PERGAMON Mathematical and Computer Modelling 38 (2003) 691–708

MATHEMATICAL
AND
COMPUTER
MODELLING

www.elsevier.com/locate/mcm

The Evolution of Human Communication and the Information Revolution— A Mathematical Perspective

A. BENCZÚR
Department of Information Systems
Eötvös Loránd University
Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary
abenczur@ludens.elte.hu

Abstract—In my paper, human communication is placed into the center of informatics. The typical schema of elementary communication and communication in information systems is discussed by natural and formal schemas. The schemas are used to explain the main characteristics of the epochs of the evolution of human communication and the key effect of dramatic change in communication due to the very fast development of information technology.

The main lows of two mathematical theories, namely Shannon's information theory and the Kolmogorov algorithmic entropy, are explained together with their roles in communication. The coincidence of the two entropies on very large objects is proved. The difference between large, algorithmically generated, compressible objects and the typical, uncompressible random objects is visually demonstrated by black and white colorings. The interesting self-interference of random colorings is shown and explained. © 2003 Elsevier Ltd. All rights reserved.

Keywords—Communication, Information systems, Information theory, Kolmogorov complexity, Randomness.

1. INTRODUCTION

One of the most characteristic phenomena at the beginning of the third millennium is the speed of the development of the information technologies that exceeds the pace of every former technological development and leads to the information revolution.

In my paper, I first show the place of the new world of communication in the process of the evolution of human communication. Then some formal and mathematical models follow, which show mathematical boundaries and possibilities. The general theorems do not give keys for the solution; we are facing up to an inexhaustible world of algorithmic problems.

What makes the communication run is basically the uncertainty, the randomness. To the contrary, the digital world is based on computability. The computable world cannot replace the reality; it can help in the modeling and cognition of the real world within some limit, and it may help in understanding the past and in the prediction of the future.

My starting point to approach the aggregate of what is covered by the terms computing profession, information technology, information engineering, software engineering, computer science,

0895-7177/03/$ - see front matter © 2003 Elsevier Ltd. All rights reserved. Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX
doi: 10.1016/S0895-7177(03)00272-3

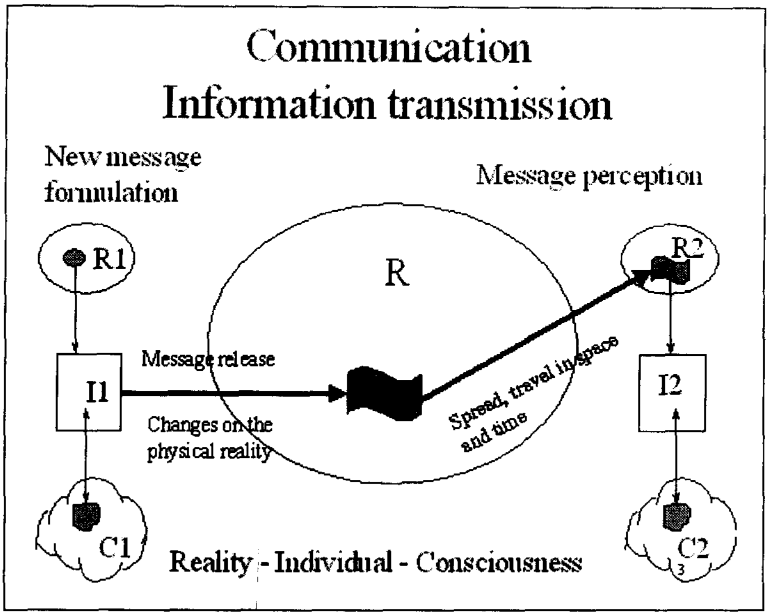


Figure 1.

information science, telecommunication, business information management, and many other related terms in the U.S. (see in [1]) is human communication. I prefer to cover all of these by the term "informatics", which is missing from the English vocabulary, but exists in German, French, and Hungarian. On the way of approaching informatics, my dominating experiences came from the introduction to probability theory and information theory by Rényi, from our common work with Arató in statistical analyses of stochastic processes, and from many database and information system development projects. In my paper, "informatics" has a wider meaning than in the definition of Denning in [1]:

> *"Today, most computer scientists understand computer science as a discipline that studies the phenomena surrounding computers. These phenomena include design of computers and computational processes, representation of information objects and their transformations, theoretical and practical problems in hardware and software, efficiency, and machine intelligence. In Europe the discipline is called 'informatics' and in the U.S. 'the discipline of computing' or information technology."*

I consider that the most important novelty in the information revolution is the new way of communication, not the computers.

2. THE BASIC MODELS OF COMMUNICATION

Communication covers all means of interaction between two human brains.

The communication or interaction of two human beings can only be realized via intermediate physical changes between them. Messages and thoughts cannot be directly exchanged between the consciousnesses of human beings. Figure 1 shows one of them as the source of emitting a new message and the other one as the perceiver, the destination of the message. The selection and formulation of a message is the result of the interaction between observation of the surrounding real world and the internal conscious processes. The source individual releases the sign of communication by appropriate change in the real physical world. This change then is spreading, traveling through time and space, reaches the observable reality of the destination individual, who perceives the sign, and by his/her conscious activity the message reaches his/her consciousness.

The small darker areas on the figure show the possible effect of randomness in this elementary communicational process. The uncertainty of the selection of the message comes from the ob-

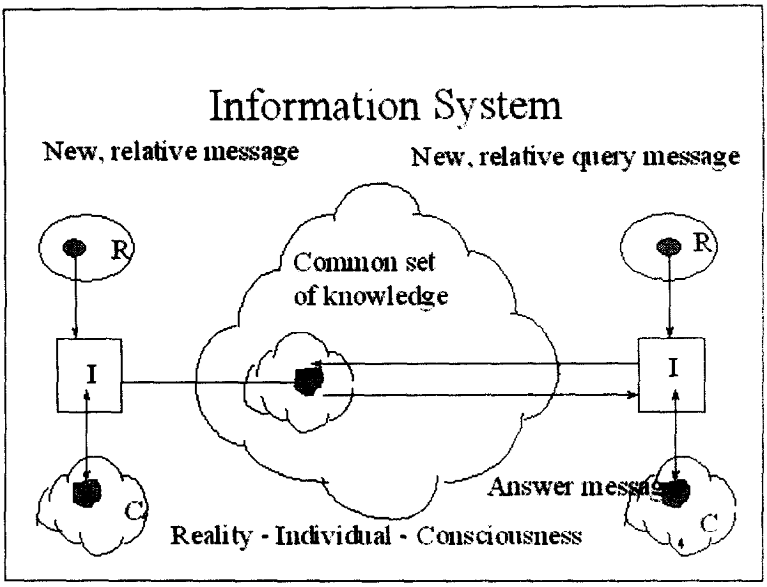


Figure 2.

servation of the real world as well as from the internal process of thinking. The change caused by the release of the message in the real world goes through random distortions and it adds a noise to the sign which reduces the accuracy of the observation. Finally, the observation of the receiving individual has same external and internal random effect too.

The necessary condition of the success of the communication is that for different messages the source should release different changes in the physical reality so that the receiving individual could distinguish them.

Turning our attention from elementary communication to more complicated communication inside communities of human individuals, we get the general model of information systems as summarized in Figure 2.

The communication system of human communities makes use of the evolution of common knowledge which is the result of previous communications and common observations, and remains accessible either in human minds or in durable physical signals. (The most important common knowledge is the language.)

Many of the new messages are not directed to definite receiving individuals; instead they are relative to the common knowledge, and they contribute to its development. From the collective knowledge the relevant, necessary information as answer messages are retrieved by means of special query messages. I call this kind of collective communication an information system. Naturally, an information system should contain some solution to "remember" the collective knowledge, something like the institution of wizards in ancient tribes, clergymen in ancient Egypt, libraries, vital registers, educational system, and among many others the most important, the language. Still, the first step of the communication, the release of a new message, is done by some changes in physical reality, and the ending step, the recognition of a message, is an observation of some changes in the physical reality by the senses of a human being.

The storage, access, distribution, and secrecy of the collective knowledge need organized solutions.

Now, the evolutionary process of human communication reaches a new epoch. After the epoch of talking and speech, then the epoch of writing and printing, the Guttenberg Galaxy, we are in the middle of the revolutionary evolving new epoch of the information society, where the computer, electronic, information, telecommunication, and multimedia technologies are integrated into a new communicational infrastructure and a worldwide digital information system. The space and time limitations of communication are vanishing, the capacity of storing, processing, accessing, and transmitting data is still growing exponentially, and a new medium with services of ever evolving intelligence is being built for human to human, human to nature, and nature to nature interaction.

3. FORMAL MODELS

It was in the middle of the 20th century when the development of formal, mathematical models of communication became a must for the utilization of new telecommunication technologies. The milestone in mathematical treatment of communication was the seminal paper [2] of Shannon. The introduction of entropy brought the possibility to mathematically analyze the information transmission process. We use Shannon's model to characterize the evolution of the communication and to point out the great qualitative differences of the present state from all the previous solutions.

Figure 3 shows a slightly modified version of Shannon's original schematic diagram of a general communication system. We stress that artificial channels are the new technical solutions to transmit the signal from the transmitter (sender) to the receiver. Both the transmitter and receiver are artificial equipment and their encoding and decoding function play very important role.

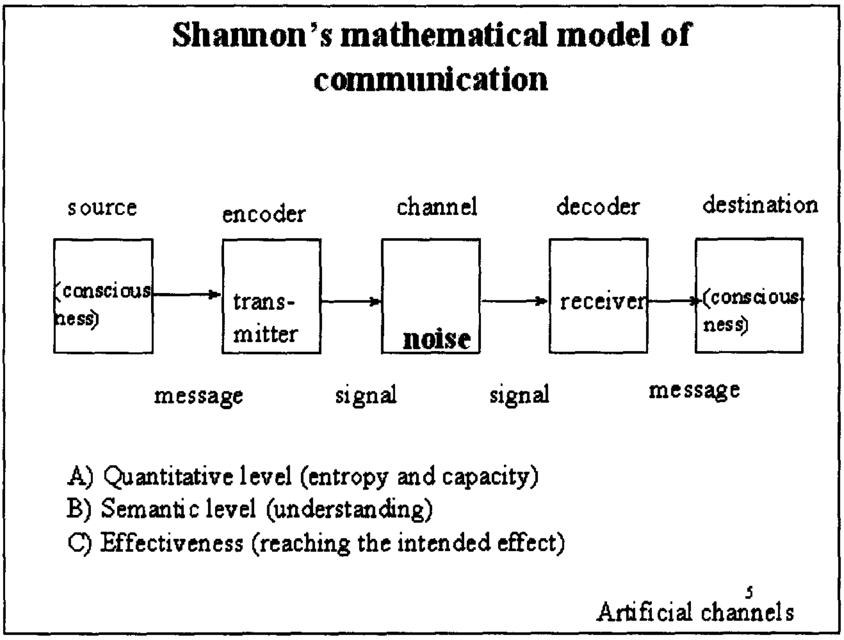
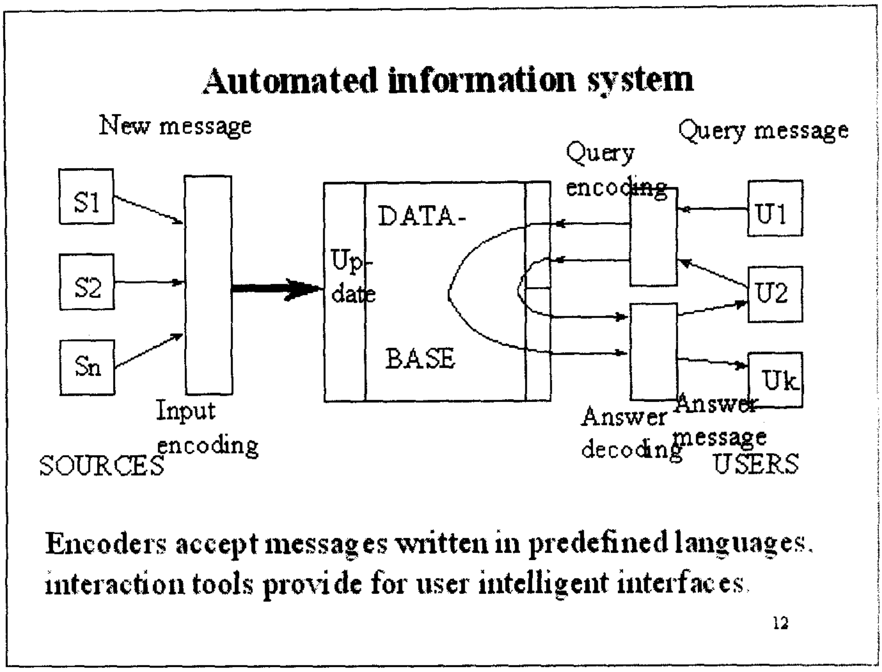

Shannon's mathematical model of communication
source
encoder
channel
decoder
destination
(consciousness)
trans-mitter
noise
receiver
(consciousness)
message
signal
signal
message
A) Quantitative level (entropy and capacity)
B) Semantic level (understanding)
C) Effectiveness (reaching the intended effect)
5
Artificial channels

Figure 3.

Similarly to Shannon's schema, the automated information system consists also of five main components as shown in Figure 4. The possibility of storing the common knowledge in the database gives new capabilities and restrictions as well. The new capabilities are in the high power and capacities in storing, accessing, processing, updating of the data representing the common knowledge, and the restrictions are that we are closed into the formal, algorithmic world of the signals. The selection of the messages should comply with the structure of the database, and formulations of the source, query, and answer messages are restricted to predefined formal languages. The evolution of information technologies widens the barrier supplying human interaction—the ensignaling and designaling phase—with more and more powerful intelligent interfaces, extending the graphical, visual, voice-based interaction and via the virtual reality up to the possibilities of teleimmersion. All these lead to new complex languages of high expressing power.



Automated information system

Encoders accept messages written in predefined languages, interaction tools provide for user intelligent interfaces.

12

Figure 4.

The fast transformation of our common knowledge into the world of signals hides serious dangers: after our messages become signals, they start an independent life. They can undergo to any computation. Decoding and designaling the result of the computation may lead to meaningless or misleading results in the real world. The human race genetically has with thinking the capability of counting and computing, that differentiates us from other behaviors of nature and life. Computers amplify this capability, and moreover, we trust too much in the correctness of computer calculations. Our messages reflect in some way the reality, and encoding them leads to the world of signals where we can introduce computations over them which are not true in their original world. Let me give a very primitive example.

Let us encode apple by 1, pear by 2, and peach by 3. Then, using addition over our codes we get apple plus pear is peach! The mistake is the consequence that the operation is not generic on the encoding; it is not invariant under the permutation of codes.

4. THE HISTORY OF COMMUNICATION IN BRIEF

The basic and formal models of human communication allow us to characterize the important possibilities of the main communicational epochs. The figures use Shannon's schema for visualization and I hope they are self-explanatory.

The typical channel is presented in the figures, and at the bottom the characteristic solution of managing the common knowledge is mentioned.

Figure 5 shows the communication schema of speech and talking. The state of the human minds is symbolized by the gray clouds with a rectangular shaped message. The encoder/transmitter and the receiver/decoder boxes are our biologically built-in "equipment", and the speed of the communication is not limited by the channel capacity, but by our limited capabilities in pronunciation and speech recognition. Simultaneity and limitation in distance also give restrictions. The only recording solution is the memory of living individuals. The birth of languages is the main outcome of the prehistoric era.

Writing and printing, the next epoch as shown on Figure 6, brought the possibility of recording messages of the present for the future. It also introduced new message releasing solutions by hand movement, and involved our most powerful sense, the vision of message recognition. New writing, which is not only a copy of an older one, needs the release of new human messages. Addition,

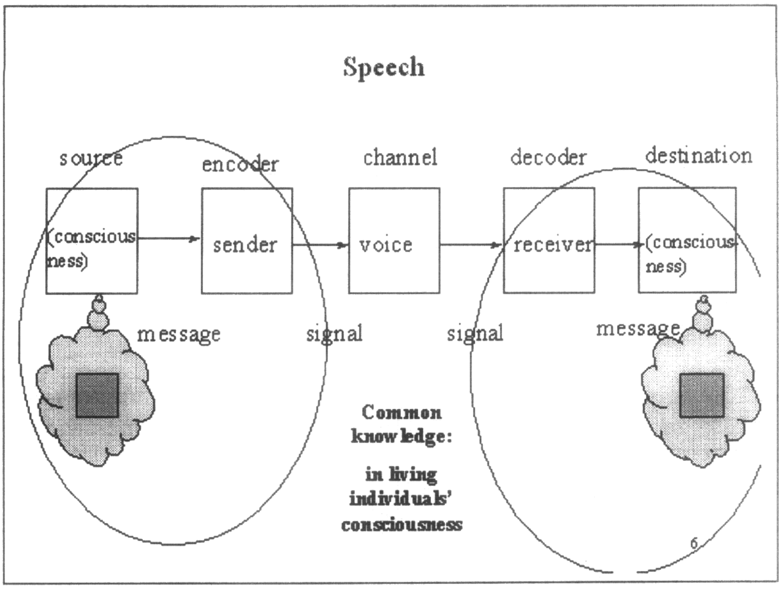


Figure 5.

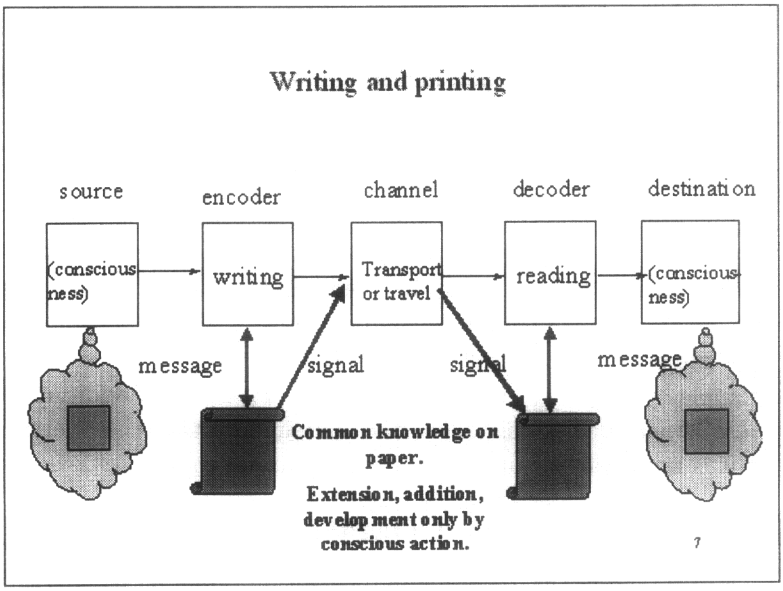


Figure 6.

processing, development, and change of the recorded knowledge can be done only by conscious action.

Although a lot of interesting questions and solutions can be studied on speech and writing based communication and information systems, they do not really need mathematical, formal models.

Now, skipping over some smaller steps in the evolution of communication we arrive at the present epoch of computers.

Figure 7 demonstrates elementary communication between persons where general purpose computers are used for encoding, transmitting, receiving, and decoding signals. Analyzing the schema from top down, the first novelty is the disconnection of the source and destination from encoder and decoder marked with question mark. Since signals used by computers cannot be directly manipulated, produced, and observed by humans, we have to construct interfaces, ensignaling and designaling devices for human/computer interaction.

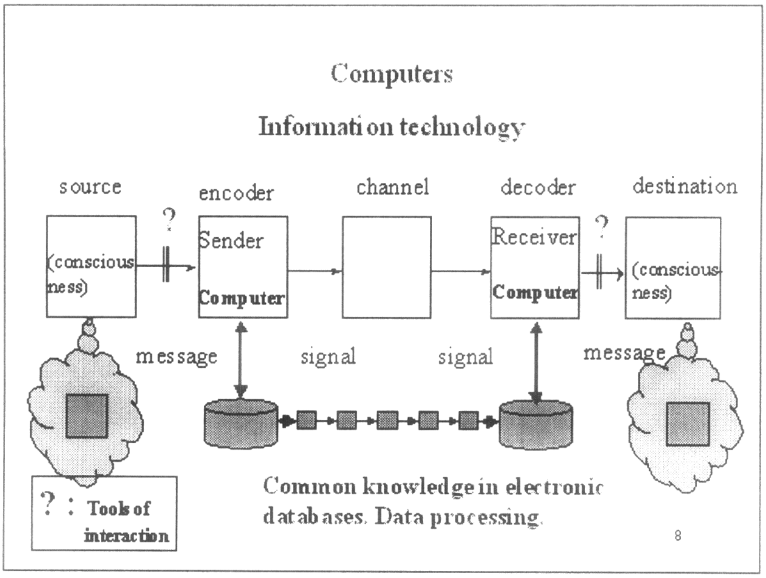


Figure 7.

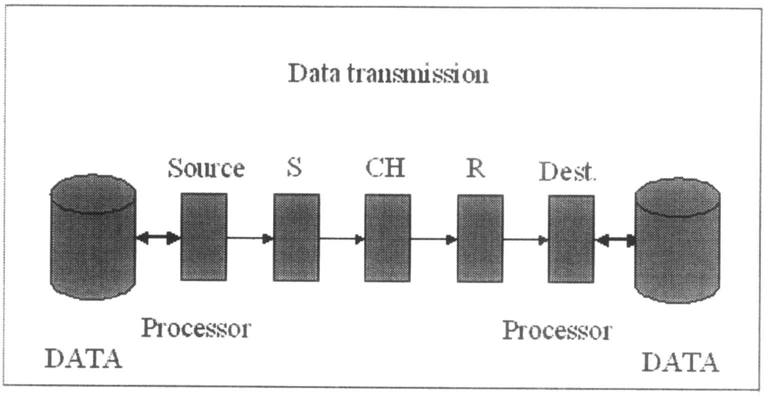
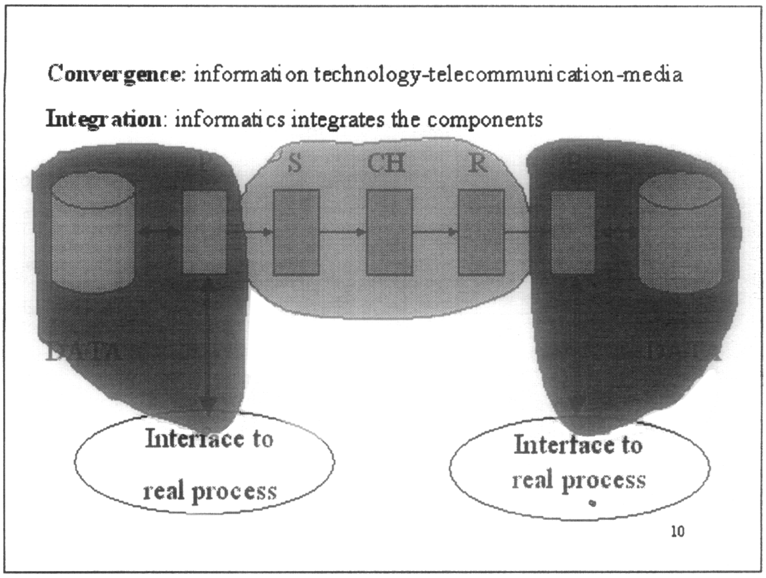


Figure 8.

Devices, like keyboard, mouse, touchpad, touch-screen, electronic gloves, microphone, camera, and a lot more, are for primary release of human messages, and transforming them immediately into electronic signals. In the opposite direction the corresponding devices, the displays, printers, loudspeakers, radio and TV sets, and virtual reality environments, are transforming and amplifying electronic signals to physical changes recognizable by human senses. (In the same way as physical changes originated by a human message release can be ensignaled and designaled, many natural and technically designed physical processes became observable by electronic devices which produce instrumental data source, and in the opposite direction from the electronic signals amplifiers produce physical processes. This is the new way of interaction, "communication" with nature.)

The next new components in the figure are the two storage boxes below the computers. Once a message became signals for the computer, it was possible to record it and store it for future communications that may have independent timing from the persons' primary communication.

Between two electronic storage systems we can build an electronic data transmission connection as you see magnified in Figure 8. So we get the basic building boxes of electronic data transmission networks. Both the source and the destination are processors, or computers. They supervise the data storage and use it for selecting messages for transmission and for recording received messages. The processors are working in an artificial physical environment and use signals that are unobservable or unrecognizable for human senses.



Convergence: information technology-telecommunication-media

Integration: informatics integrates the components

Figure 9.

Processors may have other connections to real world processes as Figure 9 shows. Persons can be connected to them via ensignaling/designaling devices. They may receive data from observation of physical processes via measuring, recording instruments, or can send signals to interfere with physical processes. We can observe the three main components of the new world of communication.

- The Sender—Channel—Receiver boxes represent the main area of *telecommunication*, including wired, satellite, optical, radio wave data transmission, broadcasting, mobile communication, and other solutions.
- The Data—Processor—Interfaces represent the area of *information technologies*.
- The *applications with their specific interfaces* and systems of services around them are expanding from automation, process control, office automation, CAD/CAM systems, e-business, e-government, e-learning, and media and culture.

I would prefer to call this whole "informatics", which integrates the three main components.

The popular concept of the convergence of information technology, telecommunication, and media reflects one of the largest effects of the evolution, the effect on mass communication. This concept puts the processor, the computer, in the center. In my approach—following Weaver [3]—communication covers all the means one human brain can influence the consciousness of some other person. Placing the communication into the center, the integrating character of informatics better fits to the matter.

The elementary building blocks of data transmission in Figure 8 can be connected to each other; they can build a network of interconnected, communicating processors, and storage systems. The communication can be interrupted, postponed, delayed, recorded, repeated, and automated. This is the basis of the internet, of the new worldwide communication infrastructure.

Storage devices, processors, transmitters, channels, receivers, and devices of interactions are all technical, engineered tools, utilizing very deep new discoveries in physics. What makes this infrastructure work is the large mass of programs running on processors. The programs connect the infrastructure to the applications, and in consequence program design, development, and maintenance deeply depend on application area. All these lead to a new system of professionals. There will be a need for an information technology specialist for application areas, and the most important general profession will be—according to Denning—the profession of computing. Ending this part, we cite his ending paragraph from [1].

"Computer scientists, software engineers, computational scientists, and other information technologists have a marvelous opportunity to transform their academic disciplines into the profession of computing. They will have to face, and cross, the chasm between their practices as inventors and visionaries, and the pragmatic interests of their many clients and customers. It will not be easy. They have shown they can do it before, and they can do it again."

5. MATHEMATICAL PROBLEMS

Let us start with the problem of one-to-one mapping from the set of possible messages to the set of available channel sequences, or signals in Shannon's model. The problem is to associate different channel sequences to different messages. According to the principle of packing in boxes, if the number of the boxes is less than the number of balls, then there will be at least one box packed with two balls. So we can solve the problem of communication if and only if the number of channel sequences is not less than the number of different messages.

The next problem is how to count them. It is not easy at all; information theory is responsible for doing that. We shall show it only for the discrete, elementary case where both the messages and the signals consist of finite grade elements, from symbols and signs, respectively. The parameter or dimension which influences the size of the problem is time.

Let us take a channel with $N(T)$ different allowable sequences of signals that can be transmitted in time T. Clearly, two channels are equivalent if their $N(T)$ functions are equal.

The capacity of the channel with function $N(T)$ is defined by the formula

$$C = \lim_{T \to \infty} \frac{\log_2 N(T)}{T}. \tag{1}$$

This definition gives for symmetric binary channels the intuitively clear C bit/sec capacity. Channels are designed and constructed to achieve a given capacity, so C is given.

Determining the number of possible messages that can be selected in time T is a bit harder. Shannon's entropy is not a direct measure for that. It measures the uncertainty of the selection from the set of possible messages. Nevertheless, for very long messages it gives the basis for an estimate of the number of typical messages. Shannon's famous formula assigns to the probability distribution $(p_1, p_2, \dots, p_n)$ the entropy

$$H(p_1, p_2, \dots, p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i. \tag{2}$$

The mathematical model of the message selection is given by a sequence of random variables, $\xi_1, \xi_2, \dots, \xi_N$, where the possible value of the variables comes from the set of symbols $(x_1, x_2, \dots, x_n)$. The average source entropy per symbol is defined as

$$H = \lim_{N \to \infty} N^{-1} H(\xi_1, \xi_2, \dots, \xi_N), \tag{3}$$

where $H(x_1, x_2, \dots, x_n)$ is the Shannon entropy of the joint distribution.

From laws of probability theory, it follows under very general assumptions (ergodicity) for the source that for very large N the selection falls to a typical set with probability $1 - \beta$, and the probability $p_{\vec{x}}$ of the selection of a given element $\vec{x}$ from the typical set satisfies

$$2^{-N(H+\rho)} < p_{\vec{x}} < 2^{-N(H-\rho)} \qquad (\beta \text{ and } \rho \text{ small with } N \text{ large}). \tag{4}$$

From this inequality we get the following estimation for $M(N)$, the number of elements in the typical set:

$$(1-\beta)2^{N(H-\rho)} < M(N) < 2^{N(H+\rho)}. \tag{5}$$

Now we can return to our original problem. Given the channel capacity C (bit/sec), and the source entropy H (bit/symbol), we shall answer the question of what the limit is for the rate V (symbol/sec) of the source where there exists one-to-one mapping from messages to sequences of channel signals for large time interval T.

From (1) and (5), it follows that for arbitrarily small positive α, β, χ there exists T_0, such that for any $T > T_0$ the following holds:

$$2^{T(C-\alpha)} < N(T) < 2^{T(C+\alpha)}; \tag{6}$$

and with probability greater than $1-\beta$, the selected message belongs to the typical set containing $M(VT)$ elements satisfying inequality

$$(1-\beta)2^{VT(H-\chi)} < M(VT) < 2^{VT(H+\chi)}. \tag{7}$$

Comparing the upper bound in (6) and the lower bound in (7), we get the necessary condition $V(H-\chi) < T(C+\alpha)$ for encoding, which means V cannot exceed C/H. In order to get sufficient conditions, we first estimate the expected size of the possible message set for sufficiently large message length N. With probability $1-\beta$ the size is $M(N)$, and with probability β the size is n^N. Using two different code lengths, T_1 and $T_2 = T_1 + T_3$ for the typical messages and for untypical messages, respectively, the sufficient time values are with arbitrarily selected small δ, $T_1 = N(H/C + \delta)$ and $T_3 = N((\log_2 n)/C + \delta)$. So the expected time sufficient to transmit messages of length N is $T = T_1 + \beta T_3 = N(H/C + \varepsilon)$, where ε can be arbitrarily small when N is large. This means that for any $\gamma > 0$ operating the source at rate $V = C/H - \gamma$, then it is possible to encode the messages of the source in such a way as to transmit them through the channel without any loss. This is the proof of Shannon's fundamental theorem for noiseless channel and ergodic sources.

From this introductory mathematical reasoning we get the basic limits and possibilities of the elementary communication, but it does not give direct help in solving real communicational problems.

Entropy formula (2) is invariant under the permutation of the probabilities, and does not depend on the possible set of the message symbols. All this information should be available in some form at the encoding and decoding processors. Without computational and storage capacities, we were not able to use the new digital channels.

The utilization of high speed (large capacity) channels by low bit/symbol sources needs very high symbol rate and, as suggested from the proof of the fundamental theorem, encoding of blocks of very long messages. Instead of collecting messages in time, we can collect them in space, which means the multiplexing of messages coming from a large number of sources.

The capacity of high speed digital channels is growing a bit faster than predicted; today 40 Gigabit/sec is the highest capacity of a wide area network in the civil sphere; it connects two supercomputing centers of two universities in the U.S. The predicted available channel capacity in 2008 will be 1 Terabit/sec, which is 10^{12} bit/sec. This allows multiplexing simultaneous typing of ten billion ($= 10^{10}$) persons using typical keyboards of 128 keys at a 13 keystrokes/sec rate.

The mathematical model for the random behavior of the source is the probability distribution. A very important practical condition for the approximating distribution is that its entropy should be greater than the entropy of the real source. For example, assuming identical distribution for selecting a message symbol, among the possible joint distributions of selecting a sequence of symbols, the independent distribution has the highest entropy. In good models, the typical set of very long real messages should be a subset of the typical long sequences of the approximating distribution.

We arrive at the second mathematical problem if we investigate very long and very complex messages as they are, without any distribution of the message selection, and from the possibility of compression. This leads to the realm of algorithms and to another notion of entropy, Kolmogorov's algorithmic entropy, or the Kolmogorov complexity.

The theory of Kolmogorov complexity has its roots in the beginning of the Sixties after the works of Kolmogorov, Solomonov, and Chaitin. The three mathematicians independently of each other and nearly the same time found the fundamental theorem of the theory. The origin of Kolmogorov complexity is the paper [4] of Kolmogorov. A good summary can be found in the book [5] and in the special edition [6] of *The Computer Journal.* Kolmogorov's original question was: what is the length of the shortest program that prints the book *War and Peace*?

The measure of compressibility is related intuitively to short codes. Without loss of generality, we suppose that the codes are finite binary words. Moreover, sorting binary words by lengths and by their binary value, we have a one-to-one mapping from the set of finite binary words ($\boldsymbol{\Omega}$) onto the set of nonnegative integers ($\boldsymbol{N}$), and therefore, we do not differentiate the n^{th} binary word and the integer n.

Let us consider decoding function $f : \boldsymbol{N} \to \boldsymbol{N}$. Without loss of generality we suppose that all the decoding functions we use in the sequel are computable functions (partial recursive functions).

DEFINITION 1. *The complexity of $x \in \boldsymbol{N}$ relative to the partial recursive function f is*

$$C_f(x) = \min\langle l(p) \mid f(p) = x\rangle$$

if such p exists, and $C_f(x) = \infty$ otherwise.

FUNDAMENTAL THEOREM. *(See Kolmogorov, Solomonoff, Chaitin.) There exists optimal partial recursive function $f_0(p)$, such that for every partial recursive function $f(p)$*

$$C_{f_0}(x) \leq C_f(x) + k_f, \tag{8}$$

for all $x \in \boldsymbol{N}$, where k_f depends only on f.

(The proof uses a universal $U : \boldsymbol{N} \times \boldsymbol{N} \to \boldsymbol{N}$ partial recursive function, for which there exists for every $f(p)$ partial recursive function n_f, such that $U(n_f, p) = f(p)$. From an appropriate encoding function of the ordered pair (n_f, p) we can construct f_0. The optimal function minimizes the sum of the length of the decoding program and the length of the code.)

Clearly, the complexities relative to optimal functions differ only in an additive constant. So we agreed to fix from now on the optimal function f_0.

DEFINITION 2. THE KOLMOGOROV COMPLEXITY. *The Kolmogorov complexity of $x \in \boldsymbol{N}$, denoted by $I(x)$, is defined as*

$$I(x) = C_{f_0}(x).$$

The function $I(x)$ is an objective measure of compressibility, but it has an important drawback: it is not computable.

In spite of that, $I(x)$ has two elementary properties which are very useful in analyzing its properties and in its application.

PROPERTY 1. According to the Fundamental Theorem and formula (8), $C_f(x)$ gives an upper estimate of $I(x)$ up to an additive constant.

Application: taking $f(x) = x$ proves $I(x) < l(x) + k_1$.

PROPERTY 2. The Kolmogorov complexity of a given number x is the length of an existing and unique number; that is, $I(x) = k$ means that for a unique p, $f_0(p) = x$ and $l(p) = k$.

Application: there are less than 2^k numbers of less than k Kolmogorov complexity, just because there are no more code words shorter than k.

From the many special versions of Kolmogorov complexity, one with the greatest importance is the prefix Kolmogorov complexity. We get the prefix complexity if we use only prefix-free decoders. The fundamental theorem holds for prefix-free decoders, and we agreed to fix from now on the g_0 optimal prefix-free decoder. The prefix-free Kolmogorov complexity is defined by

$$K(x) = C_{g_0}(x).$$

The next important concept is the conditional Kolmogorov complexity. Intuitively, the conditional complexity answers the question: what is the length of the shortest program that can compute x from y? First, similarly to the unconditional complexity, we define the conditional complexity by a given partial recursive function $f : \boldsymbol{N} \times \boldsymbol{N} \to \boldsymbol{N}$.

DEFINITION 3. *The conditional complexity of x given y according to function f is $C_f\langle x \mid z\rangle = \min\langle l(p) \mid f(p,y) = x\rangle$, if such p exists and ∞ otherwise.*

THEOREM 2. FUNDAMENTAL THEOREM. *There exists an optimal partial recursive function $f_0(p,y)$ (prefix-free in p $g_0(p,y)$) such that for any partial recursive functions $f(p,y)$ $C_{f_0}\langle x \mid y\rangle \leq C_f\langle x \mid y\rangle + k_f$, for all x, y where k_f depends only on f.*

Fixing from now on the optimal functions $f_0(p,y)$ and $g_0(p,y)$, the conditional Kolmogorov complexity is defined by

$$I\langle x \mid y\rangle = C_{f_0}\langle x \mid y\rangle,$$

and for prefix-free complexity

$$K\langle x \mid y\rangle = C_{g_0}\langle x \mid y\rangle.$$

All the properties of the conditional Kolmogorov complexity are similar to the unconditional complexity.

An important advantage of the conditional complexity is in the possibility to measure the conditional complexity of an element given the set containing it. The set, naturally, should be finite or should have finite representation.

Let $\boldsymbol{A}$ be a finite set of m finite binary words, and let its finite representation be the finite binary word a. Then for $x \in \boldsymbol{A}$, we can define the conditional complexities of x given the set $\boldsymbol{A}$ as $I\langle x \mid A\rangle = I\langle x \mid a\rangle$ and $K\langle x \mid A\rangle = K\langle x \mid a\rangle$. From Property 2 we easily get that at most $2^{-k}m$ elements of $\boldsymbol{A}$ can have smaller conditional complexity than $\log_2 m - k$ and for the prefix-free complexity $2^{-k-1}m$ elements of $\boldsymbol{A}$ can have smaller complexity than $\log_2 m - k$. (The elements of $\boldsymbol{A}$ with $I\langle x \mid A\rangle \geq \log_2 m - \Delta$ are called by Kolmogorov Δ-random elements of $\boldsymbol{A}$.)

An important system of finite sets is defined by any effectively enumerable set $B \subset \boldsymbol{N} \times \boldsymbol{N}$, where the set B_a with parameter a is $B_a = \langle x \mid (x,a) \in B\rangle$, and we suppose that the number of elements in B_a is $m_a < \infty$.

LEMMA 1. *Under these conditions*

$$I\langle x \mid a\rangle \leq \log_2 m_a + k_B,$$

and

$$K\langle x \mid a\rangle \leq \log_2 m_a + 2\log_2\log_2 m_a + c_B,$$

for all a and $x \in B_a$, where k_B and c_B depends only on the enumeration function of B.

The practical meaning of the lemma is that given B_a, there is no better way to encode its elements than using the uniform code of length m_a, and this encoding is computable.

The world of the large databases and automated information systems suggests the feeling that computers produce new information. It is natural to ask, what is the information quantity we can retrieve from a database? What does it mean, that from a database that stores the father-son relationships we can retrieve the name of the grandfather of a given person? Do we really gain information? The conditional Kolmogorov complexity gives the exact answer: the information quantity which we can retrieve from a database cannot exceed the information quantity we entered to the database. We state it formally as the *law of information conservation (nongrowth), or information balance.*

In order to give the exact mathematical formulation, let y denote the data in the database (including all the program codes for managing, accessing, processing, retrieving data), q denote the prefix-free encoding of the query, and v the answer to the query from the database. Since the answer is a computable function $f(q, y) = v$, using the inequality of the fundamental theorem we get

$$K\langle v \mid y\rangle = C_{g_0}\langle v \mid y\rangle \leq C_f\langle v \mid y\rangle + n_f \leq l(q) + n_f. \tag{9}$$

In practice, inequality (9) means that the information quantity of the new information in the answer message produced from a database for the query message q under the condition that the content of the database is given cannot be greater than the information quantity of the query message. In other words, this means that inside the automated information system's database there is no information source. It is important to note that the real gain in getting the answer $v = f(q, y)$ is not its conditional information quantity, but rather the algorithmic work of evaluation of function f that typically is impossible for human thinking. The other important value of answer message v for the destination is in the fact that the destination typically has only partial knowledge about the content of the database, so the conditional complexity of v given the knowledge of the destination can be much greater than the conditional complexity given the content of the database.

Investigation of the effect of the growth of the database on query and answer complexity is beyond the scope of the paper.

After this short introduction to Kolmogorov complexity it is natural to raise the question what is the relationship between the two notions of information quantity. The essence of the answer is that the larger the messages are, the closer the two entropy quantities are to each other.

For very large random sequences (messages) coming from an ergodic source the selection falls with probability greater than $1 - \beta$ into a set of typical sequences, $B(N)$, and the probability distribution of the selection from $B(N)$ converges to the uniform distribution. (β can be taken small with the length of the message becoming large.) Inequalities (4) and (5) say that for arbitrarily small β, ρ the number of typical sequences, $M(N)$, satisfies for N large enough the inequality $(1 - \beta)2^{N(H-\rho)} < M(N) < 2^{N(H+\rho)}$.

Using the notations for formula (3), let us choose an arbitrary prefix-free binary encoding of the symbols $x_1, x_2, \ldots, x_n$, and denote by $y(\vec{x})$ the encoding of the sequence $\vec{x} = x_{i_1}x_{i_2}\ldots x_{i_N}$. Whenever the membership of $B(N)$ can be checked by a computable statistical predicate, the system of sets of binary strings $A_N = \langle y(\vec{x}) \mid \vec{x} \in B(N)\rangle$ satisfies the condition of Lemma 1. From Lemma 1, we get for any $y \in A_N$ and N large that

$$N(H - \rho) + \log_2(1 - \beta) < K\langle y \mid A_N\rangle < N(H + \rho) + 2\log_2 N(H + \rho) + k_B \tag{10}$$

The lower estimate in (10) holds only on the overwhelming majority of A_N.

Dividing the sides of inequality (10) by N, we get

$$H - \rho_1 < N^{-1}K\langle y \mid A_N\rangle < H + \rho_2, \tag{11}$$

which proves that for large objects the two entropy measures give the same value in limit.

Summarizing, we conclude in both mathematical theories that in our effort to solve efficiently the encoding of messages and knowledge, the best approach is to find the possible smallest set that covers all the possible and typical situations we should face to in the future, and there is no better way for encoding the elements than using the uniform length encoding. In both ways we arrive at the same limitations and possibilities of computable encodings.

A strong suggestion comes from the discussion above that there must be some connection between randomness and algorithms. Paradoxically, the randomness of infinite sequences can be characterized in algorithmic way. It is worth mentioning that the last paper [7] of Kolmogorov (with Uspenskii) gives an exciting summary on the strong relationship of algorithms and randomness where Kolmogorov complexity plays a central role.

In the case of finite objects the situation is quite different. Imagine a random string of zeroes and ones coming from the experiment of 1000 tossings of a fair coin. Start replacing the ones in

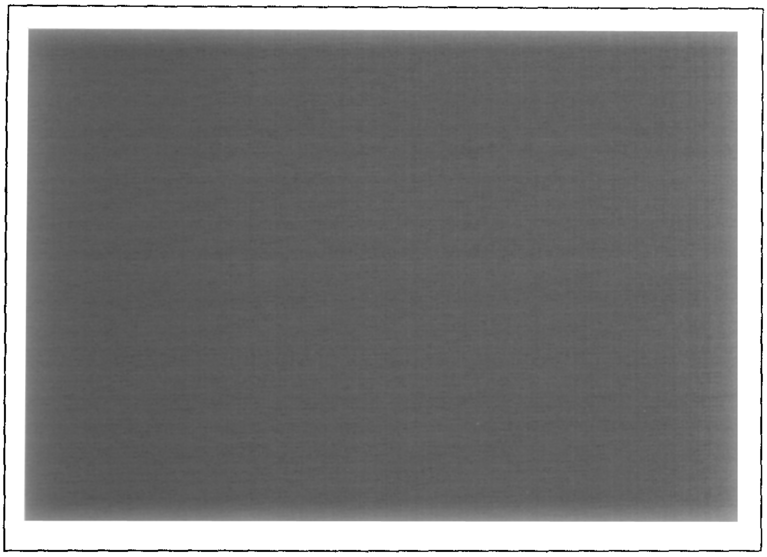


Figure 10.

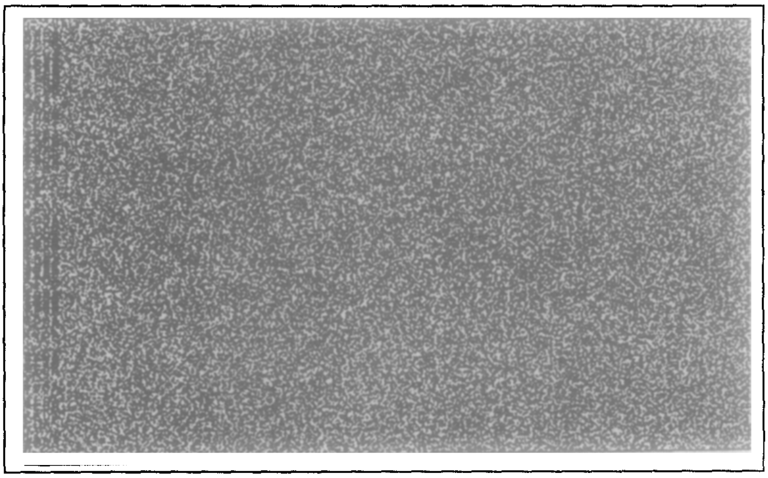


Figure 11.

the string with zero in some random order. Finally, we get the nonrandom string of 1000 zeroes. Is it possible to say when we crossed the border between random and nonrandom strings?

For a finite string from a finite set the randomness cannot be defined, but following Kolmogorov's approach from [7], we can define the defect of the randomness as the difference of the base 2 logarithm of the number of elements of the set and the conditional Kolmogorov complexity of the element given the set. This is the motivation of Δ-randomness mentioned before Lemma 1.

The typical elements of a set are those which are Δ-random for some small Δ. The full characterization of a large set consists of the description of the set of its typical elements via some computable characterization and from the individual encoding of typical elements. This suggests to avoid the use of special shorter encoding for simpler elements and to treat the few possible very complex elements as exceptions.

For the visual demonstration of the above principle let us take the black and white pictures of uniformly 1/2 gray level, that is, where half of the pixels are colored black, and the others white. (These pictures are visual representations of fuzzy sets with constant 1/2 membership function.) Figure 10 shows the chessboard or checkered coloring, and the color of the left upper pixel uniquely determines the picture. Figure 11 is a coloring generated by a random number generator. (I am still using the first printing I got in 1990. The darker strip on the left side is

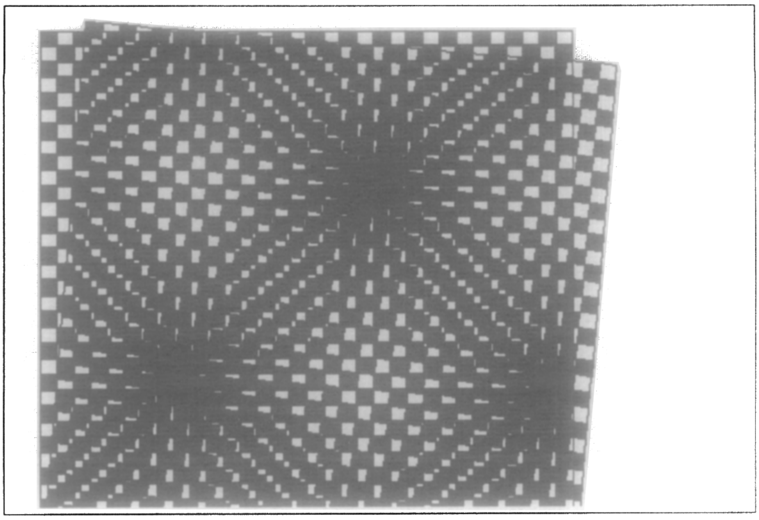


Figure 12.

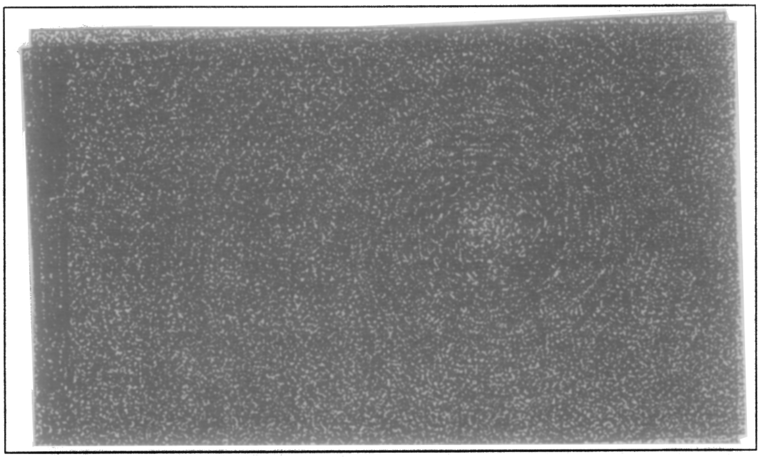


Figure 13.

the effect of error messages of the laser printer.) For those who do not know the generator, it is impossible to make a dense encoding of that picture. Clearly, it is a typical element of our set of pictures.

Thanks to chance, when I was preparing a presentation about algorithmic definition of fuzziness (see [8]) I put the transparency on the printing of the pictures. The interference of checkered colorings as shown magnified in Figure 12 produces a picture of squared symmetric darker and lighter areas. It is a nice Moiré pattern.

The self-interference of the random coloring was surprising. As Figure 13 shows, concentric circles are observable around a light center. The spectacle is more impressive while the transparency is moved and rotated over the printing. Reducing the size of the transparency, the interference lines form either radiated lines (no rotation) or spiral curves winding to the light center. See Figures 14 and 15.

The explanation of this phenomenon is in the strong regularity of randomness: in a random coloring every colored pattern of a small number of pixels should occur with relative frequency close to its probability. From this, it follows that there must be many black patterns, such that they fit the invariant curves of the transformation of the plane that corresponds to the placement of the transparency on top of the printing. These patterns tend to become longer when placed on each other along the invariant curves. We can visually observe it within some area around the fixed point of the transformation.

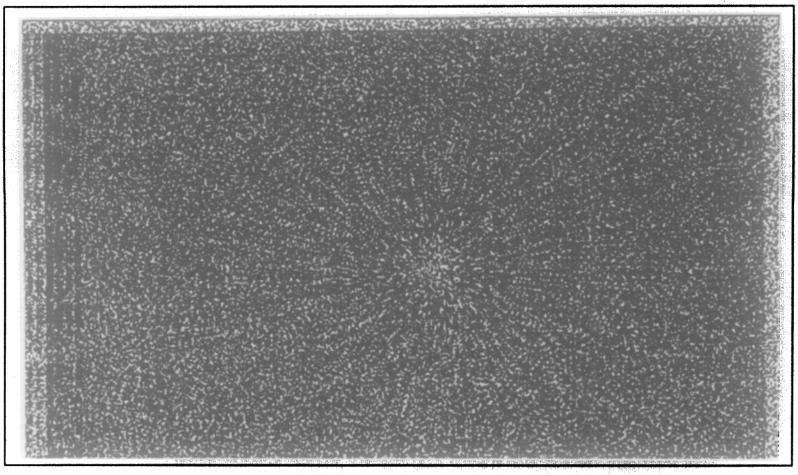


Figure 14.

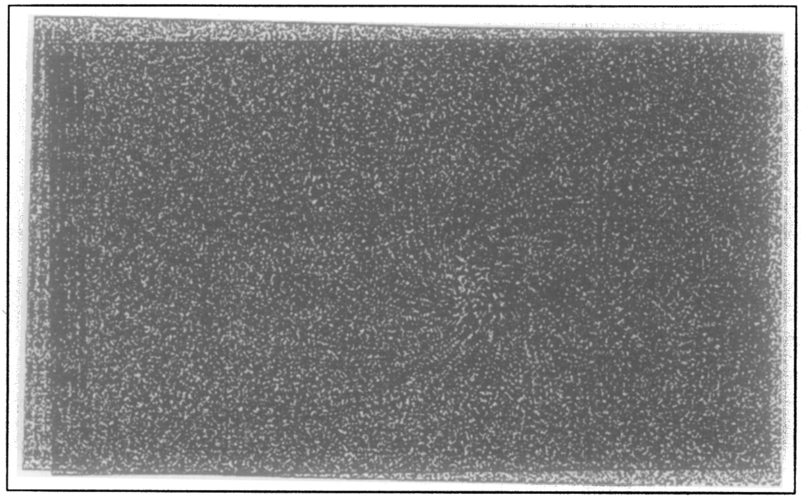


Figure 15.

Another more formal explanation is that the dispersion of the numbers of black pixels along the invariant curves are greater than along any other area when we put the transparency on the printing. Placing the transparency of a line of independently colored black and white pixels shifted by k pixels over the original line, the dispersion of the number of black pixels in n consecutive pixels is $np(2-p)(1-p)^2 + 2(n-k)p(1-p)^3$, where the black color is chosen with probability p. In the case of placing independent colorings on each other, the dispersion of the number of black pixels in the same configuration is $np(2-p)(1-p)^2$.

Kolmogorov complexity has a strong connection to inductive inference, to the general theorems of machine learning, and to computable statistical inference. The basics of this theory were developed by Solomonoff [8] and Levin [9].

The main point is the proposal for a universal *a priori* probability distribution on the set of nonnegative integers. The construction of the universal *a priori* distribution uses an analogy with Shannon entropy. The information quantity of a probability value p in the Shannon entropy is $-\log_2 p$. The information quantity of x assigned by the prefix Kolmogorov complexity is $K(x)$. Using the analogy, the universal *a priori* distribution (the Solomonoff-Levin distribution) is defined by

$$\pi_x = 2^{-K(x)}. \tag{12}$$

Since the system of code-words used by the optimal function $g_0(p)$ is prefix-free it follows from the Kraft inequality that

$$\sum_{x \in N} \pi_x < 1.$$

Unfortunately, the distribution π_x is not full and not computable. But it is semicomputable from below, and it multiplicatively dominates every distribution which is semicomputable from below. (A real valued function f is semicomputable from below iff there exists a recursive function $g(x,k)$ with rational values, nondecreasing in k, with $\lim_{k\to\infty} g(x,k) = f(x)$.)

The dominating property of π_x is the following: for any measure m_x which is semicomputable from below there exists a constant c such that $m_x < c\pi_x$ for all x. The intuitive meaning of the dominating property of π_x is that it is the possible most uniform probability distribution on the nonnegative integers. So in Bayesian inference if there is no information on *a priori* distribution on the countable set of hypotheses the best choice is the universal *a priori*. This is the generalization of Bayes's principle to infinite countable hypotheses and algorithmic (machine) inference. See detailed discussion in [10].

6. CONCLUSION

All the above mathematical models give laws of possibilities and limitations of our artificial, digitized communication technology. These laws are typically valid in limit, and as many ideal concepts of mathematics like infinity, continuity, probability, and universal algorithms, they approximate finite systems from the infinity. Now, the new achievements of information technologies have enlarged the possibility to solve very large finite problems. The effects of the above mathematical lows are getting stronger on these large problems, but they do not absolve us of searching for good approximating, compromised finite solutions. It is important to note that after messages or knowledge have been encoded to signals, they are not random at all; they were random before or during the selection or observation. What is still random in this new worldwide networked signal collection is the noise in it. Paradoxically, programs are the most significant source of noise that may corrupt the stored data. The main task, responsibility, and mission of computing professionals (or forming the term informaticians from the corresponding Hungarian term "informatikusok") are the reduction of this noise.

It was my definite intention to give a central role in my paper to randomness. Communication has its role only in random, uncertain situations. We can trust in randomness as the essential property of nature and life, and therefore, human communication will remain among human beings.

I close my paper with the words of Kolmogorov about randomness:

> *"In everyday language we call random those phenomena where we cannot find a regularity allowing us to predict precisely their results. Generally speaking, there is no ground to believe that random phenomena should possess any definite probability. Therefore, we should distinguish between randomness proper (as absence of any regularity) and stochastic randomness (which is the subject of probability theory). There emerges the problem of finding reasons for the applicability of the mathematical theory of probability to the real world."*

REFERENCES

1. P.J. Denning, Computing the profession, In *Computer Science Education in the 21st Century*, (Edited by T. Greening), pp. 27–46, Springer-Verlag, New York, (2000).
2. C.E. Shannon, A mathematical theory of communication, *Bell System Technical Journal* **27**, 379–423 and 623–656, (July and October, 1948).
3. C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL, (1949).

4. A.N. Kolmogorov, Three approaches to the quantitative definition of information, *Problems of Information Transmission* **1** (1), 1–7, (1965).
5. M. Li and P. Vitanyi, *An Introduction to Kolmogorov Complexity and Its Applications*, Second Edition, Springer-Verlag, (1997).
6. A. Gammerman and V. Vovk, Editors, *"Kolmogorov Complexity"*, Special Issue of *The Computer Journal* **42** (4), (1999).
7. A.N. Kolmogorov and V.A. Uspenskii, Algorithms and randomness, *SIAM J. Theory of Probability and Applications* **32**, 389–412, (1987).
8. A. Benczúr, An attempt to the algorithmic definition of fuzziness, *Annales Univ. Sci. Budapest.*, *Sect. Comp.*, 19–33, (1991).
9. R.J. Solomonoff, A formal theory of inductive inference I, II, *Information and Control* **7**, 1–22, 224–254, (1964).
10. L.A. Levin, Universal search problems, *Problems of Information Transmission* **9**, 265–266, (1973).
11. M. Li and M.B. Vitányi, *Kolmogorov Complexity and its Applications, Handbook of Theoretical Computer Science*, Chapter 4, (Edited by J. van Leeuwen), Elsevier Science, (1990).