Key words – Science of information, spatio-temporal, semantic and structural information, Darwin channel, noisy constrained capacity.

Wojciech SZPANKOWSKI
Department of Computer Science
Purdue University
W. Lafayette, IN 47907, U.S.A.
spa@cs.purdue.edu

FACETS OF INFORMATION IN COMMUNICATIONS*

Abstract

Information permeates every corner of our lives and shapes our universe. Understanding and harnessing information holds the potential for significant advances. Information is communicated in various forms: from business information measured in dollars, to chemical information contained in shapes of molecules, and to biological information stored and processed in living cells. Our current understanding of the formal underpinnings of information date back to Claude Shannon's seminal work in 1948 resulting in a general mathematical theory for reliable communication in the presence of noise. This theory enabled much of the current day storage and communication infrastructure. However, a wide spread application of information theory to economics, biology, life science and complex networks seems to be still awaiting us. We shall argue that a new *science of information* is to rekindle for extraction, comprehension, and manipulation of *structural, spatio-temporal* and *semantic* information. We conclude this essay with a list of challenges for future research.

1 Introduction

Information is still the distinctive mark and arguably the basic commodity of our era, so that the need for deeper reflection and study is intensifying. The notion and theory of *information* introduced by Claude Shannon in 1948 have served as the backbone to a now classical paradigm of digital communication. Television, internet, voyage to the Moon, unbreakable ciphers, computers, NASA's planetary probes, home compact disk audio systems, they all benefited from Shannon theory. Shannon's notion of information quantifies the extent to which a recipient of data can reduce its statistical uncertainty when "semantic aspects of communication are irrelevant" [23]. Unfortunately, that formalization of information hardly captures all of the needed nuances, and the accompanying theory has not lent itself to non-trivial applications outside the native context. We yet have to develop theory that provides a satisfactory formalism and overarching answers for extraction, comprehension, and manipulation of *structural, spatio-temporal* and *semantic* information in scientific and social domains. Consequently, one can argue the necessity for the next information revolution and perhaps launching a new *Science of Information* that integrates research and teaching activities from all angles: from the

*This work was supported in part by the NSF Grants CCF-0513636, DMS-0503742, DMS-0800568, and CCF -0830140, NIH Grant R01 GM068959-01, NSA Grant H98230-08-1-0092, EU Project No. 224218 through Poznan University of Technology, and the AFOSR Grant FA8655-08-1-3018.

fundamental theoretical underpinnings of information to the science and engineering of novel substrates, biological networks, chemistry, communication networks, economics, physics, and complex social systems.

Surprisingly enough, advances in information technology and widespread availability of information systems and services have largely obscured the fact that *information* remains undefined in its generality, though considerable collective effort has been invested into its understanding [2, 3, 18, 22, 24, 27]. Shannon wrote in his 1953 paper [24]: "The word "information" has been given many different meanings ... it is likely that at least a number of these will prove sufficiently useful in certain applications and deserve further study and permanent recognition." C. F. von Weiszsäcker argued [27] against an absolute definition of information, claiming that: "Information is only that which produces information" (relativity) and "Information is only that which is understood" (rationality). One also observes that information extraction depends on available resources (e.g., think of guessing an integer with one bit storage available [21, 26]). It follows therefore, that in its generality, *information is that which can impact a recipient's ability to achieve the objective of some activity in a given context within limited available resources* [17]. We shall adopt this definition for the purpose of this essay.

In passing we should add that we view here "communication" very broadly. Living cells do communicate and process information as well as users in communication/wireless networks, information is communicated in arbitrage of financial markets, spike trains between neurons carry information to the brain, and so forth. We are aiming at understanding "communicated information" in its generality.

2 Science of Information

Information, understood broadly, is capable of unifying seemingly unrelated areas such as information theory, physics of information, value of information, Kolmogorov complexity and information flow in life sciences [19]. This is illustrated in Figure 1. One can argue, however, as the participants of the first workshop of *Information Beyond Shannon*, Orlando, FL, 2005 did, that the following aspects of information were never adequately addressed in the past and therefore threaten to raise severe impediments to diverse applications of science of information:

- **Delay**: In many communication problems, especially when information is transmitted over a network, the amount of delay incurred is an important and nontrivial factor [10]. For example, complete information arriving late maybe useless to the receiver, whereas incomplete information arriving early may be valuable (e.g., in a signaling cascade associated with a specific cell function, delay or loss of signals can be lethal, timeliness is the key for arbitrage of financial markets, timely intelligence is essential to virtually all security and defense related applications). This is not simply a question of understanding the classical delay-rate trade-off (via the reliability function [3]), but a complex issue involving our choice of how and what to transmit, as well as the actually utility of the information being transmitted.
- **Space**: In interacting systems, spatial localization often limits information exchange, with obvious disadvantages as well as benefits. These benefits typically result from reduction in interference as well as ability of system to modulate and react to stimulus (common examples range from wireless systems to immune response).

Figure 1: Information Synergy.

- **Information and Control**: In the above delay-bandwidth example we have a conflict between two objectives: Attaining a high transmission rate, and sending the information with small delay. But information is exchanged in space and time for decision making, thus timeliness of information delivery along with reliability and complexity constitute basic objectives. More generally, there are cases where we have some control not only over the coding part, but also about other design aspects of the communication setting (e.g., the network topology, power distribution, routing, etc). How can the two tasks be optimally combined?
- **Utility**: We often find that the utility of what is transmitted depends on different factors, such as the time at which it arrives, and perhaps even the actual contents of the message. How can such utility considerations be incorporated into the classical coding problem?
- **Semantics**: In many scientific contexts experimenters are interested in signals, without knowing precisely what these signals represent. Examples of this situation are very common in biology: DNA sequences and spike trains between neurons are certainly used to convey information, but little more than that can be assumed a priori. Often, one of the first steps in the analysis is to try and estimate the amount of information contained in these signals - how should that be done? Estimating the entropy is typically not appropriate: It offers a measure of the structural complexity of the signal, but it does not measure its actual information content. For example, it ignores the fact that there may be noise present and it does not take into account that certain parts of the signal may be irrelevant to the receiver. Is there a general way to account for the actual "meaning" of signals in a given context?
- **Optimal versus Real Communication Systems**: In the last example of observing, say, a spike train between two neurons in an animal's brain, the difficulty in extracting information from the signal (or even measuring the amount of information present) in

part arises from the fact that the scientist is not in Shannon's position of designing an optimal system given known communication constraints. Instead, she is analyzing an existing and typically sub-optimal communication system. We can ask whether there is a general methodology for the study of such systems, so that the actual context within which they operate is taken into account

- **Dynamic information**. In a complex network, information is not just communicated but also processed and even generated along the way (e.g., response to stimuli is processed at various stages – with immediate response processed at site of stimulus, higher-level response processed in the brain, response to emergency events is coordinated at various levels, ranging from first responders to command and control centers). How can such considerations of dynamic sources be incorporated into an information-theoretic model?
- **Learnable Information**: One can argue (and some have) that in all scientific endeavours, the only task is to extract information from data. How much information can actually be learnt from a given data set? In Shannon theory, one starts from a (possibly unknown) model for the data-generating mechanism and calculates its entropy, but in practice the starting point is only the data. Is there a general theory that provides natural model classes for the data at hand? What is the cost of learning the model, and how does it compare to the cost of actually describing the data? Risannen's MDL theory offers guidelines in this direction [20].
- **Structure and Organization**: We still lack measures and meters to define and quantify information embodied in structure and organization (e.g., information in nanostructures, biomolecules, gene regulatory and protein interaction networks, social networks, networks of financial transactions, etc.). Typically, these measures must account for associated context, and incorporate diverse (physical, social, economic) dynamic observables and system state).
- **Limited Resources**: In many scenarios, information is limited by available resources (e.g., computing devices, bandwidth of signaling channels). How much information can be extracted and processed with limited resources? This relates to **complexity and information** where different representations of the same distribution may vary dramatically when complexity is taken into account (e.g., computing a number from its prime factors is easy but factoring it is known to be much harder).

It is our belief that further advances in science of information and its diverse applications depend on answering the above challenges.

3 Two Examples

In this section we discuss two examples: one illustrating spatio-temporal aspect of information in wireless massive networks, and the other related to information transfer in biological systems (the so called Darwin channel).

3.1 Space-time Paradox in Stable Wireless Networks

Wireless local area networks (WLANs), multihop mobile ad hoc networks (MANETs), and social networks are a logical next step towards an ubiquitous computing environment. The related technological challenges, e.g., volatile connectivity, power awareness, and increased node autonomy, have also become scientific ones. Recent research [8, 9] in MANETs has led to the discovery of the "space capacity paradox" and "time capacity paradox." The theoretical capacity of a multihop wireless network increases with node density and node mobility in spite of the apparently devastating effect of transmission interference. A deeper information-theoretic understanding of these properties is likely to bring about a breakthrough in MANET technology and deployment, on condition that more realistic network operation models are adopted.

Classical information theory studies capacity of channels connecting two endpoints. This approach is hard to adopt to mobile nodes which relay information in a multihop manner and time-varying topology. Therefore, some authors (cf. [14]) introduce the concept of the *spatio-temporal relaying*. A relay in a space-time situation carries information from a mobile transmitter (space) in its past (time) to a mobile receiver (space) in its future (space). Here, the past and future are defined with respect to the causal physical trajectory of nodes that forms a path in a spatio-temporal space of information transfer. The quality of the transmission depends on the respective spatio-temporal positions of the transmitter and receiver. Thus the concept of a space-time relay transcends classical information theory.

The challenge is to extend the celebrated Shannon capacity formula $\log_2(1 + \frac{S}{N})$ per Hertz (S signal, N noise) to multi-source wireless networks. Recently, Jacquet [13] proved that the maximum information rate I per second per Hertz in a network of dimension d (e.g., $d = 1$ for a line, $d = 2$ for a plane, $d = 3$ for a three-dimensional space) with the Rayleigh factor α is

$$I = \frac{\alpha}{d}(\log 2)^{-1}.$$

This formula is remarkable since it connects three main ingredients: space (d), physics of wave propagation (α), and information theory ($\log 2$).

Furthermore, in [9] it is shown that the theoretical capacity of a multihop wireless network is proportional to the square root of the network size (number of nodes). This remarkable result promises enormous wireless capacity for ultra-dense networks (e.g., one million nodes with available bandwidth of 1 Mb/s can reach a total capacity on order of gigabits per second, unprecedented for mobile networks). However, attempts to verify these predictions in a network with WiFi nodes bring unsatisfactory results: the space capacity has a tendency to decrease with the number of nodes, rather than increase as theoretically predicted. This reflects the well-known fact that the WiFi medium access protocol, primarily designed for wireless LANs, does not scale to multihop networks. Several analytical models using random placement of nodes were subsequently developed. By using wave attenuation and packet capture rules, minimal (slotted Aloha-like) medium access protocols were designed to fit into the theoretically predicted scaling property. Closed formulae for the probability of packet capture versus distance were found that can be used to identify the performance bottleneck for any routing protocol via certain equilibrium equations [12]; this promises further advances in the Science of Information. For example, it turns out that the Optimized Link State Routing (OLSR) protocol in its basic version cannot support more than a few thousand nodes. For networks this large the bottleneck is the *overhead*, since every node must inform all other nodes about its local connectivity, creating control traffic flows that percolate through the whole of the network (even if broadcast optimally, as prescribed by OLSR). Of interest become routing strategies whereby only significant connectivity changes are broadcast.

Recently, a time counterpart of the space paradox was demonstrated. Under the hypothesis of ergodic node mobility it was proved that the capacity of a mobile wireless network can be linear in the number of nodes. Thus, one million nodes can reach a capacity on the order of a thousand Gb/s. However, this comes with a price tag of growing delay. One can therefore ask how much useful information is really passed.

More generally there are fascinating relations between spatio-temporal properties and information propagation (speed) in wireless networks. In [14] an upper bounds is derived for the propagation speed of one bit in a wireless mobile network embedded in a map of dimension d. The question still remains how to estimate the propagation speed as function of the information rate. Clearly, the speed decreases with the requested capacity so that information theoretical upper-bounds need to be extracted and realistic models must be developed.

3.2 Information Transfer in Biological Systems

We now switch to communications in biology and discuss information flow in biological systems by introducing the *mutation channel* and the *Darwin channel*. The mutation channel is a classical insertion/deletion channel [5, 4], while the Darwin channel is a novel information-theoretic channel, described in details below, that models preferential Darwinian selection.

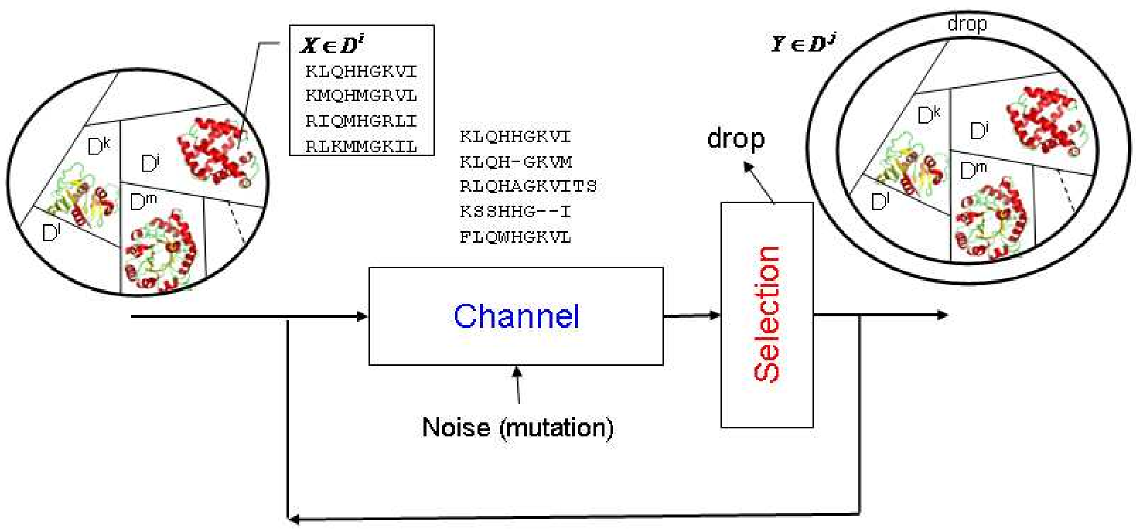

Figure 2: Darwin Channel

Darwin channel described in Figure 2 is designed to model the flow of genetic information through *temporal constrained* channel with feedback (i.e., surviving sequences/genes re-enter the evolutionary process in time and space preserving functionality). More precisely, the original input (biological) sequence $X_1^n = X_1 \dots X_n$ is restricted to a *constrained* (Darwinian preselected) set $\mathcal{D}_n$ that is a proper subset of all possible input sequences. This set is partitioned into subsets $\mathcal{D}^i$ of sequences of the same *functionality* (e.g., by using a scoring function often used in biology). This partition is represented by a function

$$F : \mathcal{D}_n \to \{0, 1, \dots, M-1\}$$

for some M (that may depend on n) such that all $X_1^n \in \mathcal{D}^I$ are assigned to index $I := F(X_1^n)$. In the Darwin channel *mutation* constitutes the *noise*, and it can take the form of insertion, deletion or substitution, where is the latter is the most dominant factor. The output sequence $Z_1^m = Z_1 \dots Z_m$ is in general of random length m. An output sequence Z_1^m is either assigned to one of the subsets $\mathcal{D}^i$ through F or erased (declare "dead"). In such an information transfer scenario an input sequence is declared to be "functionally surviving" *if both X_1^n and its corresponding output sequence Z_1^m belong to the same functional subset $\mathcal{D}^I$*; otherwise an error occurs leading to a non-surviving sequence. Furthermore, to model temporal behavior, we follow Eigen's observations (i.e., "there are correlations between error rate and genome length") and allow error rate to be a function of n. The main open problem is how to determine the amount of "information" being transfer through the channel. As a matter of fact, we first need to address more fundamental questions, namely what is information in such systems and how to measure it?

Let us now consider two special cases to illustrate some difficulties we may encounter analyzing the Darwin channel. Assume for now there is no feedback, noise is just a substitution with probability of error ε being fixed and very small, say $\varepsilon = 10^{-8}$. Furthermore, to focus we postulate that the constrained set $\mathcal{D}_n$ consists of (d, k) $(d < k)$ binary sequences in which any run of zeros must be of length at least d and at most k (e.g., such sequences model spike trains of neurons). We make one more important assumption, namely we measure the information transfer in such a system by the statistical dependency between the output sequence Z^n and the input sequence X^n through the *mutual information* $I(X; Z)$. Then one may ask what values $I(X; Z)$ takes and perhaps what is the maximum value of $I(X; Z)$ over all possible input distribution. The latter question leads to the *noisy constrained capacity* problem that has been unsolved since Shannon [7].

Let us first look at $I(X; Z)$. Observe that

$$I(X; Z) = H(Z) - H(Z|X)$$

where $H(Z)$ and $H(Z|X)$ are entropy and conditional entropy, respectively. But, as easy to see, $H(Z|X)$ is the entropy of the noise, that is, $H(Z|X) = H(\varepsilon) = -\varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon)$. Just we are left with the problem of estimating the entropy of Z.

In our case, X is a (d, k) sequence and Z a noisy version of X. But a (d, k) sequence can be generated as an output of an automaton, thus X is a Markov sequence and then Z is a *hidden Markov process* (HMP). Unfortunately, entropy of a HMP process is not easy to estimate [6]. Fortunately, recently in [11, 15] it was proved that for $\varepsilon \to 0$ the following holds

$$H(Z) = H(P) - f_0(P)\varepsilon \log \varepsilon + f_1(P)\varepsilon + o(\varepsilon)$$

for some explicitly computable coefficients $f_0(P)$ and $f_1(P)$ where P is the distribution of the underlying Markov process of X and $H(P)$ its known Markov entropy [11, 16]. Even more interestingly, if we are interested in the maximum mutual information, that is, the noisy constrained capacity

$$C(\mathcal{D}, \varepsilon) = \sup_{X \in \mathcal{D}} I(X; Z) \lim_{n\to\infty} \frac{1}{n} \sup_{X_1^n \in \mathcal{D}_n} I(X_1^n, Z_1^n)$$

then the situation becomes more complicated. Recently, in [11, 16] it was proved that

$$C(\mathcal{D}, \varepsilon) = C(\mathcal{D}) - (1 - f_0(P^{\max}))\varepsilon \log \varepsilon + (f_1(P^{\max}) - 1)\varepsilon + o(\varepsilon)$$

where $C(\mathcal{D})$ is the noisless capacity, that is, $C(\mathcal{D}) = -\log \rho_0$, where ρ_0 is the smallest real root of (cf. [25])

$$\sum_{\ell=d}^{k} \rho_0^{\ell+1} = 1.$$

Thus, even under these simplifying (and rather biologically naive) assumptions, the Darwin channel is a "hard nut to crack".

Let us finally consider the temporal aspect of the Darwin channel. To illustrate our point, we make the most simplifying assumptions, namely the Darwin channel is plainly a simple binary symmetric channel in which each bit incurs a random delay T before it reaches the receiver [17]. A bit that reaches the destination after a given deadline τ is dropped. Furthermore, let ε be the probability of error. We assume that the longer a bit takes to reach the receiver, the lower the probability of a successful transmission (which is an accurate model in certain biological situations as observed by M. Eigen). For $t \le \tau$ the probability of a successful transmission is $\Phi(\varepsilon, t)$ (e.g., $\Phi(\varepsilon, t) = (1-\varepsilon)^t$), and hence the probability of error is $1 - \Phi(\varepsilon, t)$ for some $\varepsilon > 0$. What is the capacity of such a *temporal channel*? If the delay is exponentially distributed, then one easily finds for $\Phi(\varepsilon, t) = (1-\varepsilon)^t$ that

$$C(\tau) = [(1 - P(T > \tau))][1 - H(\rho)],$$

where $\rho = P(x|x)/(1-e^{-\tau})$ with $P(x|x) = (1-(1-\varepsilon)^\tau e^{-\tau})/(1-\ln(1-\varepsilon))$ being the probability of a successful transmission. Observe that with a stringent delay bound, the capacity of the channel is adversely affected by frequent erasure and the capacity drops due to temporal errors.

4 Concluding Remarks

In the upcoming workshop *Information Beyond Shannon*, Venice, Italy, December 29-30, 2008 (`http://mobilfuture.com/venice/`) the organizers list the following challenges facing us:

- Frederick P. Brooks, Jr., wrote in "The Great Challenges for Half Century Old Computer Science" [1]: "Shannon performed an inestimable service by giving us a definition of Information and a metric for Information as communicated from place to place. We have no theory however that gives us a metric for the Information embodied in structure. ... This is the most fundamental gap in the theoretical underpinning of Information and computer science. A young information theory scholar willing to spend years on a deeply fundamental problem need look no further."

- Information accumulates at a rate faster than it can be sifted through, accessed and digested by humans, so that the bottleneck, traditionally represented by the medium, is drifting towards the receiving end of the channel.

- In a growing number of situations, the overhead in accessing Information prevails over that of fruition, which makes information itself practically unattainable or obsolete.

- Capabilities akin to contents addressing and semantic access and transmission are not even in sight, while computing and communication infrastructures of the new Millennium induce drastic mutations on the conventional notions of Knowing and Learning, Guessing and Discovering.

- Microscopic systems seem not to obey Shannon postulates of information [2]. In the quantum world and on the level of living cells traditional Information often fails to accurately describe reality.

The cross-disciplinary research in science of information advocated here hopefully will lead to the development of an active and thriving community of students and scholars to pursue its goals. As the first step, we have launched recently at Purdue the *Institute for Science of Information* `http://www.isi.purdue.edu` that should serve as home for such activities. We certainly hope that similar centers will soon emerge in Europe and around the world.

5 Acknowledgment

The author thanks Drs. J. Konorski (TUG), C. Jedrzejek (PUT), and participants of the first *Workshop Information Beyond Shannon*, Orlando, 2005 for sharing their views on information.

References

[1] F. Brooks. Three great challenges for half-century-old computer science. *J. the ACM*, 50:25–26, 2003.

[2] C. Brukner, A. Zeilinger, Conceptual Inadequacy of the Shannon Information in Quantum Measurements. *Phys. Rev.* A 63, 2001.

[3] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Second Edition, John Wiley & Sons, New York, 2006.

[4] R. L. Dobrushin. Shannon's theorem for channles with synchronization errors. *Problems of Infformation Transmissions*, 3:18–36, 1967.

[5] E. Drinea and M. Mitzenmacher. On lower bounds for the capacity of deletion channels. *IEEE Transaction of Information Theory*, 52:4648–4657, 2006.

[6] Y. Ephraim and N. Merhav, "Hidden Markov processes," *IEEE Trans. Inform. Theory*, 48, 1518–1569, 2002.

[7] J. Fan, T. L. Poo, and B. Marcus. Constraint gain. *IEEE Transaction of Information Theory*, 50:1989–1999, 2004.

[8] M. Grossglauser and D. Tse, Mobility Increases the Capacity of ad-hoc Wireless Networks, *IEEE/ACM Trans. Networking*, 48, 477-486, 2002.

[9] P. Gupta and P.R. Kumar, Capacity of Wireless Networks, *IEEE Trans. Information Theory*, 46, 388-404, 2000.

[10] B. Hajek and A. Ephremides, Information Theory and Communication Networks: An Unconsummated Union, *IEEE Trans. Information Theory*, 44, 2416-2434, 1998.

[11] G. Han and B. Marcus, Capacity of Noisy Constrained Channels, *Information Theory Symposium*, Nice, 2007.

[12] P. Jacquet, Space-time Information Propagation in Mobile ad hoc Wireless Networks, `http://ee-wcl.tamu.edu/itw2004/program/jacquet_inv.pdf/`, *ITW 2004*, San Antonio, 2004

[13] P. Jacquet. Realistic wireless network model with explicit capacity evaluation. Technical Report INRIA RR-6407, INRIA, 2008.

[14] P. Jacquet, Bernard Mans and G. Rodolakis, Information Propagation Speed in Delay Tolerant Networks: Analytic Upper Bound. *Proc. ISIT*, 6-10, Toronto, 2008.

[15] P. Jacquet, G. Seroussi, and W. Szpankowski. On the Entropy of a Hidden Markov Pprocess. *Theoretical Computer Science*, 395, 203-219, 2008.

[16] P. Jacquet, G. Seroussi, and W. Szpankowski. Noisy constrained capacity. In *Information Theory Symposium*, volume IEEE, pages 986–990, 2007.

[17] J. Konorski and W. Szpankowski, What is Information?, Festschrift in Honor of Jorma Rissanen, 154-172, 2008.

[18] D.G. Luenberger, *Information Science*, Princeton Univ. Press, 2006.

[19] P. Nurse, Life, Logic, and Information, *Nature*, 454, 424-426, 2008.

[20] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific, Singapore, 1998.

[21] J. Roche. Gambling the Mnemonically Impaired, *IEEE Trans. Information Theory.*, 48:1379–1392, 2002.

[22] J. Seidler, *The Science of Information*, WNT, Warszawa, 1982 (in Polish).

[23] C. Shannon, A Mathematical Theory of Communication, *Bell System Technical Journal*, 27, 379-423 and 623-656, 1948.

[24] C. Shannon. The Lattice Theory of Information. *IEEE Transaction on Information Theory*, 1:105–107, 1953.

[25] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences.* New York: John Wiley & Sons, Inc., 2001.

[26] S. Venkatesh and J. Franklin. How Much Information Can One Bit of Memory Retain About Bernoulli Sequence, *IEEE Trans. Information Theory.*, 37:1595–1604, 1991.

[27] C.F. von Weiszsäcker and E. von Weiszsäcker, Wideraufname der begrifflichen Frage: Was ist Information?, Nova Acta Leopoldina, 206, 1972.