second Law. Boltzmann was the first to attempt to *derive* the Second Law from the dynamics of the particles. In my opinion, this, as well as other attempts, will inevitably fail in principle. First, because it is impractical to solve the equations of motion for some 10^{23} particles. Second, because one cannot get *probabilities* from the *deterministic* equations of motion. Third, and perhaps most important, because of the indistinguishability of the particles. It is well known that whenever we write the equation of motions of any number of particles, we must first *label* the particles. This is true for classical as well as for the quantum mechanical equations of motion. However, the very act of *labeling* the particles violates the principle of ID of the particles. Therefore, one cannot derive the second law of thermodynamics which applies to indistinguishable (or unlabel-able in principle; see also appendices J & M) particles from the equations of motion of labeled particles. To put it differently, if we are strictly forbidden from labeling the ID particles, not even temporarily, then we have no equations of motion, hence there exists no conflict.

I have pondered long and hard over the choice of notation. On the one hand, I wanted to use a new notation for the “entropy” and the “temperature,” to stress the difference in the meaning and the units of these quantities. Yet, I did not want to change to a completely unfamiliar notation. During the time of writing the book, I was considering a sort of compromise to use \bar{S} and \bar{T} instead of S and T , but in different units. Finally, I came to the conclusion that the best way is to keep the same notation for S and T , but to emphasize, quite frequently that we mean S/k and kT for the new quantities S and T . There is one modified notation however. We use H , following Shannon, for the more general measure of information in the sense of $-\sum p_i \log p_i$, when applied to *any* distribution. Whenever we discuss a thermodynamic system at equilibrium, H becomes identical with S . This conserves the notations of S and T

on the one hand, and makes the distinction between the more general concept of information and the specific application of the same concept for thermodynamic systems, on the other. For instance, in Chapter 4, we shall discuss various problems of one or two particles (or coins) bound to M sites (or hidden in M boxes). These are purely informational problems, and we use H in this context. However, for $N \rightarrow \infty$ and $M \rightarrow \infty$, we get into the realm of a thermodynamic system of N ligands absorbed on M sites. In this case, we shall switch from the informational measure of information H to the informational measure that previously was referred to as entropy, and is denoted by S . Thus, the notation S is identical, both formally and conceptually, with Shannon's measure of information H , whenever it applies to a thermodynamic system at equilibrium.

As for the justification of this book: there are a few books which include in their titles both "statistical mechanics" and "information theory." In fact, all of these, including Jaynes' pioneering work use "information theory." They use the maximum-entropy principle to predict the most plausible distributions in statistical mechanics. None of these *base* statistical mechanics on the concept of *information*, which is the aim of this book. Moreover, all the examples given in Chapter 6 are presented, discussed and analyzed in terms of the changes in the amount of missing information in spontaneous processes. I found this point of view very fruitful, illuminating and worth publishing.

The book is written in textbook style but it is not a textbook of statistical thermodynamics.

If you are reading this preface and pondering whether or not to read the rest of the book, I suggest that you take a simple "test," the result of which should help you in making the decision.

Consider the following chain of reasoning:

- (i) Mixing is conceived as a process that increases disorder in the system.
- (ii) Increase in disorder is associated with increase in entropy. Therefore, from (i) and (ii), it follows that:
- (iii) Mixing increases the entropy.

If you have no idea what I am talking about in this “test,” you should first study one of the standard textbooks on statistical thermodynamics, then come back to read this book.

If you *do not agree* with conclusion (iii), i.e., if the result of the “test” is *negative*, then I am sure you can read, understand and hopefully enjoy the book.

If you *agree* with conclusion (iii), i.e., the result of the “test” is *positive*, then you have a problem. You should take medication, and read this book until you test negative! I hope that after reading the book, you will understand why I suggest this test.

ACKNOWLEDGMENT

I would like to thank my friends and colleagues Eshel Ben-Jacob, Max Berkowitz, Diego Casadei, João Paulo Ferreira, John Harland, Ken Harris, Art Henn, Harvey Leff, Azriel Levy, Robert Mazo, Mihaly Mezei and Andres Santos. They all did a wonderful job of reading the manuscript and detecting many errors and mistakes. I wish to thank once again the assistance and the inspiration I got from my wife Ruby. As always, without her help, writing this book would not have been possible.

This book was written while I was a visiting Scholar at the center of theoretical biological physics, the University of California, San Diego, CA.

I am very grateful to Peter Wolynes for his hospitality while visiting UCSD.

Arieh Ben-Naim
Department of Physical Chemistry
The Hebrew University of Jerusalem
Jerusalem, Israel