PROCEEDINGS OF THE ROYAL SOCIETY A | MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The physical character of information

Mahesh Karnani, Kimmo Pääkkönen and Arto Annila

References	**This article cites 50 articles, 5 of which can be accessed free** http://rspa.royalsocietypublishing.org/content/early/2009/04/17/rspa.2009.0063.full.html#ref-list-1
P<P	Published online 22 April 2009 in advance of the print journal.
EXiS Open Choice	This article is free to access
Subject collections	Articles on similar topics can be found in the following collections biophysics (80 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to *Proc. R. Soc. A* go to: **http://rspa.royalsocietypublishing.org/subscriptions**

The physical character of information

BY MAHESH KARNANI[1], KIMMO PÄÄKKÖNEN[2] AND ARTO ANNILA[1,2,3,*]

[1]*Department of Biosciences,* [2]*Institute of Biotechnology, and*
[3]*Department of Physics, University of Helsinki, 00014 Helsinki, Finland*

The mathematical theory of communication defines information in syntax without reference to its physical representation and semantic significance. However, in an everyday context, information is tied to its representation and its content is valued. The dichotomy between the formal definition and the practical perception of information is examined by the second law of thermodynamics that was recently formulated as an equation of motion. Thermodynamic entropy shows that the physical representation of information is not inconsequential in generation, transmission and processing of information. According to the principle of increasing entropy, communication by dissipative transformations is a natural process among many other evolutionary phenomena that level energy-density differences between components of a communication system and its surroundings. In addition, information-guided processes direct down along descents on free energy landscapes. The non-integrable equation for irreversible processes reveals that there is no universal analytical algorithm to match source to channel. Noise infiltration is also regarded by the second law as an inevitable consequence of energy transduction between a communication system and its surroundings. Communication is invariably associated with misunderstanding because mechanisms and means of information processing at the receiver differ from those at the sender. The significance of information is ascribed to the increase in thermodynamic entropy in the receiver system that results from execution of the received message.

1. Introduction

Information theory has its roots in statistical mechanics (Szilard 1929; Fisher 1935; Shannon 1948; Kullback 1959; Shannon & Weaver 1962; Jaynes 2003). This foundation is apparent in the central theorems of Shannon (1948) and Shannon & Weaver (1962). The source coding theorem defines average information per message as entropy H,

$$H = k \sum_{j=1}^{N} p_j \log\left(\frac{1}{p_j}\right), \tag{1.1}$$

where p_j is the probability of a message j normalized by all messages N and $k > 0$ is a mere constant of a unit of measure, in analogy to Boltzmann's constant k_B that relates to the average energy T in units of kelvin. Shannon's noisy-channel

* Author and address for correspondence: Department of Physics, University of Helsinki, 00014 Helsinki, Finland (arto.annila@helsinki.fi).

1

coding theorem says that the message j, eventually represented in redundant numbers $N_j > 1$, can be communicated reliably when encoded so that the rate of communication, including noise, is below the channel's capacity. These two well-known theorems, however, speak neither about the means to represent the message nor about the significance of the message. At first sight, our concern for the representation may appear downright secondary and our interest in the significance categorically specific. Is information not abstract by its nature and its content associated with context?

At least genetic information is chemically coded in nucleic acid genomes, just as data are deposited on magnetic or optic media, but, in general, we suspect that there is no information without representation. Consistent with the representation requisite of information, energy densities are known to set the ultimate bounds for entropy (Brillouin 1960). These physical bounds, as we will show, influence structuring of information on any medium or communication channel. Therefore, natural data structures are skewed, nearly lognormal distributions (Kapteyn 1903; Limpert *et al.* 2001), that sum up sigmoidally, approximately in a power-law manner, e.g. lengths of genes (Zhang 2000), words of natural languages (Naranan & Balasubrahmanyan 1992; Sigurd *et al.* 2004) and edges in the World Wide Web (Albert & Barabasi 2002). The physical nature of information can be examined by the second law using its recent formulation (Sharma & Annila 2007) that explicitly links increasing thermodynamic entropy with decreasing free energy. The formalism reveals that communication is a natural process, one among many others to disperse energy. The holistic tenet is known as the maximum-entropy production principle (Ziegler 1983; Swenson 1989, 2000; Mahulikar & Harwig 2004; Martyushev & Seleznev 2006) that has been used to describe various evolutionary phenomena (Ulanowicz & Hannon 1987; Brooks & Wiley 1988; Salthe 1993; Schneider & Kay 1994; Matsuno & Swenson 1999; Brooks 2000; Lorenz 2002; Dewar 2003). Specifically, the maximum-entropy production principle allows one to address physical information in evolution (Swenson & Turvey 1991).

In this study, we find from the principle of increasing entropy, when it is given as an equation of motion, that information is affected by its physical representation and communication is directed by thermodynamic driving forces via mechanisms of energy transduction. The differential equation of motion for natural processes can also be given as the principle of least action (Kaila & Annila 2008) and associated with Newton's second law (Tuisku *et al.* 2009) to rationalize diverse evolutionary courses (Grönholm & Annila 2007; Annila & Annila 2008; Annila & Kuismanen 2008; Jaakkola *et al.* 2008*a*,*b*; Würtz & Annila 2008; Karnani & Annila 2009; Sharma *et al.* 2009). Although the following results concerning the physical nature of information and thermodynamic imperatives in information processing are not conceptually new, they are derived directly from the second law given in the form of an equation of motion. We will first define the thermodynamic entropy associated with physical representations of information. Subsequently, the definition is inspected to show that it qualifies as a measure of information. Then, the principle of increasing entropy is used to understand evolution of data structures and imperatives in information processing and transmission. Finally, the thermodynamic theory of communication allows one to associate the significance of a message with the entropy increase in the receiver system resulting from the execution of the message.

2. Thermodynamic information

We are by no means the first to claim that information is physical (Szilard 1929; Landauer 1961, 1991, 1996*a*,*b*; Swenson 1989; Lloyd 2000), but wish to underline that all information processing, i.e. generation, encoding, transmission, decoding and interpretation, belong to the class of natural processes (Kondepudi & Prigogine 1998; Sharma & Annila 2007) where entropy increases by dispersal of energy. The role of thermodynamics in communication cannot be ignored because a message must be constructed for it to exist and executed for it to signify. For example, any observation, as a means of collecting information, is an irreversible energy transduction process (Brillouin 1960; Tuisku *et al.* 2009). Our objective is to use the physical portrayal of information and processing to resolve the dichotomy between the formal mathematical definition of information limited to the syntax by equation (1.1) and the practical association of information with its significance.

Let us consider a medium where all bits are initially set to 0. It will take a quantum of energy in a transition to turn a bit from 0 to 1, i.e. to write the simplest message. The energy input is the necessary ingredient that distinguishes the 0-state from the 1-state. In a biological context, this transformation operation from one state to another could correspond to absorbance of a photon in a retinol molecule or to a synthesis of a nucleic acid residue. Thus, according to the adopted thermodynamic viewpoint, there is information in deviations, 'up or down', from the average energy density, i.e. in the free energy. Conversely, there is no flow of information when all thermodynamic gradients have vanished. We will proceed to establish the thermodynamic entropy as a measure of information.

In a thermodynamic system (e.g. an organism or a file server), the amount of energy that is required to build and maintain a data structure (e.g. a genome or an electronic archive) may be minute in comparison with other maintenance costs (e.g. of cellular metabolism or air conditioning), but it is still necessary. The energy density that is invested to represent a message j in identical copies of N_j is given by $\Phi_j = N_j \exp(G_j/k_B T)$, where the Gibbs free energy G_j is relative to the average energy $k_B T$ (Gibbs 1993–1994). For example, the energy density $\int d\Phi_j$ integrated over the medium relates to temperature T via Boltzmann's constant k_B. Owing to the physical representation requisite, all forms of information processing must comply with thermodynamics. Thus, according to the thermodynamic standpoint, any form of density-in-energy holds information, irrespective of whether it is explicitly referred to as a message or not. Information ascribed to a message depends on the surroundings because the free energy in deviations is relative to the surrounding average density. The dependence is apparent in information-guided processes, e.g. gene expression and protein folding, which direct in different ways in differing surroundings.

The natural process of energy dispersal (Carnot 1824; Boltzmann 1905; Kondepudi & Prigogine 1998) advances most rapidly when the most voluminous flows of energy funnel from high densities to low densities down along the steepest available descents of the free energy (Darwin 1859; Sharma & Annila 2007; Kaila & Annila 2008; Tuisku *et al.* 2009). This is the principle of maximal entropy production. The thermodynamic entropy $S = k_B \sum \ln P_j$ used here is derived from statistical physics (Sharma & Annila 2007). It is an additive logarithmic probability measure of an open system where each $\ln P_j$ term denotes the energy requirements to represent a message j in numbers N_j on a medium
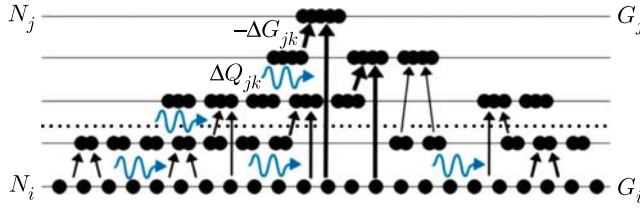
Figure 1. Information is physically represented by densities-in-energy $\Phi_j = N_j \exp(G_j/k_B T)$ (filled objects). A message j in numbers N_j is represented by the density Φ_j at an energy level G_j given in relation to the average energy density $k_B T$ (dotted line). The physical representations of information result from dissipative processing via jk-transitions (vertical arrows) that couple to changes in Gibbs free energy $-\Delta G_{jk}$ and possibly also recruit external energy ΔQ_{jk} (wavy arrows) in the assembly process. During the transformations of representations, flows of energy channel via various mechanisms σ_{jk} (depicted as varying thicknesses of arrows). According to the principle of least action, the most voluminous transforming flows direct along the steepest descents in the free energy landscape to increase entropy most rapidly. In this way, information processing and communication are formulated as thermodynamic processes that are connected to the surrounding processes.

$$
\begin{aligned}
S &= k_B \sum_{j=1} \ln P_j = k_B \sum_{j=1} \ln \left(\prod_k \frac{N_k}{g_{jk}!} \exp \frac{\Delta Q_{jk} - \Delta G_{jk}}{k_B T} \right)^{N_j} \Bigg/ N_j! \\
&\approx k_B \sum_{j=1} N_j \left(\sum_k \ln \frac{N_k}{N_j} + \frac{\Delta Q_{jk} - \Delta G_{jk}}{k_B T} - \ln g_{jk}! + 1 \right) \\
&= k_B \sum_{j=1} N_j \left(\frac{\sum_k \mu_k + \Delta Q_{jk} - \mu_j}{k_B T} + 1 \right),
\end{aligned}
\tag{2.1}
$$

where N_k denotes the physical representatives of symbols that are available to construct the messages N_j. The energy required for the coding is given by the difference $G_j - \sum G_k = \Delta G_{jk}$ between the message j and its constituents k that are eventually used in degenerate numbers g_{jk} for each message j. The amount of external energy ΔQ_{jk} that is required in the synthesis is also denoted. The form of S is self-similar, i.e. densities are composed of densities, describing hierarchical organization of energy transduction networks (Salthe 1985, 2002). When S is multiplied by T, the resulting formula of energy expresses, for each message j, how much energy $N_j k_B T$ is bound in its representations and how much energy $N_j(\Sigma \mu_k + \Delta Q_{jk} - \mu_j) > 0$ is still free to increase the number of representations N_j. The representations N_j result from statistically independent actions, denoted by $\prod_k$, when constituents N_k react with each other and bring about concomitant changes in energy $\Delta Q_{jk} - \Delta G_{jk}$. The natural process aims at balancing the potential $\mu_k = k_B T \ln \Phi_k / g_{jk}!$ associated with the physical representations N_k and the available energy ΔQ_{jk} in the surroundings with the potential $\mu_j = k_B T \ln \Phi_j$ that is associated with the physical representations N_j (figure 1). Equation (2.1), unlike earlier statistical physics entropy formulae, is valid for open systems. It is consistent with Carnot's definition of entropy and the basic maxim of thermodynamics that the entropy maximum coincides with the free energy minimum.

The rate of entropy increase is obtained from equation (2.1) by differentiation and using the chain rule (Sharma & Annila 2007)

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \sum_{j=1} \frac{\mathrm{d}S_j}{\mathrm{d}N_j} \frac{\mathrm{d}N_j}{\mathrm{d}t} = k_{\mathrm{B}} \sum_{j=1} \frac{\mathrm{d}N_j}{\mathrm{d}t} \left(\sum_k \ln \frac{N_k/g_{jk}!}{N_j} + \frac{\Delta Q_{jk} - \Delta G_{jk}}{k_{\mathrm{B}} T} \right)$$

$$= k_{\mathrm{B}} \sum_{j=1} \frac{\mathrm{d}N_j}{\mathrm{d}t} \left(\frac{\sum_k \mu_k + \Delta Q_{jk} - \mu_j}{k_{\mathrm{B}} T} \right) \geq 0. \qquad (2.2)$$

It reveals that entropy is increasing when flows $\mathrm{d}N_j/\mathrm{d}t$ via various jk-pathways are creating and destroying representations of messages by levelling various potential energy differences $\sum \mu_k + \Delta Q_{jk} - \mu_j$. In practice, the jk-indexing defines algorithms (physical processes) that yield the representations N_j from the representations N_k via syntheses or degradations (e.g. dissipative transformations that flip bits). The inequality characteristic of the second law merely states that the information processing system will always move towards more probable states. For example, when free energy is available $N_j(\sum \mu_k + \Delta Q_{jk} - \mu_j) > 0$, the system will generate the j-messages. Otherwise (when less than 0), the j-messages will be corrupted, as their representations will be consumed in other transformations. The rate criterion $\mathrm{d}S/\mathrm{d}t \geq 0$ implies maximal increase that is achieved when the system evolves at maximal rates $(\mathrm{d}N_j/\mathrm{d}t)(\sum \mu_k + \Delta Q_{jk} - \mu_j)$, i.e. via the most voluminous flows $\mathrm{d}N_j/\mathrm{d}t$ of representations (e.g. communication) down along the steepest gradients $\mu_k + \Delta Q_{jk} - \mu_j$. The natural process aims at the free energy minimum partition by abolishing all energy-density differences, i.e. $N_j(\sum \mu_k + \Delta Q_{jk} - \mu_j) = 0$.

The connection between entropy and free energy provided by equation (2.1) and between their rates by equation (2.2) means that it is impossible to create or destroy information without a change in free energy. Information defined by thermodynamic entropy is embodied in existing representations $k_{\mathrm{B}} \sum N_j$ and in the free energy $k_{\mathrm{B}} N_j(\sum \mu_k + \Delta Q_{jk} - \mu_j)/k_{\mathrm{B}} T$, which is available to make more representations. At the maximum-entropy state, the information processing system houses the free energy minimum distribution (partition) of messages. Importantly, the surrounding densities-in-energy, denoted by ΔQ_{jk}, influence the structuring and processing of information. Eventually, in the expanding Universe, all energy-density differences in material forms will dissipate and, since the free energy will disappear, all information will also vanish (Tuisku *et al.* 2009). The thermodynamic description links information and its processing systems seamlessly with their surroundings.

The flow of information is the flow of its representations $\mathrm{d}N_j/\mathrm{d}t$. The flow is proportional to the free energy per $k_{\mathrm{B}} T$ (Sharma & Annila 2007),

$$\frac{\mathrm{d}N_j}{\mathrm{d}t} = \sum_k \sigma_{jk} \left(\ln \frac{N_k/g_{jk}!}{N_j} + \frac{\Delta Q_{jk} - \Delta G_{jk}}{k_{\mathrm{B}} T} \right) = \sum_k \sigma_{jk} \left(\frac{\sum_k \mu_k + \Delta Q_{jk} - \mu_j}{k_{\mathrm{B}} T} \right),$$

$$(2.3)$$

where the coefficient σ_{jk} denotes the mechanistic capacity (conductance) of a channel for the particular message j in its material representations that transform from N_k. When there is more than one algorithm to construct a particular message, or more than one channel to transmit it, the flows of

information distribute among diverse processes and channels (figure 1) so as to maximize the overall rate dS/dt. This is just to say that receiver systems depend on their channel capacities when acquiring information from their surroundings. This implies that the receiver systems are competing for information, as they are competing for free energy in general. The thermodynamic bias, known as the free energy, will direct evolution towards increasingly more effective mechanisms to detect, acquire and process free energy from the surrounding sources.

It is worth pointing out that, when the law of mass action or the logistic equation is used to model flows, instead of using equation (2.3), the kinetics is decoupled from the thermodynamics. Consequently, the inaccurate models direct investigations to various kinetic mechanisms, hoping to find those that would reproduce the observed characteristics, rather than realizing that functional structures and hierarchical organizations emerge naturally to consume free energy (Annila & Annila 2008).

3. Properties of information

To justify the adopted physical formalism, we will proceed to relate equations (2.1)–(2.3) to the basic concepts of information theory and to the known mathematical properties of information (Fisher 1935; Mathai & Rathie 1975). The part $k_B \sum N_j \ln(1/N_j)$ of the first term on the second line of equation (2.1) resembles the average information per message given by equation (1.1). However, the physical representation of the message j in numbers N_j, given by equation (2.1), is not normalized by all messages N so that it would correspond exactly to the probability $p_j = N_j/N$ in equation (1.1). The reason for not normalizing in the thermodynamic context is that the total number of messages may not be known *a priori* when the physical medium and the amount of energy to deposit and transform information on the medium are not defined. In other words, the information processing system is evolving. Therefore, the upper index of j-summation in equations (2.1) and (2.2) has also been left open. For example, a genome as a repository of information may grow larger or shrink down during evolution (Gregory 2005). Likewise, increasingly larger hard disks are currently acquired to deposit more information in electronic form. The reason for not normalizing probabilities is more fundamental than a mere unawareness of the available load and capacity. The principal reason is that all information processing systems are open to their surroundings in order to carry out dissipative transformations that distinguish identities from each other (Tuisku *et al.* 2009). Thus, there is no conserved quantity to qualify for a norm. When the driving forces and flows are inseparable from each other, the information processing system is non-Hamiltonian.

Thermodynamic entropy S representing information is *non-negative*, as is information entropy H. To show that S, like H, is *convex*, we insert equation (2.3) into equation (2.2) to obtain the quadratic form $dS/dt = k_B \sum \sigma_{jk} [(\sum \mu_k - \mu_j + \Delta Q_{jk})/k_B T]^2 \geq 0$, which is non-negative because a communication machinery is inevitably realized from non-negative Φ_j, hence all conductance $\sigma_{jk} > 0$. The second law of thermodynamics $dS \geq 0$ underlies the mathematically established convexity of information. The correspondence between the physical and mathematical definitions of information is further strengthened because the

thermodynamic entropy does not depend on the order of terms that represent messages, hence S of equation (2.1), in analogy to H of equation (1.1), is an *additive* measure of information. However, as will be described later, the rate of entropy increase dS/dt at any given moment during the information processing at the receiver system does depend on the order in which messages are received. In other words, two messages that associate with equally large total dissipation can be distinguished from each by their different rates of dissipation.

Since thermodynamic driving forces are not all independent of each other (Kondepudi & Prigogine 1998), a change in a potential μ_k will affect any other potential μ_j in the communication system. For example, when the storage capacity cannot rapidly adjust in response to a demand, the memory space that is used for one message is away from the capacity available to other messages. Dissipative transformations among the interdependent physical representations of information are invariably coupled to the entropy change. Hence S, similar to H, is *continuous*. This is in accordance with the Radon–Nikodym theorem that tells us how to change from one probability measure to another.

The rate of change in thermodynamic entropy, given by equation (2.2), clarifies that emergence of a nested hierarchical data structure of 'messages within messages' is a probable evolutionary scenario. For example, a genome contains chromosomes that, in turn, are composed of gene regulatory networks, which house genes written by codons that are made of bases (Gregory 2005). Likewise, a book contains chapters that, in turn, are composed of paragraphs that house sentences written by words made of letters. Such an organization of information is a natural outcome, as it results from evolution towards the free energy minimum state. The optimum partition is most effective in energy dispersal. At every level of hierarchy, the jk-indexing in equation (2.2) denotes how to construct each message j from k by dissipative processes, with concomitant changes in energy. This self-similar branching principle is familiar from diverse evolutionary processes that manifest themselves, e.g. as phylogenic trees. In information theory, the self-similar branching characteristic is referred to as *recursivity*.

Differences in dissipative transformations are necessary for the message j to distinguish from all other representations corresponding to other messages $k \neq j$ in the same information processing system (Kaila & Annila 2008; Tuisku *et al.* 2009). For example, entropy of a word will change when a letter j is exchanged to another symbol k, but S remains intact in a conserved exchange of j for the identical symbol j or for another symbol $\hat{\jmath}$ that the processing system fails to distinguish from j. It is worth noting that a given letter in upper case is a different symbol from its lower case. They are associated with different energies, and the processing system is said to understand this when it is able to distinguish energetically one from the other. Likewise, a mutation in a nucleic acid codon associates inherently with a change in energy. Bases are distinguished from each other precisely by energy differences in pairing interactions (Almlöf *et al.* 2007). However, a change at the molecular level is not necessarily distinguished at a cellular level. Not every genetic mutation will lead to a phenotype manifestation.

The information measure, referred to as *directed divergence*, is the first term $N_j \ln(N_k/N_j)$ in equation (2.1) and its sum is the *divergence* (Kullback 1959), also known as the Kullback–Leibler divergence or relative entropy, which satisfies Gibbs' inequality. Accordingly, thermodynamics is formulated in terms of
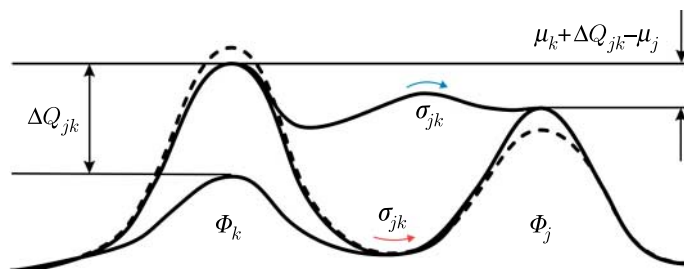
Figure 2. Information processing is dissipative transformations of physical representations from an energy density Φ_k to another Φ_j that together form an energy-density landscape. The data structure is formed and maintained by an influx of energy ΔQ_{jk} from surroundings (e.g. via current from a power chord or by combustion of metabolites) that couples to the information processing. The flows of energy (red and blue arrows) channel down along differences in energy $\mu_k + \Delta Q_{jk} - \mu_j$ by various mechanisms, e.g. σ_{jk}. Since potentials μ_j are in relation to each other, the flows of information keep changing with the changing landscape (present, solid curve; prior, dashed curve), leaving no invariants of motion to predict precise trajectories. There is, therefore, no universal deterministic algorithm that would enable the use of full capacity at all times to transmit *a priori* unknown data streams.

relative measures; differences between μ_j are important, not their absolute values, and G_j are given relative to $k_B T$. In the same relative sense, a change in entropy S, or equivalently a reduction in free energy by an energy flow, is the thermodynamic measure of discriminating between representations of messages (Kullback 1959). For example, two messages j and $\hat{j}$ are distinguished from each other when the processor finds an energy difference in their physical representations, otherwise the processor regards the two as indistinguishable (Kaila & Annila 2008; Tuisku *et al.* 2009). Specifically, when no energy-density differences between the two representations are found during encoding, the messages are, according to equation (2.1), indistinguishable at the source. Likewise, when no energy differences between the two representations are found during decoding, the receiver system regards the corresponding messages as identical. However, a communication system may evolve in its mechanisms to distinguish finer and finer energy differences, e.g. to improve in concept analysis. In all cases, when there is no change in entropy according to equation (2.2), there is no information processing or communication.

Thermodynamic entropy and the Kullback–Leibler divergence share the topological properties of distance, except for the symmetry and triangle inequality (Kullback 1959). Thermodynamic divergence is the difference $\sum \mu_k - \mu_j$ between potentials j and k. The physical reason for divergence not being a proper distance is that dissipative transformations are not symmetric under time reversal (Tuisku *et al.* 2009), i.e. the exoergic *jk*-reaction releases a quantum of energy, whereas the endoergic reaction consumes ΔQ_{jk} (figures 1 and 2). Also, the non-Euclidian characteristic means that the directional distance traversed during a *jk*-transition in a free energy landscape will be altered when other densities-in-energy begin to interact with Φ_j or Φ_k. Owing to the continuous interdependency among all Φ_j in a communication system, the 'distance' in energy among any two constituents is affected by a change in any other. The mutual

relationships are not invariants of motion because a change in identity is invariably associated with a gain or loss of energy. In the algorithmic context, the topological properties of the manifold mean that, when there are alternative ways to produce the representations N_j of message j from representations N_k, those with minimal changes in energy, i.e. least dissipation are favoured (figures 1 and 2). Although the principle itself is simple, it may be tricky to find the most optimal way of generating information when the transformation pathways are coupled to each other. A change in state will affect the set of accessible states in future.

In many cases, it is apparent that information is represented in deviations from the average energy $k_B T$. For example, information in a speech is represented in pressure perturbations about ambient atmospheric pressure, the average energy density. Likewise, a chemically coded data structure will collapse when the ambient energy density (i.e. temperature) is increased enough. In the derivation of equations (2.1) and (2.3), we have implicitly assumed that entropy is a *sufficient statistic* (Kullback 1959) because the $k_B T$ term emerged as a constant from Stirling's approximation. The system is sufficiently statistical (steady) when the energy influx or efflux does not change its average energy markedly, i.e. $\sum k_B N_j (\sum \mu_k + \Delta Q_{jk} - \mu_j) \ll k_B T$. The assumption fails when the system fails to redistribute received influx rapidly enough to maintain internal balance. For example, when a processor does not cool rapidly enough in response to a work load and overheats, the computational architecture will soon be ruined. The assumption of being sufficiently statistical also fails when a transmission line itself does not remain invariant, but begins to evolve. When the conduction system is evolving, the channel capacity depends on the transmitted message. For example, close to the high-frequency cut-off, dispersion relationships of conducting media are nonlinear and hence all messages are not transmitted equally well. The flow exceeds the channel's capacity for physical representations, and the linearity of equation (2.3) is broken down. Then, the transmission is a function of the driving force, rather than being a constant σ_{jk}. In those cases, when the system is not sufficiently statistical, underlying evolutionary processes of the constituent systems at a lower level of hierarchical organization of matter can be described by the same scale-independent thermodynamic formalism.

All in all, thermodynamic entropy shares, by the aforementioned properties, the well-known mathematical characteristics of information entropy. Thus, we conclude that the form of equation (2.1) qualifies as a genuine measure for information. Moreover, we claim that the terms in equation (2.1), which are additional to those in equation (1.1), are not of secondary importance but are essential in understanding what information is. The thermodynamic entropy distinguishes states in energetic terms, whereas information entropy dis-tinguishes by configurations. Sometimes the difference is communicated as a distinction between functional and configurational organization (Majerník 2002). The universal function is that of the energy dispersal. We accept that coherent and incoherent configurations of states are different from each other; however, the difference is recognized in detection first by a dissipative transformation. The value in understanding the physical nature of information is that it allows us to rationalize consequences of thermodynamic imperatives in communication, as well as giving, by the entropy increase, a functional meaning to a message.

4. Natural data structures

Natural data structures, such as genomes, books, file systems and data servers, are repositories of information that share common characteristics. Most notably, they display skewed distributions and hierarchical organization. The physical representation of information allows us to understand that these ubiquitous characteristics are consequences of the second law.

External energy is necessary to maintain a data structure, i.e. different energy densities. The physical representation of information is a non-equilibrium state that will begin to corrupt towards the equilibrium when external energy is diminished. Thus, communication systems are no different from other open systems that evolve towards stationary states in their respective surroundings (Bertalanffy 1953; Nicolis & Prigogine 1976; Sharma & Annila 2007; Kaila & Annila 2008; Tuisku *et al.* 2009). Under a steady influx of energy, the distribution of messages on any medium will evolve towards a stationary-state distribution that is obtained from equation (2.2) when imposing the familiar steady-state condition (Sharma & Annila 2007)

$$\frac{\mathrm{d}S}{\mathrm{d}t} = 0 \Leftrightarrow \mu_j = \sum_k \mu_k + \Delta Q_{jk} \Leftrightarrow N_j = \prod_k \frac{N_k}{g_{jk}!} \exp\left[\frac{(\Delta Q_{jk} - \Delta G_{jk})}{k_\mathrm{B}T}\right]. \qquad (4.1)$$

The recursive formula for the steady-state maximum-entropy partition implies a nested hierarchical organization of information, i.e. messages are within messages. For example, a eukaryotic genome, a message itself, is organized in chromosomes; each of them in turn, with its own organization, houses coding and non-coding sequences. Genes are flanked by intergenic regions, and a gene is fragmented by non-coding DNA sequences similar to files on a computer hard disk that fragment due to numerous input–output operations. Genomic operations, such as mutations, insertions, deletions, duplication, gene transfer, exon shuffling, intron gain and loss, as well as polyploidy, all maintain the genomic information in the skewed steady-state maximum-entropy partition (Grönholm & Annila 2007; Jaakkola *et al.* 2008*b*; Würtz & Annila 2008; figures 3 and 4). For example, frequency of amino acid versus its rank is similar to letter frequency versus its rank. Discrete power-law probability distributions, such as Zipf's law in linguistics (Zipf 1935, 1949; Altmann 2002) and Lotka's law (Lotka 1926), we recognize as approximations of equation (4.1).

The characteristically skewed distribution of information is particularly easy to illustrate when the physical representations N_j of message j are constructed from base constituents in non-degenerate ($g_{jk}=0$) numbers N_0 and using quanta $j\Delta Q_{10}$. Then, equation (4.1) yields the formula for the maximum-entropy distribution,

$$N_j = N_0^j \exp\left[\frac{j(\Delta Q_{10} - \Delta G_{10})}{k_\mathrm{B}T}\right] = \exp(-j\varepsilon), \qquad (4.2)$$

where $\varepsilon = (\Delta G_{10} - \Delta Q_{10} - k_\mathrm{B}T \ln N_0)/k_\mathrm{B}T$. The total number of base constituents $N = \sum jN_j$ is distributed with a skewed probability density $P(j) = jN_j$ and a sigmoid cumulative curve is dominated by the ubiquitous power-law region (Balasubrahmanyan & Naranan 2000; Naranan & Balasubrahmanyan 2000; figure 3).
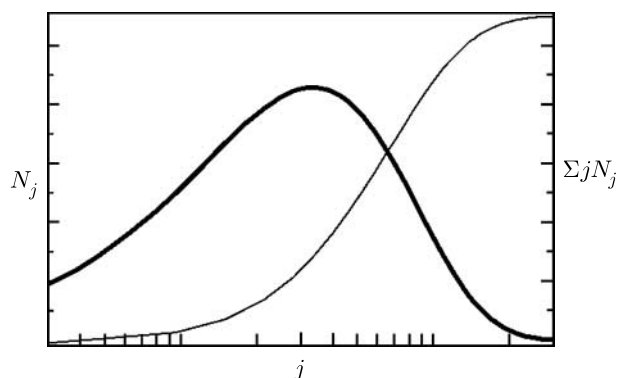
Figure 3. Maximum-entropy distribution of physical representations of information N_j versus j and cumulative curve (thin curve).
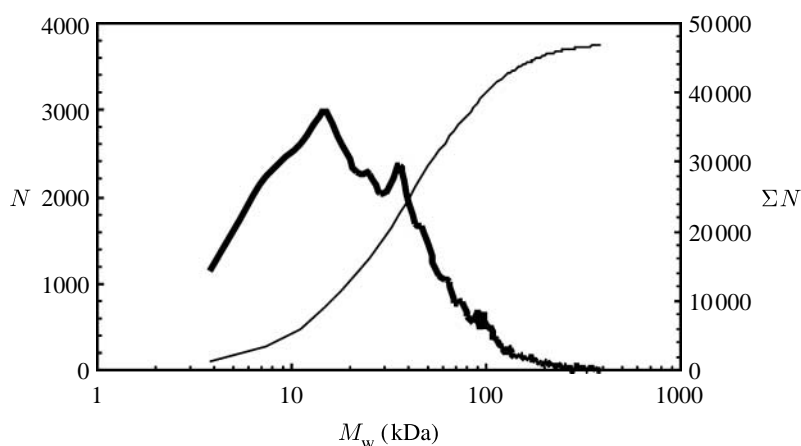

Figure 4. Distribution of known protein three-dimensional structures N versus molecular weight M_w (in unified atomic mass units) and cumulative curve (thin curve) determined by X-ray crystallography and nuclear magnetic resonance spectroscopy. The thermodynamic imperative of energy dispersal manifests itself as skewed distributions and sigmoidal cumulative curves.

   Natural data structures emerge from non-integrable processes given by the equation of motion (equation (2.2)). An initial thermodynamic driving force, e.g. for making a data structure such as a genome, may be large but means and mechanisms of information processing limit the initial growth of data. For example, when the numbers of physical representations for bits in a message are increasing, combinatorial choices among them are increasing rapidly, but soon every additional bit contributes to S less and less. Later, the rate of increase levels off when thermodynamic potentials are consumed in making all those diverse messages that do not differ all that much from each other in functional characteristics. When the physical resources are simply running out in making diverse messages, it becomes important for the system (e.g. an organism) what is deposited and transmitted to its surroundings. We claim that these material considerations are not insignificant but essential for understanding the nature of information and its processing according to the second law.

It is worth noting that the maximum-entropy distribution (equation (4.1)) was obtained directly from the condition $dS = 0$, not ad hoc by introducing and solving for two Lagrange multipliers $\alpha = \prod N_k$ and $\beta = 1/k_B T$ as usual. Lagrange's method has previously been used to determine the optimal rate of transmission relative to the fidelity evaluation, although it was understood that the matching of information flow is a non-deterministic process similar to tuning several coupled electric circuits for optimal total transmission (Shannon 1948; Shannon & Weaver 1962). Since the usual objective of statistical mechanics is to determine only the stationary partition, it is easy to miss that information processing is driven by the free energy because it has vanished at the steady state. By contrast, when the probabilities of messages are associated with matter and energy, the characteristics of information (equation (2.1)) are obtained without imposing further constraints as natural consequences of thermodynamics governing their representations.

The maximum of equation (2.1) (i.e. natural data structure of equation (4.1)) is not the maximum of equation (1.1) when all p_j are equal. Such an equipartition does not correspond to reality, which displays skewed distributions. Nevertheless, if one so wishes, the mathematical and physical information measures can be related to each other by an exponential transformation at the stationary state where neither H nor S contain free energy terms. It removes from S all medium dependence (i.e. matter and energy constraints) by equalizing energy densities for diverse representations of messages. However, the resulting fallacy of abstract and medium-independent information makes it difficult to understand properties of information, imperatives in processing and communication, as well as emergence of hierarchically organized information networks. Alternatively, one may attempt to tailor boundary conditions in a deterministic way to reproduce skewed distributions from equation (1.1) hoping to match natural data structures. However, such a deterministic approach does not yield understanding of non-deterministic evolution of data structures, archives and networks.

5. Information in transmission

The transmission of a physical representation of message j at the rate dN_j/dt via a channel with a specific capacity σ_{jk} for the particular message j devours free energy. Although contemporary biotic and electronic information processing systems consume only little energy per bit in encoding, transmitting and decoding, even these minute thermodynamic costs (Zurek 1989) in processing have non-negligible consequences. The rate of entropy increase is the criterion of natural selection and evolution naturally selects for the maximal energy dispersal by the most voluminous flows that channel along the steepest gradients in the free energy landscape (Kaila & Annila 2008). This is also pointed out by the constructal law (Bejan 1997). We emphasize that the notion of evolution by natural selection is not limited to biotic systems. For example, increasingly more effective world-wide communication systems are currently emerging. In this context, natural selection is often communicated as proportional growth or preferential attachment (Adamic *et al.* 2001; Albert & Barabasi 2002), but not provided directly by the second law of thermodynamics in the scale-independent formulation (equation (2.2)) to describe evolution of complex networks (figure 1).
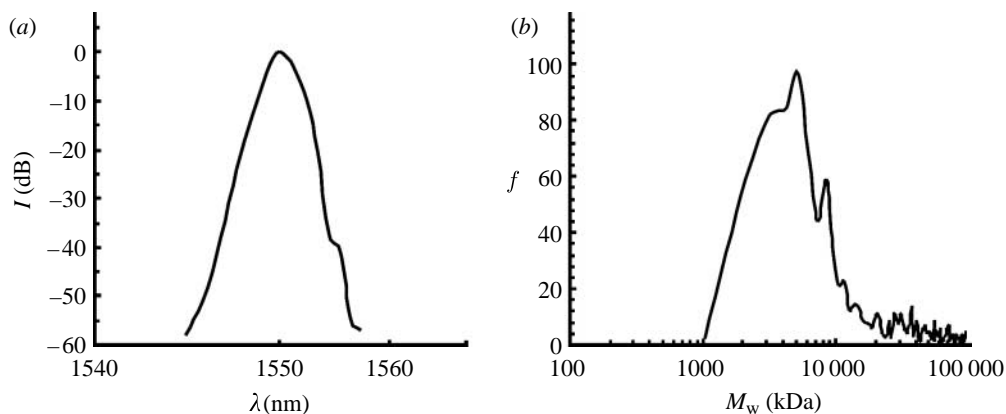
Figure 5. Transmissions of (*a*) electromagnetic signals via optical fibre (optical fibre data, see http://www.fiber-optics.info/glossary-s.htm#SM_Fiber) and (*b*) hereditary messages in plasmids (plasmid data, see http://www.ncbi.nlm.nih.gov/genomes/static/o.html) of molecular weight M_W (in unified atomic mass units) display approximately lognormal dispersion relationships that stem from the universal imperatives to represent information physically. In both cases, the central region of the band is most amenable for communication.

The idea of encoding is to transform the original energy-density representation of a message to a transmissible energy density. The encoding aims to limit the flow of physical representations to the maximum transmission rate provided by the dS/dt-limited channel capacity. To exploit limited capacity more effectively, data are compressed by removing redundancy. However, the compression is not without drawbacks because the non-redundant data do not tolerate errors that may filter in during noisy transmission (Shannon 1948; Shannon & Weaver 1962). Achieving error-free communication is further compromised when a receiver perceives the degree of redundancy differently from the transmitter. According to the noisy-channel coding theorem, transmission of a message as a physical representation over a noisy channel is analogous to the aforementioned noiseless transmission, as noise is represented in the same way as information (Shannon 1948; Shannon & Weaver 1962). In the physical picture, the noise is also an energy density, with typically random fluctuations about the average energy. Noise infiltration is a consequence of the second law. Communication is deteriorating when the density-in-energy, representing the encoded message, leaks from the channel or medium to the surroundings, which is typically lower in density. By the same token, when the representation of a message as a deviation from $k_B T$ is almost negligible, i.e. close to the noise level, the transmission is unreliable. For example, high-resolution devices are cooled to reduce the thermal contribution in a faint signal. Owing to the channel's physical properties, the typical transmission band is a skewed distribution and the dispersion relationship of transmission is sigmoid (figure 5). Its power-law region (linear on a log–log plot) is most amenable for transmission. Protected channels, such as bacterial pili, myelinated axons and shielded cables, enhance isolation from the surrounding densities-in-energy to ensure transmission by limiting noise infiltration. Likewise, speech is encoded for a long-distance transmission on an electromagnetic carrier wave, which does not,

unlike sound waves, couple markedly with surrounding densities-in-energy, most notably the atmospheric medium, but couples first to electrons of the receiver antenna.

Encoding for redundancy is also a way to ascertain communication (Shannon 1948; Shannon & Weaver 1962). Since all natural data structures will eventually, without dissipative maintenance, corrupt towards k_BT, the message, or parts of it, are often encoded redundantly, i.e. in copies of N_j to ascertain the transmission. For example, bacteria that are exposed to radioactivity maintain redundant copies of their entire genomes (Hansen 1978; Samoĭlenko 1983; Cox & Battista 2005). Apparently, the genetic code is unambiguous and redundant to secure high fidelity over long periods of deposition, as well as over translation and transcription (Sonneborn 1965; Woese 1967; Copley *et al.* 2005). Likewise, words of natural languages are redundant in letters (Shannon 1948; Shannon & Weaver 1962) to secure communication despite impaired hearing or reading, and data files are backed up to avoid losses. The problem of optimal coding is a fundamental one because the second law of thermodynamics, given as the equation of motion (equation (2.2)), has, in general, no analytical solution. The fact that there are no invariants of motion (Sharma & Annila 2007) in communication was already recognized by Shannon in the problem of matching (Shannon 1948; Shannon & Weaver 1962). Probabilities keep changing during processing. For example, a message in transmission takes capacity from other messages. How much is in use depends on the particular message. In terms of physics, trajectories are non-integrable and the irreversible process associated with degrees of freedom is non-deterministic. Hence, when the number of messages and their order in transmission are not known *a priori*, it is, in general, impossible to predict the optimal usage of capacity, but algorithms that are able to adapt their encoding on the basis of statistical analysis of input data streams are expected to provide the highest transmission rates (MacKay 2003). In practice, the expected data stream can often be compressed highly efficiently using a frequency-sorted binary tree (Huffman 1952). Finally, it is of interest to note that the best match of S at the source to the channel capacity S_c gives transmission efficiency in analogy to thermodynamic efficiency $\eta = Q_c/Q = T_c S_c / TS \leq 1$. The equality holds only for reversible transmission, since all other processes are dissipative. However, reversible transmission is not communication because both the transmitter and the receiver are sending and acquiring the very same messages that they already possess. The entropy status of the receiver does not change.

6. The meaning of message

Recently, the difficulty of quantifying meaning and context, two critical aspects of biological information, has been seen as a major obstacle for developing a mathematical formalism to describe information in living systems (Corning & Kline 1998; Gatenby & Frieden 2007). Customarily, the mathematical theory of communication is considered to be limited to syntax in information. However, as was pointed out by Weaver, the mathematical theory bears potential for also addressing semantic questions (Shannon 1948; Shannon & Weaver 1962). Using the thermodynamic definition for information, the significance of a message is attributed to the ensuing increase in entropy in the receiver system. The idea to

quantify meaning by the receiver's entropy status is, of course, not new as such (Bennett 1988; Elitzur 1994), but formulated here by the second law as an equation of motion.

The significance of a message for a receiver system is exemplified by considering the transmission of an antibiotic resistance gene from one bacterium to another. Owing to the common genetic code, the standard in communication using chemically represented information, the received message can be decoded by transcription and executed by translation to yield functional agents. The thermodynamic significance of the message is that the newly acquired resistance allows the bacterium to produce entropy by rapid reduction of energy gradients. However, if the bacterium happened to have the resistance gene already before the transmission, the acquired additional copy has less value in retaining the metabolic state that is quantified by the thermodynamic entropy. Further copies would contribute less and less to the free energy reduction. Finally, if the bacterium is never exposed to the particular antibiotic, the corresponding resistance gene would have no apparent thermodynamic value. In this case, the piece of information required to raise the particular resistance will eventually be lost in order to free the associated maintenance to power some other functions that are potentially more rewarding in terms of dS/dt.

In general, the more a message will increase entropy, the more meaningful it is for the receiver system. The acquired information will allow the receiver to access and consume more free energy. A message that does not lead to a marked change in S is considered indifferent. Obviously a message, such as a viral genome, can also be harmful for the receiving host system, i.e. lead to decreased dissipation by curtailing access to the free energy. For the virus system, on the other hand, this corresponds to increased dissipation and access to free energy. Therefore, the message, consisting of the host system, is thermodynamically meaningful for the virus system, which might equally well be considered a receiver. An overall local ecosystem might also be the receiver. We conclude that the transmitted syntax obtains its significance in the dissipative process executed by the particular receiver system. The viewpoint of subjective information interpretation for increased entropy is inherent in the physics of open evolving systems (Tuisku *et al.* 2009).

Owing to differences in mechanisms and driving forces between transmitter and receiver systems, the processing at the two sites differs accordingly. The discrepancy between the evolutionary course that is anticipated by the transmitter and that actualized by the receiver causes misunderstandings. When the transmission capacity and free energy are limited, it may not be possible to communicate effectively enough to remove all doubts. Therefore, the receiver system will have no alternatives but to decode transmitted information and generate understanding by *its* own mechanisms and driving forces. Generated data structures by the receiver may not map one-to-one to those instructed by the source (figure 6). This concept of receiver-biased interpretation is well understood in communication theory (Dobkin & Pace 2006); here, it is found as a thermodynamic consequence.

As the thermodynamic theory of communication requires any message to have a physical representation, the thermodynamic cost of information processing may be quantified. Channel capacities are often insufficient to maintain incessant communication to level all density differences. The desire to keep all informed is,
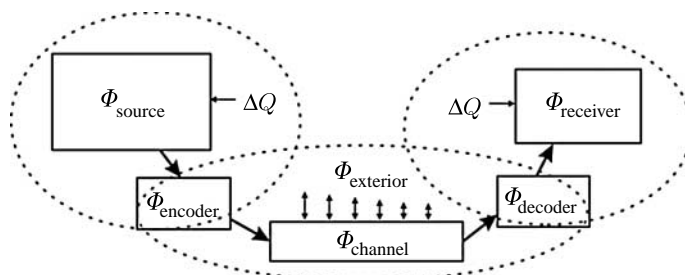
M. Karnani et al.

Figure 6. Diagram of a communication system as an energy transduction system. The physical presentation Φ_{source} of a message at the source is typically encoded for transmission to a low-energy-density representation Φ_{encoder} to minimize losses during the transmission. The representation Φ_{channel} carried by a channel may corrupt during transmission due to the second law of thermodynamics, which tends to equalize differences with respect to external energy density Φ_{external}. The received signal is decoded to the representation Φ_{decoder} and interpreted at the receiver Φ_{receiver}. All components of the communication system bathe in various external energy densities. Typically, the source and the receiver are coupled to an external density of energy ΔQ that is used to generate, maintain and interpret information, whereas the representations of information typically tend to leak to the surroundings during the transmission.

nevertheless, motivated as it articulates the imperative to diminish any emerging potential difference as soon as possible by processes that are statistically independent of each other. Communication is a means of interacting, and interactions define systems. Frequent interactions diminish density differences rapidly among the system's constituents, whereas differences tend to build up when communication is sporadic among weakly coupled components. Perfect understanding would mean a state where all potential differences are abolished by frequent communication. Because the free energy resources are finite, the state of perfect understanding is difficult to attain. On the other hand, when all energy is dispersed evenly, there is no free energy and hence no information to be conveyed among the transmitter and receiver systems that have evolved to become indistinguishable.

7. Discussion

According to the common contemporary consent, the second law of thermo-dynamics is perceived to drive disorder. Therefore, it may appear, at first sight, inconceivable that this universal law could possibly account for the existence and orderly characteristics of information, as well as for its meaningful content. However, the second law, or equivalently the principle of increasing entropy, merely states that differences among energy densities tend to vanish. When the surrounding energy density is high, the system will evolve towards a stationary state by increasing its energy content, e.g. by devising orderly machinery for energy transduction to acquire energy. Accordingly, communication is regarded in this study, among many other processes, as a way to disperse energy, and communication systems are viewed, among many other mechanisms, as machinery for energy transduction. The view of evolving nature as a flattening energy landscape forms the core of the presented argumentation that

communication systems are machineries that facilitate rapid dispersal of energy. Means and methods to communicate are perceived to rise naturally when surrounding densities-in-energy are high. These conclusions are supported by the fact that the second law is found to account for the physical representation of information, characteristics of data structures, imperatives in transmission and for significance in interpretation.

The strictly material and operational tenet perceives information exclusively via its physical presentations. The adopted standpoint strikes a contrast with the mathematical theory of communication, which considers probabilities as mere numbers that denote relative frequencies of messages and disregards the physical quantities, referred to as densities-in-energy, that represent information. Thermodynamic entropy distinguishes states, i.e. distinct representations, from each other in energetic terms. This is in contrast to the informational entropy that distinguishes energetically identical configurations on the basis of relative phases. The configurations may be in the same phase to exhibit orderly coherence or in different phases to display disordered incoherence. Either way, the probability is the same (Griffiths 1995; Sharma & Annila 2007; Tuisku *et al.* 2009); hence, thermodynamic entropy is also the same. The definition of information via thermodynamic entropy does not anticipate gains in information processing from quantum computing because changes in computational states are irrevocably coupled with dissipation. We do recognize that a particular information processing system may also execute reversible computation (Bennett 1982). However, such a system, corresponding to $dS/dt = 0$, is closed, i.e. there are no net fluxes to or from it. Such a computational circuit cannot communicate, e.g. to deliver answers or take up questions from its surroundings, because all observations are dissipative processes (Fisher 1935; Brillouin 1960; Tuisku *et al.* 2009).

Syntax of information, when described by thermodynamics, is associated with the entropy of the physical representation, and significance of information is associated with the entropy increase in the receiver system when it executes the encoded instruction. For example, a microbe may benefit from acquired genetic information simply by digesting the physical representation or richly by executing the encoded instruction. A nutrient as such is a high-energy-density deviation with respect to the surroundings, but it is also a chemotactic signal of additional free energy for the receiver to consume. Messages are valued by the entropy increase that is realized by consuming the free energy. The acquired densities-in-energy that subsequently enable enhanced dissipation beyond mere combustion of their physical representations are regarded as highly informative. The thermodynamic analysis thus sheds light on the origin of information and information-guided processes. Apparently, the immediate value in the free energy contained in the physical representation transformed during evolution to instruct an access to even more free energy (Gibson 1966; Swenson & Turvey 1991; Bejan 1997). In addition, the idea that meals are messages has been an inspiration to many anthropologists (Harris 1979).

The thermodynamic theory of communication is self-consistent by relating all information generation, transmission and interpretation to the principle of increasing entropy. Nevertheless, it may be difficult to see that ultimately all activities do disperse energy. However, without detailed knowledge of all processes, it follows, from the second law of thermodynamics, that no process can be driven without free energy and without transduction mechanisms. Energy

transduction networks organize themselves at all scales in nested hierarchies to increase the rate of energy dispersal (Annila & Kuismanen 2008). According to the thermodynamic theory of communication, information networks that share the same self-similar architecture are no different from other energy transduction networks. Therefore, skewed distributions and their sigmoid cumulative curves that are dominated by power-law regimes are also characteristics of data structures. Moreover, communication standards, such as alphabet and grammar, became established as means to disperse energy effectively, just as amino acid chirality consensus emerged as a means to attain states higher in entropy (Jaakkola *et al.* 2008*a*). The definition of information by thermodynamic entropy is transparent in the description of information-guided processes. For example, information in an amino acid sequence is, by the thermodynamic definition, the free energy with respect to the surrounding energy density. Thus, the quest to diminish free energy directs protein folding towards a stationary-state partition that depends on the surrounding energy density (Sharma *et al.* 2009). Likewise, genetic information is associated with free energy that directs the development of an organism towards a stationary state in its specific surroundings. Although these evolutionary trajectories are non-deterministic, hence unpredictable in detail, the overall course takes towards the free energy minimum where the driving forces expire.

Generation of information and its subsequent transmission are motivated by an anticipated increase in energy dispersal that the sender system aims to accomplish together with the informed receiver systems. Alternatively, the source may signal to decoy receivers in its surroundings, hoping to continue its own energy transduction by exploiting the surrounding systems. Importantly, the surroundings must be wealthier in total than the system to motivate such behaviour. Consequently, we reason, on the basis of the theory of evolution by natural selection founded on thermodynamics, that ultimately all communication aims to enhance energy transduction to level differences among energy densities. During continual communication, common coding protocols and other standards will be established, just as they were and are being established to facilitate energy transduction by contemporary economic systems. For example, the primordial world settled for the common genetic code and today the global communication systems are settling for international Internet protocols. We expect that the overall course towards the increasingly more integrated global communication system is also sigmoid, the ubiquitous characteristic of natural processes. The initial phase was slow, when communication systems struggled to establish common protocols, basic concepts and language. Now, later, the natural process has gained speed and is levelling to the state of maturity where all that is new is rapidly distributed.

Undoubtedly, there are many more aspects in the theory of information and communication than we have been able to address. For example, it seems that evolution of memes (Dawkins 1976) and their propagation could be rationalized using the physical representation requisite. Nonetheless, we expect that the study conveys an obvious but clarifying thought, namely that information is nothing but physical, to imply that the thermodynamic, rather than the mathematical, theory accounts for communication.

References

Adamic, L. A., Lukose, R. M., Puniyani, A. R. & Huberman, B. A. 2001 Search in power-law networks. *Phys. Rev. E* **64**, 046 135. (doi:10.1103/PhysRevE.64.046135)

Albert, R. & Barabasi, A.-L. 2002 Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–94. (doi:10.1103/RevModPhys.74.47)

Almlöf, M., Andér, M. & Åqvist, J. 2007 Energetics of codon–anticodon recognition on the small ribosomal subunit. *Biochemistry* **46**, 200–209. (doi:10.1021/bi061713i)

Altmann, G. 2002 Zipfian linguistics. *Glottometrics* **3**, 19–26.

Annila, A. & Annila, E. 2008 Why did life emerge? *Int. J. Astrobiol.* **7**, 293–300. (doi:10.1017/S1473550408004308)

Annila, A. & Kuismanen, E. 2008 Natural hierarchy emerges from energy dispersal. *Biosystems* **95**, 227–233. (doi:10.1016/j.biosystems.2008.10.008)

Balasubrahmanyan, V. K. & Naranan, S. 2000 Information theory and algorithmic complexity: applications to linguistic discourses and DNA sequences as complex systems. Part II. Complexity of DNA sequences, analogy with linguistic discourses. *J. Quant. Linguist.* **7**, 153–183. (doi:10.1076/0929-6174(200008)07:02;1-Z;FT153)

Bejan, A. 1997 *Advanced engineering thermodynamics.* New York, NY: Wiley.

Bennett, C. H. 1982 The thermodynamics of computation. *Int. J. Theor. Phys.* **21**, 905–940. (doi:10.1007/BF02084158)

Bennett, C. H. 1988 Logical depth and physical complexity. In *The universal Turing machine: a half century survey* (ed. R. Herken), pp. 227–258. Oxford, UK: Oxford University Press.

Bertalanffy, L. 1953 *Physik des Fliessgleigewichts.* Braunschweig, Germany: Vieweg.

Boltzmann, L. 1905 Populäre Schriften (partially transl. Theoretical physics and philosophical problems by B. McGuinness, Reidel, Dordrecht, 1974). Leipzig, Germany: Barth.

Brillouin, L. 1960 *Science and information theory.* New York, NY: Academic Press.

Brooks, D. R. 2000 The nature of the organism: life takes on a life of its own. *Proc. NY Acad. Sci.* **901**, 257–265.

Brooks, D. R. & Wiley, E. O. 1988 *Evolution as entropy: toward a unified theory of biology.* Chicago, IL: The University of Chicago Press.

Carnot, S. 1824 Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance. Paris, France: Bachelier.

Copley, S. D., Smith, E. & Morowitz, H. J. 2005 A mechanism for the association of amino acids with their codons and the origin of the genetic code. *Proc. Natl Acad. Sci. USA* **102**, 4442–4447. (doi:10.1073/pnas.0501049102)

Corning, P. A. & Kline, S. J. 1998 Thermodynamics, information and life revisited, part II: thermoeconomics and control information. *Syst. Res.* **15**, 453–482. (doi:10.1002/(SICI)1099-1743(199811/12)15:6<453::AID-SRES201>3.0.CO;2-U)

Cox, M. M. & Battista, J. R. 2005 Deinococcus radiodurans—the consummate survivor. *Nat. Rev. Microbiol.* **3**, 882–892. (doi:10.1038/nrmicro1264)

Darwin, C. 1859 *On the origin of species.* London, UK: John Murray.

Dawkins, R. 1976 *The selfish gene.* London, UK: Oxford University Press.

Dewar, R. 2003 Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states. *J. Phys. A: Math. Gen.* **36**, 631–641. (doi:10.1088/0305-4470/36/3/303)

Dobkin, B. A. & Pace, R. C. 2006 *Communication in a changing world.* New York, NY: McGraw-Hill.

Elitzur, A. C. 1994 Let there be life: thermodynamic reflections of biogenesis and evolution. *J. Theor. Biol.* **168**, 429–459. (doi:10.1006/jtbi.1994.1123)

Fisher, R. A. 1935 The logic of inductive inference. *J. R. Stat. Soc.* **98**, 39–82. (doi:10.2307/2342435)

Gatenby, R. A. & Frieden, B. R. 2007 Information theory in living systems, methods, applications, and challenges. *Bull. Math. Biol.* **69**, 635–657. (doi:10.1007/s11538-006-9141-5)

Gibbs, J. W. 1993–1994 *The scientific papers of J. Willard Gibbs.* Woodbridge, CT: Ox Bow Press.

Gibson, J. 1. 1966 *The senses considered as perceptual system.* Boston, MA: Houghton Mifflin.

Gregory, T. R. (ed.) 2005. *The evolution of the genome.* San Diego, CA: Elsevier.

Griffiths, D. 1995 *Introduction to quantum mechanics.* Englewood Cliffs, NJ: Prentice Hall.

Grönholm, T. & Annila, A. 2007 Natural distribution. *Math. Biosci.* **210**, 659–667. (doi:10.1016/j.mbs.2007.07.004)

Hansen, M. T. 1978 Multiplicity of genome equivalents in the radiation-resistant bacterium *Micrococcus radiodurans. J. Bacteriol.* **134**, 71–75.

Harris, M. 1979 *Cultural materialism—the struggle for a science of culture.* New York, NY: Random House.

Huffman, D. A. 1952 A method for the construction of minimum-redundancy codes. *Proc. Inst. Radio Eng.* **40**, 1098–1101.

Jaakkola, S., Sharma, V. & Annila, A. 2008a Cause of chirality consensus. *Curr. Chem. Biol.* **2**, 53–58. (doi:10.2174/187231308784220536)

Jaakkola, S., El-Showk, S. & Annila, A. 2008b The driving force behind genomic diversity. (http://arxiv.org/abs/0807.0892)

Jaynes, E. T. 2003 *Probability theory. The logic of science.* Cambridge, UK: Cambridge University Press.

Kaila, V. R. I. & Annila, A. 2008 Natural selection for least action. *Proc. R. Soc. A* **464**, 3055–3070. (doi:10.1098/rspa.2008.0178)

Kapteyn, J. C. 1903 *Skew frequency curves in biology and statistics.* Groningen, The Netherlands: Astronomical Laboratory, Noordhoff.

Karnani, M. & Annila, A. 2009 Gaia again. *Biosystems* **95**, 82–87. (doi:10.1016/j.biosystems.2008.07.003)

Kondepudi, D. & Prigogine, I. 1998 *Modern thermodynamics.* New York, NY: Wiley.

Kullback, S. 1959 *Information theory and statistics.* New York, NY: Wiley.

Landauer, R. 1961 Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **5**, 183–191.

Landauer, R. 1991 Information is physical. *Physics Today*, May.

Landauer, R. 1996a Minimal energy requirements in communication. *Science* **272**, 1914–1918. (doi:10.1126/science.272.5270.1914)

Landauer, R. 1996b The physical nature of information. *Phys. Lett. A* **217**, 188–193. (doi:10.1016/0375-9601(96)00453-7)

Limpert, E., Stahel, W. A. & Abbt, M. 2001 Log-normal distributions across the sciences: keys and clues. *Bioscience* **51**, 341–352. (doi:10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2)

Lloyd, S. 2000 Ultimate physical limits to computation. *Nature* **406**, 1047–1054. (doi:10.1038/35023282)

Lorenz, R. D. 2002 Planets, life and the production of entropy. *Int. J. Astrobiol.* **1**, 3–13. (doi:10.1017/S1473550402001027)

Lotka, A. J. 1926 The frequency distribution of scientific productivity. *J. Wash. Acad. Sci.* **16**, 317–323.

MacKay, D. J. C. 2003 *Information theory, inference, and learning algorithms.* New York, NY: Cambridge University Press.

Mahulikar, S. P. & Harwig, H. 2004 Conceptual investigation of the entropy principle for identification of directives for creation, existence and total destruction of order. *Phys. Scr.* **70**, 212–221. (doi:10.1238/Physica.Regular.070a00212)

Majerník, V. 2002 A conceptualization of the configurational and functional organization. *Glottometrics* **3**, 110–135.

Martyushev, L. M. & Seleznev, V. D. 2006 Maximum entropy production principle in physics, chemistry and biology. *Phys. Rep.* **426**, 1–45. (doi:10.1016/j.physrep.2005.12.001)

Mathai, A. M. & Rathie, P. N. 1975 *Basic concepts in information theory and statistics.* New Delhi, India: Wiley Eastern.

Matsuno, K. & Swenson, R. 1999 Thermodynamics in the present progressive mode and its role in the context of the origin of life. *Biosystems* **51**, 53–61. (doi:10.1016/S0303-2647(99)00013-1)

Naranan, S. & Balasubrahmanyan, V. K. 1992 Information theoretic models in statistical linguistics. 2. Word frequencies and hierarchical structure in language statistical tests. *Curr. Sci.* **63**, 297–306.

Naranan, S. & Balasubrahmanyan, V. K. 2000 Information theory and algorithmic complexity: applications to linguistic discourses and DNA sequences as complex systems. Part I. Efficiency of the genetic code of DNA. *J. Quant. Linguist.* **7**, 129–151. (doi:10.1076/0929-6174(200008)07:02;1-Z;FT129)

Nicolis, G. & Prigogine, I. 1976 *Self-organization in non-equilibrium systems*. New York, NY: Wiley.

Salthe, S. N. 1985 *Evolving hierarchical systems: their structure and representation*. New York, NY: Columbia University Press.

Salthe, S. N. 1993 *Development and evolution: complexity and change in biology*. Cambridge, MA: MIT Press.

Salthe, S. N. 2002 Summary of the principles of hierarchy theory. *Gen. Syst. Bull.* **31**, 13–17.

Samoĭlenko, I. I. 1983 Radioresistance mechanisms of *Micrococcus radiodurans. Nauchnye Doki Vyss. Shkoly Biol. Nauki.* **10**, 4–17.

Schneider, E. D. & Kay, J. J. 1994 Life as a manifestation of the second law of thermodynamics. *Math. Comp. Model.* **19**, 25–48. (doi:10.1016/0895-7177(94)90188-0)

Shannon, C. E. 1948 The mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423. See also pp. 623–656.

Shannon, C. E. & Weaver, W. 1962 *The mathematical theory of communication*. Urbana, IL: The University of Illinois Press.

Sharma, V. & Annila, A. 2007 Natural process—natural selection. *Biophys. Chem.* **127**, 123–128. (doi:10.1016/j.bpc.2007.01.005)

Sharma, V., Kaila, V. R. I. & Annila, A. 2009 Protein folding as an evolutionary process. *Physica A* **388**, 851–862. (doi:10.1016/j.physa.2008.12.004)

Sigurd, B., Eeg-Olofsson, M. & van Weijer, J. 2004 Word length, sentence length and frequency—Zipf revisited. *Stud. Linguist.* **58**, 37–52. (doi:10.1111/j.0039-3193.2004.00109.x)

Sonneborn, T. M. 1965 Degeneracy of the genetic code: extent, nature, and genetic implications. In *Evolving genes and proteins* (eds V. Bryson & H. J. Vogel), pp. 377–397. New York, NY: Academic Press.

Swenson, R. 1989 Emergent attractors and the law of maximum entropy production: foundations to a theory of general evolution. *Syst. Res.* **6**, 187–198.

Swenson, R. 2000 Spontaneous order, autocatakinetic closure, and the development of space–time. *Ann. NY Acad. Sci.* **901**, 311–319.

Swenson, R. & Turvey, M. T. 1991 Thermodynamic reasons for perception-action cycles. *Ecol. Psychol.* **3**, 317–348. (doi:10.1207/s15326969eco0304_2)

Szilard, L. 1929 Über die Entropieverminderung in einem thermodynamischen System bei Eingrien intelligenter Wesen. *Z. Phys.* **53**, 840–856. (doi:10.1007/BF01341281)

Tuisku, P., Pernu, T. K. & Annila, A. 2009 In the light of time. *Proc. R. Soc. A* **465**, 1173–1198. (doi:10.1098/rspa.2008.0494)

Ulanowicz, R. E. & Hannon, B. M. 1987 Life and the production of entropy. *Proc. R. Soc. Lond. B* **232**, 181–192. (doi:10.1098/rspb.1987.0067)

Woese, C. R. 1967 *The genetic code: the molecular basis for genetic expression*. San Francisco, CA: Harper and Row.

Würtz, P. & Annila, A. 2008 Roots of diversity relations. *J. Biophys.* (doi:10.1155/2008/654672)

Zhang, J. 2000 Protein-length distributions for the three domains of life. *Trends Genet.* **16**, 107–109. (doi:10.1016/S0168-9525(99)01922-8)

Ziegler, H. 1983 *An introduction to thermomechanics*. Amsterdam, The Netherlands: North-Holland.

Zipf, G. K. 1935 *The psychobiology of language*. Boston, MA: Houghton-Mifflin.

Zipf, G. K. 1949 *Human behaviour and the principle of least-effort*. Cambridge, MA: Addison–Wesley Press.

Zurek, W. H. 1989 Thermodynamic cost of computation, algorithmic complexity and the information metric. *Nature* **341**, 119–124. (doi:10.1038/341119a0)