On Cybernetic Epistemology

Valentin Turchin

Address – *The City College, the City University of New York*

Introduction

This paper is an outline of what can be called *cybernetic epistemology* developed in the course of the work on the *Principia Cybernetica* project[2,6,7]. As a continuation of this work, a paper on cybernetic metaphysics and the semantics of basic concepts of science is being prepared for publication. The motivation, naturally, includes the wish for an up-to-date philosophical system, but also an attempt to create a basis for the semantics of natural language to be used in the work on artificial intelligence. It is the author's belief that to succeed in transferring to the machine our ability to understand natural language it is necessary to break down the meaning of the language into some elementary units – not just small, but really minimal. There is work being done in this direction, e.g.[3]. In the *Principia* project we attempt to go further in the minimization of the number of primitive concepts. We base our efforts on the concepts of *prediction* in epistemology, and *action* in metaphysics, used as atoms into which to decompose our perception of complex phenomena.

Philosophy is called to answer such basic for every intelligent creature, questions as: Who am I? Where do I come from? Where am I going? How true is my knowledge? What is, in the last analysis, the nature of things? What is good and what is evil?

Each time gives its own answers to these questions. These answers are strongly influenced by the current state of knowledge and production. Our philosophy is a product of the emergence of evolutionary theory at the end of the 19th century and cybernetics in the 20th century. This can be seen both in the method with which we approach philosophical problems, and in the answers we give.

We understand cybernetics in a very wide sense proposed by Norbert Wiener as the science of control, communication and organization. Seeing the world in these terms is, to some extent, complementary to seeing it through the eyes of a physicist, for whom the world is space, time, and matter. Physics of the 19th century led one naturally to the philosophy of mechanisticism. The world was believed to be "really" a three-dimensional space with time somehow flowing uniformly everywhere and small particles moving along definite trajectories which are determined by the laws of mechanics and the forces acting between the particles. In this world there was no place for indeterminacy and freedom.

Physics of the 20th century dismissed this picture of the world as a naive projection of our theories into reality. It had to, under the pressure of experimental facts. We now regard the world as much more mysterious than it was for the scientists of the 19th century. We refrain from identification of our theories with reality. We see theories as no more than certain ways of organizing and partially predicting the flow of our perception. We have found that indeterminacy is in the nature of things. With the advent of cybernetics and computers, we started creating models of human perception and systematically exploring various ways of the organization of sensory data. Thus the emphasis shifted from matter to organization, from hardware to software. We now regard organization as more fundamental and primary aspect of things than their material content, which, after all, is nothing else but one of the notions we use to organize sensory data. Therefore, we call our basic philosophy cybernetic.

We retain in this paper the format adopted in the Principia Cybernetica project where the text is broken into a hierarchical structure of *nodes*, roughly according to the rule: one concept, one node. Nodes are opened by their names in boldface and closed by ■, as in the use of parentheses. When after the end of a node we return to the parent node, an occasional *back-to* headline helps the reader to know where he is. ■

Progressive formalization

Our method of developing philosophy is a progressive gradual formalization. This is the method universally used in science. We first rely on an intuitive understanding of simple concepts, then on the basis of this understanding we convey the meaning of more formal and exact, but also more complex, concepts and ideas.

This statement itself is an illustration of our method. We used in it the words "understanding", "meaning" and "formal". In due course, we shall analyze these notions and give to them, in their turn, "more formal and exact" meanings. These new meanings, however, will not come to *replace* the original meanings, but *make an addition* to them.

Compare this with the situation in physics. We start this branch of science speaking about bodies and their masses, measuring distances in space by applying rulers, etc. Later, when we study the structure of matter, we find that those bodies and rulers, are nothing else but certain structures consisting of huge numbers of atoms. This concept of a ruler is, however, a new concept, even though it refers to the same thing. To come to the concept of a ruler as an atomic structure, we must pass a long path, at the beginning of which a ruler is a simple thing the usage of which is easy to explain.

We come to philosophy with the standards and methods of science. We try to define and explain such basic things as "meaning", "understanding", "knowledge", "truth", "object", "process" etc. But to explain, e.g., understanding, we must rely on understanding in its usual intuitive sense, because otherwise we will not know if we ourselves understand what we are saying; so, there will be little chance for our words to be meaningful.

Or take the concept of an object. We have a *node* devoted to it. But we cannot do without speaking about objects long before we come to that node – in a close analogy with the two concepts of a ruler in physics.

Relations between things in this world are very often circular, so we are often at a loss when trying to start and finish definitions. Using various levels of formalization allows us to avoid vicious circles in definitions. Suppose we define concept A using informally concepts B and C. Then we want to make B more exact. We define it, and find that it depends on new concepts D and E, but also on the already defined A. Hence if we were to postpone the formalization of a concept until the time when all the concepts on which it depends are formally defined, we would never finish the job, or limit the scope of the work to a strictly hierarchical subset of the total set of concepts we actually need. Instead, we recognize that there are various levels of formalization of essentially the same concept, and we allow them to coexist. Thus after defining A with the use of B, we define B', which is a formal version of B; the old, less formal, concept B is not discarded. Now we may or may not redefine A; if we do redefine it using B', then we have the concept A'. Whenever we want to understand a definition, we start unwinding the chain of dependent definitions until we come to basic intuitive notions about which there should be no disagreements. The purpose of physical theory is to give us a picture of the world different from our common perception, in order to learn more about things and to create new things which dramatically change our life. Similarly, we create formal versions of our common notions in order to understand better how our language and mind work, and to create artificial languages and minds, which will imitate our mental processes, and one day, probably, go beyond what is possible for us. ■

Back to **Introduction**

Traditionally, philosophy is divided into three parts, and we have the corresponding three nodes: Epistemology, Metaphysics, and Ethics. Epistemology answers the question: What is knowledge and its relation to reality? Metaphysics struggles with the problem: What is the ultimate nature of things? Ethics gives answers to the questions: What is good and what is evil? What is it that we ought to do, or to strive for?
We deal with these problems in the Principia Cybernetic project. Also included are foundations of logic, mathematics, natural sciences and sociology. Evolution is included as a separate node, and it precedes ethics, because our ethics is strongly based on the concept of evolution. ■

Epistemology

Before answering a question, the question must be understood; after receiving an answer, the answer must be understood. Philosophy too often provokes the response: "I do not understand. For me this is just words devoid of any meaning". We make efforts to avoid this situation. As the first stages of the development of our philosophy, we establish a certain criterion to decide whether a given text is meaningful or meaningless. Specifically, we associate meaning with knowledge: the meaning of a statement is understood if the statements can add to our knowledge (no matter whether this addition may be true or false). We, therefore, begin our philosophy with epistemology, to discuss what is knowledge.

Knowledge

In cybernetics we say that a purposive cybernetic system S has some *knowledge* if it has a *model* of some part of the world in which S finds itself. Informally, a model of a process is another process which somehow mimics, or simulates, the primary process; thus by using a model it becomes possible to know something about the modeled process without actually running it, or predict developments before they actually happen.

Model

The formal definition of a model follows. We start with a definition as it is most commonly used, and then give it a more general form.

The common notion of modeling, further referred to as *the modeling scheme* (see Fig.1) is as follows.

Figure 1 *The modeling Scheme*
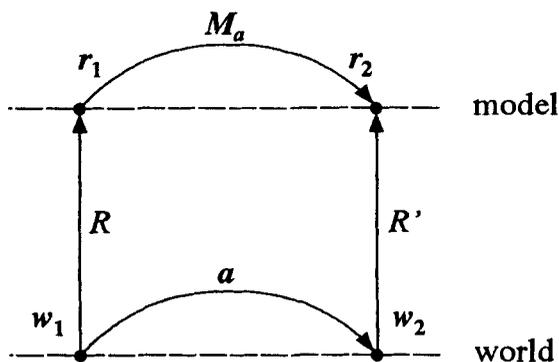
A cybernetic system S has a *model* of (some part of) the world if the following conditions take place.

1 S has a subsystem which we shall call its *brain* and can execute a procedure R called *representation function*, such that if performed when the state of the world is w it causes a certain state r of the brain; r is referred to as the *representation* of w. (Note that in higher animals what we here call brain corresponds to sense organs and the brain proper). Representation functions are also referred to as *abstractions*.

2 Let the world be in a state w_1, and suppose that system S makes an action a. We shall denote as $w_2 = F_a(w_1)$ the ensuing state of the world. We do not assume here that F_a is a function in the usual sense which uniquely determines w_2 when w_1 is given. We do not assume that the world is deterministic. $F_a(w_1)$ denotes any of the possible states of the world after the action a. We can think of it as a non-deterministic function. Mathematically, it is a relation between w_1 and w_2. In addition to representation function R, the system S must be able to perform another procedure, let us denote it as M_a and call *the modeling procedure* (the

term *simulation* procedure is also used), which mimics function F_a, so that when applied to r_1 it produces the representation r_2 of the state w_2 resulting from the action a:

$$r_2 = M_a(r_1) = R(w_2)$$

or, expanding denotations:

$$r_2 = M_a(R(w)) = R(F_a(w))$$

Thus by applying M_a to r_1 system S can predict, to some extent, the development of events in the world resulting from its action a.Trying, in the model, various actions a, the system can make an informed selection of the action which is the best for its purposes.(We assume here, of course, that the processes of representation and modeling end before the action a is made, or at least completed).

This definition describes equally well the case of a dog catching in flight a piece of sugar, and an astronomer who computes the position of a planet in the sky. In the first case, the model is built in the material of nervous cells of the dog's brain, in the second case from the signs that the astronomer writes on paper when he makes computations.

Mathematically, a model is defined by the representation function $R(w)$ and the family of modeling functions $M_a(r)$, for possible actions a of the system. When we fix an action a, we have what is known as a *homomorphism* from the relation F_a to M_a. Thus a model is a family of homomorphisms.

Some remarks on our definition of a model.

- As a special case, the action a may be an empty action – just doing nothing, waiting, or observing.
- What we call the world is usually a very small part of it.
- The representation function used on the starting state w_1 must not necessarily be the same as the function used on the end state w_2. We shall denote the former as $R(w)$ and the latter, when necessary, as $R'(w)$.

Very often a model of some process yields predictions on a continued basis; for instance, we predict weather hours and days in advance (however imperfectly). Thus we have a row w_1, w_2, w_3,... etc. of the states of the world, and the corresponding row of the states of the model: r_1, r_2, r_3,... etc. This extended model is a device which implements one modeling scheme after another: the prediction of r_2 for w_2 on the basis of r_1, then r_3 for w_3 on the basis of r_2, etc.

So, using the term *prediction* informally, we can say that a model is a *generator of predictions*. We shall give a formal definition of prediction, thereby formalizing the generalized definition of model as a generator of predictions, but first we want to show that this generalization is necessary in order to overcome a certain limitation of the modeling scheme.

The definition of the modeling scheme includes an important assumption of *non-interference*: that the procedure R (which is a sequence of actions), when it acts on a state w of the world, *does not change* that state. But we know from quantum mechanics that this assumption, strictly speaking, never holds. When we deal with macroscopic phenomena we can, nevertheless, use the non-interference assumption, because the impact of the instruments we use in R on the state w can usually be

made insignificant. But on the atomic and subatomic levels, we cannot abstract from the impact made by the instruments we use to get information about the objects studied. This may also be true in psychology and sociology, e.g. when a person makes a prediction about the behavior of a group to which he or she belongs.

There is one more non-interference assumption: the modeling procedure M_a imitating the action a does not alter in any way the course of the action, including the final result w_2. Usually, this assumption is easily true; the most salient case is that of the astronomer. His computations, certainly, do not change the motion of celestial bodies. But when we try to apply the modeling scheme to self-description, we may have problems.

So, a generalization of the modeling scheme is necessary, which would be applicable even in situations where the assumption of non-interference may not be true. Consider the following two, two-staged processes which we see in the modeling scheme, and let them run in parallel, so that mutual influence is not excluded:

The first process:

1 $R(w_1)$ results in r_1;
2 $M_a(r_1)$ results in r_2.

The second process:

1 a results in w_2;
2 $R'(w_2)$ results in r_2'.

The prediction produced by the model is that $r_2 = r_2'$. We can go further, however, in the formalization of the concept of prediction and making it more universal. Wait for the time when both processes end and start the third process C which compares r_2 and r_2'. All three processes together constitute one combined process, which is nothing else but the process of testing the model. If our model is correct, then the process C, and with it the whole combined process that tests the model, will succeed, that is r_2 and r_2' will be found identical. The model *predicts* that its test will succeed. Therefore we introduce the formal concept of a *prediction* ■

Prediction

We define a *prediction* as the statement that a certain process, referred to as a *test* comes to a *successful end*, i.e. to a certain, defined in advance, stage after which we simply do not care what happens to the process. The prediction that a test T will come to a successful end is further denoted as $T!$.

To make the concept of a test more precise, we associate with each test T a *termination subtest* $E(t_i)$, which is applicable to every state t_i of T and always ends, resulting in one of the following three symbols: *Go-on* if t_i is not yet final, so that T is going on; *Succeed* if t_i is the final state of T and T succeeds; *Fail* if t_i is the final state of T and T fails.

The statement that T ends in failure will be denoted as $\bar{T}$. It is also a prediction. It states the successful end of the following process: run T; at each stage apply the termination subtest E; stop and declare success when the result of E is *Fail*.

We define formally a generalized model as anything that produces one or more predictions. When we speak of producing predictions, we have in mind, of course, some objects that represent predictions, e.g. texts in a certain language which

enable us to reproduce the process that the prediction is about. The objects representing processes are referred to as their *objectifications*.

Formally, we can fit our general concept of prediction into the frame of the modeling scheme, if we even further extend the range of possible actions a, namely, allow a to be an arbitrary process which may include both actions of the system S which carries the model, and any other actions and processes. Let the *brain* of S be found always in one of only two states, and let them have the names *Succeed* and *Fail*. The representation function $R'_a(w_2)$ will result in *Succeed* if w_2 is the end state of the process a which succeeded, and *Fail* otherwise. The modeling function $M(r)$ will be universal and very simple: it immediately produces the state *Succeed*. Now the model we built makes exactly one prediction: that the process a ends in success.

∎

Back to Knowledge

However, even the definition of knowledge as a generator of predictions, is not general enough to cover the whole concept. Pieces of our knowledge (statements, or propositions) may not necessarily produce verifiable predictions, but only some objects that will be used to produce prediction. Moreover, they may produce objects which produce objects which in turn produce predictions, and so forth to any depth of the hierarchy of knowledge objectifications.

A simple example from mathematics: the equation $x + y = y + x$ is not immediately verifiable, but it produces – this is a shortcut for 'we know how to use it to produce' – such equations as $7 + 4 = 4 + 7$. This object, in its turn, is still too abstract for a direct verification. We can, however, verify the prediction that four apples and seven apples can be added in either order with the same result.

We come, therefore, to the following definition: a piece of knowledge is a generator which we can use to produce (generate) predictions or other pieces of knowledge. This definition is *recursive*: it defines a piece of knowledge through a basic case (prediction) and itself. The result is a kind of *hierarchy* of objects, which, unlike the more familiar types of hierarchy, unfolds from the top downwards. We start running such a generator, and it produces some objects. If an object is a prediction, we take it as one of the ultimate results of the process. If it is a generator, we run it further in parallel with the running of all the generators produced up to the moment (they may include the object's parent, grand-parent, etc.). We shall say that the predictions obtained in this way are *hierarchically produced* by the original generator. Thus we can define a (piece of) knowledge as a hierarchical generator of predictions.

In the simplest case a generator may produce exactly one prediction. Thus a prediction can be considered as a special case of a hierarchical generator. Also, the possibility is not excluded that an object legitimately regarded as a piece of knowledge will infinitely produce other pieces without ever producing a single prediction: call this piece of knowledge empty. This is because with our very general definition of recursive production we cannot always say in advance if a given piece will ever produce predictions. So we formally allow the existence of an empty piece of knowledge. Of course, if we can prove that a given piece of knowledge is empty, we need not regard it as knowledge at all.

A cybernetic system makes predictions in order to achieve certain goals, ultimately – survival and proliferation. True knowledge is an instrument of survival. Knowledge is power. ■

Back to Epistemology

When we compare our cybernetic epistemology with other philosophies, two features present themselves immediately: the dynamic character of knowledge as compared with the static concept of the classical reflection-correspondence theory; and a new look at the concept of the subject of knowledge, or the concept of 'I'

Reflection-correspondence theory

Compare our cybernetic epistemology with the classical reflection-correspondence theory of meaning and truth. One of the oldest questions of philosophy is: What is the meaning of words and phrases of a language? The naive answer is: those things which the words denote. This is known as the reflection theory of language. Language, like a mirror, creates certain images, reflections of the things around us. With the reflection theory of language we come to what is known as the correspondence theory of truth: a proposition is true if the relations between the images of things correspond to the relations between the things themselves. Falsehood is a wrong, distorted reflection. In particular, to create images which correspond to no real thing in the world is to be in error.

With this concept of meaning and truth, any expression of our language which cannot be immediately interpreted in terms of observable facts, is meaningless and misleading. This viewpoint in its extreme form, according to which all unobservables must be banned from science, was developed by the early nineteenth-century positivism (Auguste Comte). Such a view, however, is unacceptable for science. Even force in Newton's mechanics becomes suspect in this philosophy, because we can neither see nor touch it; we only conclude that it exists by observing the movements of material bodies. Electromagnetic field has still less of reality. And the situation with the wave function in quantum mechanics is simply disastrous.

The history of the Western philosophy is, to a considerable extent, the history of a struggle against the reflection-correspondence theory. We now consider language as a material to create models of reality. Language is a system which works as a whole, and should be evaluated as a whole. The job the language does is organization of our experience, which includes, in particular, some verifiable predictions about future events and the results of our actions. For a language to be good at this job, it is not necessary that every specific part of it should be put in a direct and simple correspondence with the observable reality.

Unlike our dynamic concept of modeling as production of predictions, the classical concept of reflection is static. It immediately raises questions like what does it actually mean that one thing "reflects" another? Also, how do we know that reflection takes place? To confuse things further, a distinction between mind and matter was made, which produced the question: how can our ideas, belonging to mind, reflect objects belonging to the realm of matter?

The cybernetic understanding of knowledge is much more precise. This precision is achieved by introducing dynamics into the picture. The mapping from the world to language present in the modeling scheme is not required to be a "reflection"; we need not compare these two strata of reality. To see that the model works, we only have to compare things from the stratum of language.

All that has been said about language can be applied also to human thought. In cybernetic view, thought works because it implements some models of the world, not because it somehow statically reflects it. The difficult questions of the correspondence between thought and its objects simply do not arise. ∎

Subject of knowledge: 'I'

Knowledge is both objective and subjective because it results from the interaction of the subject (the cybernetic system) and the object (its environment). Knowledge about an object is always relative: it exists only as a part of a certain subject. The subject of knowledge makes use of it in making decisions. Thus we shall also refer to it as the *user* of knowledge.

We can study the relation between knowledge and reality: first of all, whether a piece of knowledge is true or false. For that purpose the subject of knowledge, a cybernetic system S, must become part of a metasystem S' in whose terms the knowledge of S and its relation to reality can be expressed. But this only makes the knowledge relative to the system S'. This kind of metasystem transition may be repeated many times, resulting, possibly, in ever deepening knowledge, but at any time the knowledge will be still relative to a definite cybernetic system, for which it is a model used for obtaining predictions: nothing more, nothing less. The cybernetic concept of knowledge features an organic unity of the object and the subject of knowledge.

Some philosophers of the 19th century started reasoning with the distinction between 'I' and 'not-I'as the most primordial and self-evident fact of being. This sounded metaphysical and unnecessary to contemporary scientists, because they thought about the world as existing independently of my 'I'; they also thought about their descriptions of the world's phenomena as completely objective. My 'I' was simply part of the world, as well as your 'I', his 'I', and all other objects, animate and inanimate.

The development of physics, however, demanded a careful attention to the process of obtaining our knowledge. The 'I' crept into science as the observer of Einstein's relativity theory and the experimenter of quantum mechanics. We often speak of a reference system instead of the observer, and a measuring device instead of the experimenter, but this is a futile attempt to ban 'I'. A reference system and a measuring device are only ways to describe the activity of 'I'.

With the cybernetic notion of knowledge and meaning, 'I' is firmly built into knowledge. The following metaphor may help understand it. Knowledge is a tool. Each tool has a handle which you grasp when you use it. Different people (or even creatures from other planets) may use the same tool, but whenever it is used, somebody grasps the handle, and this is the 'I' of knowledge, meaning and truth. Semantic analysis discovers 'I' in some of our most important notions. Take causality. What do we mean by saying that A causes B? We cannot observe this

relation between phenomena. We can observe only a time sequence '*B* after *A*', but this may be two consequences of a common cause, or just a coincidence. The real meaning of the statement of causality includes my freedom to allow or not to allow *A* to happen, to choose between *A* and 'not *A*'. Then the meaning of '*A* causes B' is: if I allow *A* to happen, *B* will take place too, but if I do not allow *A*, *B* will not take place. This is a meaningful proposition, a model of reality; but this model inextricably includes 'I'. To test causality, there must be a freedom to allow or not to allow *A*. 'I' is the only source of this freedom. ∎

Intuition

We often use the words "intuition", "intuitively", meaning some act of our thought which cannot be decomposed into any constituting elements and, in particular, cannot be further explained or justified. This is the result of activating some non-verbal model of the world implemented in our brain. Sometimes we are very certain about what our intuition tells us; on other occasions we find ourselves in the state of *doubt*.

Doubt

Doubt is a result of having in the mind more than one model applicable to the current situation. Some of our brain models are built-in from birth, or develop in a short time after birth. We can hypothesize that such models tend to be applicable in a unique way, so they do not invoke doubt. Other models develop in the process of life and are determined by our experience. It may well be that at different times two or more non-verbal models were developed which can be applied in the current situation. We perceive this as doubt. This meaning of the word "doubt" is reflected in its etymology. The English "doubt" is, apparently, descending from "double", meaning a double opinion. In the Russian language the word for doubt is "so-mnenie", which breaks down as "so-mnenie", literally *co-thought*, a parallel thought. ∎ ∎

Language

A *language* is a system which, if properly controlled, can produce objects called *messages*.

Message

A *message* is an object which has some *meaning* (see below) for some agent. ∎

Syntax

By the *syntax* of a language we mean a method of distinguishing between objects that are *messages* of this language and those that are not. This method may be informal, when it requires a human agent to perform, or formal, when it can be performed by a machine.

Parsing

Often the syntax of a language is defined by a non-deterministic procedure G which produces all legitimate messages. Then the procedure which for any given object decides whether it could be produced by G, and if it could then by which steps of G, is known as *parsing*. ■

Back to Syntax

Various languages have widely different rules of syntax, which for each specific language, if it is formal, can usually be readily described. So there is no point in discussing syntax in general. We only want to mention two syntax units which are almost universally present in languages, even if called by different names: *symbols* and *sentences*.

Symbol

A *symbol* is the maximal unit (subsystem) of a language which can never be broken into parts. For instance, with natural human languages the morphemes (such as word roots, prefixes, prepositions etc.) play the role of symbols. In programming languages the symbols are usually referred to as *tokens*. ■

Sentence

A *sentence* is the minimal unit of a language which can have a meaning of its own, in separation from the rest of the message. Very often a message of a language is a sequence or a tree of sentences. Sentences of natural human languages are, with some reservations (e.g. concerning the use of pronouns) *sentences* in our formal sense. In programming languages sentences are usually called *statements*; this is contrary to the wide use of the term *statement* as denoting some description of a situation, like a *proposition* in logic. It is this wider usage that we catch in our formalization of *statement*. Statements of programming languages are, in our terminology, *commands* (see below). ■ ■

Semantics

Our definition of knowledge allows us to further define meaning and truth.

Meaning

By the definition of a message, it must have a *meaning*. But what is meaning? First of all, messages are always addressed to a certain agent. which in a particular case may be the sender of the message itself. Also, they are supposed to influence the actions performed by this agent, otherwise there would be no sense in producing and passing the message.

We know of exactly two ways of changing the actions of an agent. One way is to change the state of the world as far as the agent is concerned. The other way is to change the agents model of the world (if any). Thus by the type of their semantics, messages of a language break into two categories: *commands* and *statements*.

Command

A *command*, or an *imperative message*, is a message which directly changes the state of the world as far as the receiver of the message is concerned, i.e. the set of possible actions for it. By changing the state of the world the sender of a command can prevent the receiver from doing something or open a possibility of doing something new. The meaning of a command is the way it changes the state of receiver. In the classicall imiting case a command leaves the receiver no choice at all, so that the action of the receiver is uniquely determined by the command. Computer languages can serve as an illustration of this kind of semantics. Another example is the control of a rocket from the earth. It involves a certain language of electromagnetic signals sent to the rocket. Here the command leaves almost no choice for the rocket, *almost*, because there is no absolute precision of the implementation of continuous quantities. This situation is typical; the degree of *almost* is, of course, varying from case to case.

Statement

For primitive cybernetic systems which have no model of the world, the command language is the only language that can be understood. If a system has a model of the world, as in the case of a human being or a robot, the message received may not restrict the actions of the system directly, but change the model of the world used by it. This will, obviously, influence its actions in an indirect way, and, possibly, not immediately. We call such messages *statements*. The content and the meaning of a statement is in the *knowledge* it passes to the receiver. It must be noted that when we simply say "knowledge" we mean only its impact on the generation of predictions and do not distinguish between true and false (erroneous) knowledge. Thus to be meaningful, a sentence must meet the same requirement as a piece of knowledge: we must know how to be able to produce predictions from it, or produce tools which will produce predictions, or produce tools to produce such tools, etc. If we can characterize the path from the statement to predictions in exact terms, the meaning of the statement is exact. If we visualize this path only vaguely, the meaning is vague. If we can see no path from a statement to predictions, this statement is meaningless. ■

Instrumental meaning

Our definition of meaning as potential knowledge assigns meaning only to messages, but not to their parts. Yet parts also have some meaning; we do not say that the word 'table' is meaningless. It has a meaning, but of its own kind, which we shall designate as *instrumental*, because it is used in assigning meaning to such sentences as 'This is John's table', or 'My table is is made of metal'. The preposition 'in' also has an instrumental meaning: it is used to construct such sentences as 'There is an apple in this box'. Usually it is clear from the context whether we speak of meaning proper (full meaning), or of instrumental meaning. One exception is our definition of a message as the minimal unit of the language that has a definite meaning. We must make it clear that we have in mind full, not instrumental, meaning.

If we understand the instrumental meaning of a word, we can describe how precisely it is used in models. Since to describe models we have to describe processes which they involve, a description of instrumental meaning will, probably, also involve the description of certain processes. For example, the meaning of such a noun as 'table' is in the abstraction process which recognizes an object as a table. This process must necessarily be used in the meaning of any sentence including the word 'table'. ∎

Concept

A *concept* is any part of language which has an instrumental meaning. In human language we distinguish *neuronal* and *logical* concepts (see *Human language* below). ∎ ∎

Abstraction

The vertical lines in the modeling scheme are functions, or mappings R from the states of the world w_i to their representations r_i (the states of the brain of the subject system S). This mapping is always an *abstraction*: some aspects of the state of the world are necessarily ignored, abstracted from, in the jump from reality to its description. The role of abstraction is very important: it reduces the amount of information the system S has to process before making decisions.

We often use the same word to denote both a process and its result. Thus all representations r_i resulting from mapping are also referred to as abstractions. We may say, for instance, that *triangle* is an abstraction from all specific triangles. The term *abstract concept* or simply *concept* is often used, and will be used here, as a synonym of abstraction. It should be kept in mind, however, that abstraction, or an abstract concept, is not just a specific representation (such as a linguistic object), but the procedure which defines what is ignored and what is not ignored in the mapping. Obviously, the object chosen to carry the result of abstraction is more or less arbitrary; the essence of the concept is in the procedure that transforms w_i into r_i.

Given a representation function R and a specific representation r, we can construct a specialized function which will recognize the fact that a given state of the world w belongs to the abstraction r. We define this function as follows:

$$P_r(w) = \begin{cases} True & \text{if } R(w) = w \\ False & \text{if } R(w) \neq w \end{cases}$$

Here *True* and *False* are two symbols chosen to represent two possible results of the function. Such functions as P_r are known as *predicates*. The predicate $P_r(w)$ takes the value *True* for those and only those states w that are represented by r.

Whenever we speak of a predicate, we can also speak of a set, namely, the set of all states of the world w for which the predicate is true (takes the symbol *True* as its value). We shall call this set *the scope* of the abstract concept. Sometimes it is convenient to define an abstraction through its scope, e.g., a member of John Smith's family can be defined by simply listing the qualifying persons. But the universal way is the definition by property, in which case a reformulation through a set does not clarify the concept. If we "define" abstract triangle as the set of all specific triangles, it still remains to define what a specific triangle is. The real definition of an (abstract) triangle is a predicate, procedure which can distinguish a triangle from everything else.

A cybernetic system may be interested, depending on its current purposes, in different parts, or aspects, of reality. Breaking the single all-inclusive state of the world w_i into parts and aspects is one of the jobs done by abstraction. Suppose I see a tea-pot on the table, and I want to grasp it. I can do this because I have in my head a model which allows me to control the movement of my hand as I reach the tea-pot. In this model, only the position and form of the tea-pot is taken into account, but not, say, the form of the table, or the presence of other things, as long as they do not interfere with grasping the tea-pot. In another move I may wish to take a sugar-bowl. Also, I may be aware that there are exactly two things on the table: a tea-pot and a sugar-bowl. But this awareness is a result of my having two distinct abstractions: an isolated tea-pot and an isolated sugar-bowl.

Abstraction breaks the world into parts by ignoring something, while including something else. This refers to processes, not only to objects. In fact, the only possibility to define what is a process (informally – a part of the world) is to refer to an abstraction mechanism: what is left is a process, what is abstracted away is not this process, but "other processes".

The degree of abstraction can be measured by two parameters: its *scope* and *level*; correspondingly, there are two ways to increase abstraction.

The scope was defined above. The wider the scope of a concept, the more abstract it is. The classical example is: cats, predators, mammals, etc. up to animal. With this method, the content of the concept (i.e. the amount of specificity in its procedural definition) decreases as the degree of abstraction increases. We come finally to such universal abstractions as *object* or *process* which carry almost no information.

The level of an abstraction is the number of *metasystem transitions* involved. In the modeling scheme, the subject-of-knowledge system S is a metasystem with respect to the world. Indeed, S *controls* the world: it takes information from the world as input to create a representation, processes the choice of a certain action a, and executes it at the output, changing thereby the state of the world. The brain of S, as the carrier of world representations, is on the meta-level, a "meta-world", so to say. Assigning to the world level 0, we define the abstractions implemented in the brain as abstractions of the first level.

However, the creation of models of reality does not necessarily end with the first level abstractions. The brain of the level 1 can become the world for the new brain, the brain of the level 2. Naturally, the necessary machinery comes with it – the second metasystem transition is taking place.

Our human brain is known to have a many-level hierarchical structure. The "brain" of the first level is nothing but our sense organs. The eye contains the retina and the machinery necessary to throw light on it and create representations of the world. The retina keeps first level abstractions, which constitute the input (a new world) for the second level of the brain models, whose representations will be abstractions of the second level; these abstractions, in turn, become input for the next level of models, and so on.

The human language has also a hierarchical structure of abstractions. The concept of number can serve as an example of an abstraction of the second level in this structure. Specific numbers are first-level abstractions. We derive them directly from our experience using counting as the representation function in this model, which we know as arithmetic. The *brain* (in our technical sense) of the arithmetic model includes some physical carrier, such as fingers or pebbles or paper and pencil, to represent specific numbers. A specific number, say three, is an abstraction from three apples, three sheep, etc., whatever we can count. The concept of a number is an abstraction from one, two , three, etc., which are all abstractions themselves and exist only in the first-level model, and if not in that model, would not be found anywhere. In algebra the second-level abstraction number becomes a material for higher abstractions, and this metasystem transition can be repeated many times. ■ ■

Human language

Two cybernetic systems are involved in human language and thought.

The first system is the human brain. It creates models of reality whose material body is the nerve nets, and which we therefore call *neuronal* models. Neuronal models use neuronal, or intuitive, concepts (functional elements).

The second system is language. Its functioning is linguistic activity in society. Linguistic activity creates models of reality whose material body consists of linguistic objects. The functional elements of this system are *logical (linguistic)* concepts.

Neuronal vs. logical concepts

The brain and the languages are closely interrelated. Language is an offspring, and, in a certain sense, a continuation of the brain: using language we create new models of reality, which were not embedded in our brain by nature. Some substructures in our brain are representations of the states of the world. Some of the linguistic objects are representations of those representations: we refer to them as most concrete, or low-level, concepts. Such words as "cat", "apple", "to run" , and the concepts they fixate, are of that kind.

But human language (like the human brain) is a multilevel hierarchical system. We create *theories*, where we use abstractions of higher levels, logical concepts the

machinery of which (do not forget that concepts are *functional* units) requires something in addition to the brain: some material linguistic objects. Thus while we can regard small numbers, like two or three, as neuronal concepts because we immediately recognize them, bigger numbers, like 137, can function only using some representations external to the brain.

The concepts of the higher levels do not replace those of the lower levels, as they should if the elements of the language reflected things "as they really are", but constitute a new linguistic reality, a superstructure over the lower levels. We cannot throw away the concepts of the lower levels even if we wished to, because then we would have no means to link theories to observable facts. Predictions produced by the higher levels are formulated in terms of the lower levels. It is a hierarchical system, where the top cannot exist without the bottom.

We loosely call the lower-level concepts of the linguistic pyramid concrete, and the higher-level abstract. This is correct as long as one keeps in mind that abstraction is not always abstraction in scope from things observable, but also, and most importantly, an abstraction from abstractions of the lower levels of the same linguistic system. Pure abstraction from specific qualities and properties of things leads ultimately, as its scope increases, to the loss of contents, to such concepts as some and something. Abstractness of a concept is actually its "constructness", the height of its position in the hierarchy, the degree to which it needs intermediate linguistic objects to have meaning and be used. Thus in algebra, when we say that x is a variable, we abstract from its value, but the possible values themselves are numbers, i.e. linguistic objects formed by abstraction in the process of counting. This intermediate linguistic level of numbers must become reality before we use abstraction on the next level. Without it, i.e. by a direct abstraction from countable things, the concept of a variable could not come into being. ∎

Formalization

When language is used for comparatively narrow professional purposes there is a tendency to limit the number of terms used and to give them a more precise and constant meaning. We say that the language is being *formalized*. If this process is carried through to its logical conclusion, the language will be completely formalized, or *formal*. A language is formal if its usage relies only on the "form" of linguistic objects, and not their intuitive meanings.

To make this definition precise, we must specify a set of perceptions (that is abstractions) and actions which are registered and performed in the same way by all members of the society whom the languages serves. We shall refer to these perceptions and actions as *universally defined*. A language is formal if all the processes involved in its usage, namely the representation function $R(w)$, the modeling function $M(r)$, and (for command languages) the set of possible actions are expressed in terms of universally defined perceptions and actions.

We usually assume that universally defined perceptions and actions can be relegated to a machine. The question is still open whether this is a realistic assumption. We accept it with a qualification that if there is a doubt about a specific abstraction or action, it must be excluded from the universally defined set. Then a formal language is a language usable by a properly constructed machine. A machine of that

Valentin Turchin

kind becomes an *objective* model of reality, independent from the human brain which created it. This makes possible a series of consecutive metasystem transitions, where each next level deals with a well-defined, objective reality of the previous levels. Thus the language becomes an *ultrametasystem*, exhibiting the *stairway effect*[4] and an explosive growth in volume and power. Just as mastering the general principle of using tools to make better tools gives rise to consecutive metasystem transitions and the creation of industrial system, so mastering the principle of describing (modeling) by means of a formal language gives rise to the creation of the hierarchical system of formal languages on which the modern science is based. ■

Four types of linguistic activity

Classification of languages by these two parameters of abstraction and formalization leads to the following four types of language-related activities[4] (see Fig.2).

Art is characterized by unformalized and concrete language. Words and language elements of other types are important only as symbols which evoke definite complexes of mental images and emotions. Philosophy is characterized by abstract informal thinking. The combination of high-level abstract constructs used in philosophy with a low degree of formalization requires great effort of intuition and makes philosophical language the most difficult type of the four. Philosophy borders with art when it uses artistic images to stimulate intuition. It borders with theoretical science when it develops conceptual frameworks to be used in construction of formal scientific theories. The language of descriptive science must be concrete and precise; formalization of syntax by itself does not play a large part, but rather acts as a criterion of the precision of semantics. ■ ■ ■

Figure 2 *Types of linguistic activity*

	Concrete language	Abstract language
Unformalized language	Art	Philosophy
Formalized language	Descriptive sciences	Theoretical sciences, mathematics

Truth

A piece of knowledge is *true* if the predictions made by the subject (the user) of knowledge on the basis of this knowledge are true. A prediction that test T is successful, denoted as $T!$, is true if the test T is, in actual fact, successful.

Let us examine the systems aspect of this definition. Statements of a language represent some pieces of knowledge which can be reconstructed from statements by the recipient of the message (which may be the sender of the message itself). These pieces of knowledge produce, ultimately, some predictions: $T_1!$, $T_2!$, ... etc. that certain tests are successful. Knowledge about the world constitutes a metasystem with respect to the world in which the tests T_1, T_2, ... take place. When we speak of statements and determine their various properties, we create one more metasystem level; this is a metasystem transition. While the first level is the level of the *use* of knowledge, the second level is that of the *analysis and evaluation* of knowledge. We shall denote as $\gamma(S)$ the process of examination and analysis of the statement S with an aim of finding whether S is true. We shall view it as a *test*. Its success means that S is true, its failure means nothing, because it could result either from S being false, or from our inability to establish that S is actually true.

Hidden in our definition of the process γ is the assumption of its *correctness*, by which we mean that if $\gamma(S)$ for some S is finite, $\gamma(S)!$, then S is true:

$$\gamma(S)! \implies S$$

How can we devise a correct procedure γ? Ideally, we would like to establish the truth of statements by direct *verification*.

Verification

Verification is a direct use of the definition of a statement to decide if the statement is true.

Predictions are, in principle, verifiable. You only have to initiate the test that the prediction is about and wait until it comes to the final state. Therefore, if a prediction is correct, there is, in principle, a way to verify this, though it may take a very long time, thus making it impossible in practice.

The situation is worse with general statements, which may generate an infinite set of predictions. It is impossible to verify them all. We have to rely on some kind of *proof.* ■

Proof

By *proof* of a statement S we mean any test process $\gamma(S)$ the success of which the subject of knowledge accepts as the evidence that S is true; this implies that the subject is prepared to make decisions on the basis of the predictions generated by S.

When a statement cannot be directly verified, it is still possible that we take it as true relying on our intuition. For example, consider a test process T the stages of which can be represented by single symbols in some language. Let further the process T develop in such a manner that if the current stage is represented by the symbol A, then the next stage is also A, and if the current stage is a different

symbol, the process comes to a successful end. This definition leaves us absolutely certain that T is infinite, even though this cannot be verified. In *logic* we call some, most immediate, of such statements *axioms* and use them as a basis for establishing the truth of other statements. In natural sciences some of the self-evident truths serve as a starting point for the construction of theories.

Our definition of proof identifies the truth of S with the prediction $\gamma(S)!$, where the process γ is, as defined before, *some* process of examination and analysis of statements. The new element here is that we have related γ explicitly to the user (the subject) of knowledge. To justify the use of a certain γ we can now make reference to the intuition or the practice of the user. But is this sufficient? How do we know that a given process γ is correct? In our definition of proof, can we limit processes γ to correct processes only?

Unfortunately, we cannot. To say that γ is correct is to say that for all S, $\gamma(S)!$ is true only if S is true. This is, obviously a general statement, not a prediction. It cannot be directly verified. Thus we have to use another, or the same, procedure γ' in order to define what its truth means:

$$\gamma'(\forall S\ (\gamma(S)! \Rightarrow S))!$$

But how do we know that γ' is correct? This is a vicious circle. To define formally the correctness of a process γ we have first to define the correctness of itself, or that of a similar process.

Thus the correctness of proof cannot be reduced to anything more reliable. There is no absolute criterion of truth. Verification could be one such thing, but, unfortunately, we cannot verify an infinite number of facts. So, we simply *assume* that our process γ is correct, and the statements we prove are indeed true. We *define* the meaning of the statement 'S is true' as the prediction: $\gamma(S)!$. If we discover an error (see *contradiction* below), we change something in γ and try again, assuming that our new γ is correct. This is done both in natural sciences and in mathematics, and there is no other road to truth.

It might seem that this whole discourse is only the matter of notation, but this would be a wrong impression. In order to assign meaning to a statement in our epistemology, this statement must be interpreted as a prediction or a hierarchical generator of predictions. We have given an analysis of the meaning of the statement 'S is true' and have shown that it necessarily includes a reference to a certain process γ.

It could also be said that there was not much point in that analysis, because for all practical purposes we could identify 'S is true' with the statement S itself. But this would be avoiding the problem, not solving it. The problem would pop up anyhow in the analysis of the logical connectives, to which we pass over now. ■

Conjunction

Let S_1 and S_2 be some statements. Consider the statement 'S_1 and S_2', which is known as the *conjunction* of the two statements. What is its meaning? Since the

meaning of a statement is the set of predictions it produces, when we put forward two statements, we take responsibility for both of them, and for all predictions produced by either of them. We come to the following definition.

First-level definition of conjunction

If S_1 and S_2 are two statements then the generator which produces S_1 and S_2, and nothing else, is its *conjunction*, and is denoted $S_1 \wedge S_2$. ∎
We call this definition *first-level* because it directly constructs out of two statements over some domain a new statement over the same domain, without any reference to the concept of truth or any other analysis of statements. We can also define conjunction as an operation of the second level, the level of proofs.

Second-level definition of conjunction

Let two statements, S_1 and S_2 be given. Define a summary process (test) T as follows. Run the proof process $\gamma(S_1)$. If it ends in failure, end T in failure. If it ends in success, run $\gamma(S_2)$. If this second process ends in success, end T in success. If it ends in failure, end T in failure. (Obviously, if any of the two processes is infinite, T is also infinite). Then the prediction $T!$ is the *conjunction* of S_1 and S_2, i.e. the statement that both are true. ∎
The conjunction thus defined is, of course, the statement that both S_1 and S_2 are true, and this is how conjunction is usually defined. When the discourse is informal, no distinction is made between the two definitions, the same as in the case of S and $\gamma(S)!$. In a formal treatment we should give preference to the first-level definition, because it formalizes how we actually use conjunction. If the second-level definition is adopted, then to use it we must first derive from it the first-level generator, in a manner similar to when we conclude from 'S is true' that we can use S. ∎

Disjunction

The *disjunction*, (inclusive or) of S_1 and S_2, denoted as $S_1 \vee S_2$, is the statement that at least one of the two statements is true. The remarkable thing is that disjunction does not allow a first-level definition; only the following second-level definition is possible.
Let two statements, S_1 and S_2 be given. Run in parallel two processes: $\gamma(S_1)$ and $\gamma(S_2)$; call this summary process T, and end it successfully when (and if) at least one of the two parallel processes comes to a successful end. Then the disjunction $S_1 \vee S_2$ is the prediction $T!$. ∎

Implication

If-then statement, or implication, $S_1 \Rightarrow S_2$ is a conditional statement setting forward S_2 in the case if S_1 is true. S_1 is referred to as *antecedent*, or *premise*; and S_2 as *consequent*, or *conclusion*.

Two kinds of implication are distinguished by the reason for which they are true. *Logical* implication takes place when S_2 is a part of the meaning of S_1, thus when we say S_1 we thereby also say S_2. *Material* implication connects two statements which may be in no way related by their meanings. It establishes the connection by force, so to say, declaring it as an empirical fact, a kind of law of nature. In Kant's terminology, a logical implication makes an *analytic* judgment, while a material implication is a *synthetic* judgment.

For a simple example, consider two implications:

(1) (x is an apple) $\rightarrow$ (x is a fruit)
(2) (x is an apple) $\Rightarrow$ (x is edible)

The first implication is logical, because being a fruit is a part of being an apple. The second implication is material, because an apple, by its usual definition as the fruit of a certain class of trees, is not necessarily edible. There are inedible apples, after all, so the truth of the second statement may depend on the domain of the possible values of the variable x.

Other examples:

$$(S_1 \wedge S_2) \Rightarrow S_1$$

is a logical implication, while

$$2 + 3 = 3 + 2$$

is a material implication (this is Kant's famous type of synthetic judgments a priory).

Logical implication

In our formalism, a logical implication registers the fact that S_1 is a generator which produces S_2. Consider two statements S_1 and S_2. Form the following test T. Run S_1, and compare each produced statement (not only predictions) with S_2. If they compare successfully, end T in success. Logical implication $S_1 \Rightarrow S_2$ is the prediction $T!$. ∎

Material implication

To formalize material implication we construct a machine which tries to prove S_1 and if succeeds produces S_2. Consider two statements S_1 and S_2. Construct the generator which runs the test $\gamma(S_1)$ and if it succeeds produces S_2. This generator is the material implication $S_1 \Rightarrow S_2$. ∎

In mathematical logic there is no difference between material and logical implication, but according to our definition, the implication used there is material. Indeed, mathematical logic is purely formal, syntactic; no reference to the meaning of statements is made, only the usage of the compound statement is defined, and the usage is: if the premise is true, the conclusion can also be taken as true. This is a material implication.

The implication of formal logic has an odd property: if the premise S_1 is false, then no matter what the conclusion S_2 is, the implication is true. This strikes everyone who starts studying formal logic as contradicting our intuition and common sense. In our interpretation a material implication with a false premise is not true in the same sense as a prediction, or a generator of predictions, can be true. Neither is it false. It is just empty. This, we believe, is in perfect agreement with our intuitive expectation.

The definition of logical implication is of the second metasystem level, because it speaks of a relation between statements. Yet it does not require the concept of truth. The definition of material implication does require it. It cannot be understood without a reference to the process γ. ■ ■

Negation and Falsehood

The statement that a statement S is false, 'not S', is the *negation* of S.

The law of the excluded middle

According to the common understanding of truth and falsehood, what is not true is false, and vice versa. The following statement is known as *the law of the excluded middle*: For any statement S, either S, or its negation is necessarily true. In other words, every statement is either true or false; there is no third, or middle, possibility. ■

The principle of progressive formalization allowed us to use *informally* negation and the law of the excluded middle, as well as other logical connectives. The time has come, however, to make the concept formal. We shall then see that the law of the excluded middle is not that obvious (even though ultimately it will be, in our philosophy, completely restored in its rights).

What does it mean that a statement is false? To be meaningful, the statement of falsehood must be, as any other statement, interpreted as a hierarchical generator of predictions. Let us start with the simplest case of a statement, which is just one prediction.

Consider a prediction $T!$. If it is not true, then test T either ends in failure, or is infinite, i.e. never ends at all. We have already defined what does *or* mean. (Actually, our natural understanding of this 'or' is exclusive, but we can replace it by inclusive). Therefore, we must determine what is:

(1) the meaning of T ending in failure;

(2) the meaning of T being infinite.

The first is easy. It is, obviously the prediction that T stops, and the comparison of the result with the symbol *Fail* succeeds.

Now consider the statement that T is infinite. We shall denote it as $T?$. It is not a prediction. But it is a generator of predictions. Indeed, T is a *test*, and our definition of a test includes the existence of a *termination subtest* $E(t_i)$ which is applied to the stages t_i, $i = 1, 2, \ldots$ etc. of the test T, and for each stage determines, in a finite time, whether it is *Succeed*, *Fail*, or *Go-on*. Therefore, the statement that the state t_i is *Go-on* is a prediction, namely, the success of the test that runs $E(t_i)$ and

compares the result with *Go-on*. Now we can define what the meaning is of the statement that T is infinite: it is the generator which produces the following series of predictions:

t_1 is *Go-on*
t_2 is *Go-on*
t_3 is *Go-on*
... etc., infinitely

Thus a statement $T?$ that a prediction is false is a general statement, a generator of an infinite set of predictions. It cannot be verified directly.

We can now sum up as follows the three cases that can take place when we try to verify a prediction (see Fig.3).

Figure 3 *Verifying predictions*

	Testing process:	Prediction is:
1	ends in succes	verifiably true
2	ends in failure	verifiably false
3	never ends	false, but this is not directly verifiable

Now consider an arbitrary generator. It must be pronounced false if at least one prediction produced by it is false. We observe, however, that we may contemplate two contradicting predictions and be in doubt as to which of them is false. Yet it is clear that it is wrong if both are produced. Therefore we can base our definition on the fact of contradiction alone. However, to establish that a prediction is false we may, again, need to define the falsehood of a generator, because the negation of a prediction may be, as we have seen above, an infinite generator of predictions. Therefore we base our definition on an undeniable fact of a *verifiable* contradiction.

Contradiction

The situation when there is a prediction $T!$ that a certain test T ends in success, and also there is a prediction T that test T ends in failure is referred to as a (verifiable) contradiction. ∎

Back to Negation and Falsehood

The statement $T?$ that test T is infinite contradicts the statement that it is finite. Indeed, from the finiteness of T we conclude that there exists a stage t_n, such that the subtest $E(t_n)$ results in either *Succeed* or *Fail*. But from $T?$ it follows that for all t_i, including t_n, $E(t_n)$ is *Go-on*. Thus for the always finite test which compares the result of $E(t_n)$ with *Go-on* we have two contradicting predictions.

Let T be a test. The statements $T!$, T, and $T?$ will be called *atomic*. Any pair of these statements, put forward simultaneously, constitutes a contradiction. It must be kept in mind that $T?$, though atomic, is not a prediction.

We can now define what is falsehood. It is contradiction to truth.

More formally, the negation of S means that there exists such a true statement X (a context) that when used together, X and S produce a contradiction.

Therefore, in addition to the truth-proving process $\gamma(S)$ we introduce a process $\bar{\gamma}(S)$ which tries to find such a provable true statement X, that S together with X produce a contradiction. It should be stressed that at this stage of the development of our basic definition, the definition of the fundamental processes g and gn remains informal. The reader can even have legitimate doubt that a consistant and formal definition of these processes is possible at all. It is possible, however, and we refer the reader to[5] where this matter is considered in detail.

We see that after we have defined the exact meaning of logical connectives, the law of the excluded middle, which now takes the form:

$$\forall S\ (\gamma(S)!\ \vee\ \bar{\gamma}(S)!)$$

does not seem as obvious as it seemed before formalization. It requires that for each statement S one of the processes $\gamma(S)$ and $\bar{\gamma}(S)$ ends successfully. Therefore, it strongly depends on the nature and specific character of these two processes. It is not immediately clear that one of the two must be necessarily finite.

Moreover, it can be shown that with the most common understanding of these processes, when $\gamma(S)$ is proof of S, and $\bar{\gamma}(S)$ that of the negation of S, in some non-contradictory axiomatic system, there is always a statement for which both processes are infinite (Goedel's theorem). This seems a very good reason for the rejection of the law of the excluded middle, which is actually done by the school of thought known as *intuitionism*. In our philosophy, however, the interpretation of γ and $\bar{\gamma}$ is such that the law of the excluded middle, together with the whole classical logic, remains valid. ∎

Foundation of logic and mathematics

The main semantic principle in the Principia Cybernetica project is that all statements must have a clearly defined (formal) meaning, and that the meaning of statements is in hierarchical generation of predictions. We saw that in order to assign a formal meaning to the concepts of truth and falseness, as well as the basic logical connectives, we had to introduce certain processes γ and $\bar{\gamma}$ which represent, respectively, the proof and refutation of statements by the user of mathematical knowledge.

On this basis a theory has been further developed, which answers the questions concerning the foundation of mathematics, including logic and set theory (see a brief exposition in[5]). Here are some highlights of the theory.

1 Mathematics is seen as a branch of science. The objects of mathematical knowledge are of the same nature as those of other sciences: the abstracted phenomena of the world we live in. As with every branch of science, mathematics has its own type of objects, namely, symbolic models of the world constructed by other ("natural") sciences, and the models of the models mathematics creates itself. There are no differences of principle between mathematics and natural sciences with regard to the methods of acquisition of knowledge and their reliability.

2 Our theory is concerned with symbolic mathematics, which is the official language of all mathematicians at present. While the importance of three-dimensional geometric intuition is acknowledged, no attempt is made to analyze its role.

3 In addition to mechanical processes of computation and proof, which play so important a role in contemporary mathematics, we introduce *metamechanical* processes, which cannot be modeled on such devices as Turing machines, or defined through recursive functions. Metamechanical processes differ from mechanical processes in that they are controlled by the *subject* of knowledge, i.e. its user and developer. In the conceptual apparatus of modern physics, the inseparability of the phenomenon and the *observer*, i.e. the object and the subject of knowledge, is well known. The user of our theory is a close analogue of the observer of physics.

4 Within this conceptual framework we interpret the fundamental aspects of mathematics and logic. We show that there are two ways of defining the concept of truth in our formalism, i.e. the definition of the processes $g(S)$ and $gn(S)$. One leads to intuitionist logic, the other to classical logic.

5 Sets of set theory are interpreted as processes which generate objects. Metamechanical processes provide for not recursively-enumerable and non-denumerable sets. The axioms of the Zermelo-Fraenkel system, except the axiom of choice, are proven as theorems. (The interpretation of the axiom of choice is left open for the time being.) The paradoxes of logic and set theory are explained. We show that those objects and propositions which in informal reasoning lead to paradoxes are, in our theory, uninterpretable, have no meaning.

6 The consistency of set theory is proved. This does not contradict Goedel's result of the impossibility of proving the consistency of a theory in the theory itself, because our theory cannot be formalized in set theory. ∎

Theory

We distinguish between factual statements and theories. If the path from a statement to verifiable predictions is short and uncontroversial, we call it factual. A theory is a statement which can generate a wide scope of predictions, but only through some intermediate steps, such as reasoning, computation, the use of other statements. Thus the path from a theory to predictions may not be unique and often becomes debatable. Between the extreme cases of statements that are clearly facts and those which are clearly theories there is a whole spectrum of intermediate cases.

Top-level theories of science are not deduced from observable facts; they are constructed by a creative act, and their usefulness can be demonstrated only afterwards. Einstein wrote: "Physics is a developing logical system of thinking whose foundations cannot be obtained by extraction from past experience according to some inductive methods, but come only by free fantasy"[1].

Since theories usually produce an infinite number of predictions, they cannot be directly verified. But they can be refuted. For this end it suffices to find one false prediction.

The statement of the truth of a theory has essentially the same meaning as that of a simple factual judgment: we refer to some experience which justifies, or will

justify, the decision-making on the basis of this statement. When this experience is in the past we say that the truth is established. When it is expected in the future we say it is hypothetical. There is no difference of principle between factual statements and theories: both are varieties of models of reality which we use to make decisions. A fact may turn out to be an illusion, or hallucination, or a fraud, or a misconception. On the other side, a well-established theory can be taken for a fact. And we should accept critically both facts and theories, and re-examine them whenever necessary. The differences between facts and theories are only quantitative: the length of the path from the statement to the production of predictions. ■ ■

Acknowledgement

References

1. Einstein, A., Physics and reality, *Franklin Institute Journal* vol.221, pp. 313-357, 1936.
2. Heylighen, F., Joslyn, C., and Turchin, V. A short introduction to the Principia Cybernetica project, *Journal of Ideas*, Vol.2, No 1, pp.26-29, 1991.
3. Schank,R.C. and Abelson, R.P. *Scripts, Plans, Goals, and Understanding.* Lawrence Erlbaum, Hillside, N.J. 1977.
4. Turchin,V.F. *The Phenomenon of Science*, Columbia University Press, 1977.
5. Turchin, V.F. A constructive interpretation of the full set theory, *The Journal of Symbolic Logic*, vol.52, pp.172-201, 1987.
6. Turchin, V.F. and Joslyn, C., The Cybernetic Manifesto, *Kybernetes*, vol.19, Nos.2 (pp.63-64) and 3 (pp. 52-55), 1990.
7. Turchin, V.F. Cybernetics and Philosophy, *The Cybernetics of Complex Systems - Self-organization, Evolution and Social Change, Proceedings of the 8-th International Congress of Cybernetics and Systems*, Axel Duwe, Intersystem Publications, pp.61-74 (1991).