to Problems of Automatic Control. Fizmatgiz, 1960.
17. Sveshnikov A. A. Applied Methods of the Theory of Random Functions. Sudpromgiz, 1961.

R. L. Dobrushin and Ya. I. Khurgin

Problems of the Information Theory

(Part of this material has been borrowed from a book prepared for publication by R. L. Dobrushin in coauthorship with L. Ye. Filippova and B. S. Tsybakov, Teoriya Optimal'nogo Codirovaniya Informatsii (The Theory of Optimum Coding of Information)).

1. The concept of "information" is utilized extensively. An exact definition of it, like any of the other basic natural scientific concepts, is difficult, if only because such concepts evolve in the course of time, but everyone understands its objective content. Abstraction, transmission, reception, storage and transformation of information represent some of the essential phenomena of life around us. Because any controlling system deals with information, information is one of the basic concepts of cybernetics.

Study of information, its properties and characteristics offers considerable difficulties chiefly because of the subjectivity of the determination of the value or usefulness of the same information for different people. Thus, for example, information about the birth of a child is received absolutely differently by the father of this child, the relatives, neighbors, friends and strangers.

The development of communication engineering (telephone, telegraph, radio, television and others), engineering designed for the transmission of information, has led to distinguishing one aspect of this concept which has proved to be well adapted for study on mathematical and technical levels. This study has produced abundant fruit at the present time. Therefore, we shall give our main attention to the theory of transmission of information and shall only briefly touch on the other aspects of this theory.

2. Attention should be directed to the fact that information is not a physical characteristic of objects. In the transmission, storage, transformation of information energy is expended, but this energy in no way characterizes the information itself. Thus, for example, information about the weather can be transmitted over the telephone, radio, telegraph, can be published in the newspaper, and some energy is everywhere expended for this transmission. However, the quantity

of energy expended for the transmission of a certain message depends only on the level of engineering and can decrease infinitely; it does not depend on the nature of the message, its meaning or its importance nor even on how probable this message is.

In contrast to the objectsstudied, for example, in physics, which do not depend on man, information in the human society is basically created by the people themselves. Therefore, in constructing the information theory it should be taken into consideration that the object of study -- information -- and the theory for studying it are interrelated; the theory created in a certain way also transforms the very object of study. This is a unique feedback between the theory and the object of study, and this feedback creates great opportunities for the development of information theory and its applications.

To date there is no precise definition of the concept "information" which might completely cover this concept and serve as the basis for the construction of the general information theory. Attempts have already been made along this line. The most serious of them should be considered the work by A. N. Kolmogorov, in which the concept of "information" has been introduced axiomatically. However, this is a very complex mathematical construct and is far from objectivity.

In the technical literature usually "information theory" means the entire totality of applications of mathematical methods to problems of description, processing, storage and transmission of information in the very broadest sense of this word. With this approach the subject of the information theory proves to be very broad, and the mathematical methods which are applied are exceedingly heterogeneous. Actually, the applied nature of the work published in this field does not introduce very much that is specific into the mathematical apparatus used therein. From the viewpoint of mathematicians the majority of these studies can be classified under the theory of probability processes, mathematical statistics, theory of games, and others.

3. It should be emphasized that the information theory is essentially of a statistical nature. Actually, we acquire new information when we do not know beforehand which messages will be received; when the expected messages may be received, but they may or may not be received, that is, when the entire situation is of a random character. As is well known, random events can be of two kinds: they can be subordinate to stable statistical regularities (like the various mass phenomena) or may not be subordinate to such rules and regulations, they may not be repeated many times under the same conditions.

Situations which are subordinate to stable statistical rules and regulations can be studied by means of experiments and subsequent statistical treatment of them. We do not know of any methods of mathematical study of random phenomena which are of a non-statistical nature.

Therefore, we are dwelling only on the aspect of the information theory which pertains to phenomena of a statistical character.

4. Phenomena subordinate to stable statistical rules and regulations are studied in the theory of probability, which is the basis of mathematical statistics. Mathematical statistics is a science of how to extract information from existing observations, from experiments. Its content is the development of methods of statistical observation and analysis of statistical data, the data of experiments. Mathematical statistics arose from the requirements of demography, economics, biology and other experimental sciences. At the present time, it is a separate field of science with a very large set of topics and rich mathematical apparatus. (Many of the basic principles of mathematical statistics and particularly the basis of it -- the theory of probability -- were created in Russia and in the Soviet Union. However, at the present time it is widely used in the West and to a lesser degree in the Soviet Union because of an insufficient understanding of the importance of mathematical methods in the fields of economics and biology by many economists and biologists and the associated neglect of methods of mathematical statistics).

What has been stated permits us to conclude that if we understand the information theory to a great extent it should include mathematical statistics.

5. However, when we speak of the information theory, frequently consciously or unconsciously we have in mind the new ideas advanced a little more than 10 years ago by the well-known American scientist, C. Shannon (Leningrad, 1). These ideas apply to the problem of the possibility or impossibility of transmission of given information over a communications channel under given specific conditions if optimal methods of coding and decoding of the message being transmitted are utilized. We shall call the scientific trend founded by Shannon the "theory of optimum coding of information" (in the broad sense of the word); it represents an integrated scientific discipline which is clearly distinguishable by its specific mathematical methods.

6. The general schema analyzed by Shannon, if we describe it without striving for mathematical accuracy of presentation, is the following:

With the creation of a system of communication it is impossible to consider it known beforehand specifically which message will have to be transmitted with the aid of this system. However, it is natural to suppose that the possible messages belong to some given class of messages, X. This class of messages can be of a very heterogeneous nature. For example, the multitude of messages transmitted by means of the telegram represents the totality of all-possible texts in the Russian language. In the radio broadcasting system the message at the input is a sound wave which is mathematically described by a certain time function. A situation in which it is necessary to transmit a certain message known beforehand (let us say, alarm signals) but where it is unknown specifically at what moment it will be transmitted can constitute another example. Here, we must take the totality of all the moments of time at which the need for transmission of this message can arise as the totality, X.

Evidently, giving the totality of messages, X, which can be transmitted far from exhausts all the properties of messages of importance to us. As a matter of fact, the problem of information transmission is facilitated considerably if we take into consideration the fact that some messages appear much more often than others. With creation of a transmission method we can orient ourselves in such a way that the transmission of the frequent messages be most rapid and uninterrupted; at the same time, we can reconcile ourselves to the difficulties arising from the transmission of the least frequent messages. Therefore, it is useful to know the "frequency" (that is, the probability) of occurrence of one possible message or another beforehand. The problem of definite calculation of probabilities in practical cases is usually complicated, but the introduction of such characteristics is inevitable since it is dictated by the nature of the problems under analysis. Therefore, in the future we shall always assume that the statistical properties of the message have been given, that is, the probability distribution of the possible messages, X, in space has been given.

In addition, certain requirements for accuracy of transmission of the message are presupposed. They are also of a statistical nature. For example, naturally an attempt is made to keep the average number of errors from exceeding a set limit in telegrams transmitted from one city to the next. From Shannon's theory it follows that in those cases where the message at the input assumes an incalculable number of values (for example, if the message at the input is a sound wave or a two-dimensional image in television transmission) it is impossible to achieve a complete correlation of messages being sent and received with transmission over any real communications channel. However, the requirement of such complete correlation is in many cases unreasonable. Let us say that it is irrational to require from a constructor of a radio broadcasting system that the accuracy of reproduction of a sound signal in the loudspeaker be greater than the possibilities of the human ear, which receives far from all the sound waves. A reasonable condition for accuracy of transmission might be, for example, the requirement that the functions describing the sound waves at the input and output of the communications system be different from each other by no more than a set constant value. To be sure, it would be better to formulate the requirement for accuracy of transmission based on the physiological characteristics of the human ear, but here we would use as a basis physiological problems which have not yet been solved.

Mathematically, requirements for accuracy are usually formulated as certain limits or compatible probability distributions of the transmitted and received messages.

7. Let us now proceed with a description of the communication channel. First of all, let us suppose that a certain totality, Y is set, the elements of which, y, we shall call signals at the input of the channel. Just as in the case of messages, the nature of the totality, Y can be most different. As the result of transmission along the communication channel the input signal y is converted into a certain signal at the output of the channel, $\tilde{y}$. In the simplest case of non-erroneous transmission of information the signal at the output, $\tilde{y}$ simply coincides with the signal at the input of the channel, y. However, in any real physical communications channels errors, or as they say in radio-engineering, noise, slips in for various reasons, which lead to the fact that the signal at the output is different from the signal given at the input. In order to describe such a channel with noise mathematically it is necessary to indicate the statistical properties of every set input signal occurring at the output with transmission of these signals or, in the more accurate language of the theory of probability, the conditional probability distribution of the signal at the output with a set signal at the input.

Therefore, the totalities Y and $\tilde{Y}$ and the conditional probability distribution over the space $\tilde{Y}$ given for each element y of the totality Y define the communication channel as a whole. Sometimes, in defining the channel certain limitations of the statistical properties of the signal at the input are included (for example, frequently the natural requirement is made that the average power of the signal at the input not exceed a given constant value).

8. In order to convert the message into a signal which is to be transmitted over the communication channel it is necessary to carry out an operation which is called "coding" of the message. It consists of assigning a certain signal to each of the possible messages at the input of the channel. When the message at the input assumes a fixed value, a signal corresponding to this value is transmitted over the channel. With the aid of the decoding operation for the signal at

the output of the channel obtained as the result of transmission a certain content of the message is restored which is called the "output message". The use of coding and decoding is inevitable if the class of messages, X is different in nature from the totality of signals at the input of the channel, Y. However, the most important consideration on behalf of the introduction of coding and decoding is the fact that the quite complex and reasonably selected methods of coding and decoding make it possible to improve essentially the quality of transmission and increase its speed. Therefore, the abstract schema of information transmission presented in Fig 1 is obtained.

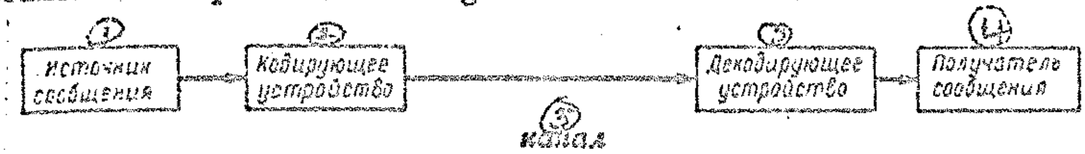

Fig 1. Block Diagram of Information Transmission System. 1. Source of message; 2. Coding apparatus; 3. Decoding apparatus; 4. Receiver of the message; 5. Channel.

Let us analyze several specific examples. In the system of telegraph communication which needs to be transmitted the message consists of text written on the telegraph forms. The statistical properties of these texts can be found experimentally. The message at the output is the telegram delivered to the addressee. The natural requirement of transmission accuracy lies in the fact that the probability of the addressee's receiving a telegram different from that transmitted is not very great. Signals at the input and output of the telegraph communication channel are a sequence of dots, dashes and spaces represented by the appropriate electrical impulses. Coding is a method which makes it possible to assign a sequence of dots, dashes and spaces to every phrase in the Russian language, while decoding is a method which describes the reverse operation. Coding by means of the Morse alphabet is widely known (incidentally, as has been noted frequently, the Morse code is not very close to being the best one).

The ordinary statistical model of the radio communications channel consists of the following. Time functions, $y(t)$ describing the strength of the corresponding electromagnetic field represent the signals at the channel input. At the channel output a signal arises, $\tilde{y}(t) = y(t) + \xi(t)$, where $\xi(t)$ is a random addition which in radioengineering is called "additive noise". The statistical properties of the channel are determined by the statistical properties of the stochastic process, $\xi(t)$

thereby, usually, a requirement for limitation of the mean power of the signal at the input is included in the definition of the channel; this is associated with the fact that the energy possibilities of any real transmitter are limited. The coding methods used for such a situation are customarily called modulation methods in radioengineering. For example, with amplitude modulation of the signal, signals of the type $y(t)=[ax(t)+1] \cos \omega_0 t$ are assigned to the input message, $x(t)$ for transmission over the channel; here, ω_0 is the frequency at which the given transmission is carried out. With phase modulation the signal $y(t)=A \cos [\omega_0 t+ax(t)]$ and others are given to the message, $x(t)$. In this case the decoding is called "rectification" of the signal.

The schema presented in Fig 1 is applicable not only to the description of information transmission but also to the description of storage of information. The situation is that the storage of information can be represented like the transmission of information but not spatially (from one place to another) but rather in time (from one moment to the next). Thereby, all-possible factors which lead to the forgetting and distortion of the information being stored correspond to noise. From this viewpoint we can consider the book a communication channel in the sense of the schema described above. The writing and printing of the book will be the coding; the reading of it, the decoding.

Complex codes, recommended by the information theory, are utilized, for example, for recording information in the memory of the machine in particularly important cases (for example, as has been reported in the American scientific press, for recording programs in the plans of human space flight). Defects in the magnetic tape, on which the record is frequently made, lead to the fact that various symbols are not reproduced during the reading. Nevertheless, the use of codes for correcting errors assures correct reproduction of the initial information. The theoretical formulation of the problem of storing information becomes of particularly great current importance in connection with the creation of specialized apparatus for prolonged storage of information.

9. Now, we can formulate the basic problem of the theory of optimum coding of information. The message with set conditions for accuracy of transmission and for the communication channel is considered unknown and fixed, but a possible variation of coding and decoding methods within broad limits is assumed. The two following questions are being studied: 1) under what conditions do methods of coding and decoding exist so that the given message can be transmitted over a given communications channel with the fulfillment of set conditions for accuracy of transmission; 2) how shall we find these methods if they exist?

The important service of Shannon lies in the fact that he found the asymptotic solution of this problem and indicated the statistical parameters of the message and communication channel on which the answer depends. These parameters are the entropy of the message under given transmission conditions and the capacity of the channel and are expressed by the characteristics of the probability distribution which they give, called "entropy" and "quantity of information".

It should be emphasized that in Shannon's theory no definitions are given for understanding "information", and this concept is not utilized; only the concept quantity of information is utilized, which is expressed in a uniform way by the probability characteristics of the messages being studied. In this theory the semantic content is of no significance; essential only is how frequently various messages appear at the communication channel input and how probable are the distortions of their carrier signals in the channel.

It has been found that in the theory of information transmission constructed in this way, which does not utilize the semantic content of messages but is based only on their probability characteristics, a reasonable measure of the quantity of information can be introduced in only one way. Specifically, if we introduce entropy axiomatically based on its simplest and most obvious properties, the mathematical expression for entropy is uniquely determinate (with an accuracy within insignificant constant values).

The first work by Shannon contributed to the occurrence of a vigorous stream of work by his followers, in which along with serious achievements associated with further development of Shannon's ideas, many uncritical and sometimes simply thoughtless attempts were made to use the concept introduced in the most varied fields of knowledge. To a considerable degree such work was based on attempts to utilize the mathematical concept of the quantity of information introduced by Shannon in all those cases where they were dealing with information in the general treatment of this term, although from the mathematical viewpoint the rationale of the magnitudes introduced by Shannon as measures of the quantity of information is justified only in the situation described above with transmission of information over the communication channel. In the justification of their conceptions many referred to Shannon's axioms for entropy, but, as has been noted by A. N. Kolmogorov (Leningrad, 5) such a complicated and varied concept as information can hardly be characterized mathematically by means of a single numerical magnitude. Shannon's axioms therefore show only that no other numbers will be better characteristics of the information. Nevertheless, the possibility has not been ruled out that entropy which satisfies these axioms can prove to be an unfortunate characteristic for describing the indefiniteness of an experiment or at least it may be successful only in certain respects. Justification of the reasonableness of Shannon's definitions for a case of transmission of information is the fact that these factors give us an answer to the basic problem posed above of the theory of optimum coding of information, but the degree to which these magnitudes are suitable for application in the analysis of other situations is so far unclear, by and large. In 1956, Shannon wrote the following: "I personally believe that many of the concepts for the information theory will prove to be useful in other fields and, certainly, some of the results already obtained appear to be very promising. However, the detection of such applications is not the trivial process of translating terms into a language for the new field but rather is a fatiguing process of advancing hypotheses and checking them experimentally" (Leningrad, 17).

Naturally, many of the premature attempts at application of Shannon's ideas showed no results, and after the wave of enthusiasm there was a certain degree of disillusionment. Simultaneously, mathematicians did and are continuing to do considerable work on the theoretical interpretation and clarification of the basic results of Shannon, demonstration of the limits of action and precise formulation of his theory. Gradually, the field of application of Shannon's ideas is becoming better known and better understood by those who take up the applications of this theory, while those who compromise it with thoughtless applications are becoming progressively fewer every year.

10. In the theory of transmission and storage of information, that is, in the field where there is no doubt of the adequacy of Shannon's ideas some disillusionment has also occurred associated with the fact that so far the ideas of Shannon's theory of coding of information have not given much for practical construction of communications systems. The optimum methods of coding and decoding recommended by Shannon's theory are very complicated, and realization of them in practice is a costly measure. In addition, the basic theorem of Shannon only guarantees their existence without giving recipes for practical realization of these methods. Then, it was shown that the methods of coding and decoding already in existence, relatively simple and well known to specialists in communications engineering, in many important cases give results which are qualitatively comparable with those which can be given by the optimum methods recommended by Shannon's theorem. In such cases the utilization of optimal methods makes it possible to increase only a little the quantity of information to be transmitted at the expense of an essential complexification of methods of coding and decoding.

From what has been stated, however, it does not follow that Shannon's theory is of little practical importance. No one doubts the fundamental importance of the law of conservation of energy, but at the same time no one gets the idea of complaining of the fact that by using this law only it is impossible to give recipes for converting one type of energy into another with a high degree of efficiency. The very fact that Shannon's theorem shows that in some situations it is impossible to transmit a given class of messages over a given communications channel with a certain low degree of error is an important practical conclusion which keeps the constructor from making fruitless attempts.

Recently, through the efforts of many specialists on information theory it has been possible, at least for some very simple situations, to point out the relatively available and effective methods of establishing optimum methods of coding and decoding of information. In connection with this, in recent years interest in the practical applications of Shannon's information theory has increased again sharply. The theoretical achievements in the field of construction of optimum methods of transmission and the progress in the engineering of computers necessary for practical realization of such methods, on the one hand, and the development of communications engineering, which utilizes progressively more complex and costly equipment for the transmission of information and the increase in requirements for the range and reliability of transmission (for example, in connection with the problems of space radio communication being posed and of current importance), on the other hand -- all this has already led to the fact that the utilization of very complex optimum methods of coding and decoding is now becoming economically and technically justified. The first attempts at actual realization of such optimum communication systems are already being made. Undoubtedly, the development of cybernetics will lead to a much broader incorporation of new methods, based on the results of the theory of optimum coding of information.

11. It is interesting to note some fruitful attempts to utilize the information theory in a narrow sense (that is, the theory of optimum coding) in psychology. With the aid of simple and clever experiments it has been determined, for example, that the mechanism of some psychological reactions (of the selection reaction type) is described by the same numerical characteristics as the optimum processes in communications systems. Specifically, the reaction time to the simplest stimuli (light, sound) is a linear function of the quantity of information which is carried in a message unknown beforehand. On the basis of these specific experiments one can hardly draw a general conclusion as to the optimum structure (from the viewpoint of the theory of optimum coding) of all the reactions of the living organism, but these

experiments indicate a new approach to the study of the body's reactions. These problems are not only of cognitive but also of purely applied significance. Specifically, now considerable attention is being given to the problem of participation of man in a cybernetic system (man-automat system), and thereby the knowledge of many characteristics of the human body as components in this cybernetic system is of first importance -- without this knowledge it would be impossible optimally to construct the entire system.

12. Let us dwell, in conclusion, on some of the specific problems of current importance in the information theory (in the narrow sense). a) In the fundamental work by C. Shannon (1948) (Leningrad, 1) almost all the ideas of this theory which have been worked out at present are contained in a relatively complete form. However, from the mathematical viewpoint presentation of them has been inadequately satisfactory. The haziness of certain definitions and the lack of distinctness of the proofs have led to perplexity and even to misunderstandings and erroneous conclusions.

The problem of the clear-cut mathematical substantiation of Shannon's theory has been very difficult, but at the present time this problem may be considered basically solved. The main work on this substantiation was done in Moscow (Leningrad, 2, 3, 4, 5, and 6).

b) The definition of such basic characteristics of the various communication channels as the speed of transmission and capacity represent an important task. In the simplest cases, for gaussian channels and additive noise these characteristics have been calculated. However, in more complicated cases, for example, for channels with non-additive noise or for channels with random and variable parameters, calculation of the transmission rate and capacity give rise to considerable difficulty. Definite progress along this line has already been made, but there is still a very large number of various unsolved problems (Leningrad, 7, 8, 9).

c) We have already mentioned the fact that Shannon's basic results indicate, under certain conditions, only the existence of an optimum code in a certain asymptotic sense, with the utilization of which transmission is possible with as low a frequency of error as is suitable as well as the fact that in some cases it has been possible to find adequately active and physically accomplishable methods of coding and decoding. However, the tremendous variety of communications channels utilized and conceivable requires considerable further work along the line of constructing specific mathematical systems and actual engineering systems of coding and decoding which are sufficiently similar to the optimal (Leningrad, 10, 11, 12, 13, 14).

d) The study of methods of coding and decoding in information transmission channels with feedback is particularly important for cybernetics. Along this line so far very little has been done; only the simplest cases have been studied (Leningrad, 15, 16).

The corresponding problems require at times not only solutions but also a new approach and a clear-cut mathematical formulation. Apparently, considerable attention will have to be given to this group of problems in the immediate future.

Bibliography

1. C. E. Shannon, Weaver W. The Mathematical Theory of Communications. The University of Illinois Press, 1949; for an incomplete translation see the collection: Theory of Transmission of Electrical Signals in the Presence of Noise. Foreign Literature Publishing House, 1953.
2. A. Ya. Khinchin. The Concept of Entropy in the Theory of Probability. Uspekhi Matematicheskikh Nauk (Progress of Mathematical Sciences), 1953, Vol VIII, No 3, (55) page 3.
3. A. Ya. Khinchin. The Basic Theorems of the Information Theory. Uspekhi Matematicheskikh Nauk, XI, No 1 (67), 1956, page 17.
4. A. Finestein. Fundamentals of the Information Theory. Translation from the English. Foreign Literature Publishing House, Moscow, 1960.
5. A. N. Kolmogorov. Theory of Information Transmission. Materials of the Session on Scientific Problems of Automation of Production, Plenary Meeting. Publishing House of the Academy of Sciences USSR, Moscow, 1956.
6. R. L. Dobrushin. General Formulation of the Fundamental Shannon Theorem in the Information Theory. Uspekhi Matematicheskikh Nauk, 1959, XIV, No 6 (90), page 3.
7. I. M. Gel'fand, A. M. Yaglom. Calculation of the Quantity of Information about a Random Function Contained in Another Such Function. Uspekhi Matematicheskikh Nauk, 1957, XII, No 1, page 3.
8. R. L. Dobrushin, Ya. I. Khurgin, B. S. Tsybakov. Approximate Calculation of the Capacity of Radio Channels with Random Parameters. Works of the All-Union Conference on the Theory of Probability and Mathematical Statistics, 1958. Yerevan'. Publishing House of the Academy of Sciences of the ArmSSR, 1960, page 164.
9. Chang Sze-Hou. Capacity of a Certain Asymmetrical Channel with Finite Memory. IRE Trans., Information Theory,

1958, Vol IT-4, No 4, page 152.
10. Peterson W. W. Error-Correcting Codes. MIT Press, New York-London, 1961.
11. Wozencraft J. M., Reiffeu B. Sequential Decoding. MIT Press, New York-London, 1961.
12. R. R. Varshamov. Evaluation of the Number of Signals in Codes with Correction of Errors. DAN SSSR (Reports of the Academy of Sciences USSR), 1957, 117, No 5.
13. L. F. Borodin. Equidistant and Other Codes Optimal and Close to Optimal. Radiotekhnika i Elektronika (Radioengineering and Electronics), 1960, No 6.
14. P. Elias. Coding for Two Channels with Noise, in the Collection Theory of Transmission of Messages. Foreign Literature Publishing House, 1957.
15. C. E. Shannon. Channels with Side Information at the Transmitter. IBM J. Research and Development, 1958, Vol 2, No 4, page 289.
16. R. L. Dobrushin. Transmission of Information over a Channel with Feedback. Theory of Probability and Its Application, 1958, 3, No 4, page 495.
17. C. E. Shannon. IRE Trans., Information Theory, 1956, Vol IT-2, No 1.

K. B. Arutyunov and D. V. Svecharnik

Sampling, Initial Processing, Storage and Transformation of Information about the Course of Production Processes

Introduction

The life of living organisms and the life of man proceeds in a continuous interaction with the environment, adaptation of the organisms to it, in protection against the dangerous effects of the environment and, at the highest level of consciously acting organisms, in actively influencing the environment, transforming it in the necessary direction.

In the world in which the organisms live a multitude of disturbances (external influences) arises which threaten to take the values of the variables essential for normal functioning and life of the organism (for example, the moisture content of the body, the inner temperature of a warm-blooded animal, the glucose content in the blood, acidity and others) beyond certain permissible limits.

Many deviations are incompatible with life. Therefore,