Article

Information Entropy As a Basic Building Block of Complexity Theory

Jianbo Gao [1,2,*], **Feiyan Liu** [3], **Jianfang Zhang** [3], **Jing Hu** [1,2] **and Yinhe Cao** [2]

[1] School of Mechanical Engineering and China-Asean Research Institute, Guangxi University, 100 Daxue Road, Nanning 530005, China; E-Mail: jing.hu@gmail.com
[2] PMB Intelligence LLC, Sunnyvale, CA 94087, USA; E-Mail: yinhec@yahoo.com
[3] Management School, University of Chinese Academy of Sciences, Beijing 100190, China; E-Mails: liufeiy09b@mails.ucas.ac.cn (F.L.); zjf@ucas.ac.cn (J.Z.)

* Author to whom correspondence should be addressed; E-Mail: jbgao.pmb@gmail.com.

Abstract: What is information? What role does information entropy play in this information exploding age, especially in understanding emergent behaviors of complex systems? To answer these questions, we discuss the origin of information entropy, the difference between information entropy and thermodynamic entropy, the role of information entropy in complexity theories, including chaos theory and fractal theory, and speculate new fields in which information entropy may play important roles.

Keywords: information entropy; thermodynamic entropy; emergence; chaos; fractal; complexity theory; multiscale analysis

1. Introduction

What are complex systems? What are emergent behaviors? What role does information entropy play for quantitatively studying them? These are the important questions a thinking person will naturally ask when s/he comes across the terms, complex systems and emergent behaviors.

A complex system is often defined as a system composed of a large number of interconnected parts that, as a whole, exhibit one or more properties that are not obvious from the properties of the individual parts. Some researchers are aware of the caveat of requiring a complex system to be a large system with

many interconnected parts, realizing that a small system, such as a pendulum, may also exhibit complex chaotic behavior. However, many other researchers favor an even more elaborated definition, following the thought-provoking attributes of Earth identified by Kastens *et al*. [1]: nonlinear interactions, multiple stable states, fractal and chaotic behavior, self-organized criticality and non-Gaussian distributions of outputs. While complex systems cannot readily be studied with a reductionist paradigm, in principle, it does not matter much whether one wishes to adopt a simpler or more complicated definition for complex systems, so long as one gives oneself a proper setting to study interesting universal behaviors of complex systems, besides researching specific features of a system.

It is generally accepted that an emergent behavior is a non-trivial or complex collective behavior when many simple entities (or agents) operate in an environment. Nature, as well as life are full of emergent behaviors. On the very large scale, we have the famous example of a spiral galaxy, whose formation may be explained by the density wave theory of Lin and Xu [2]. On a smaller, but still a gigantic, scale, we have another fascinating example—the great red spot of Jupiter [3], which is a high pressure, anti-cyclonic storm akin to a hurricane on Earth, with a period of about six days and a size of three Earths. It has persisted for more than 400 years. Of course, a hurricane (also called a typhoon or a tropical cyclone) is certainly a well-known example of emergent behavior. Other often cited examples of emergence include phase transitions and critical phenomena [4], bird flocking [5,6], fish schooling [7–10] and sand dunes [11]. Note that nonlinearity is an indispensable condition for observing emergent behaviors, but hierarchy is not. To appreciate the latter, it suffices to note that simple models prescribing local interactions can simulate bird flocking and fish schooling quite well [5–10].

It must have been aeons since mankind was first fascinated by emergent behaviors in complex systems. It is only in recent decades that researchers have attempted to quantitatively and systematically study them. As a result, a few powerful new theories have been created, including chaos theory and fractal theory. They are collectively called complexity theories, and information entropy has been playing an essential role in these theories.

To better solve emerging scientific, technological and environmental problems, it is necessary to discuss the origin of information entropy, identify key differences between information entropy and thermodynamic entropy, understand the role of information entropy in the complexity theories and anticipate new fields in which information entropy may play critical roles. These will be the main themes of this essay. In order for the presentation to be readily accessible by a layman, we shall focus on conceptual discussions. However, we will not shun mathematical discussions, in order for the material to also be useful for experienced researchers.

2. The Origin of Information Entropy

Information entropy was originally created by Claude Shannon as a theoretical model of communication, *i.e.*, the transmission of information of various kinds [12]. There are two technical issues in communications: (1) How may the information at the source be quantified and represented? (2) What is the *capacity* of the system—*i.e.*, how much information can the system transmit or process in a given time?

In communications, the first critical observation is that messages have to be treated as random, *i.e.*, unknown by the receiver before the messages are received. Indeed, a conversation becomes meaningless if the listener always knows exactly what the speaker may say next. This observation naturally leads to the following scheme for communication: (i) collect all the potential messages to be sent over a communication channel as a set of random events, $(A_1, A_2, \cdots, A_n)$; (ii) assign a probability, $p_i \geq 0$, where $\sum_{i=1}^{n} p_i = 1$, to the i-th message, which measures the likelihood of the occurrence of the i-th message.

In probability theory, $(A_1, A_2, \cdots, A_n)$ is called a complete system of events [13]. They correspond to $(1, 2, 3, 4, 5, 6)$, when throwing a die, or $(head, \ tail)$, when tossing a coin. If the die or the coin is unbiased, then $\{p_i = 1/6, \ i = 1, \cdots, 6\}$ for die throwing, and $\{p_i = 1/2, \ i = 1, 2\}$ for coin tossing. When the die or coin is biased, however, the probabilities will take different values. In communications, the coin tossing may be associated with binary questions: yes or no, black or white, red or blue, and so on. The average amount of information one obtains from the scheme:

$$A = \left(\begin{array}{cccc} A_1 & A_2 & \cdots & A_n \\ p_1 & p_2 & \cdots & p_n \end{array} \right)$$

when one of the messages is received is given by the information entropy defined by:

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

By convention, $p_j \log p_j = 0$ if $p_j = 0$. While Equation (1) has many desired properties, the logarithm, in particular, provides a convenient unit for quantifying the amount of information. This unit is called bit when the base of the logarithm is two—when a binary problem, yes or no, true of false, with equal probability of 1/2, is considered, the information is one bit, whenever an answer is given. Bit is the unit of data stored and processed in any computing machine.

Note that if there is only one of p_i's that is one, while all others are 0, then $H = 0$. In this case, we have a deterministic scheme and gain no knowledge at all by reading the messages sent by a communication devise. At the other extreme, when the events occur with equal probability of $1/n$, H attains the maximum value of $\log n$. A DNA sequence, consisting of four nucleotides, A-adenine, T-thymine, C-cytosine and G-guanine, is close to the uniform distribution case and, thus, on average, contains close to two bits of information for each base [14].

Using the idea of redundancy, decades of hard work have lead to many excellent error-correcting codes to efficiently represent messages to be transmitted over communication channels. This paves the way for the wide-spread use of computers to deal with everything in life. In particular, among the most important schemes developed is the Lempel-Ziv (LZ) complexity [15,16], which is the foundation of a commonly used compression scheme, *gzip* (more discussions on LZ complexity will be presented in Section 7.1). Therefore, the first problem, how may the information at the source be quantified and represented, has been fully solved. (Peter Shor, an eminent mathematician at MIT, has extended the redundancy idea in an ingenious way to quantum computation and developed a quantum error correction scheme [17].)

The answer to the second question, what is the capacity, C, of a channel, is also given by Shannon in his epic paper. The precise answer, formulated using *mutual information*, which is a natural extension of the concept of information entropy, is given by:

$$C = B \log_2(1 + S/N) \tag{2}$$

where B is the bandwidth of the channel in Hertz and S/N is the signal-to-noise ratio. Mutual information essentially measures how the messages received compare with messages sent over a communication channel.

Albeit not a suitable place to prove Equation (2), we partially justify the theorem to enhance the understanding of communications. Given a signal and noise power of S and N, the total power is $P = S + N$. In the case of an analog signal, we may partition the signal waveform into many bins, with each bin representing one of the messages. Here, one has to consider the worst case for the channel—all messages are equally likely, so that the channel is continuously transmitting new information. The largest number of bins possible is given by

$$2^b = \sqrt{P/N} = \sqrt{1 + S/N}$$

In this case, each message may be represented by b bits. If we make M measurements of the b-bit level in a time, T, then the total number of bits of information collected will be

$$M \cdot b = M \log_2(1 + S/N)^{1/2}$$

and the information transmission rate, I, in unit of bits per unit time, is

$$I = \frac{M}{T} \log_2(1 + S/N)^{1/2}$$

Recognizing that the largest $\frac{M}{T}$ possible is the highest practical sampling rate, $2B$, Equation (2) naturally follows.

It is important to note that as $B \to \infty$, the capacity, C, does not become infinite. This is because noise power, N, is also proportional to B. Denoting $N = \eta B$, where η is the noise power per unit bandwidth, and utilizing

$$\lim_{x \to 0}(1 + x)^{1/x} \to e$$

we have

$$C(B \to \infty) = \frac{S}{\eta} \log_2 e = 1.44\frac{S}{\eta}$$

3. Entropy in Classical Thermodynamics

The word entropy apparently arose first in classical thermodynamics, which treats state variables that pertain to the whole system, such as pressure, volume and the temperature of a gas. A mathematical equation that arises constantly in classical thermodynamics is:

$$dH = dQ/T \tag{3}$$

where dQ is the quantity of heat transfer at temperature T and dH is the change in entropy. The second law of thermodynamics asserts that dH cannot decrease in a closed system. Classical thermodynamics makes no assumptions of the detailed micro-structure of the materials involved.

Classical statistical mechanics, in contrast, tries to model the detailed structure of the materials and from the model to predict the rules of classical thermodynamics. For example, the pressure of the gas can be explained as gas molecules, treated as little, hard, perfectly elastic balls, in constant motion against the walls. Even a small amount of gas will have an enormous number, N, of particles. In fact, as a crude order of magnitude, N may be taken as the Avogadro constant, $N_A = 6.022 \times 10^{23}$. If we imagine a phase space, whose coordinates are the position and velocity of each particle, then the phase space for the gas particles is a subregion of a $6N$-dimensional space. Assuming that for a fixed energy, every small region in the phase space has the same probability as any other, Boltzmann found that the following quantity plays the role of entropy:

$$H = k_{\scriptscriptstyle B} \ln \frac{1}{P} \tag{4}$$

where P is the probability associated with any one of the equally likely small regions in the phase space with the given energy, and $k_{\scriptscriptstyle B}$ is the Boltzmann constant.

Gibbs, in trying to deal with systems that did not have a fixed energy, introduced the "grand canonical ensemble", which is, in essence, an ensemble of the phase spaces of different energies of Boltzmann. Gibbs deduced an entropy of the form:

$$H = \sum_i p(i) \ln \frac{1}{p(i)} \tag{5}$$

where $p(i)$ are the probabilities of the phase space ensembles. Expression-wise, this is identical to Equation (1). It is thus no wonder that some researchers would consider information entropy to be a redundant term.

However, identity in mathematical form does not imply identify of meaning, as Richard Hamming emphasized in his interesting book, *The art of probability* [18]. The most fundamental difference between information entropy and thermodynamic entropy is that information entropy works with a set of events with arbitrary probabilities, while in thermodynamics, it is always assumed that gas particles occupy any region of a container with equal probability. Therefore, information entropy is a broader concept than thermodynamic entropy. To further help understand this, it is beneficial to note that Myron Tribus was able to derive all the basic laws of thermodynamics from information entropy [19]. More importantly, while thermodynamic entropy may not be very relevant to the description of genomic and proteomic sequences and many emerging complex behaviors, information entropy is an essential building block of complexity theory [20] and can naturally quantify the amount of information in biological sequences [14,21].

4. Entropy Maximizing Probability Distributions

One of the most important applications of entropy is to determine initial distributions relevant to many phenomena in science and engineering by maximizing entropy. Uniform distribution is one such distribution. However, it is not the only one. Depending on constraints, there are other distributions that

can maximize entropy. To facilitate this discussion, we first need to extend information entropy based on discrete probabilities to that based on the probability density function (PDF). This is given by the *differential entropy*, defined by:

$$H = - \int f(x) \log f(x) dx \qquad (6)$$

For simplicity, here, we only list two elementary distributions that maximize entropy under appropriate constraints. This discussion will be resumed briefly in the next section.

(1) Exponential distribution, with PDF given by:

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0 \qquad (7)$$

maximizes entropy under the constraint that the mean of the random variable, X, which is $1/\lambda$, is fixed.

This entropy maximization property is perhaps one of the main reasons why we encounter exponential distributions so frequently in mathematics and physics. For example, a Poisson process is defined through exponential temporal or spatial intervals, while the sojourn times of Markov processes are exponentially distributed [20]. Exponential distribution is also very relevant to ergodic chaotic systems, as recurrence times of chaotic systems follow exponential distributions [22,23]. Exponential laws play an even more fundamental role in physics, since the basic laws in statistical mechanics and quantum mechanics are expressed as exponential distributions, while finite spin glass systems are equivalent to Markov chains.

(2) When mean, μ, and variance, σ^2, are given, the distribution that maximizes entropy is the normal distribution with mean, μ, and variance, σ^2, $N(\mu, \sigma^2)$.

The fundamental reason that normal distributions maximize entropy is the Central Limit theorem—a normal distribution may be considered an attractor, since the sample mean of a sufficiently large number of independent random variables, each with finite mean and variance, will be approximately normally distributed.

5. Entropy and Complexity

Broadly speaking, any behavior that is neither completely regular (or dumb) nor fully random may be called an emerging complex behavior. Representative complex behaviors include chaotic motions and fractal behaviors. The latter include random processes with long-range correlations, which is a subclass of the fascinating $1/f$ phenomena [24,25].

To facilitate the following discussions, we first note different forms of motions, in increasing order of complexity: fixed point solutions, periodic motions, quasi-periodic motions, chaotic motions, turbulence and random motions. Interestingly, a similar sequence is observed in solid materials: crystal, quasicrystal, fractal and aperiodic random form. In particular, Dan Shechtman won the Nobel Prize in Chemistry in 2011 for discovering quasicrystals. One would have anticipated the existence of quasicrystals by the similarity between these two sequences, noticing that quasi-periodic motions have been known in the dynamics community for a long time. In terms of abundance, quasicrystals are much fewer than fractal shapes.

A. Fractal: Euclidean geometry is about lines, planes, triangles, squares, cones, spheres, *etc*. The common feature of these different objects is regularity: none of them is irregular. Now, let us ask a question: are clouds spheres, mountains cones and islands circles? The answer is obviously, no. In pursuing answers to such questions, Mandelbrot has created a new branch of science—fractal geometry [26–29].
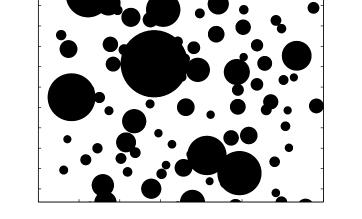
For now, we shall be satisfied with an intuitive definition of a fractal: a set that shows irregular, but self-similar, features on many or all scales. Self-similarity means that part of an object is similar to other parts or to the whole. That is, if we view an irregular object with a microscope, whether we enlarge the object by 10 times or by 100 times or even by 1,000 times, we always find similar objects. To understand this better, let us imagine that we were observing a patch of white cloud drifting away in the sky. Our eyes were rather motionless: we were staring more or less in the same direction. After a while, the part of the cloud we saw drifted away, and we were viewing a different part of the cloud. Nevertheless, our feeling remained more or less the same.

Mathematically, self-similarity or a fractal is characterized by a power-law relation, which translates into a linear relation in the log-log scale. To understand how power-law underlies the perception of self-similarity, let us imagine a very large number of balls flying around in the sky, where the size of the balls follows a heavy-tailed power-law distribution:

$$p(r) \sim r^{-\alpha} \tag{8}$$

See Figure 1.

Figure 1. Random fractal of discs with a Pareto-distributed size: $P[X \geq x] = (1.8/x)^{1.8}$.

Being human, we will instinctively focus on balls whose size is comfortable for our eyes—balls that are too small cannot be seen, while balls that are too large block our vision. Now, let us assume that we are most comfortable with the scale, r_0. Of course, our eyes are not sharp enough to tell the differences between scales r_0 and $r_0 + dr$, $|dr| \ll r_0$. Nevertheless, we are quite capable of identifying scales, such as $2r_0$, $r_0/2$, *etc*. Which aspect of the flying balls may determine our perception? This is essentially given by the relevant abundance of the balls of sizes $2r_0$, r_0 and $r_0/2$:

$$p(2r_0)/p(r_0) = p(r_0)/p(r_0/2) = 2^{-\alpha}$$

Note that the above ratio is independent of r_0. Now, suppose we view the balls through a microscope, which magnifies all the balls by a scale of 100. Now, our eyes will be focusing on scales, such as $2r_0/100$, $r_0/100$ and $r_0/200$, and our perception will be determined by the relative abundance of the balls at those scales. Because of the power-law distribution, the relative abundance will remain the same—so does our perception.

B. Tsallis non-extensive entropy and powerlaw behavior: One attractive means of explaining the ubiquity of power-laws and fractal behaviors is through maximization of Tsallis entropy, named after a brilliant Brazilian physicist, Tsallis [30,31]. To explain the idea, we first extend Shannon's information entropy to the *Renyi entropy*, defined by:

$$H_q^R = \frac{1}{1-q} \log\Big(\sum_{i=1}^{m} p_i{}^q\Big) \tag{9}$$

The purpose of introducing a spectrum of q in *Renyi entropy* is to amplify larger or smaller probabilities. For example, when $q \gg 0$ or $q \ll 0$, large or small probabilities dominate the right side of Equation (9), respectively. *Tsallis entropy*, defined by:

$$H_q^T = \frac{1}{q-1} \Big(1 - \sum_{i=1}^{m} p_i{}^q\Big) \tag{10}$$

is related to the Renyi and Shannon entropies through simple relations:

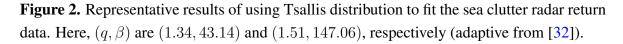$$H_q^R = \frac{\ln\Big[1 + (1-q)H_q^T\Big]}{1-q}, \quad \lim_{q \to 1} H_q^R = \lim_{q \to 1} H_q^T = -\sum_{i=1}^{m} p_i \ln p_i$$
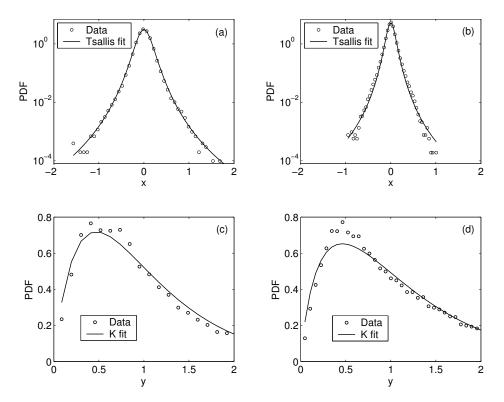
However, Tsallis entropy has a different focus—it aims to find a specific q, often different than one, that best characterizes a phenomenon that is neither regular nor fully chaotic/random. Tsallis entropy is non-extensive in the sense that for a compound system comprising two independent subsystems, Tsallis entropy is not the summation of Tsallis entropies for the two subsystems. So far, a number of workshops and conferences have been organized to discuss Tsallis non-extensive statistics.

By maximizing the form of Tsallis entropy for continuous PDFs, one can obtain Tsallis distribution [32]:

$$p(x) = \frac{1}{Z_q}[1 + \beta(q-1)x^2]^{1/(1-q)}, \ \ 1 < q < 3 \tag{11}$$

where Z_q is a normalization constant and β is related to the second moment. When $5/3 < q < 3$, the distribution is a heavy-tailed power-law described by Equation (8). In particular, when $q = 2$, the distribution reduces to the Cauchy distribution, which is a stable distribution with infinite variance [20]. As an application to the analysis of real world data, we have shown in Figure 2 the analysis of sea clutter radar return data. As one may expect, q is distinctly different from one.

Figure 2. Representative results of using Tsallis distribution to fit the sea clutter radar return data. Here, (q, β) are $(1.34, 43.14)$ and $(1.51, 147.06)$, respectively (adaptive from [32]).

C. Chaos: Fractal behaviors are not limited to geometric objects. They can also manifest themselves as temporal variations, such as stock market price variations and chaotic motions. While the meaning of chaos is consistent with intuitive understanding, here, we shall confine ourselves to its strict mathematical meaning, *i.e.*, exponential divergence,

$$d(t) \sim d(0)e^{\lambda_1 t} \tag{12}$$

where $d(0)$ denotes a small separation between two arbitrary trajectories at time 0, $d(t)$ is the average separation between them at time t and $\lambda_1 > 0$ is the largest positive Lyapunov exponent. This property is called sensitive dependence to initial conditions and is the origin of the fascinating butterfly effect: sunny weather in New York could be replaced by rainy weather sometime in the near future after a butterfly flaps its wings in Boston. This property is vividly shown in Figure 3: initially close by points in the chaotic Lorenz attractor rapidly diverge and, soon, are everywhere on the attractor.

To better appreciate the concept of sensitive dependence to initial conditions, let us consider the map on a circle:

$$x_{n+1} = 2x_n \ \mathrm{mod} \ 1 \tag{13}$$

where x is positive and $\mathrm{mod} \ 1$ means that only the fractional part of $2x_n$ will be retained as x_{n+1}. This map can also be viewed as a Bernoulli shift or a binary shift. Suppose that we represent an initial condition, x_0, in binary:

$$x_0 = 0.a_1 a_2 a_3 \cdots = \sum_{j=1}^{\infty} 2^{-j} a_j \tag{14}$$
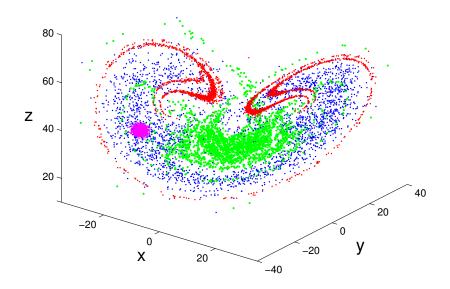
where each of the digits, a_j, is either one or zero. Then,

$$x_1 = 0.a_2 a_3 a_4 \cdots$$

$$x_2 = 0.a_3a_4a_5\cdots$$

and so on. Thus, a digit that is initially far to the right of the decimal point, say the 40th digit (corresponding to $2^{-40} \approx 10^{-12}$), and, hence, has only a very minor role in determining the initial value of x_0, eventually becomes the first and the most important digit.

Figure 3. Ensemble forecasting in the chaotic Lorenz system: 2,500 ensemble members, initially represented by the pink color, evolve to those represented by the red, green and blue colors at $t = 2, 4$ and 6 units.

A chaotic motion is generally characterized as a strange attractor. By strange, we mean the exponential divergence. By attractor, we have finiteness in motion. This incessant stretching and folding back in phase space often leads to a fractal structure for the underlying attractor. The fractal or capacity dimension of this attractor may be determined as follows: Partition the phase space containing the attractor into many cells of linear size, ϵ. Denote the number of nonempty cells by $n(\epsilon)$. Then,

$$n(\epsilon) \sim \epsilon^{-D_0}, \;\; \epsilon \to 0$$

where D_0 is called the box-counting dimension.

The concept of the box-counting dimension can be generalized to obtain a sequence of dimensions, called the generalized dimension spectrum. This is obtained by assigning a probability, p_i, to the i-th nonempty cell. One simple way to calculate p_i is by using n_i/N, where n_i is the number of points within the i-th cell and N is the total number of points on the attractor. Let the number of nonempty cells be n. Then:

$$D_q = \frac{1}{q-1} \lim_{\epsilon \to 0} \left(\frac{\log \sum_{i=1}^{n} p_i^q}{\log \epsilon} \right) \tag{15}$$

where q is real. In general, D_q is a non-increasing function of q. D_0 is simply the box-counting or capacity dimension, since $\sum_{i=1}^{n} p_i^q = n$. D_1 gives the information dimension, D_I:

$$D_I = \lim_{\epsilon \to 0} \frac{\sum_{i=1}^{n} p_i \log p_i}{\log \epsilon} \tag{16}$$

The above consideration can be extended to monitor the detailed temporal evolutions of a chaotic attractor. Again, all we need to do is to partition the phase space into small boxes of size ϵ, compute the

probability, p_i, that box i is visited by the trajectory and calculate Shannon entropy. For many systems, when $\epsilon \to 0$, information linearly increases with time [33]:

$$I(\epsilon, t) = I_0 + Kt \tag{17}$$

where I_0 is the initial entropy, which may be taken as zero for simplicity, and K is the *Kolmogorov-Sinai (KS) entropy*.

To deepen our understanding, let us consider three cases of dynamical systems: (i) deterministic, non-chaotic; (ii) deterministic, chaotic; and (iii) random. For case (i), during the time evolution of the system, phase trajectories remain close together. After a time, T, nearby phase points are still close to each other and can be grouped into some other small region of the phase space. Therefore, there is no change in information. For case (ii), due to exponential divergence, the number of phase space regions available to the system after a time, T, is $N \propto e^{(\sum \lambda^+)T}$, where λ^+ are positive Lyapunov exponents. Assuming that all of these regions are equally likely, then $p_i(T) \sim 1/N$, and the information function becomes:

$$I(T) = -\sum_{i=1}^{N} p_i(T) \ln p_i(T) = (\sum \lambda^+)T \tag{18}$$

Therefore, $K = \sum \lambda^+$. More generally, if these phase space regions are not visited with equal probability, then:

$$K \leq \sum \lambda^+ \tag{19}$$

Grassberger and Procaccia [34], however, suggest that equality usually holds. Finally, for case (iii), we can easily envision that after a short time, the entire phase space may be visited. Therefore, $I \sim \ln N$. When $N \to \infty$, we have $K = \infty$.

The above discussions make it clear that, albeit thermodynamic entropy may not be very relevant for describing fractal and chaotic behavior, Shannon's information entropy is always a basic building block. As one can easily perceive, a precise definition for the *KS* entropy will again be based on Shannon's information entropy, by replacing $p_i(T)$ in Equation (18), to be the joint probability that the trajectory is in boxes, $i_1, \cdots, i_d$, at d subsequent times. In order not to overwhelm readers with mathematical equations, we will not go into the details here.

D. Distinguishing chaos from noise: For a long time, a finite Kolmogorov entropy has often been thought to indicate deterministic chaos. This practice is still being followed in many applications. This aspect of chaos research may be summarized by the following analogy: many researchers were chasing the beast of chaos on a wild beach. One was yelling, "Here is a footprint". Another was echoing, "Here is another" $\cdots$. After a long while, some careful minds pointed out that those may just be their own footprints. Among the most convincing counter-examples are the $1/f$ random processes, which have fractal dimensions and finite Kolmogorov entropies and, thus, may be misinterpreted as deterministic chaos [35,36].

If one can think a little deeper, one can readily realize that it is impossible to obtain an infinite Kolmogorov entropy with a finite amount of random data. This is why the issue of distinguishing chaos from noise has been considered a classic and difficult one [37–39]. To fundamentally solve the problem, one has no choice but to resort to multiscale approaches. One of the most viable approaches to tackle

this issue is the scale-dependent Lyapunov exponent (SDLE). SDLE is a function of the scale parameter and, thus, is entirely different from the conventional concept, the Lyapunov exponent, which is a number. Among the multiscale complexity measures, SDLE has the richest scaling laws. For example, for chaotic motions, SDLE is a constant, indicating truly exponential divergence. However, SDLE is a power-law for $1/f$ processes. Therefore, distinguishing chaos from noise is no longer a problem. Moreover, through an ensemble forecasting approach, SDLE can tie together many different types of entropies in dynamical systems. For more details, we refer to [40–42].

E. Statistical complexity: Finally, we note that information entropy is a deterministic complexity measure, since it quantifies the degree of randomness. Sometimes this is considered not ideal for characterizing a type of behavior that is neither regular nor completely random. An alternative, statistical complexity, has been proposed that can be maximized for neither high nor low randomness [43,44]. Interestingly, information entropy still is a significant building blocking in this actively evolving field [45,46].

F. Multiscale analysis: Albeit chaotic dynamics have fractal properties, there is an important subset of fractal behaviors, random fractal behaviors, that are entirely different from deterministic chaotic dynamics. Recognizing that the foundations of random fractal behaviors are random and many non-chaotic, but random behaviors may be modeled by random fractals, Gao *et al.* [20] have advocated to: (1) use chaos and random fractal theories synergistically to solve a broad range of problems of real world impact and (2) use multiscale approaches to simultaneously characterize the behaviors of complex signals on a wide range of scales.

There exist a number of multiscale approaches. Among them is the random fractal theory, whose key element is *scale-invariance*, *i.e.*, the statistical behavior of the signal is independent of a spatial or temporal interval length. With scale-invariance, only one or a few parameters are sufficient to describe the complexity of the signal across a wide range of scales where the fractal scaling laws hold. Because of the small number of parameters, fractal analyses are among the most parsimonious multiscale approaches [20]. Other multiscale methods include SDLE, which has been briefly discussed earlier, the finite-size Lyapunov exponent [47–49], (ϵ, τ) entropy [50] and multiscale entropy [51]. For analysis and modeling of a single time series data, these approaches may be considered to often be adequate. What are significantly lacking are the tools for studying the detailed interactions between two or more systems, *i.e.*, involving two or more time series data.

6. Time's Arrow

Although the basic laws of physics are time reversible (*i.e.*, that if time t in all equations were substituted with $-t$, the relations would still hold), time irreversible processes are ubiquitous. From the mixing of cold and warm water, to the burning of a match and the breaking of glass, common real-world experience tells us that regardless of the mathematical formulation of basic physical laws, time's arrow only points in one direction.

To resolve this paradox, Ludwig Boltzmann developed the concept of Boltzmann entropy and the H-theorem. As we have discussed, Boltzmann entropy is the logarithm of the number of all possible

microscopic states (or phase space volume). The H-theorem governs the time evolution of the (negative) entropy, and implies that entropy has to be constant or increase in time. Taken together, the two concepts provide a constraint upon the directionality of time.

Boltzmann's scheme, albeit very successful, has been controversial [52,53]. Historically, a decisive objection was made by Ernst Zermelo, who pointed out that on the basis of the Poincare recurrence theorem, a closed dynamical system must eventually come back arbitrarily close to its initial state. Thus, eventually, every system will be "reversible", and entropy cannot always increase.

Intuitively, the Poincare recurrence time for a closed system to return to its initial state must be inconceivably long. Richard Feynman asserted, "It would never happen in a million years" [54]. Vladimir Arnold thought it may be longer than the age of the solar system [55]. Boltzmann himself reportedly replied:"You should wait that long!" [56]. A quantitative estimate of the length of this time recently has been given by Gao as [22]:

$$T(r) \sim \tau \cdot r^{-(D_I-1)} \tag{20}$$

where τ is the sampling time, r is the size of a subregion in the phase space to be re-visited and D_I is the information dimension already discussed. For fully random gaseous motions, as we have discussed, D_I may be taken as on the same order of $6N_A$, where N_A is the Avogadro constant. Therefore, the recurrence time is on the order of $10^{36 \times 10^{23}}\tau$, if we take $r \sim 1/10$. This time is too long to be relevant to reality!

While objection to Boltzmann's scheme is not quite relevant to reality, research on the opposite line is more productive—Cédric Villani, a genial French mathematician and a Fields medalist of 2010, is able to compute entropy production from the Boltzmann equation and find the rate of convergence to equilibrium [57].

It was unfortunate that Boltzmann committed suicide. However, a sober great mind, Willard Gibbs, who was a contemporary of Boltzmann, would not attach himself too much to the depressing meaning of the second law of thermodynamics. Gibbs serenely concluded that recurrence time had to be very long. However, there might be a chance to observe processes that would violate the second law of thermodynamics. Indeed, such violations can be readily observed in small, nanoscale systems over short time scales [58]. This possibility, in fact, has much to do with Equation (20)—the exponent is $D_I - 1$, not D_I. When D_I is large, essentially, there is no difference between D_I and $D_I - 1$. However, when a system is small, $D_I - 1$ instead of D_I will make the recurrence time much shorter.

While in closed physical systems, violations of the second law of thermodynamics may not be easy to observe, negative entropy flow is a rule rather than an exception in life—life feeds on negative entropy, as asserted by Erwin Schrödinger [59]. Lila Gatlin argues that entropy reduction within living systems occurs whenever information is stored [21]. More precisely, we may say that life functions when genetic codes are executed precisely and external stimuli are properly processed by neurons. The source of this negative entropy is the Sun, as discussed by the eminent mathematician and theoretical physicist, Roger Penrose, in his best selling popular science book, *The Emperor's New Mind* [60].

7. Entropy in an Inter-Connected World: Examples of Applications and Future Perspectives

As expected, information entropy has found interesting applications in almost every field of science and engineering. In this closing section, we explain how entropy can be applied to analyze complex data and speculate on the frontiers where information entropy may play critical roles.

7.1. Estimating Entropy from Complex Data: The Use of the Lempel-Ziv Complexity

When the probability distribution for a complex system is known, using Equation (1), information entropy can easily be computed. If all that is known is time series data, how can entropy be computed? The answer lies in the Lempel-Ziv (LZ) complexity [15,16].

The LZ complexity and its derivatives, being easily implementable, very fast and closely related to the Kolmogorov complexity [61,62], have found numerous applications in characterizing the randomness of complex data.

To compute the LZ complexity, a numerical sequence has to be first transformed into a symbolic sequence. The most popular approach is to convert the signal into a 0–1 sequence by comparing the signal with a threshold value, S_d [63]. That is, whenever the signal is larger than S_d, one maps the signal to one, otherwise, to zero. One good choice of S_d is the median of the signal [64]. When multiple threshold values are used, one may map the numerical sequence to a multi-symbol sequence. Note that if the original numerical sequence is a nonstationary random-walk type process, one should analyze the stationary differenced data instead of the original nonstationary data.

After the symbolic sequence is obtained, it can then be parsed to obtain distinct words and the words be encoded. Let $L(n)$ denote the length of the encoded sequence for those words. The LZ complexity can be defined as:

$$C_{LZ} = \frac{L(n)}{n} \tag{21}$$

Note, this is very much in the spirit of the Kolmogorov complexity [61,62].

There exist many different methods to perform parsing. One popular scheme is proposed by the original authors of the LZ complexity [15,16]. For convenience, we call this Scheme 1. Another attractive method is described by Cover and Thomas [65], which we shall call Scheme 2. For convenience, we describe them under the context of binary sequences.

- **Scheme 1:** Let $S = s_1 s_2 \cdots s_n$ denote a finite length, 0–1 symbolic sequence; $S(i,j)$ denote a substring of S that starts at position i and ends at position j, that is, when $i \leq j$, $S(i,j) = s_i s_{i+1} \cdots s_j$ and when $i > j$, $S(i,j) = \{\}$, the null set. Let $V(S)$ denote the vocabulary of a sequence, S. It is the set of all substrings, or words, $S(i,j)$ of S, (*i.e.*, $S(i,j)$ for $i = 1, 2, \cdots, n$; $j \geq i$). For example, let $S = 001$, we then have $V(S) = \{0, 1, 00, 01, 001\}$. The parsing procedure involves a left-to-right scan of the sequence, S. A substring, $S(i,j)$, is compared to the vocabulary that is comprised of all substrings of S up to $j-1$, that is, $V(S(1, j-1))$. If $S(i,j)$ is present in $V(S(1, j-1))$, then update $S(i,j)$ and $V(S(1, j-1))$ to $S(i, j+1)$ and $V(S(1, j))$, respectively, and the process repeats. If the substring is not present, then place a dot after $S(j)$ to indicate the end of a new component, update $S(i,j)$ and $V(S(1, j-1))$ to $S(j+1, j+1)$ (the single symbol in the $j+1$ position) and $V(S(1, j))$, respectively, and the process continues. This parsing

operation begins with $S(1, 1)$ and continues until $j = n$, where n is the length of the symbolic sequence. For example, the sequence, 1011010100010, is parsed as $1 \cdot 0 \cdot 11 \cdot 010 \cdot 100 \cdot 010\cdot$. By convention, a dot is placed after the last element of the symbolic sequence. In this example, the number of distinct words is six.

- **Scheme 2:** The sequence, $S = s_1 s_2 \cdots$, is sequentially scanned and rewritten as a concatenation, $w_1 w_2 \cdots$, of words, w_k, chosen in such a way that $w_1 = s_1$ and w_{k+1} is the shortest word that has not appeared previously. In other words, w_{k+1} is the extension of some word, w_j, in the list, $w_{k+1} = w_j s$, where $0 \leq j \leq k$ and s is either zero or one. The above example sequence, 1011010100010, is parsed as $1 \cdot 0 \cdot 11 \cdot 01 \cdot 010 \cdot 00 \cdot 10\cdot$. Therefore, a total of seven distinct words are obtained. This number is larger than the six of **Scheme 1** by one.

The words obtained by Scheme 2 can be readily encoded. One simple way is as follows [65]. Let $c(n)$ denote the number of words in the parsing of the source sequence. For each word, we use $\log_2 c(n)$ bits to describe the location of the prefix to the word and one bit to describe the last bit. For our example, let 000 describe an empty prefix, then the sequence can be described as $(000, 1)(000, 0)(001, 1)(010, 1)(100, 0)(010, 0)(001, 0)$. The total length of the encoded sequence is $L(n) = c(n)[\log_2 c(n) + 1]$. Equation (21) then becomes:
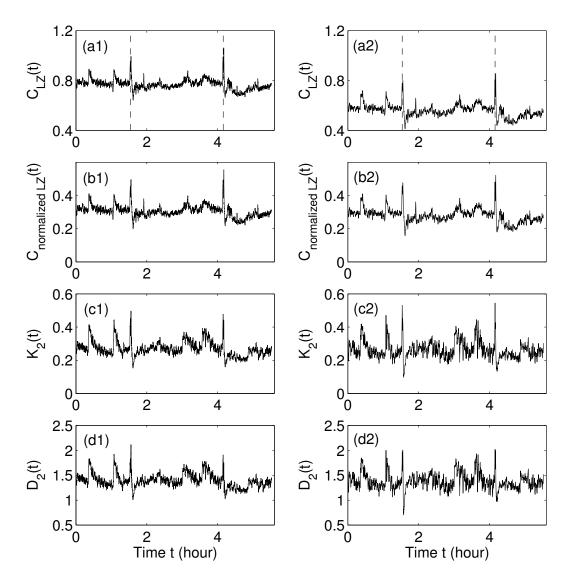
$$C_{LZ} = c(n)[\log_2 c(n) + 1]/n \tag{22}$$

When n is very large, $c(n) \leq n/\log_2 n$ [15,65]. Replacing $c(n)$ in Equation (22) by $n/\log_2 n$, one obtains:

$$C_{LZ} = \frac{c(n)}{n/\log_2 n} \tag{23}$$

The commonly used definition of C_{LZ} takes the same functional form as Equation (23), except that $c(n)$ is obtained by Scheme 1. Typically, $c(n)$ obtained by Scheme 1 is smaller than that by Scheme 2. However, encoding the words obtained by Scheme 1 needs more bits than that by Scheme 2. We surmise that the complexity defined by Equation (21) is similar for both schemes. Indeed, numerically, we have observed that the functional dependence of C_{LZ} on n (based on Equations (22) and (23)) is similar for both schemes.

For infinite length sequences, the LZ complexity is equivalent to the Shannon entropy. In particular, the LZ complexity is zero for periodic sequences with infinite length. However, when the length of a periodic sequence is finite, the LZ complexity is larger than zero. In most applications, a signal is of finite length. It is therefore important to find a suitable way to ensure that the LZ complexity is zero for a finite periodic sequence and one for a fully random sequence. This issue was first considered by Rapp *et al.* [66]. It is taken on again by Hu *et al.* [67], recently, using an analytic approach. Specifically, they have derived formulas for the LZ complexity for random equiprobable sequences, as well as periodic sequences with an arbitrary period, m, and proposed a simple formula to perform normalization. An application to the analysis of electroencephalography (EEG) data for epileptic seizure detection is shown in Figure 4. We observe that the LZ complexity, albeit simple, is as effective as the correlation entropy and the correlation dimension from chaos theory for detecting seizures. For more details, we refer to [67]. Furthermore, for a deep understanding of the connections among different complexity measures for EEG analysis, we refer to [68].

Figure 4. The variation of (**a1,a2**), the Lempel-Ziv (LZ) complexity, (**b1,b2**), the normalized LZ complexity, (**c1,c2**), the correlation entropy, and (**d1,d2**), the correlation dimension with time for the EEG signal of a patient. (a1–d1) are obtained by partitioning the EEG signals into short windows of length, $W = 500$ points; (a2–d2) are obtained using $W = 2,000$. The vertical dashed lines in (a1,a2) indicate seizure occurrence times determined by medical experts.

7.2. Future Perspectives

Where will information entropy be most indispensable? We believe those fields that must interface with human behavior. In those situations, as with the first step, the use of information theory is not so much about providing a formula to quantify the amount of uncertainty. Rather, it helps researchers to comprehensively understand a significant problem, *i.e.*, to define a complete system of events needed for applying information entropy. Comprehensiveness is the guiding principle for data-driven multiscale analysis of complex data [20] and is the prerequisite for making long-lasting impacts.

Before we make speculations, we note that, as science advances, what is currently uncertain or unknown may become partially or fully known in the future, and then, information entropy will decrease.

The situation is similar to what Professor Yuch-Ning Shieh of Purdue University has contemplated during our personal conversation: "The amount of dark matter may decrease when some explicit form of dark matter is found in the future". To understand this aspect better, we may take the case of estimating information entropy from DNA sequences as a concrete example. While entropy is close to two bits, based on the distribution of nucleotide bases [14], it becomes significantly lower than two bits when sequential correlations are taken into account [69] (for recent studies on DNA sequences, we refer to [70–75]).

A more complicated situation is provided by the use of information theory in psychology. After an initial fad of information theory in psychology during the 1950s and 1960s, it no longer was much of a factor, due to increasingly deeper understanding of information processing by neurons [76]. However, recently, there has been a resurgence of interest in information theory in psychology, for the purpose of better understanding uncertainty-related anxiety [77]. This new model focuses on the weighted distribution of potential actions and perceptions as subjectively experienced by a human being and assigns lower entropy to stronger goals. In essence, this model is hierarchical, with neural science working at the bottom layer and subjective decision making involving multiple layers. A hierarchical model, in essence, is what "order based on order", using Erwin Schrödinger's words [59].

Let us now consider the potential use of information theory in environmental science and engineering. For concreteness, let us start with the widespread PM2.5 pollution in China. First, let us consider the physics of PM2.5 pollution. It is known that PM2.5 sulfates reside 3–5 days in the atmosphere. With an average wind speed, say, 5 m/s, the residence time of several days yields a "long range transport" and a more uniform spatial pattern. On the average, PM2.5 particles can be transported as far as 1,000 or more km from the source of their precursor gases. This is the main mechanism responsible for Hong Kong's PM2.5 pollution in winter [78]. Therefore, physics-wise, it would be interesting to study PM2.5 pollution systematically, including the rate of generation in the sources, the variation of the PM2.5 level with wind speed, atmospheric transportation, and so on, to determine what should be done to limit PM2.5 pollution to a certain range. There will be many uncertainties in such an effort. Information entropy, and even thermodynamic entropy, should be able to play some important roles.

Next, let us consider the consequences of severe PM2.5 pollution. The first may be the adverse health effects, which have been widely discussed in the media. Published medical studies on the effects of PM2.5 pollution only considered moderately high concentrations of PM2.5 particles. Since, in many cities of China, PM2.5 concentration is simply off the scale for many consecutive days, one clearly has to ask whether the adverse effects of PM2.5 particles on health depend linearly or non-linearly on the concentration of PM2.5 particles. If nonlinear, then some kind of bifurcation, *i.e.*, dramatically increased health problems when PM2.5 concentration exceeds a certain level, may occur. Then comes the question of medical costs on such pollution-induced health problems. Furthermore, one may ask how much PM2.5 pollution may harm outdoor animals, especially birds, since they do not have any means to protect themselves against harmful PM2.5 particles constantly present in a huge spatial range.

This cost does not simply stop at the medical level. Severe pollution makes foggy and hazy weather appear more frequently. Foggy and hazy weather decreases visibility, forces the closure of roads, worsens traffic congestion, causes more traffic accidents and casualties, discourages shopping, forces cancellation of thousands of flights, and so on. Clearly, this "messiness" can be associated with dramatic increase in

entropy. Such considerations will be critical for convincing responsible governmental agencies to take decisive actions to reduce PM2.5 pollution.

Next, let us discuss the relationship between economic development and entropy production. While many countries have made significant progress in GDP, it is important to note that, so far, little attention has been paid to the notion of entropy when developing economic growth models. This includes Marx's economic theory [79] and the economics Nobel prize-winning model, the Solow-Swan neo-classical growth model [80,81]. With our living environment so gravely endangered, it is the very time to seriously address the critical issue of sustainable growth.

At this point, we should discuss what entropy really means in economics. A general interpretation is to associate entropy with some distributions of economic data [82]. Indeed, entropy for the distribution of negative incomes can predict economic downturns remarkably well, including the recent gigantic financial crisis [83]. In the emerging new field, econophysics, which tries to develop a thermodynamic analogy for the economy, energy and entropy are associated with capital and production function, respectively [84]. Such a view is too rigid, however, since a fixed amount of money, when used differently, can lead to entirely different consequences. For example, in 2012, the ex-wife of the billionaire golfer, Tiger woods, ordered her newly bought mansion, in $12 million, to be torn down and re-built, for the reason that the mansion was too small for her. Again in 2012, right after the super hurricane Sandy in New York, some poor people in New York city were struggling for survival, due to lack of food. Yet, some rich people were struggling for a different purpose—busy consuming wines costing $1,000 per bottle, due to the flooding of their basements by Sandy. Our view is, to aptly discuss entropy in economics, one has to comprehensively evaluate all the possibilities and their positive and negative consequences. Waste and disruptive consequences of certain growth processes ought to be associated with a large increase in entropy.

This is an information exploding age. The key here is to aptly use big data—as estimated by Mckinsey [85], if US healthcare were to use big data creatively and effectively to drive efficiency and quality, the sector could create more than $300 billion in value every year. In the developed economies of Europe, government administrators could save more than $149 billion in operational efficiency improvements alone by using big data, not including using big data to reduce fraud and errors and to boost the collection of tax revenues. Additionally, users of services enabled by personal-location data could capture $600 billion in consumer surplus. It is critical to maximally utilize the available information from big data for the general well-being of mankind.

Acknowledgments

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Kastens, K.A.; Manduca, C.A.; Cervato, C.; Frodeman, R.; Goodwin, C.; Liben, L.S.; Mogk, D.W.; Spangler, T.C.; Stillings, N.A.; Titus, S. How geoscientists think and learn. *Eos Trans. Am. Geophys. Union* **2009**, *90*, 265–272.

2. Lin, C.C.; Shu, F.H. On the spiral structure of disk galaxies. *Astrophys. J.* **1964**, *140*, 646–655.

3. Vasavada, A.R.; Showman, A. Jovian atmospheric dynamics: An update after Galileo and Cassini. *Rep. Progr. Phys.* **2005**, *68*, 1935–1996.

4. Zhang, G.M.; Yu, L. Emergent phenomena in physics (in Chinese). *Physics* **2010**, *39*, 543–549.

5. Hemelrijk, C.K.; Hildenbrandt, H. Some causes of the variable shape of flocks of birds. *PLoS One* **2011**, *6*, e22479.

6. Hildenbrandt, H.; Carere, C.; Hemelrijk, C.K. Self-organized aerial displays of thousands of starlings: A model. *Behav. Ecol.* **2010**, *21*, 1349–1359.

7. Shaw, E. Schooling fishes. *Am. Sci.* **1978**, *66*, 166–175.

8. Reynolds, C.W. Flocks, herds and schools: A distributed behavioral model. *Comput. Graph.* **1987**, *21*, 25–34.

9. D'Orsogna, M.R.; Chuang, Y.L.; Bertozzi, A.L.; Chayes, L.S. Self-propelled particles with soft-core interactions: Patterns, stability, and collapse. *Phys. Rev. Lett.* **2006**, *96*, e104302.

10. Hemelrijk, C.K.; Hildenbrandt, H. Self-organized shape and frontal density of fish schools. *Ethology* **2008**, *114*, 245–254.

11. Kroy, K.; Sauermann, G.; Herrmann, H.J. Minimal model for sand dunes. *Phys. Rev. Lett.* **2002**, *88*, e054301.

12. Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; The University of Illinois Press: Champaign, IL, USA, 1949.

13. Khinchin, A.I. *Mathematical Foundations of Information Theory*; Courier Dover Publications: Mineola, NY, USA, 1957.

14. Erill, I. Information theory and biological sequences: Insights from an evolutionary perspective. In *Information Theory: New Research*; Deloumeayx, P., Gorzalka, J.D., Eds.; Nova Science Publishers: Hauppauge, NY, USA, 2012; pp. 1–28.

15. Lempel, A.; Ziv, J. On the complexity of finite sequences. *IEEE Trans. Inf. Theory* **1976**, *22*, 75–81.

16. Ziv, J.; Lempel, A. Compression of individual sequences via variable-rate coding. *IEEE Trans. Inf. Theory* **1978**, *24*, 530–536.

17. Shor, P. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A* **1995**, *52*, R2493–R2496.

18. Hamming, R. *The Art of Probability for Scientists and Engineers*; Addison-Wesley: Boston, MA, USA, 1991.

19. Tribus, M. *The Maximum Entropy Formalism*; Levine, R.D., Tribus, M., Eds.; MIT Press: Cambridge, MA, USA, 1978.

20. Gao, J.B.; Cao,Y.H.; Tung, W.W.; Hu, J. *Multiscale Analysis of Complex Time Series—Integration of Chaos and Random Fractal Theory, and Beyond*; Wiley Interscience: New York, NY, USA, 2007.

21. Gatlin, L.L. *Information Theory and the Living System*; Columbia University Press: New York, NY, USA, 1972.

22. Gao, J.B. Recurrence time statistics for chaotic systems and their applications. *Phys. Rev. Lett.* **1999**, *83*, 3178–3181.

23. Gao, J.B.; Cai, H.Q. On the structures and quantification of recurrence plots. *Phys. Lett. A* **2000**, *270*, 75–87.

24. Gao, J.B.; Hu, J.; Tung, W.W.; Cao, Y.H.; Sarshar, N.; Roychowdhury, V.P. Assessment of long range correlation in time series: How to avoid pitfalls. *Phys. Rev. E* **2006**, *73*, e016117.

25. Gao, J.B.; Hu, J.; Tung, W.W. Facilitating joint chaos and fractal analysis of biosignals through nonlinear i adaptive filtering. *PLoS One* **2011**, *6*, e24331.

26. Mandelbrot, B.B. *The Fractal Geometry of Nature*; W.H. Freeman and Company: New York, NY, USA, 1982.

27. Gouyet, J.F. *Physics and Fractal Structures*; Springer: New York, NY, USA, 1995.

28. Feder, J. *Fractals*; Plenum Press: New York, NY, USA, 1988.

29. Falconer, K.J. *Fractal Geometry: Mathematical Foundations and Applications*; John Wiley & Sons: Chichester, UK, 1990.

30. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487.

31. Tsallis, C.; Levy, S.V.F.; Souza, A.M.C.; Maynard, R. Statistical-mechanical foundation of the ubiquity of levy distributions in nature. *Phys. Rev. Lett.* **1995**, *75*, 3589–3593.

32. Hu, J.; Tung, W.W.; Gao, J.B. A new way to model non-stationary sea clutter. *IEEE Signal Process. Lett.* **2009**, *16*, 129–132.

33. Atmanspacher, H.; Scheingraber, H. A fundamental link between system theory and statistical mechanics. *Found. Phys.* **1987**, *17*, 939–963.

34. Grassberger, P.; Procaccia, I. Estimation of the Kolmogorov entropy from a chaotic signal. *Phys. Rev. A* **1983**, *28*, 2591–2593.

35. Osborne, A.R.; Provenzale, A. Finite correlation dimension for stochastic-systems with power-law spectra. *Phys. D Nonlinear Phenom.* **1989**, *35*, 357–381.

36. Provenzale, A.; Osborne, A.R.; Soj, R. Convergence of the K2 entropy for random noises with power law spectra. *Phys. D Nonlinear Phenom.* **1991**, *47*, 361–372.

37. Gao, J.B.; Hu, J.; Tung, W.W.; Cao, Y.H. Distinguishing chaos from noise by scale-dependent Lyapunov exponent. *Phys. Rev. E* **2006**, *74*, e066204.

38. Gao, J.B.; Hu, J.; Mao, X.; Tung, W.W. Detecting low-dimensional chaos by the "noise titration" technique: Possible problems and remedies. *Chaos Solitons Fractals* **2012**, *45*, 213–223.

39. Gao, J.B.; Hwang, S.K.; Liu, J.M. When can noise induce chaos? *Phys. Rev. Lett.* **1999**, *82*, 1132–1135.

40. Gao, J.B.; Tung, W.W.; Hu, J. Quantifying dynamical predictability: The pseudo-ensemble approach (in honor of Professor Andrew Majda's 60th birthday). *Chin. Ann. Math. Ser. B* **2009**, *30*, 569–588.

41. Gao, J.B.; Hu, J.; Tung, W.W.; Zheng, Y. Multiscale analysis of economic time series by scale-dependent Lyapunov exponent. *Quant. Financ*. **2011**, *13*, 265–274.

42. Gao, J.B.; Hu, J.; Tung, W.W. Entropy measures for biological signal analysis. *Nonlinear Dyn*. **2012**, *68*, 431–444.

43. Feldman, D.P.; Crutchfield, J.P. Measures of statistical complexity: Why? *Phys. Lett. A* **1998**, *238*, 244–252.

44. Feldman, D.P.; Crutchfield, J.P. Structural information in two-dimensional patterns: Entropy convergence and excess entropy. *Phys. Rev. E* **2003**, *67*, e051104.

45. Emmert-Streib, F. Statistic complexity: Combining Kolmogorov complexity with an ensemble approach. *PLoS One* **2010**, *5*, e12256.

46. Zunino, L.; Soriano, M.C.; Rosso, O.A. Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach. *Phys. Rev. E* **2012**, *86*, e046210.

47. Torcini, A.; Grassberger, P.; Politi, A. Error propagation in extended chaotic systems. *J. Phys. A Math. Gen.* **1995**, doi:10.1088/0305-4470/28/16/011.

48. Aurell, E.; Boffetta, G.; Crisanti, A.; Paladin, G.; Vulpiani, A. Growth of non-infinitesimal perturbations in turbulence. *Phys. Rev. Lett.* **1996**, doi:10.1103/PhysRevLett.77.1262.

49. Aurell, E.; Boffetta, G.; Crisanti, A.; Paladin, G.; Vulpiani, A. Predictability in the large: An extension of the concept of Lyapunov exponent. *J. Phys. A Math. Gen.* **1997**, *30*, 1–26.

50. Gaspard, P.; Wang, X.J. Noise, chaos, and (ϵ, τ)-entropy per unit time. *Phys. Rep.* **1993**, *235*, 291–343.

51. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of biological signals. *Phys. Rev. E* **2005**, *71*, e021906.

52. Prigogine, I. *From Being to Becoming*; W.H. Freeman and Company: New York, NY, USA, 1980.

53. Lebowitz, J.L. Microscopic origins of irreversible macroscopic behavior. *Phys. A Stat. Mech. Appl.* **1999**, *263*, 516–527.

54. Feynman, R. *The Character of Physical Law*; MIT Press: Cambridge, MA, USA, 1967; p. 112.

55. Arnold, V.I. *Mathematical Methods of Classical Mechanics*; Springer-Verlag: Berlin, Germany, 1978; p. 71.

56. Kac, M. *Probability and Related Topics in Physical Sciences*; Interscience Publishers: Hoboken, NJ, USA, 1959; Lectures in Applied Mathematics, Series, Volume 1A, p. 62.

57. Villani, C. A review of mathematical topics in collisional kinetic theory. *Handb. Math. Fluid Dyn.* **2002**, *1*, 71–74.

58. Wang, G.M.; Sevick, E.M.; Emil Mittag, Searles, D.J.; Evans, D.J. Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales. *Phys. Rev. Lett.* **2002**, *89*, e050601.

59. Schrodinger, E. *What is Life?* Macmillan Publishers: London, UK, 1946.

60. Penrose, R. *The Emperor's New Mind: Concerning Computers, Minds, and The Laws of Physics*; Oxford University Press: Oxford, UK, 1989.

61. Kolmogorov, A.N. Three approaches to the quantitative definition of "information". *Probl. Inf. Transm.* **1965**, *1*, 1–7.

62. Kolmogorov, A.N. Logical basis for information theory and probability theory. *IEEE Trans. Inf. Theory* **1968**, *IT-14*, 662–664.

63. Zhang, X.S.; Roy, R.J.; Jensen, E.W. EEG complexity as a measure of depth of anesthesia for patients. *IEEE Trans. Biomed. Eng.* **2001**, *48*, 1424–1433.

64. Nagarajan, R. Quantifying physiological data with Lempel-Ziv complexity: Certain issues. *IEEE Trans. Biomed. Eng.* **2002**, *49*, 1371–1373.

65. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley: New York, NY, USA, 1991.

66. Rapp, P.E.; Cellucci, C.J.; Korslund, K.E.; Watanabe, T.A.A.; Jiménez-Montaño, M.A. Effective normalization of complexity measurements for epoch length and sampling frequency. *Phys. Rev. E* **2001**, *64*, e016209.

67. Hu, J.; Gao, J.B.; Principe, J.C. Analysis of biomedical signals by the Lempel-Ziv complexity: The effect of finite data size. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 2606–2609.

68. Gao, J.B.; Hu, J.; Tung, W.W. Complexity measures of brain wave dynamics. *Cogn. Neurodyn.* **2011**, *5*, 171–182.

69. Loewenstern, D.; Yianilos, P.N. Significantly lower entropy estimates for natural DNA sequences. *J. Comput. Biol.* **1999**, *6*, 125–142.

70. Fernandez, P.R.; Munteanu, C.R.; Escobar, M.; Prado-Prado, F.; Martn-Romalde, R.; Pereira, D.; Villalba, K.; Duardo-Snchez, A.; Gonzlez-Daz, H. New Markov-Shannon entropy models to assess connectivity quality in complex networks: From molecular to cellular pathway, Parasite-Host, Neural, Industry, and Legal-Social networks. *J. Theor. Biol.* **2012**, *293*, 174–188.

71. Cattani, C. On the existence of wavelet symmetries in Archaea DNA. *Comput. Math. Methods Med.* **2012**, doi:10.1155/2012/673934.

72. Cattani, C.; Pierro, G. On the fractal geometry of DNA by the binary image analysis. *Bull. Math. Biol.* **2013**, *1*, 1–27.

73. Ramakrishnan, N.; Bose, R. Dipole entropy based techniques for segmentation of introns and exons in DNA. *Appl. Phys. Lett.* **2012**, doi:10.1063/1.4747205.

74. Cattani, C. Uncertainty and Symmetries in DNA Sequences. In Proceedings of the 4th International Conference on Biomedical Engineering in Vietnam, Ho Chi Minh, Vietnam, 10–12 January 2012; Springer: Berlin/Heidelberg, Germany, 2013; pp. 359–369.

75. Jani, M.; Azad, R.K. Information entropy based methods for genome comparison. *ACM SIG Bioinform.* **2013**, doi:10.1145/2500124.2500126.

76. Luce, R.D. Whatever happened to information theory in psychology? *Rev. Gen. Psychol.* **2003**, *7*, 183–188.

77. Hirsh, J.B.; Mar, R.A.; Peterson, J.B. Psychological entropy: A framework for understanding uncertainty-related anxiety. *Psychol. Rev.* **2012**, *119*, 304–320.

78. Shi, W.Z.; Wong, M.S.; Wang, J.Z.; Zhao, Y.L. Analysis of airborne particulate matter (PM2.5) over Hong Kong using remote sensing and GIS. *Sensors* **2012**, *12*, 6825–6836.

79. Georgescu-Roegen, N. *The Entropy Law and Economic Process*; Harvard University Press: Cambridge, MA, USA, 1971.

80. Solow, R. A contribution to the theory of economic growth. *Q. J. Econ.* **1956**, *70*, 65–94.

81. Swan, T. Economic growth and capital accumulation. *Econ. Record* **1956**, *32*, 334–361.

82. McCauley, J.L. Thermodynamic analogies in economics and and finance: Instability of markets. *Phys. A Stat. Mech. Appl.* **2003**, *329*, 199–212.

83. Gao, J.B.; Hu, J.; Mao, X.; Zhou, M.; Gurbaxani, B.; Lin, J.W.B. Entropies of negative incomes, Pareto-distributed loss, and financial crises. *PLoS One* **2011**, *6*, e25053.

84. Mimkes, J. Stokes integral of economic growth: Calculus and Solow model. *Phys. A Stat. Mech. Appl.* **2010**, *389*, 1665–1676.

85. Manyika, J.; Chui, M.; Brown, B.; Bughin, J.; Dobbs, R.; Roxburgh, C.; Byers, A.H. Big data: The next frontier for innovation, competition, and productivity. 2011. Available online: http://www.mckinsey.com/insights/business technology/big data the next frontier for innovation (accessed on 26 March 2013).