Back to Contents
Next to Chapter 9

Chapter 8

The Problem of Complexity in Describing the World Scientifically[1]

A Formal Analysis of the Difficulties in Constructing Theoretical Biology

Psychological Grounds for Judging Whether a Statement Is True

Great efforts have been directed at developing the concept of the *logical truth* of statements. The approaches of Tarski and Carnap are well known (see, e.g., Gastev and Finn, 1964). However, in science, as in our everyday life, when we acknowledge the legitimacy of judgments, we are prone to proceed not so much from their logical truth as from premises of a psychological nature.

In my experience, a statement acquires the right to be called scientific when at least one of the following conditions is fulfilled.

Condition 1. A compact set of concepts is formed. This allows the making of judgments of the phenomenon under study in a compact form. In the natural sciences, we then speak of having created a theory. Theory, in some formal meaning of word, is a type of logical structure which allows us to describe the phenomenon observed in an essentially briefer manner than could be done without any theoretical considerations, after immediate observation. Even a compact presentation that does not contain any theoretical considerations is regarded as a type of theory. Mendeleev's periodic table of the elements, at the moment of its

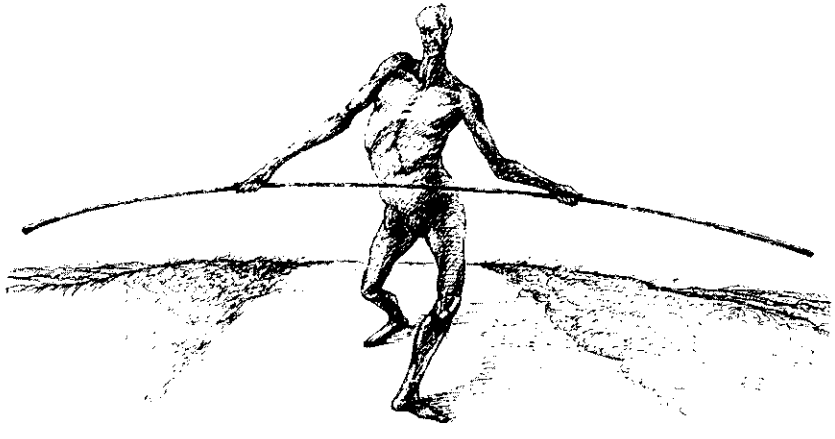
[1] A brief version of this chapter was published in the abstracts of the VIIth All-Union Symposium on Logic and Methods of Science, Kiev, 1976, pp. 234–236. This chapter was translated by A. V. Yarkho.

An Attempt to Grasp the Reality with a Set of Models

appearance, did not contain any theoretical reasoning, yet it was part of chemical theory. The same is true of Linnaeus's classification.

In mathematics, a compact presentation of judgments is embodied in mathematical structures. According to Bourbaki (1950), mathematics differs from other sciences in that judgments made within it can be reduced to mathematical structures—compact constructions rich with logical consequences. Strictly speaking, we should call mathematical science not all groups of statements written in mathematical language but only those which display mathematical structures. For example, the theory of probability acquired the status of a mathematical subject only after Kolmogorov gave its axiomatic structure. If we hold to the concepts developed by Bourbaki, we should designate as mathematical models of the external world not every description expressed in mathematical symbols but only those with rich structural content. Theoretical physics, whose structure I shall discuss in detail below, has proved to be arranged so that its content is given by fairly compact statements made in the language of mathematics. By the way, here lies the reason for the oft-repeated statement that the structures of physics are the model for all other sciences.

The construction of compact statements in the form of mathematical models seems possible in other branches of knowledge, including the humanities. As an illustration, I mention my study of language semantics. I suggested that the mechanism of speech comprehension be described by means of the Bayesian theorem. Later, I myself was amazed by



The Struggle Against Complexity

the great theoretical purport of this approach, which allowed various peculiarities of verbal behavior to be easily explained. Moreover, certain concepts of the nature of thinking itself arose easily (see Nalimov, 1974*b*, 1979, 1981). At the same time, expanded mathematization of knowledge in some cases resulted in the construction of the so-called "portrait models," which do not have any new content but simply express in mathematical language what can just as well be expressed in common language. It is only too clear why such models annoy representatives of concrete branches of knowledge: in these cases, one language is unjustifiably substituted for another one. As an example, one can ask how biology was benefited when some of its ideas were reformulated in terms of information theory.

Nonetheless, compact presentation makes possible many rich consequences. If a complicated phenomenon is expressed by a compact model, we always have a basis (psychological rather than strictly logical) for believing that such a description might include things which we have not yet observed but will be able to observe by making special efforts. In this way theories acquire their predictive power—the possibility that they may predict yet unobserved phenomena. Here lies the principal difference between modern science and alchemy. The latter described its phenomena in a mytho-poetical language, the way they were observed, and for this reason it could not predict new phenomena. The progress of European culture is largely due to linguistics, since therein was developed a symbolic language of compact ideas.

Compact presentation allows us to simulate all possible evolutions of a phenomenon. In this case it will suffice to vary the initial conditions. More than that, compact presentation allows us to control the phenomenon, which always gives the illusion of complete comprehension, i.e., of cognition. The possibility of controlling something as it is cannot be acknowledged to be the criterion of truth. Man had learned to control some technological processes (e.g., metallurgical ones) long before any scientific concepts of metallurgy appeared. A more detailed critical analysis of the concept "cognition" was given in Chapter 1 of this book. There I did not give an optimistic answer to the question of what cognition is. Here I propose the following answer: formally speaking, cognition of the world is a possibility of recording in a compact form the observed phenomena. But this answer will hardly satisfy everyone. Again, this would mean that cognition is made equal to mastery of the world, since compact presentation is just what allows us to forecast and control.

Condition 2. Past knowledge is being strengthened. In Chapter 2, I spoke of our knowledge always being probabilistically weighted: we are apt to ascribe greater probability to some judgments and smaller proba-

bility to others. New assertions obtained as a result of the analysis of newly observed phenomena acquire the features of scientific knowledge only if they strengthen the previously existing theoretical structures. The mechanism of placing new statements in correspondence with old knowledge may be described by a probabilistic model built in the system of neobayesian notions, of which I already spoke in Chapter 2.

The theorem of Bayes,[2] well known in mathematical statistics, provides us with a mathematical model of the way previously accumulated knowledge (probabilistically weighted) mixes in our mind with that newly acquired.

The development of scientific schools can be understood easily from this standpoint. First, the leader creates a fuzzy and insufficiently confident concept accompanied by doubts and reservations. In this fuzzy field of elementary events, the probability ascribed to separate judgments is determined in some way. The leader's pupils try to get new knowledge that will narrow the initial set of fuzzy concepts. At a certain moment, all the probability proves to be concentrated in a very narrow subset of the set of initial statements. The distribution becomes almost needle-shaped. Further strengthening of the correctness of the narrowly concentrated statements becomes unnecessary, and the school becomes inactive. And if we look at the publications of the members of the school, we see that their articles always end with the following incantations: our results "are in good agreement with . . . ," "support previously formulated hypotheses . . . ," "they do not contradict . . . ," "strengthen the previously obtained results." Everyone knows how pleasant it is to refer to the fact that the idea you are developing has already been mentioned somewhere, though its formulation was then rather weak. Within some paradigms, this form of presenting ideas is the only one possible, since only in this way do they acquire a scientific character.

Condition 3. Unexpected new statements evoke wonder. Wonder is a term used by Plato and Aristotle. In his *Theaetetus* Plato (1953) wrote, ". . . for wonder is the feeling of a philosopher, and philosophy begins in wonder." Aristotle's *Metaphysics* (1966) includes the words, ". . . for it is because of wondering that man began to philosophize and does so now." Bohr uses the expression "a mad hypothesis." The possibility of explaining the old in a new way, of observing it with new eyes, immensely attracts some people and arouses opposition in others. Its attractiveness lies in the fact that the new, unexpected view of things is a new guessing which might help us to see what has been concealed from us by the old system of concepts. Opposition is caused by the fact that a novel view is

[2] In my earlier books (Nalimov, 1974*b* 1981) I used it to explain the mechanism of comprehending phrases built over words with fuzzy prior semantics.

always a heresy—the refutation of what has been recognized as scientifically true.

The two above-mentioned positions are mutually exclusive, and their supporters are people with different genotypes. (On behavioral genetics, see, for example, Davis, 1975.) The struggle of opinions in science is very often only a conflict between people with different psychological characteristics.

If conditions 2 and 3 presented above are characteristics not only of scientific activity but also of all other human intellectual activities, the first condition—compact presentation—is an exclusive prerogative of scientific activity. European culture is primarily an abstract-symbolic, i.e., compact, record of our knowledge of the world.

Compactness of Theoretical Constructions in Physics[3]

In an earlier book (Nalimov, 1981), I presented the results of a study designed to determine the degree of saturation of physics texts by mathematical symbols. This I referred to as the "symbol complexity" of publications in physics. For my purposes, "words" were represented by mathematical operations such as the derivative dy/dt, the inverse matrix $\mathbf{M}^{-1}$, etc., and "phrases" consisted of symbols and the operations performed with them that are separated from the rest of the text. For example, a phrase is a mathematical expression such as

$$m = \sum_{i=1}^{n} a_i m_i$$

The study was carried out from an historical perspective. It included six books on general physics published in 1760, 1797, 1833, 1874, 1933, and 1948. In addition, four books on field theory and quantum mechanics published in 1931, 1946, 1948, and 1949 were considered. By calculating the average number of "words" and "phrases," as defined above, on a standard page of text, I was able to show that the symbol complexity in books on general physics remained more or less constant from the beginning of the nineteenth century through the rest of the period considered. No increasing trend toward saturation of the texts with symbols was observed. For publications in field theory and quantum mechanics, the situation was slightly different in that the texts were

[3]This section and the next one were published in 1977 in the Polish journal *Zagadnienia Naukoznawstwa* and in Ukranian in the journal *Avtomatika*. The translation used here is borrowed from the journal *Industrial Laboratory*, which is regularly translated in the United States.

richer in mathematical symbols, although the difference between the individual books was small.

These results led me to conclude that differences in the complexity of the contents of books do not manifest themselves by a difference in symbol complexity. Although the language of physics is undergoing a steady evolution, the text becomes more complicated not as a result of an increase in the number of mathematical symbols or phrases but as a consequence of the greatly expanded meaning carried by these symbols and phrases. Thus, the same symbol can denote a scalar, a vector, a matrix, or some other quantity. The development of physics, i.e., the expansion of its scope and meaning, takes place in such a way that the compactness of its symbols is preserved.

Let us write down some well-known formulas of modern physics.

The Lorentz transformation:

$$x'_1 = \frac{x_1 - vt}{\sqrt{1 - v^2/c^2}}; \quad t' = \frac{t - x_1 v/c^2}{\sqrt{1 - v^2/c^2}}; \quad x'_2 = x_2; \quad x'_3 = x_3$$

The Schrödinger equation:

$$i\hbar \frac{\partial \psi}{\partial t} = \mathbf{H}\psi$$

Heisenberg's uncertainty principle:

$$\Delta p \Delta x \geqslant \hbar/2$$

These very short formulas carry a very large amount of information and have given a considerable impetus to the development of modern physics; in particular, the first of these formulas expresses the theoretical feasibility of an atomic bomb, with all its consequences.

The search for compact formulas has prompted physicists to use probabilistic considerations in their theoretical constructions. In deterministic systems it is not possible to describe by a compact formula the behavior of molecules that are in a gaseous state.

In contrast to the opinion held by the philosophers of the past, the necessity of describing phenomena in probabilistic language was not a result of lack of knowledge but rather of the need to express knowledge in a compact form. Here it seems appropriate to recall the algorithmic definition of randomness. A finite sequence of numbers is called a random sequence if it is not possible to construct for it a generating

algorithm that could be written in a form shorter than the sequence itself. [The algorithmic method of definition of randomness is considered in more detail by Fine (1973), who also presents a detailed analysis of the logical foundations of all (or almost all) the probability theories available nowadays. Here I deliberately present a somewhat simpler formulation, by proceeding from a theory not yet fully developed.]

Precisely such a situation was encountered in the construction of statistical physics and thermodynamics. The statistical approach made it possible to describe the behavior of a system with the aid of average quantities, without describing the behavior of each individual molecule.

Is a Compact Description of Knowledge Possible in Biology?

From a formal point of view, the difficulties encountered during the entire period of the development of biology are due to the need for a compact description of the huge amount of material readily accumulated as a result of observations. The first successful attempt to describe the great diversity of observations was the classification of Linnaeus, which, however, was extensively revised later on. Darwin's evolutionary theory is an attempt at a compact classification of these same data, but from an historical perspective. However, from Darwin's time to our own, we have not found any all-encompassing and compact theoretical constructions with the same scope of interpretation as, for example, the formulas of theoretical physics presented above. The discovery of the biogenetic code (the most important discovery in biology in recent times) is in fact the deciphering of a language, not an explanation of how something *new* is written in this language. Here the success achieved consists in finding in biology a structure familiar to our intellect, i.e., a language, and analyzing it formally.

Many publications have appeared recently in which mathematical models of biological phenomena are constructed. There even exist special journals devoted to mathematical methods in biology, for example, *Biometrics* and the *Journal of Mathematical Biology*. It would seem that methods have been found that lead toward a compact representation of knowledge in biology. But in fact matters do not work out as simply as that. In any case, the publication of an ever-increasing number of papers on biological subjects that contain mathematical models does not bring the structure of biological science any closer to the structure of physics. What are the reasons for this?

Mathematical models in biology, just as in many other branches of science such as psychology and sociology (Harran, 1963; Szaniawski,

1975), can be divided into two classes: descriptive and theoretical (in psychology and sociology the latter are sometimes called prescriptive models). In the first case, we have models for accumulating and compactly representing the experimental data. Such descriptive models are constructed without penetrating into the essence of the phenomena under study. In the case of observational results expressed by a matrix, it is possible, for example, to represent these data with the aid of the principal components, and if it turns out that the major part of the total variance is contained in the first components, we thus obtain a considerable compression of the observational data. It may happen that the first components can be easily interpreted theoretically, but this may also not be the case, and then we adopt the method of principal components as a formal procedure for reducing the amount of data to be studied. Even a simple calculation of the sample mean and of the confidence limits for these data requires a (possibly very simple) mathematical model. A long time ago, R. Fisher defined mathematical statistics as the science of reduction of data. Note once again that the reduction of data by statistical methods can take place without penetrating into the essence of the phenomena underlying these data. This is both the strength and the weakness of statistical methods. Statistical models of this type do not yield a compact representation of our knowledge about the phenomenon under study, but are merely a compressed form of the observational results. Such a compressed representation of data naturally sharpens the intuition of the investigator.

Theoretical models claim to reveal the "mechanism" of a given phenomenon. They are constructed as deductive structures based on clearly formulated premises. They are said to be prescriptive, since they prescribe a norm of behavior that follows from these premises. In psychology and sociology these models are not so much verified as they are compared with the actually observed forms of behavior, so as to ascertain the extent to which it differs from a strictly rational (in the words of Carnap) behavior in a fully determined situation (Szaniawski, 1975). In the vast majority of cases in biology, models of this type are never compared with actual phenomena. None of the articles devoted to mathematical simulation in biology and medicine published in *Voprosy Kibernetiki* (Problemy biomeditsinskoi kibernetiki, 1975) contains any hint concerning the verifying of models by comparing them with actually observed phenomena. Many mathematicians are inclined to believe that such verification is superfluous. This also has its good reason, because if we try to verify such a model by comparing it with observations, this will be entirely ineffective. This is due to the fact that, in verifying these models, we cannot subject them to a crucial experiment. The observed phenomena are usually structured in such a way that their actual

mechanism cannot be revealed. Thus, in studying the mechanism of a chemical reaction, it is very easy for the investigator to observe the time variation of the end product of a reaction, but it is very difficult to follow all the intermediate reactions. Matters are even more complicated in psychology, sociology, and (usually) biology. It almost always turns out that the experimental data can be adequately (in the statistical sense) described also by much simpler models. What, then, is the purpose of prescriptive models in biology?

The answer to this question is very simple. It is of interest, and often also very meaningful, to examine how a system would behave if its mechanism of behavior were as assumed by the designer of the model on the basis of premises that are to a certain extent probable, but nevertheless arbitrarily selected. Then it makes sense to compare, at least at the verbal level, this model of behavior with the considerations of the biologists dealing with this problem. Such a dialogue can be of interest. But even if it does not take place, the mathematician would still be interested in finding out what can follow from such (often very simple) premises. Indeed, sometimes the results are very interesting. For example, the periodicity of development of a population, well known to biologists, is a simple consequence of the properties of the model in the "predator-prey" problem.

Mathematical models of the first type can be understood on the basis of a fairly small amount of mathematical knowledge, and hence can be fairly easily grasped by biologists. In any case, the use of statistical methods in biological research is steadily increasing, and biometrics is a mandatory subject in biology departments. Matters are entirely different for models of the second type: most biologists are simply incapable of understanding them, since this would require a mathematical education at least equal to that possessed by physicists. At present we could not imagine physics, even experimental physics, without a mathematical foundation that would make it possible to understand the ideas of electromagnetic field theory, quantum mechanics, and relativity theory. But neither biologists, nor mathematicians dealing with the construction of mathematical models in biology, feel obliged to insist on a major extension of the teaching of mathematics to biologists. I. G. Petrovskii, one of the authors of the well-known paper "A Study of the Equations of Diffusion Accompanied by an Increase in the Amount of Matter, and Its Application to a Biological Problem," published in 1937, served for many years as rector to Moscow State University, but he did not try to raise the level of mathematics taught to biologists to equal that taught to physicists. The mathematics department has a course called mathematical biology, but no corresponding course exists in the biology department.

This has a simple explanation. Mathematical models of the second

type, which claim to describe the mechanism of phenomena, do not have such a wide scope as the mathematical models of physics. The logical applicability of such models is not sufficient to justify the intellectual effort needed for a serious study of mathematics.

A compact description of biological systems that would encompass the entire complexity of their behavior is impossible, since [as was pointed out very convincingly by Monod (1972)] the world of biological phenomena can and must be described not in terms of necessity but in terms of chance [from a slightly different point of view, we have also written about this (Nalimov and Mul'chenko, 1970)]. In other words, the complexity of biological phenomena is such that it cannot be described more concisely than by writing down all the observed phenomena. On the basis of the results of a short series of observations, we cannot write an algorithm that would express (even approximately) the subsequent evolution of the system.

Let us examine this assertion in detail. One of the peculiar features of biological systems consists in the possibility of observing such phenomena at two levels, as it were. One of them is the surface level, when the phenomena take place under certain steady-state external conditions; the second level is the deep gene–molecular level which manifests itself when the conditions of existence of the system change sharply. The phenomena taking place at the surface level are to a certain extent amenable to a brief description, but the knowledge about them is not of particular interest. Let me illustrate this with an example.

Suppose that we are studying a body of standing water, for example, a lake or even an ocean. The phenomena taking place in such a body while it is quiescent could no doubt be described in some compressed form by a system of differential equations (the parameters of these equations being the rates at which certain species eat up other species). It is true, though, that such models cannot be subjected to a crucial experiment when comparing them with the results of observations, and therefore the justification of such models is doubtful to many. But this has already been pointed out above. Here, let us note something else: if the conditions of existence of the body of water change, then a model describing the phenomena at the surface level will be useless. If sharp changes in the geological or meteorological conditions occur, or if a large amount of impurities (for example, oil) is poured into the water, then a random generator begins to act at the gene–molecular level that could not have caused major disturbances in the steady state. Processes take place at the molecular level that are due to structures of an opposite character, namely, mutations—expressions of randomness on which the rigid and unambiguous grammar of the genetic code is superimposed. Here we witness a profound analogy with the language behavior of humans, where the ran-

domness manifests itself in the fuzziness of the meaning of words; the grammar superimposed on this fuzziness is one of the forms of expression of Aristotelian logic. Note that a mathematical description of phenomena in biology is most effective in the description of the heredity laws, since in this case the object of the description is a strictly ordered grammar.

We can say that the nature of change in biology is random, since it is impossible to find an expression for a sufficiently detailed description that is considerably shorter than the "most complete" description of the observed phenomenon. In other words, it is not possible to construct a model of a generator of mutations in terms of ordinary cause-effect relations, i.e., it is not possible to find the causes that unambiguously generate the full diversity of observed mutations. Having found that the nature of change is random, we are greatly surprised that there does not exist an ordinary probabilistic description of the observed phenomena. An ordinary statistical description of phenomena is possible if, on the basis of the results of observations carried out on a small sample, we can calculate the distribution parameters which make it possible to obtain an idea of the behavior of the complete sequence of phenomena. In the case of biological changes, observations made on a small sequence of phenomena do not yield information about the subsequent behavior of the system. In such a case, averaged characteristics have no significance. The individual manifestations of the phenomena are important, irrespective of their probability of occurrence.

Since the processes take place over a prolonged period, and since they encompass a large number of biological entities (carriers of mutations), it can also happen that events of very small probability are realized. These events can have very great and entirely unforeseen consequences if, as a result of mutations, we obtain features that are adapted to new conditions. We have such a situation when bacteria appear in a stretch of sea polluted by oil that are capable of decomposing the oil, thus poisoning the water with the decomposition products. Another such situation is the appearance of microorganisms that are capable of "eating" antibiotics; there recently have appeared strains of organisms whose normal development requires streptomycin. This is an example of the many unpleasant surprises with antibiotics. The response of microorganisms to changing conditions is amazing, and it virtually takes place under our eyes. Bacteria have become resistant simultaneously to four medical drugs, namely, streptomycin, chloramphenicol, tetracycline, and sulfanilamide. The resistance of bacteria is accounted for by the so-called R-factor, which can instantaneously propagate over the entire population; it is not specific, and this is especially worrisome. The possible consequences of the danger arising from this are difficult to estimate (for more details see,

for example, Bogen 1967). In research (reported by my colleagues) carried out with patients suffering from chronic tuberculosis, phlegm cultures were treated with many drugs according to the rules of experimental design. The result was surprising, in that the combinations of drugs found to be critical were those that made no sense from a medical point of view. It is as though the microbes have guessed the thoughts of the doctors and do not perform as they are expected to do.

The difficulty in reducing our knowledge of biology to a compact form can be formalized as follows. On a field of elementary events we are given some events with a very small probability that are essential by their consequences. If one of them is realized, there appears another field of elementary events with another probability distribution. It hence follows that certain low-probability events may trigger other events with a high probability. For example, doctors assert nowadays that the appearance of even a single malignant cell is sufficient for the development of cancer. The probability of the occurrence of a single such cell in a certain organ of a person in a certain time interval is perhaps small. But once such a cell has appeared, the field of elementary events will change very rapidly, and with a high probability we can expect very well-defined ill effects. In the case of ontogeny we possess a large number of observations from repeated phenomena; therefore, we can make a probabilistic estimate of what will happen after the realization of some low-probability situation. In the case of phylogeny we do not possess any information about what will happen in a new situation realized as a result of some low-probability mutation. This can be reformulated in terms of conditional probabilities, but it will not facilitate our task. Monod (1972) has pointed out that, before the appearance of life on earth, its a priori probability of occurrence must have been equal to zero. Teilhard de Chardin (1965) attaches great importance to factors that would indicate that evolution develops toward less and less probable structures.

At present, rumblings can be heard that the existing language of mathematics is insufficient for describing biological phenomena, and it is necessary to elaborate a new (entirely separate) branch of mathematics especially suited to the simulation of biological problems. In my opinion, it is not the language that is at fault; rather, we have here a situation in which the past does not give us any information about the future. The complexity of the system is maximal, and in this sense it is a random system.

What can we say about the randomness generator that accounts for biological evolution? Where is it materially incorporated? What is its mathematical model? Is it equivalent to a one-dimensional sequence of numbers with a spectrum of the white-noise type, or is it something more complicated? We know all the difficulties encountered in designing a

random number generator for a computer; in this case we try to conceal our ineptitude by pointing out that we are dealing only with pseudorandom numbers. Anyone who simulates problems on a computer knows that special attention should not be paid to "fine" effects, i.e., effects that may be caused by a violation of the randomness in a sequence of random numbers. But how are fine effects accounted for in biology, by pure randomness or by a violation of this randomness? In this way we start an inadmissible play on words, since we know too little about the nature of chance to be able to go into details. If randomness is interpreted as maximum complexity, then all this discussion becomes meaningless.

For a long period of time, beginning with Aristotle and continuing perhaps until the end of the nineteenth century, the philosophers and many scientists were of the opinion that our desire to describe something in terms of chance is due to our ignorance. But at present, when we define chance as maximum complexity, is this not simply a reformulation of our previous assertion? It seems to me that this is nevertheless something much more essential, namely, a change in our paradigm – the acknowledgment that the impossibility of describing something is due not to our ignorance but to a complexity which does not lend itself to description in principle. Take a simple illustration. If we are going to transmit the text of the Soviet newspaper *Pravda,* we have to do it word by word, without omitting anything. The reduction of the text is not possible though its contents seem perfectly clear and familiar. Perhaps some will say that we have not advanced very far, since we still do not understand the essence of what we cannot describe but have merely found for it much more important reasons than a simple acknowledgment of our ignorance.

In any case, all the talk about the possibility of associating biological change with hard radiation and other such factors does not in fact explain anything. We are simply dealing with certain triggers that either turn on or speed up the operation of a random generator unknown to us.

For a better illustration of this analysis, consider an example of a complex structure generated by a random generator about which we know something. The well-known caves in New Afon of the Caucasus, which constitute an entire sequence of dwellings, appear to express a unity of thought, an inner harmony. Some of the caves resemble temples, and other are like antechambers of temples. One senses a frozen rhythm cut in stone. Here are also curtains that have stopped moving, like petrified streams that are fancifully entwined; then we can see stalactites and elusive beings on the walls, as on those of Notre Dame.

> The caves . . .
> They are sudden,

They are beyond expression . . .
Ah! measured now by the height of the arches,
now by long passages,
now by whimsical forms,
by their perfection
which is so fresh, complete, and mysterious,
so finely felt
that only breathing broken
when it comes deep from the bottom of your heart
and aspires aloft
might express the sensation
which does not yield to words,
an inseparable welding of ecstasy and tremor
pronounced by the soul
which has submitted to monumental
and sublime silence,
so tangible and solemn
that you begin to feel the supreme work
going on in the caves
under the cover of silence.
You begin to hear its rhythm taking shape of a hymn
and you start to make out the words
"Let the flesh keep silence" (so that it might become
the witness to the miracle).

I shall attempt no further description of these caves: words cannot interlace so whimsically and marvelously, so silently and solemnly.

These caves were generated by two different factors: on the one hand by the physicochemical processess of dissolution and crystallization which are amenable to a rigorous description in terms of cause–effect relations, and on the other hand by randomness due to the inhomogeneity of the mountain rocks, their spontaneous movement, and the long-term seasonal changes in meteorological conditions. We could say that the language of physicochemical relations is used for reading the information from a random generator which represents the behavior of the lithosphere. The result of this reading manifests itself in the text, i.e., the caves. Here we know something about the location of the random generator, though it is difficult for us to imagine its mathematical model. In any case it is obvious that it must also contain fragments of periodic components related to meteorological processes. With regard to randomness, I shall note here only its property of complexity; i.e., neither the caves nor the factors generating them can be attributed to some simple process. Our description of them could not be much simpler than the possible description of the entire manifold of phenomena. The foregoing could be interpreted also as follows: We have in front of us an artistic architecture that came about by the play of chance and, more importantly,

which appeared spontaneously, without sifting of the best versions from the set of randomly generated versions, as in biological evolution. The architectonic unity prompts one to think of teleology, but from the point of view of common sense this thought is quite absurd.

Now let us return to the biosphere. What could we actually say in this connection with regard to a generator of randomness? Is this simply the introduction of errors according to the laws of roulette into texts previously written in the language of the genetic code, or does it perhaps consist in a discrete reading of texts from information flows whose complexity is such that they could be described only in terms of randomness? If the entire evolution consists only in the introduction of errors into a previously written text, then this text must have been very complicated at the moment of its inception. In an earlier paper (Nalimov, 1979), I put forward the hypothesis that our intellectual activity takes place at two levels, namely, at a discrete-logical level of language, about which we know a lot, and at a continuous (extralogical) level which we could imagine as a continuous stream of consciousness. There exists a direct contact with this continuous stream during sleep, in certain hypnotic states, during creative activity, and in religious meditation. Our everyday language behavior is constructed in such a way that, in my opinion, the interpretation of our utterances takes place at the extralogical continuous level. All this is set forth in detail in the paper mentioned (Nalimov, 1979). Here I merely intend to draw attention to a possible analogy: If the creative process of human thought consists in discrete reading from a continual flow, then could it not be that biological evolution which leads to the appearance of new "texts" in the biosphere is a discrete reading from this same flow? If this is the case, then we have a profound analogy between the process of creative thought and the development of the biosphere.[4] Our idea that chance is maximum complexity also admits such an interpretation of the nature of a randomness generator. All this, of course, could be dismissed as belonging to the realm of fantasy, but nowadays it is customary in the philosophy of science to claim the right to put forward hypotheses as fantasies (free assumptions) that can subsequently be discarded if found unsuitable.

In any case, we have now reached the stage when it is necessary to try to formulate some (even very hypothetical) ideas about the nature of a randomness generator in biology. Human activity was found to be directed toward the creation of conditions favorable for the starting of a randomness generator. Such conditions appear, on the one hand, as a

[4] This is in agreement with the ideas of Teilhard de Chardin (1965) concerning the noosphere (the thinking layer of the universe) and evolution as an arrow of biogenesis directed toward the highest point "omega" that is the final accomplishment of everything. The contradiction between the probabilistic nature of mutation and Berg's nomogenesis (1969), assumed in the literature, disappears.

result of changes in the conditions of life over large areas of land and ocean and, on the other hand, as a result of the effect of antibiotics on humans, as well as the effect of strong chemical substances. In the near future we may also witness direct intervention in the genetic structures of life. Modern biology is not in a position to foresee all the consequences of these developments. For the time being, there is perhaps only one way to obtain an answer to these questions, and that is by carrying out direct experiments on biological systems that are subjected to unusual conditions. This could be done by extensively using the methods of the design of experiments, so as to appropriately plan research which simultaneously involves a large number of independent variables. But the trouble is that such experiments can be carried out only on a small scale (in space and time). Therefore, we cannot expect to obtain from them information about similar situations realized on a large scale, when low-probability events can manifest themselves that (as noted above) can radically change everything.

Perhaps we are now in a position to give a formal definition of life: living systems are systems which are random in their essential manifestation, but in this case the randomness (in contrast to inanimate nature) is such that extremely improbable events play a decisive role. Thus, we are very near not only to ontogeny and phylogeny but also to the creative manifestations of human activity. Modern science, including such branches as probability theory and mathematical statistics, is not yet ready to deal with these structures.

The Role of Computers in Efforts to Describe the World Scientifically[5]

The progress of computers has given rise to many an illusory hope. One of these is the conviction that it is possible to construct extremely complicated mathematical models based on an elegant analysis of crude experimental data. Computers were thought to be like "mathematical spectrographs" which possess an extremely high resolvency with which they decompose the experimentally observed data into the components which are not immediately observed in the experiment. It would be hard to indicate a paper where such an idea was explicitly formulated and seriously grounded. It seemed to emerge by itself and has been and still is shared by many. This is a *paradigm* brought forth by the unprecedented capacity for computational mathematics. But, sooner or later, any paradigm is subjected to revision.

[5] A brief version of this section was published in the journal *Industrial Laboratory* (No. 3, 1978).

I shall analyze the situation using the problems of chemical kinetics. The researcher observes time changes in the output of the final product in a chemical reaction and wishes to reconstruct the mechanism of intermediate reactions which are not immediately observed in the experiment. He proceeds from the assumption that on the basis of some prior knowledge he can write a system of differential equations which would reflect the mechanism of intermediate reactions and give the model parameters. The prior information is not formalized; it includes the researcher's previous experience and, probably, the whole progress of chemistry (which depends on the researcher's level). Prior knowledge is always subjective since it reflects the attitude either of the researcher himself or of the school to which he belongs.

A system of differential equations may be regarded as embodying initial axioms, on the basis of which one should decompose experimentally observed data into their components. The list of reactions is thermodynamically open. Their number seems to be wholly determined by the technology. At present, one may come across papers including up to twenty parameters. Some time ago, when computers were less efficient, the number of intermediate reactions used not to surpass two. I think it would be instructive to build sciencemetric graphs showing the growing number of intermediate reactions analyzed as computers become more and more intricate.

The tendency is, at least, obvious: the progress of computers gives rise to a temptation to create models whose complexity depends on the computing techniques. But since the progress of the latter may be practically infinite, the complexity of models should also increase infinitely.

We have by now accumulated a broad experience which points out the principal difficulties one faces when trying to use computational methods to decompose the initial experimentally observed data into components. The problem which arises while measuring radioactive decay is well known. In this case one deals with the function

$$f(x) = A_1 e^{-\lambda_1 x} + A_2 e^{-\lambda_2 x} + \ldots + A_m e^{-\lambda_m x}$$

where the parameters of decay λ_i and activity A_i are unknown. At first sight, the problem is similar to that of decomposing the oscillation process registered as a whole into separate periodic constituents, but the lack of orthogonality for the exponential functions results in difficulties unknown in harmonic analysis. These difficulties were long ago described in texts (see, e.g., Lanczos, 1956).[6]

[6] Lanczos (1956) gives an illustration of the case when observational results generated by the sum of three exponential functions proved to be well approximated by the sum of two exponential functions, and the decay parameters for one index decreased from 3 to 1.58, and for the other, from 5 to 4.55 while

One also has to resort to methods of "mathematical spectroscopy," analyzing nonadditive multicomponent systems according to their absorption spectra obtained by means of physical spectrographs (Vasil'ev, 1976). Here several essentially different problem formulations are possible. One of them is an analysis of multicomponent mixtures with partially known composition. In its ordinary form, the problem is solved by normal regression analysis approximating the absorption spectrum of the unknown admixtures by an algebraic polynomial; at every stage of selecting a model it is necessary to solve a system of linear equations with $n + S + 1$ unknown quantities, where n is the number of components analyzed and S is the polynomial's power. Its form turns out to be equivalent to a common least-squares method which, instead of covariance matrix, includes special matrices calculated beforehand, which possess the following projective properties:

$$\bar{\mathbf{c}} = (\tilde{\mathbf{k}}\mathbf{P}_i\mathbf{k})^{-1}\mathbf{k}\mathbf{P}_i\mathbf{D}$$

i.e., at every stage of selecting a model, one has to solve a system of linear equations with only n unknown quantities. Here, $\bar{\mathbf{c}}$ is a concentration column, $\mathbf{D}$ is an optical density column, $\mathbf{k}$ is a matrix of absorption coefficients, and $\mathbf{P}_i$ is a projection matrix built on algebraic polynomials up to the ith power inclusive. The effect of applying such a projection matrix is similar to that of a low-frequency filter which lets through only high-frequency constituents (with the number larger than i) in the expansion spectrum of a target function with respect to the given basis of algebraic polynomials in Hilbert space. All problems of this sort are fraught with troubles typical for a non-orthogonal regression analysis. For the component we are interested in, it is impossible to fix confidence limits which would not depend on the choice of all the other components. This results in the absence of a unique solution of the problem.

Of course, there are many cases when such troubles may be neglected. One of the principal problems of "mathematical spectroscopy" is to outline clearly its applicability in various real situations.

But let us return to constructing models in chemical kinetics. The process passes through the following stages. The system of differential equations is analytically integrated. There emerges the function

$$\eta = \varphi(\mathbf{x}; \Theta)$$

---

the amplitudes ratio was distorted from 1:2 to 1:7. In a simplified problem, when the number of components and the approximate value of every index are known and the amplitudes A_i and corrections to the approximate estimates λ_i are to be found, decomposition results are essentially better though they still remain disappointing. At worst, instead of the true value 1.0, λ gets the value 0.5, and it proves less adequate than the preliminary estimate with $\lambda_i = 1.2$.

nonlinear with respect to parameters. The task is to elaborate the optimal experimental design generating the matrix of independent variables $\mathbf{X}$, to realize it, and then, using the least squares method, to estimate parameters Θ, to find confidence limits for the estimates, etc. All this is naturally preceded by linearization. Assume that we confine ourselves to representing the function by a Taylor polynomial in the neighborhood of the point Θ_0. In this case, while designing an experiment and evaluating its results[7] we shall have to deal with the covariance matrix $(\mathbf{X}^x\mathbf{X})^{-1}$ obtained from the independent variables matrix $\mathbf{X}$ of $N + k$ dimension (N is the number of experiments, and k is the number of parameters estimated)

$$\mathbf{X} = \{x_{r,u}\}$$

where the element of the matrix

$$x_{r,u} = \left[\frac{\partial \varphi(\mathbf{x}_u; \Theta)}{\partial \theta_r}\right]_{\Theta = \Theta_0}$$

is a partial derivative with respect to the parameter θ_r in the point $\Theta = \Theta_0$ under the fixed values $x_{1u}, x_{2u}, \ldots, x_{ku}$ corresponding to conditions of the uth experiment.

The optimality criterion for an experimental design may be formulated as follows. Points in the space of independent variables should be situated so that the determinant $|(\mathbf{X}^x\mathbf{X})^{-1}|$ will be minimal. The functional determinant (a Jacobian) $|(\mathbf{X}^x\mathbf{X})^{-1}|$ gives the transformation of coordinates of experimental space (the space with coordinates $\eta_u = \varphi(\mathbf{x}_u; \theta)$ into the space of parameters. The minimal Jacobian would mean the approximately minimal volume of a k-dimensional ellipsoid in the space of parameters determining confidence limits of the parameters. Design efficiency depends here on the values of partial derivatives. Therefore, designing an experiment for functions nonlinear with respect to parameters requires knowledge of preliminary parameter estimates. One can apply sequential methods of experimental design when all the research is divided into separate stages, parameters are estimated for every stage, and these estimates are used as a new preliminary approximation for the next stage. The optimal design is each time calculated anew. But the calculations may prove so complicated that they will reduce to nothing the gain from the optimal sequentially improved design.[8]

[7] I remind the reader that under matrix denotation the vector column Θ is given by the relation $\Theta = (\mathbf{X}^x\mathbf{X})^{-1}\mathbf{xy}$ where $\mathbf{y}$ is a vector column of observational results.

[8] Designs minimizing the volume of the variance ellipsoid of parameter estimates are called D-optimal. Despite its obvious nature, the criterion does not completely determine behavior of parameter estimates: the volume of the ellipsoid may be minimal, but the latter may be too stretched along an axis. If the

If there exist several alternative models capable of describing the mechanism of the phenomenon, discriminating experiments may be carried out to select the best one. A new problem formulation generates new optimality criteria. One of them is based on a measure depending on the difference between sums of square deviations. For two rival hypotheses, after carrying out N observations we shall deal with the difference of two values

$$S_j(N) = \sum_{i=1}^{n} [y_i - \eta_j(\mathbf{x}_i, \Theta_j)]^2 \qquad (j = 1, 2)$$

If the difference is not large enough to give preference to one of the rival hypotheses, it is suggested that the next $(N + 1)$th experiment should be held at the point where it is expected to reach its maximal value.

The second criterion is based on the use of the modulus of the logarithm of the generalized likelihood ratio calculated for two rival hypotheses: measurements are situated so as to achieve the most rapid increase of the value. Last but not least, the third criterion exploits, as a measure for discriminating hypotheses, the Kulback measure of divergence, well known in information theory: the choice of an optimal design consists in maximizing the value. A dual problem formulation is also possible: a design may be chosen with the goal of either estimating parameters or discriminating methods. The optimality criterion is here based on the necessity to minimize a weighted sum of two addends, each

researcher wishes to minimize the maximal axis of the ellipsoid of variance, he should build a design matrix which would have a corresponding covariance matrix with a minimal eigenvalue. This will be the so-called E-optimal design. The researcher may require the average variance of parameter estimates to be minimal. To this requirement corresponds an ellipsoid with the least sum of square axes length. Corresponding designs are called A-optimal; the respective covariance matrices have a minimal trace value. The list of criteria is easily continued. If we deal with simple polynomial models where sequential designing is irrelevant, we can build designs to meet various criteria, estimate them from the standpoint of other criteria, and then find a solution satisfying various approaches. For details see, for example, Korostelyov and Malutov (1975). When models are nonlinear with respect to parameters, it is wise to follow the course of sequential designing, but in this case computation of admissible designs (with respect to criteria) becomes an extremely complicated task. So one has to give preference to a certain criterion. The highly readable review by Kafarov et al. (1977) briefly discusses the question of selecting a criterion in problems nonlinear with respect to parameters; e.g., attention is paid to the fact that the D-criterion, descreasing significantly the volume of the confidence ellipsoid, does not change correlation coefficients and, therefore, does not improve the ravine surface (unsmooth surface without a distinct extremum) of the sum of square deviations. It is also of interest that a number of illustrations have shown that the volume of a hyperellipsoid of the confidence region decreases mainly at the expense of compressing it along the minor semi-axes, which are smaller than the major ones in the same order of values. At the same time, the aim of the E-optimality criterion is to make the confidence region close to a sphere. But all these observations made by Hosten for separate cases are, naturally, insufficient to allow an ultimate conclusion for a general case. In selecting an optimality criterion, one should learn to take into account the type of a model.

of them being responsible for one problem formulation. The difficulties of problem formulation are then shifted to the choice of weights. [The reader is referred to Nalimov (1971) for a brief and popular description of the questions concerning the choice of criteria in discriminating a problem and to Fedorov (1972) for a detailed and rigorous account.]

Now we pass to considering confidence limits for parameters in the models with nonlinear parametrization. The problem of confidence limits becomes crucial since parameters are regarded as physical constants inherent to the mechanism of the phenomenon described by the model nonlinear with respect to parameters. The constants have an unambiguous physical interpretation (counterparts) and their numerical estimates are to be stored in reference books or in the memory of computers operating in informational systems.

It is exactly at this point that practically insurmountable difficulties arise which often turn the whole process of parameter estimation into an illusory activity. Having analyzed the process of studying phenomenological mechanisms by mathematically expanding the total data according to the parameters of a hypothetical model, one can discover three sources of uncertainty. The first one is the random error in measurements, the second stems from the fact that model parameters cannot be estimated in a unique way, and the third source lies in the possibility of giving diverse formulations of the chemical axiomatics determining a set of several intermediate reactions.

I would like to start the discussion with the first type of uncertainty. The essential thing is that the estimates of some parameters correlate highly (their correlation coefficients reach sometimes .99 or even .999). In such models, applying an experimental design to orthogonalize the information matrix proves futile. In this case it is of no use to record confidence limits as the familiar relations $\hat{\theta}_i \pm 2\sigma\{\hat{\theta}_i\}$, since such a step ignores the information concerning correlation of estimates. When looking through a reference book containing results of other studies, the researcher, using the theoretical data at his disposal, may wish to accept as a parameter estimate any value situated close to the boundaries of the confidence interval. In accordance with statistical theory the researcher has every right to act this way and to choose at will within the given interval of confidence. But because of highly correlated estimates, he would immediately have to recalculate confidence intervals for all other parameters. To make this possible, reference books would have to include, together with parameter estimates, the corresponding covariance matrix, which is very cumbersome. Moreover, the reference book as an instrument of visual data presentation loses any utility because one has to resort to a computer. What can be done to preserve a visual type of presentation? Even if eigenvalues of the covariance matrices which gave

the axes of the variance ellipsoid of estimates are included in the reference book, presentation of a multidimensional ellipsoid will not become more visual. It is also possible to circumscribe a multidimensional parallelepiped around the ellipsoid. The presentation will acquire a more visual character since every parameter will be given maximal confidence limits independent of how the ultimate values for other parameters are fixed. But that will cost us dear since in a multidimensional case the volume of the parallelepiped will be several times larger than that of ellipsoid, especially because in a nonlinear parametrization we deal not with ellipsoids but with an ellipsoid-like shape which in a two-dimensional section looks oddly banana-like. At last, we may introduce in the reference book two-dimensional sections of ellipsoid-like figures, though such data will be too bulky. For a problem with 20 parameters, the number of sections equals 190, and each of them should be accompanied by a development.

So it is obvious that, in the case of a nonlinear parametrization, the uncertainty stemming from the experimental error is not at all easy to present visually. And this is only one source of uncertainty.

Now let us consider the second source of uncertainty: lack of a unique computation resulting from the complexity of calculation procedures. Here we should keep in mind several factors. (1) Linearization procedures may be varied: a Taylor polynomial may be given differently—once on the basis of only the first derivative, another time introducing the second derivatives as well, etc. Linearization may also be achieved without expanding into Taylor series but for each model choosing an expansion according to some functions natural for it. (2) When estimating parameters, we can get different results because an information matrix proves poorly conditioned, i.e., its determinant is close to zero, as a result of the high correlation of several parameters. (3) Computations may yield different results because the initial parameter estimates have been chosen arbitrarily. All iteration procedures require selection of the initial point. If, as usually happens, besides the absolute minimum, there exist a lot of others, poorly chosen initial estimates may lead to convergence in an undesirable stationary point on the surface of the sum of squares, or there may be no convergence at all because calculation results will spread on the surface where the stationary point is being searched for.

The well-known book by Draper and Smith (1966) describes the troubles connected with a linear least-squares method (after linearization of the model by expansion into a Taylor series) considered in the sequence of its stages in the following manner. The computation procedure may converge very slowly; strong oscillation with partial increase and decrease of the sum of squares may arise, though, in the long run, the solu-

tion may become balanced. As a matter of fact, the procedure may not converge or it may even diverge so that the sum of square deviations will grow with each iteration, though it is possible to show this method to be always convergent (the rate of convergence depends on the model whose parameters are estimated; the process may also converge slowly, and with strong oscillation). For more details, see also Bard (1974). If the computation procedure selected to solve the problem does not converge, generally speaking, the researcher knows how to act. He may change the linearization procedure[9] or, having rejected it altogether, he may change the computation procedure; within the frame-work of any such procedure, initial approximations may be varied, and this process can be formalized to a certain degree by selecting a combinatorial lattice on the basis of experimental design. Thus, for example, initial approximation for each parameter may be varied at two levels, and then points in the factor design (k is the number of parameters) or its regular replication will be the nodes of the lattice.

In a sequential design, when the initial approximations at the $(i + 1)$th step are the values obtained at the ith step, the role of the initial approximation received from the first rough preliminary experiment seems to decrease, but it hardly is eliminated for the squares sum surface of any type.

Sometimes the researcher has to take drastic measures—either to reparametrize the model or to record it in an absolutely nontraditional measure, the way it is done by Gontar' (1976). Sometimes one can manage to unite two highly correlated parameters, modifying the model only slightly so as not to violate its physical meaning. In any case, it has been noticed that a slow convergence takes place when the contour curve for the sum of square deviations has a shape of stretched bananas.

All this is well described in numerous books, including the book by Draper and Smith (1966) mentioned above. And the research is always assumed to make the best and the only right decision. But whence does this come? Not only the answer to this question is lacking—the question itself is not asked. We dare claim nowadays that, in the problems of non-

[9] Expanding the function with respect to the powers of independent variables may also be considered as linearization. This is equivalent to presenting it as a Taylor polynomial with coefficients

$$\beta_1 = \frac{\partial \varphi}{\partial x_1}, \beta_2 = \frac{\partial \varphi}{\partial x_2}, \ldots \quad \beta_{12} = \frac{\partial^2 \varphi}{\partial x_1, \partial x_2}, \ldots \quad \beta_{11} = \frac{\partial^2 \varphi}{\partial x_1^2}, \beta_{22} = \frac{\partial^2 \varphi}{\partial x_2^2}$$

which get numerical values after the experiment. In contrast to the case considered above, here, during linearization, derivatives are taken with respect to independent variables and not to parameters. If the researcher is satisfied with approximating the function by a first-order polynomial, then all regression coefficents can be estimated with zero correlation coefficients. All troubles connected with parameter estimation and their confidence limits disappear. But the initial parameters of the model disappear too; they are replaced by pseudoparameters, that is, regression coefficients. *Hence it follows that the less we wish to learn, the more definite becomes our knowledge* (the experimental potentiality being the same).

linear parametrization, statistics has lost one of its most attractive merits. It stopped being at the same time a science setting the rules for parameter estimation and a metascience estimating the reliability of its results. In any case, confidence ellipsoids, even if they could be visually presented, reflect the uncertainty stemming from random error and in no way reflect the uncertainty related to nonstandard computational procedures. One and the same problem, nonlinear with respect to parameters, can obviously yield essentially different solutions in different and equally good computational centers. Moreover, provided a good level of critical attitude, it is possible to obtain different results within one center.

The uncertainty stemming from nonstandard computational procedures, as a matter of fact, may be estimated in a purely statistical manner using, for example, variance analysis with hierarchical classification. The lowest level would contain the uncertainty due to choosing initial approximations, the highest level would be occupied by the uncertainty related to choosing models during their reparametrization, and the intermediate levels, by all other nonstandard computational procedures. But the whole structure would have a cumbersome appearance. It seems rational to limit ourselves merely to presenting observational results in several versions—naturally, only in those which a computer-mathematician thinks fit for this purpose.

If we look through journal publications concerning models nonlinear with respect to parameters, we shall notice an amazing disregard of uncertainty estimates. A covariance matrix is only very seldom given as a whole. Usually its diagonal elements are given, and this, as I have already mentioned, would be correct only for an absolutely orthogonal information matrix. As a rule, in such papers one will not find anything about the uncertainty due to computational procedures. A welcome exception is the article by Korostelyov and Malutov (1975), but despite their criticism the authors present observational results by a single model.

And now we shall at last analyze the third source of uncertainty, initial chemical axiomatics. I remarked above that the list of differential equations giving the mechanism of intermediate reactions is thermodynamically open, and the researcher selects his own rationally limited list of possible intermediate reactions based on the micro-paradigm of the scientific school to which he belongs. A researcher with a critical mind may suggest several such lists, and this accounts for the problem of discriminating hypotheses mentioned above. A complicated apparatus for discriminating procedures implicitly introduces a new source of uncertainty into results of investigations.

The first thing to be noted here is that we are not (and generally cannot be) sure that the rival models include the "true one." But without this assumption, the formulation of a discriminating problem proves ground-

less. Discriminating procedures will yield different results if the sets of hypotheses to be discriminated vary. In any case, the problem of a discriminating procedure converging to the "true model" no longer has any sense.

The second point is that a discrimination procedure may be wholly determined by the arbitrariness in parameters estimation introduced at the stage of computation. Recent experience has shown that not all models selected in the laboratory prove fit to describe the corresponding processes at a plant, though formally the models for all processes have been selected with equal rigor.

The third feature is that a model as a whole is variant with slight modifications in the initial chemical axiomatics. Such a modification of the list of intermediate reactions results not only in the appearance of new model parameters but also, as a consequence of high internal correlation among parameter estimates, in the change of numerical values of the parameters characterizing the unmodified reactions. It is practically impossible to try all possible combinations of intermediate reactions, and besides, again due to high internal correlation, extensive change of numerical values may lead to models which prove almost the same when compared with experimental results in a fixed interval of independent variable values.

The fourth peculiarity is that extrapolations are not correct. The very essence of chemical problems often enables discrimination of hypotheses in a narrow and easily achieved interval of variation so that further significant information could be obtained by extrapolating the best model of those selected. Such an approach is hardly correct since it follows from what has just been said that, while discriminating, we only estimate the interpolational power of a model, and a model whose interpolation properties for a narrow interval are quite good may at the same time be of rather poor extrapolation power. Some examples borrowed from practice show that the best model for extrapolation has been that estimated as the worst one in a discriminating experiment carried out as rigorously as possible.

The fifth point is a question. If, discriminating in a narrow interval, we find a model to behave not in the best possible way, is this enough to reject it? A slight modification of the initial chemical axiomatics might weaken the interpolation power of the model, but this is not to say that it does not reveal certain changes in parameter estimates (again due to their high correlation) which may come out in a wider but practically unrealizable range of independent variables.

Here we may end our list of claims to discriminating procedures. Now the question can be formulated in a more general philosophical form: Whence comes the whole theory of discrimination? The answer is very

simple: in the terms of Kuhn (1970*a*), its source lies in the existing paradigm, which states that the world, on the one hand, is arranged so that everything within it is governed by the only possible laws of nature and, on the other hand, it also possesses such a property that a scientific experiment allows us to discover these laws.

The first part of the paradigm has begun to lose its supporters in modern physics. It is opposed by the "bootstrap" philosophy (Chew, 1968; Capra, 1976), which holds nature to be an interrelated dynamic web irreducible either to elementary blocks of substance or to fundamental laws, equations, or principles. The term "bootstrap" cannot be related to a single model; it can only be applied to a combination of internally consistent models among which none is more fundamental than the rest. Consider the multitude of models in the physics of elementary particles (Moravcsik, 1977). The past 20 years have witnessed the emergence of numerous theories and models, some of which are conceptually contradictory of others. None of them can be rejected because each explains a part of the observed phenomena and none of them can be accepted as the only one because none can explain everything. I draw the reader's attention also to a very interesting paper by Smirnov (1977*a*) in which the problem of plurality in models in physics is discussed.

As to the second part of the paradigm, after the well-known work of Karl Popper (1963, 1965; see also Chapter 1), it has also become philosophically clear that the role of an experiment in science is limited: a hypothesis can never be experimentally supported. The only thing that can be done is to show that the experiment does not contradict the hypothesis. But the same experiment may prove consistent with some other hypothesis, as yet unformulated, and a new experiment carried out to confirm a new theory may become crucial for a hypothesis consistent with previous experiments. Any hypothesis not refuted by an experiment remains open to further tests, and, according to Popper, here lies the source of progress of natural sciences.

But if a hypothesis is refuted by an experiment, is it always rejected immediately and unconditionally? Above, in Chapter 1, I borrowed an example from Monod (1975) which described an awkward situation with Darwin's theory. Thomson (Lord Kelvin), the physicist, demonstrated by means of exact calculations that solar energy could not suffice for the evolution of life on the Earth. Darwin was depressed by these calculations. A direct experiment—measurements of heat received by the Earth, of the dimensions of the Sun, and of fuel caloricity—came to contradict his theory. However, the latter was not rejected. Monod remarks that at present we may state that Darwin's evolutionary theory implicitly contained the concept of solar nuclear energy though nobody could have had such an idea at that time. Besides, adds Monod, Darwin's theory also im-

plied the concept of a discrete biological code, contrary to Lamarck, who assumed the continuity of heritability.

One can also cite numerous examples of an opposite kind which demonstrate that negative results have been of the utmost importance for the development of science. One of them is the famous experiment of Michelson and Morley, which gave an impetus for a new era in physics.

From all these historical contradictions, it obviously follows that negative experimental results acquire significance only when combined with a system of meaningful reasoning. Statistical methods of model discrimination as they are presented in books on experimental design are too formalized. They exclude a meaningful discussion of experimental results, and at the same time they prove inconsistent as a result of their formalism since they ignore the uncertainty caused by the computational difficulties generated by the structural peculiarities of the models. But this has already been discussed in detail.

Thus, the system of our initial concepts should evidently be modified. The researcher's new paradigm must not presuppose the existence of the one and only true model, even if he investigates the mechanism of phenomena. Why should the researcher assume the existence of what proves illusory in the long run? He might better follow the example of the physicists who, in quantum mechanics, abandoned Laplacian determinism and even its weaker forms after the illusory nature of its serviceability became obvious.

Besides, the results of studying the mechanisms of phenomena should be presented not by a single model but by several of them. The variety of models may result both from the different initial chemical axiomatics and the insurmountable absence of unique computational procedures. It seems rational to acknowledge the possibility of performing discriminating experiments since they are rather informative, but they should be given only a limited significance.

If we decide to hold this viewpoint, we shall immediately contradict the traditional view on the role of mathematical statistics in research. Like Ronald Fisher, I long believed that the task of statistics was the reduction of data. A statistically trained researcher is able to present his results in a much more compact form than if he registered experimental data directly as they were obtained. Under a new problem formulation, *mathematical statistics* will be used not to *reduce* data but to *unfold* them. Numerous models nonlinear with respect to parameters with their confidence limits development of two- or three-dimensional sections of ellipsoid-like figures will look more complex and cumbersome than the immediately observed values—an independent variables matrix **X** and a vector of observational results **Y**.

But will the researcher then be able to perceive the information about

the process he studies when it is presented in such a cumbersome way? Imagine an audience looking at the screen and seeing there the filmed variety of graphic data accompanied by certain comments. The researchers will watch all these data extracted from the experiment while varying both computation procedures and initial chemical axiomatics. This process may also be presented as a human–computer dialogue. The important thing is whether the researcher has an insight that will allow him to present the mechanism of the phenomenon in a new way and to outline further research. In other words, whereas the reduction of data used to be performed on the logical level, the application of statistical procedures has transferred it to the intuitive level. A computer unfolds the information contained in the experiment; a researcher will have to reduce it while comprehending it theoretically. The potentialities of a computer meet some unexplored human potentialities, and they switch roles: humans are now to meditate over the free information flows generated by computers.

It seems pertinent to draw an analogy with what is now happening with *the foundation of biology, the theory of evolution.* At present there exists a set of evolutionary theories which are hard to classify. They may be said to contain as a basis a list of "evolutionary factors." The variety of theories is formed by ascribing weights to these factors. Theories with similar factor weights are naturally unified into groups, and these groups are given a label. The mechanism of constructing evolutionary theories may be called "logical spectroscopy." A biologist or, to be more correct, a paleontologist observes the sections of the distant past and attempts to decompose them according to all evolutionary factors by ascribing weights to the latter. The beginning of our century, up to the 1920's witnessed the emergence of novel evolutionary theories. The process has by now slackened or stopped altogether. A theoretical biologist behaves in the following manner: after becoming familiar with the whole variety of evolutionary theories, he constructs his own, i.e., ascribes new weights to the set of known evolutionary factors. As a result, we have as many theories as there are theoretical biologists. This variety, individually reduced to compact homogeneity, opens up the possibility for individual, creative work. [The ideas concerning the theory of evolution are stimulated by an extremely interesting report, "Classification of Evolution Theories," made by S. V. Meyen at the School of Young Scientists in Theoretical Biology in Kondopoga, February 1977. See also his paper (Meyen, 1975).]

I would like to conclude this chapter with the following considerations.

One of the principal scientific tasks is to explain observational results according to the underlying factors (or mechanisms). The vast data ac-

cumulated up to the present show that this task in its general formulation cannot be unambiguously fulfilled either by the methods of "numerical spectroscopy" or by those of "logical spectroscopy." Computers have only made the task more difficult. But now a new way seems to emerge: *reducing information by means of unfolding it,* i.e., presenting it through a set of models. However, this may be another illusion.

A few words must be said about another approach to simulating complex systems by computers, which does not refer to "mathematical spectroscopy." I have in mind the grandiose program of simulating five ecosystems: the desert, coniferous and foliage forests, the tundra, and the prairie, carried out in the United States in 1969–1974. Expenses for the research of only the three latter systems exceeded 22 million dollars, 8.6 million of which were allotted directly for the simulation, synthesis, and control of the whole project; 700 researchers and postgraduates from 600 U.S. scientific institutions participated in the project; 500 papers were published by 1974, though the final report is not yet ready. Mathematical language was used in the project to give an immediate (not reduced) description of the observed phenomena. A lot of different models were used which were divided into blocks with an extremely great number of parameters (their total number reaching 1,000), and yet it was emphasized that the models described the system under study in an approximate and simplified manner. The researchers had to give up any experimental verification (or falsification) of models; instead they used "validification," which means that a model is accepted if it satisfies the customer and is particularly favorably evaluated if bought by a firm. At present, all these activities are being evaluated. Mitchell et al. (1976) conducted a thorough analysis of the material and evaluated the simulation of the three ecosystems in an extremely unfavorable way. From a general methodological standpoint, the following feature is important to emphasize: the language of mathematics, for the first time, is allowed to unify different biological trends, and this has happened without a generally novel or profound understanding of ecology. Mathematics was used not to reduce complexity but to give a detailed, immediate description. This is a new tendency in science. But where will it lead? The time is not yet ripe for a final conclusion, but scepticism is quite in order.

Concluding Remarks: Dialectics of Reduction and Expansion of Knowledge in the Development of Science

We are unintentionally witnessing an amazing phenomenon. Up to now, the development of science was directed at obtaining a form of

knowledge with maximum reduction of redundancies—a compact representation of the world. But now virtually before our eyes, there appears a tendency to present knowledge in an expanded form, by a multitude of equally legitimate models. Each individual model implies the presentation of knowledge in reduced form, but the acknowledgment of the legitimacy of many models describing one and the same phenomenon allows the presentation of our knowledge of the world in an expanded form. This is where the dialectics of the development of science is manifested.

The necessity to resort to many models first became evident long ago in the problems of multidimensional statistical analysis. The passage to the principal components is definite if the metric of the initial variables is given, but it may be given in various ways. Factor analysis is indifferent toward the initial metric, but the rotation of axes, obtained after a corresponding transformation, is performed arbitrarily (for more detail, see Nalimov, 1971). In non-orthogonal multidimensional regression analysis, the same experimental results may be equally represented by the entire diversity of models (an instance of such analysis is given in Chapter 1 of Nalimov and Chernova, 1965).

From the general methodological point of view, though, it was possible to ignore these facts, by assuming that these are merely particular cases related to situations without sufficient initial theoretical premises. However, the contents of the above section make the matter look more complicated. Even if we are provided clearly formulated initial theoretical premises of the mechanism of phenomena, we still have to resort to many models. Moreover, if we compare all this with the prevailing situation in the physics of elementary particles, the problem of an expanded representation of knowledge will acquire a threatening dimension.

Back to Contents
Next to Chapter 9