Back to Contents
Next to Chapter 4

Chapter 3

Mathematics as a Language of Science[1]

Using Mathematics to Describe the External World

> *The two elements can be distinguished by reflective thought, but cannot be rent asunder.*
>
> SATKARI MOKERJEE[2]

Introduction: Formulation of the Problem

The language of mathematics is an oft-repeated word combination, a sort of language cliché.[3] I would like to analyze this phrase and reveal its content.

It is convenient to start studying a phenomenon where it acquires its most explicit form. Mathematics as a language gets especially explicit while describing the phenomena of the external world. To my mind, *applied mathematics is use of mathematics as a language*. It is also possible to speak of the *language* of mathematics while discussing mathematical problems proper. But this delicate question is beyond our consideration. I should like only to demonstrate that mathematics, as a deductive science with its own problems, turns into a language when it starts to be used for describing external phenomena.

To succeed in this task, we must separate pure and applied

[1] The contents of this chapter in a slightly modified form entitled "Logical Foundations of Applied Mathematics" were published in Russian as a preprint of Moscow State University (no. 24, 1971) and in English in the journal *Synthèse* (27:211-250, 1974). This chapter was translated by A. V. Yarkho.

[2] From *The Jaina Philosophy of Non-Absolutism* (Calcutta, 1944), cited by Mahalanobis (1954).

[3] The linguistic nature of mathematics was well comprehended by Kant in whose system of views mathematical judgments were prior synthetic judgments preceding experience. Further, he held that, if one confined himself to "pure" rational concepts of natural science, each separate teaching of the nature of science would be as rich as the mathematics it contained.

mathematics. First of all, I should like to show that the internal contents of mathematics, *its structures,* turn merely into grammar when used to solve applied problems. Is there any sense in drawing a demarcation line between applied and pure mathematics? The question can be formulated more broadly: What is the sense of any system of classification? Is there any sense, for instance, in dividing people into men and women and in speaking of the peculiarities of male and female psychology, or is it possible to speak only of human psychology and its aspects? Classification of the phenomena observed is a mode of human thinking. It is not at all a simple task, and it may not always be worthwhile to discuss the true nature of the elements of classification. In many cases the accepted system of classification is only an idiosyncratic point of view which we ourselves choose in order to consider a complex system. A look at the system from a certain point of view may give rise to biased notions. This was wittily pointed out by S. I. Vavilov (1955) in his book *The Eye and the Sun*: on page 93 he showed a picture of a reclining man, taken with the camera near his boots. The boots showed up immensely large and the man's head excessively small. The picture of the man proved to be distorted though it had been received by means of an optical objective, which by virtue of its name should have given an objective picture. It is often very tempting to look at a system from a specific angle emphasizing those sides of phenomena which usually remain hidden. This is the aim of my attempt to draw a demarcation line between pure and applied mathematics. Many mathematicians furiously oppose such a division, saying that no special applied mathematics exists but only applications of mathematics. Such a formulation is due to a desire to avoid discussing the real problems which gave birth to the difficulties we face when we attempt to use mathematics broadly in all the variety of scientific and industrial research. A curtain of diffidence is thrown over all possible aspects of a complicated and, probably, indivisible phenomenon.

In the following two sections I shall try to give a brief account of the modern ideas of the logical foundations of pure mathematics. We need this to create a background for our further judgments on the logical structure of applied mathematics. The presentations will be mainly illustrated by examples from mathematical statistics, the branch of science with which I am most familiar.

Axiomatic–Deductive Construction of Traditional Mathematics: Logical Structures of Pure Mathematics

A desire to draw a demarcation line between pure and applied mathematics is evidently as old as mathematics itself. I shall restrict myself to citing Felix Klein (1945), a well-known mathematician:

> With the *Ancient Greeks* we find *a sharp separation between pure and applied mathematics*, which goes back to Plato and Aristotle. Above all, the well-known *Euclidean structure of geometry* belongs to pure mathematics. In the applied field they developed, especially, *numerical calculations*, the so-called *logistics* (λογος, general number). To be sure, the logistics was not highly regarded, and we know that this prejudice has, to a considerable extent, maintained itself to this day – mainly, it is true, in those persons only who cannot themselves calculate numerically. The low esteem for logistics may have been due in particular to its having been developed in connection with *trigonometry* and the needs of *practical surveying*, which to some did not seem sufficiently aristocratic. In spite of this fact, it may have been raised somewhat in general esteem by its applications in astronomy, which, although related to geodesy, always has been considered one of the most aristocratic fields. You see, even from these few remarks, that the Greek cultivation of science, had its sharp separation of different fields each of which was represented with its rigid logical articulation . . . (p. 80)

Later on, during more than two thousand years, the two branches of mathematics, pure and applied, sometimes merged in the papers of many-sided scientists and sometimes diverged again. I shall not trace this process, for such a historical analysis is not a part of my task, though certainly it is of great independent interest. It became possible to approach this problem from a new standpoint when an article appeared under the pseudonym Nicholas Bourbaki bearing the original title "L' Architecture des Mathématiques." In this section an attempt is made to cast a glance at the entire variety of mathematical subjects, whereas Bourbaki of course deals only with pure mathematics, but from a certain unified standpoint. The question is formulated in the following way: Is mathematics in all its modern versatility still one science? In the following quotation from "The Architecture of Mathematics," the author speaks of pure mathematicians:

> To give a general idea of the science of mathematics at the present time seems at first glance to be an almost insurmountable task because of the breadth and variety of the subject. As in all the other sciences, there has been a considerable increase in the number of mathematicians and works devoted to mathematics since the end of the 19th century. The articles devoted to pure mathematics published throughout the world in the course of a normal year cover several thousands of pages. Doubtless, not all of them are of equal value; but after pouring off the inevitable waste, it still remains true that each year mathematics is enriched by a throng of new results, is con-

> stantly diversified and ramified in theories which are ceaselessly modified and remolded, confronting and combining with one another. No mathematician, even if he devoted all his time to it, would be capable today of following this development in all its detail. A number of them shut themselves up in one corner of mathematics, which they never seek to leave, and are not only almost completely ignorant of everything outside their subject, but would even be incapable of understanding the language and terminology employed by their colleagues in a specialty separated from their own. There is scarcely anyone, even among the most cultivated, who does not feel lost in certain regions of the immense mathematical world. As for those such as Poincaré or Hilbert who imprint the stamp of their genius in almost every domain, they constitute a very rare exception even among the greatest.
>
> Thus there can be no question of giving to laymen a precise image of something that even mathematicians cannot conceive in its entirety. But it is possible to ask whether this luxuriant proliferation is the growth of a vigorously developing organism which gains more cohesion and unity from its daily growth, or whether on the contrary it is nothing but the external sign of a tendency toward more and more rapid crumbling due to the very nature of mathematics, and whether mathematics is not in the process of becoming a Tower of Babel of autonomous disciplines, isolated from one another both in their goals and in their methods, and even in their language. In a word, is the mathematics of today singular or plural? (Bourbaki, 1950)

To the question of whether mathematics is one science, very precisely formulated, a positive answer is given. It is stated that mathematics (I repeat that only pure mathematics is meant) is a unified science. Its unity is preconditioned by the system of its logical structure. A characteristic peculiarity of mathematics is the precise axiomatic-deductive method of constructing systems of judgments. Any mathematical paper is first of all characterized by the long chain of logical conclusions it contains. But, Bourbaki remarks, stringing syllogisms together is only a transforming mechanism. It can be applied to any system of premises. This is but an outward characteristic of a system or, if you like, of its language, and it does not yet reveal the system of logical structure given by postulates.

The system of postulates in mathematics is by no means a multicolored mosaic of separate initial statements. The peculiarity of mathematics lies in the fact that the systems of postulates form specific concepts, mathematical structures rich in those logical conclusions which can be deduced on their basis. Mathematical structures are applicable to a variety of elements whose nature remains unknown (Bourbaki, 1960). In order to give a structure, it is sufficient to define certain relations between these elements on the basis of a system of axioms. Mathematics thus turns out to be of the nature of a calculus. Systems of judgments are constructed there without any appeal to implicit assumptions, common sense, or free associations. The problem consists in testing that the results obtained are indeed consequences of the initial assumptions. The formulation of the problem of *testing the truth of the initial axioms* is senseless. But the system of axioms determining a mathematical structure should be constructed so as to be rich in logical consequences.

The ideas of universal symbolism and logical calculus can be traced back to Leibniz, though the modern precise definition of mathematics as a strictly formalized calculus became possible only after the works of Frege, Russell, and Hilbert had appeared. Kleene (1952) offers the following characterization of Hilbert's philosophical stance:

> . . . those symbols, etc. are themselves the ultimate objects, and are not being used to refer to something other than themselves. The metamathematician looks at them, not through or beyond them; thus they are objects without interpretation or meaning. (p. 64)

We may speak of algebraic structures, such as groups, bodies, and rings (Bourbaki, 1958) or, for instance, of topological structures, where the notions of neighborhood, limit, and continuity which had appeared earlier at an intuitive level, are formulated mathematically. Bourbaki (1958) defines algebra as follows:

> The object of Algebra is the study of structures determined by there being given one or several laws of composition, internal or external, between the elements of one of several ensembles. (p. 41)

I shall dwell on topological structures so that the reader can have a clearer idea of the way mathematical structures are created. Topology (the science of location), founded by Riemann, resulted from the desire to develop a study of continuous values not on the basis of measuring distances, since in this case it is always necessary to introduce a definition of measure and to trouble about scales, but on the basis of relations of mutual arrangement and inclusion. It turned out that the notion of a "neighborhood," which is of great importance in mathematics, can be defined without resorting to the notion of distance. In order to do this, it sufficed to formulate a seemingly quite simple statement that every subset containing a neighborhood of a point A is a neighborhood of this point as well, and that the intersection of neighborhoods of the point A is also a neighborhood of this point. I shall cite this statement in the refined form it has in *Topologie Générale*, one volume of *Les Structures Fondamentales de l'Analyse* (Bourbaki, 1958):

> Definition 1 — An ensemble $\mathfrak{T}$ of parts of an ensemble E defines on E a topological structure (or more briefly a topology) if it possesses the following properties (called the axioms of topological structure): (O_{I}) Each union of ensembles of $\mathfrak{T}$ is an ensemble of $\mathfrak{T}$; (O_{II}) each

intersection of ensembles of $\mathfrak{Z}$ is an ensemble of $\mathfrak{Z}$. The ensembles of $\mathfrak{Z}$ are called open ensembles of the topological structure defined on $\mathfrak{Z}$.

Definition 2 – An ensemble possessing a topological structure is called a topological space.

Definition 3 – In a topological space E, one calls a neighborhood A of E each ensemble which contains an open ensemble containing A. The neighborhoods of a part $\{X\}$ reduced to a single point are also called neighborhoods of point X.

Thus, in a fairly simple way, the axioms of topological structures and definitions connected with them can be formulated. The choice of these axioms is, of course, arbitrary to a certain extent and appeared historically as a result of a long search. These, and likewise certain other axioms and definitions formulated in an equally simple manner, gave rise to numerous consequences inferred from them by a purely deductive method.

The formalization of mathematics, its axiomatic construction, was completed to a considerable extent by the end of the nineteenth century. During that period the German mathematician Dedekind and the Italian mathematician Peano formulated an explicit system of axioms for arithmetic, and the German mathematician Georg Cantor constructed the theory of infinite sets which appeared to be the basis for the foundations of calculus. But is was only in the 1930's that Kolmogorov succeeded in constructing an axiomatic theory of probability, and only then did probability theory acquire the status of a mathematical science.

Thus, we see that pure mathematics is primarily characterized by the presence of concepts rich in their consequences, i.e., mathematical structures formulated briefly and laconically as a system of postulates. From these postulates conclusions are inferred by means of deductive reasoning which do not depend on any additional explanation of how various statements or postulates can be interpreted in the phenomena of the external world. This point of view began to be distinctly developed in mathematics toward the end of the nineteenth century in the works of Frege and Russell, although a bit earlier Boole had, in the process of creating the algebra of logic, considerably broadened the possibilities of logical deduction and had at the same time made the whole system of judgments much more formal.

The game of chess is often regarded as a model of mathematics (Weyl, 1927) or, if you like, a parody of mathematics. In chess the pieces and the squares on the chessboard are the signs of the system, the rules of the game are the rules of inference, the initial position of the pieces is a system of axioms, and the subsequent positions are the formulae inferred from the axioms. The initial position and rules of the game prove to be very rich; in skillful hands they give rise to a great number of interesting games. Whereas the aim of chess is to checkmate the opponent, the aim of mathematical proof is to obtain a number of theorems. In both cases it is important not just to achieve the aim but to do this elegantly and, certainly, without contradictions. In mathematics certain situations are regarded as contradictory in the same way as, for instance, ten queens of the same color contradict the chess calculus. But the most interesting thing about this comparison is that in chess, as well as in mathematics, logical operations are carried out without any interpretation in terms of the phenomena in the external world. For instance, we do not care what the pawns correspond to in the external world or whether the restrictions imposed on the rules of operating on them are reasonable.

Now we can pass to comparing pure and applied mathematics. But before doing this I would like to mention those limitations which are imposed on the system of deductive constructions. Without this, the reader would get an exaggerated idea of the possibilities of deductive logic.

Limitations Imposed on Deductive Forms of Thinking by Gödel's Theorem

The principle of verification is a criterion of scientific thinking. A statement becomes scientific if it is formulated so that it can be tested. In the natural sciences, the testing of hypotheses is accomplished by comparing them with the behavior of the external world with the help of specific experiments or specifically organized observations (the latter might also be called "experiments"). On the face of it, everything seems quite simple here, but in fact this is not so. The problem of verification, even in the natural sciences, is fraught with great, and in the last analysis probably insurmountable, logical difficulties. (For details, see Chapter 1.) In testing hypotheses we have to face a vexing asymmetry: on the one hand, no experiment supporting the hypothesis is sufficient for its unconditional acceptance, and on the other hand, a single negative result is sufficient for rejecting the hypothesis. A hypothesis in natural sciences is always open to a further test, and this, according to Popper (1965), accounts for the progress of natural sciences. To overcome logical difficulties connected with the acceptance of hypotheses, we have to resort to the language of probability theory when formulating certain rules of the Game against Nature. (This is discussed in greater detail in Chapter 2 of Nalimov, 1971.)

Other difficulties, perhaps more serious ones, are connected with the problem of verification in mathematics. I have already mentioned that in mathematics there is no question of testing the truth of the initial postulates by means of comparing them with observations of the external world. The principle of verification is fulfilled here in the requirement imposed on the system of axioms that they be internally consistent. This means that they must be formulated so that there is no possibility of inferring from them theorems contradicting one another. Correctness of the system of mathematical judgments lies in their consistency.

The need to test the internal consistency of axiomatic systems was a natural consequence of the growing abstractness of mathematical constructions because it was necessary to give a logical foundation to the right to existence of these mathematical structures which determined the development of the separate branches of mathematics. The question of the internal relations between axioms had been troubling mathematicians ever since ancient times, i.e., immediately after the appearance of the Euclidean axioms, the first mathematical structure well known to us. Many efforts had been wasted on attempts to derive the fifth postulate (the postulate of parallels) from the rest of the postulates. But after the non-Euclidean geometries appeared, the question was reformulated, and it became necessary to show their internal consistency. Mathematicians first confined themselves to proofs of relative consistency. The method of mathematical simulation was used. For this purpose it was necessary to construct, within the system of known or generally accepted mathematical structures, models in which the axioms of new structures would be fulfilled. One system of mathematical constructions was interpreted by means of another. When this succeeded, it was stated that the new structures were internally consistent if the old ones were consistent. Thus, for example, a plane in Riemannian geometry is simulated by the surface of a sphere in the three-dimensional Euclidean space. On this sphere arcs of a great circle correspond to the straight lines. Then, indeed, it is impossible to draw through an arbitrarily given point on the surface of the sphere any arc of a great circle which would not cross an arbitrarily chosen circumference of the great circle on the surface of the same sphere. Riemannian postulates thus turn into theorems of Euclidean geometry. The next step was made by Hilbert, who showed that the Euclidean postulates are fulfilled for a certain algebraic model and are consequently consistent if algebra is consistent.

The question of consistency became especially acute after contradictions had been found in the Cantor theory of sets, particularly the famous paradox of Bertrand Russell, which was later named after him. It is noteworthy that Russell, especially in his joint book with Whitehead (Russell and Whitehead, 1910), strives (after Frege) to present pure mathematics as a part of formal logic. Mathematics is now simulated in terms of logic. And if the axioms of arithmetic are only the expression of suitable logical theorems, then the compatibility of the corresponding axioms of logic will follow from the consistency of arithmetic.

The contradictions revealed in the theory of sets, which constitutes the basis of many branches of mathematics, created a crisis situation. In 1904, Hilbert attacked the problem of the absolute consistency of arithmetic, thus acknowledging the insufficiency of proofs of relative consistency. Later, during the decade 1920–1930, Hilbert and his school published a number of papers reporting certain results which, as it seemed at the time, implied the consistency not only of arithmetic but also of the theory of sets. But in 1931, Gödel proved his famous theorem "On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems," which implied the failure of the attempts of Hilbert and his school. (*Principia Mathematica* is the title of the above-mentioned book by Russell and Whitehead.)

There certain logical systems called recursive logics are described. The axioms are regarded as certain "strings" of symbols and the rules on inference are regarded as means of obtaining "strings" out of "strings." Two restrictions are imposed on the rules of inference: they should be finitistic and strictly deterministic. The term "finitistic method" was not strictly defined by Hilbert, and later a heated discussion of its meaning broke out. Evidently, the term should be regarded as defining a method in which transfinite induction is not used. (The latter arises when one has to resort to those transfinite numbers which appear when the notion of an ordinal number is generalized to infinite sets.) The term "deterministic method" need not cause any bewilderment. It means the use of certain unambiguous rules which, when applied to similar data, always yield coinciding results.

The proof of Gödel's theorem on undecidability is too complicated to present here. It is preceded by forty-six preliminary definitions and several auxiliary theorems. A detailed description of Gödel's theorem and its significance for modern mathematics is presented by Kleene (1952). Attempts to give a simple proof of the theorem are made by Nagel and Newman (1960) and by Arbib (1964). I shall only briefly remark that an important role in proving this theorem is played by the arithmetization of mathematics, which is traditionally called Gödel numbering. Every mathematical statement is coded here by an arithmetic formula. The study of mathematical statements is brought down to the study of arithmetic relations.

From Gödel's theorem, it follows that all generally used logical systems in which arithmetic is expressible are incomplete, if they are consistent. There exist true statements expressible in the language of these systems which cannot be proved in the system itself. Furthermore, it follows from the same theorem that it is impossible to prove the consistency of arithmetic systems by means of concepts which can be expressed in this logic. Another consequence from this theorem is that however greatly one increases the number of axioms of this logic, their number being fixed, it will not make the system complete. There will always be new truths which can be expressed in terms of this logic but not inferred from it.[4]

It is difficult to give a formal definition of the concept of "proof" in mathematics. In the process of the development of mathematics, there appear new ways of proving previously unthinkable results. Kleene (1952) expresses this idea as follows:

> . . . we can imagine an omniscient number-theorist. We should expect that this ability to see infinitely many facts at once would enable him to recognize as correct some principle of deduction which we could not discover ourselves. But any correct formal system which he could reveal to us, telling us how it works, without telling us why, would still be incomplete. (p. 303)

Consequently, when mathematics is regarded as a strictly formalized axiomatic–deductive method, the restrictions which are imposed on this statement by Gödel's proof should not be overlooked.

In concluding this section, I would like once more to draw the reader's attention to the fact that the name Hilbert is also connected with the creation of metamathematics, a discipline concerned with the theory of proofs. The object of metamathematics is mathematics, and the latter is spoken of in a language which is a metalanguage with respect to that of mathematics.

And whereas mathematics is a strictly formalized system, metamathematics appears to be intuitively meaningful (though it can be formalized as well), and its statements are formulated in ordinary language. In this connection Kleene (1952) says:

> The assertions of the metatheory must be understood. The deductions must carry conviction. They must proceed by intuitive inferences and not, as the deductions in the formal theory, by applications of stated rules. Rules have been stated to formalize the object theory, but now we must understand without rules how those rules work. An intuitive mathematics is necessary even to define formal mathematics. (p. 62)

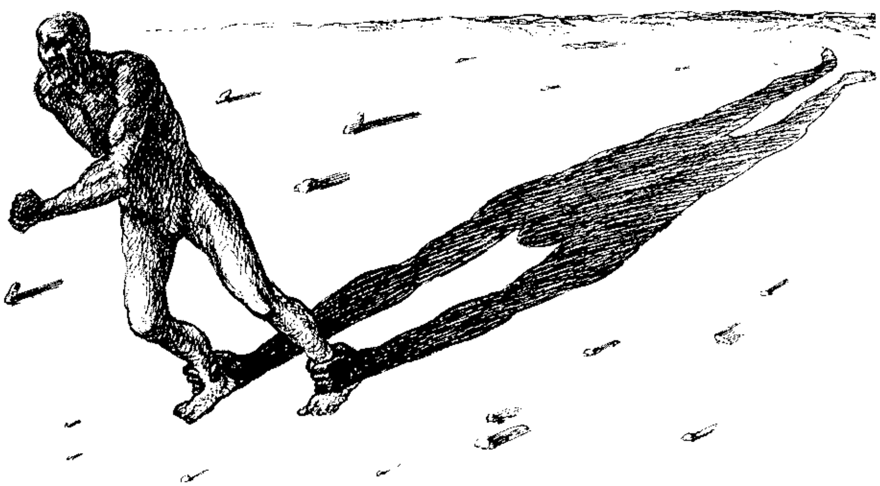
In the following section I shall try to make certain metamathematical statements concerning applied mathematics.

[4] Later, Gentzen managed to overcome Gödel's difficulty in proving the consistency of arithmetic, but in doing this he had to use transfinite induction up to the ordinal number surpassing all the numbers of an infinite sequence built in a certain way. Such a mode of proof, however, can no longer be regarded as legitimate. I cannot go into particulars in discussing this extremely complicated question and shall not consider all the other attempts to prove the consistency of arithmetic; all of them, in any case, exceed the framework of the problem formulated by Hilbert.

Mosaic Structures of the System of Judgments in Applied Mathematics

Now we can finally pass to the analysis of the system of judgments in applied mathematics. But first of all we should agree as to what we shall mean by applied mathematics. It hardly makes sense to use the term "applied mathematics" for such branches of knowledge as theoretical mechanics or hydrodynamics. They are independent branches of study built axiomatically and deductively, and they have internal logical structures fairly similar to those of pure mathematics in the sense of Bourbaki. It seems to me that it became possible to speak of applied mathematics as a specific and original phenomenon only after the beginning of the process traditionally called "mathematization of knowledge." This is the consequence of the penetration of mathematics into subjects connected with a direct evaluation and interpretation of complex systems of data. Such systems are often called diffuse: no partition can be made which would clearly single out phenomena of the same origin. In such a situation it is difficult to make any precise statement concerning the mechanism underlying the phenomena, and thus it is obviously impossible to construct axiomatic concepts. Later, it became clear that such complex systems should be not only studied and described but also controlled.

For problems of applied mathematics of this kind, the following peculiarity turned out to be typical: integrated mathematical structures



Mathematization of Knowledge

rich in logical consequences disappeared from sight. They were replaced, in some cases, by a multi-colored mosaic of criteria, and for this mosaic structure, the formulation of the question of consistency, which was so significant for the structures of pure mathematics, lost its sense. In other cases, certain statements based on uncertain intuitive considerations were expressed in mathematical language. Here the chain of syllogisms whose presence is at least an outward characteristic of mathematical constructions disappeared completely.

Let us try to illustrate this with examples. First of all, consider two examples connected with the construction of logical structures of a mosaic character. The first is a problem of response surface experimental design (Nalimov, 1971). Imagine a sufficiently complicated process dependent on a great number of variables. It may be, for instance, any actual technological process. The mechanism underlying the phenomena of such a process is unknown, and an attempt to study it is rather senseless because of the complexity of the system. It is reasonable to restrict oneself to performing experiments aimed at obtaining a simple (polynomial) model whose geometric image will be a response surface. On the response surface, a region should be found which would ensure the optimal conditions for the technological process. If the possible bounds of variation of every variable are coded with the numbers $+1$ and -1, the region of experimentation may be given by the multidimensional cube with vertices whose coordinates are given by the permutation of the numbers $\pm 1, \pm 1, \ldots, \pm 1$ or by the sphere inscribed within the cube.

The problem of experimental design is to choose in a certain optimal way the location of experimental points in the region of experimentation. If we wanted to solve the problem in a manner strictly corresponding to the general ideas of mathematical statistics, we should require such a location of N experimental points for which the matrix $\mathbf{X}$ of the independent variables would be arranged so that the determinant of the information matrix $\mathbf{X}^T\mathbf{X}/N$ would be maximal on the whole possible set of designs. Such a design is called D-optimal. In this case, the ellipsoid of variance of the regression coefficient estimators will be minimal.

From this standpoint R. Fisher, one of the founders of modern mathematical statistics, developed the concept of algorithm choice for obtaining the best estimators in evaluating experimental data. Here we shall follow the same logical conception, but extend it to solve a broader problem, that of choosing the optimal experimental design. In this case we would seem to act in the frame of the logical constructions which can be called the structure of mathematical statistics, but, in fact, the matter is much more complex. First of all, it turns out that, using the above criteria, it is impossible, strictly speaking, to build an experimental design with a reasonably small number of experimental points. True, this difficulty has been overcome: it has turned out that by using computers it is possible to construct a quasi D-optimal design with the number of points exceeding only by a small margin the number of parameters estimated (see Nalimov and Golikova, 1971). There was another problem which was not so easy to cope with, namely, the necessity of taking into account a number of other optimality criteria which were also in no way connected with the central idea of the development of mathematical statistics. Here, as previously (Nalimov and Golikova, 1971), I shall enumerate all the known criteria:

(1) D-optimality: max $| \mathbf{X}^x\mathbf{X}/N |$.
(2) G-optimality: min max $\sigma^2\{y\}$ where y is the value of the response function.
(3) The minimum of the average variance for the response function: min $\bar{\sigma}^2\{y\}$.
(4) The minimum of the maximal variance of the regression coefficient estimates: min max $\sigma^2\{b_i\}$.
(5) A-optimality: the minimal sum of the squares of the principal semiaxes of the ellipsoid of variance of the regression coefficient estimates (the minimal trace of the covariance matrix $[(\mathbf{X}^x\mathbf{X})/N]^{-1}$ of the regression equation).
(6) The minimum of the major axis of the ellipsoid of variance.
(7) The maximal precision of the estimate of the extremum coordinates.
(8) The minimal error under the condition that the response surface of order $d + 1$ is approximated by the polynomial of degree d.
(9) The best possibility to estimate the lack of the representation of observational results by a polynomial of a given degree.
(10) Proximity of the number of observations to that of parameters estimated.
(11) Rotatability: the matrix $\mathbf{X}^x\mathbf{X}$ is invariant with respect to the orthogonal rotation of the coordinates.
(12) Orthogonality: cov $\{b_i\, b_j\} = 0$.
(13) The possibility of making a definite type of nonlinear transformations of independent variables preserving the design optimality.
(14) The possibility of splitting the design into orthogonal blocks in order to eliminate the uncontrollable time drift.
(15) Composite designs: the possibility of using the points of design constructed to present the results by a polynomial of degree d as a subset of points for an optimal design of degree $d + 1$; this problem may arise when the polynomial of degree d presents the observational results inadequately.

(16) Non-sensitivity to separate flagrant errors in observational results.
(17) Simplicity of computations.
(18) Non-sensitivity to errors in independent variables.
(19) The possibility of estimating, at every step, the lack of fit.
(20) Visual representation of results.
(21) Uniformity: $\sigma^2\{y\}$ should not depend on the distance from the center of the experiment.
(22) The minimal sum of relative errors in regression coefficient estimate.

Not all the criteria are equally significant; some of them, e.g., 17–20, can be regarded as desirable properties of the design rather than independent criteria. We still have about fifteen indisputable criteria. They may be regarded as axioms of designing; the designs corresponding to these axioms may be regarded as theorems. In some cases these theorems result from using comparatively simple mathematical methods, e.g., linear algebra, in constructing a rotatable design. In other cases, e.g., in developing the concept of D-optimality, quite up-to-date mathematics is used: the theory of games, the theory of sets, and functional analysis. Here the system of axioms is obviously of a mosaic character. Of such a mosaic system of initial premises, it is senseless to ask the question concerning its internal consistency. Only in the simplest case, for linear designs, do some of the above-mentioned criteria prove to be compatible. This means that it is possible to build designs satisfying several criteria simultaneously. But everything becomes much worse for the second-order designs, especially when they are given discretely (for further details concerning the compatibility of criteria, see Nalimov and Golikova, 1971). It immediately becomes clear, without any additional analysis, that criteria 8 and 9, for example, are mutually contradictory. The mosaic of the structure of initial premises entails some difficulties (rather unusual for traditional mathematics) in systematizing and classifying designs.

The pragmatic problem is not simply solved either: it is not easy for a statistician to speak with an experimenter. It is possible, of course, to try to appeal to a hypothetical "omniscient experimenter," assuming that in the different concrete situations he will somehow be able to choose one of the criteria. But this does not seem altogether realistic. An experimenter discussing his problem with a mathematician cannot even decide unequivocally the main question: whether the design should be constructed on a multidimensional cube or on a sphere inscribed in it. For a mathematician the sphere is attractive for its natural symmetry. But for an experimenter it is difficult to imagine a region in the space limited by a multidimensional sphere; he gives the bounds of variation of each variable independently, and for him the region of experimentation is limited by a parallelepiped which, after a simple transformation of variables, naturally turns into a cube. The use of vertices of the multidimensional cube enlarges the region of the space under consideration and sharply improves the design characteristics from the standpoint of their D-optimality. However, such an asymmetrical (with respect to the center of the experiment) increase of the region studied may lead to unjustifiable difficulties, e.g., to the necessity of increasing the order of the approximating polynomial.

But now let us assume that we have somehow managed to agree upon the way the restrictions are imposed on the experimental region of the factor space and that the experimenter has informed the mathematician that he is first of all interested in finding the region of response surface where the extremum is situated. A decision has been taken to run a quasi D-optimal design. After the evaluation of experimental results, it has turned out that the extremum is situated somewhere very far out of the design region and, consequently, does not make sense physically. The purpose of the study is then changed: the experimenter says that now he only wants to estimate the contribution made by the separate regression coefficients. This problem may prove unsolvable, since in certain quasi D-optimal designs the regression coefficients are correlated, the correlation coefficient being close to unity.

The situation will become more complicated if we broaden the problem formulation and consider experimental designs aimed at singling out or, as it is usually put, screening out the dominant effect out of a fairly large set of potentially possible ones. It has appeared possible to create the concept of a screening experiment (Nalimov, 1971) and to construct an experimental design with fewer experiments than the number of "suspicious" effects. What is important here is that the estimates obtained as a result of these designs prove to be biased and inefficient.

The whole concept of screening experiments appears to be built in defiance of the main premises of mathematical statistics which I would like to regard as a certain logical structure in the sense of Bourbaki. It turns out that the real problems of mathematical statistics are solved outside the boundaries of this structure. Particularly great difficulties arise in teaching experimental design in all its diversity (Nalimov and Golikova, 1971) when one attempts to explain to an audience the whole mosaic structure of the variety of initial premises, which do not yield to systematization. The students complain that all this "creates a mess in their heads." We are tempted to call such mathematics with mosaic structure vulgar, although some fragments of this mosaic, for instance, the concept of D-optimality, seem quite respectable to mathematicians. Someone would perhaps like to compare the present-day state of things with that existing in geometry. Indeed, as is known, there are many geometries. But, as we have already seen, their axioms can be interpreted on the models of one another. And, what is probably most important, nobody tries to build geometrical constructions which would simultaneously satisfy several axiomatic systems. If, for instance, the general relativity theory has required non-Euclidean geometry, no one will strive to make it Euclidean in addition.

In the field of forecasting stochastic processes, the situation is quite similar, although there are differences. Everything is all right with forecasting stationary stochastic processes. We have here the well-known Kolmogorov–Wiener method, which has been found in the framework of the well-developed system of concepts of stationary stochastic processes. The problem is practically reduced to solving an integral equation whose kernel is the autocorrelation function of the stochastic process. But in fact all or almost all really observable stochastic processes prove to be non-stationary, at least judged by the behavior of their mathematical expectation. There is no mathematical theory of non-stationary stochastic processes. But, as it is put in one paper, there nevertheless exist myriads of publications in which various solutions of the problem are suggested. The best ones have the following structure: a certain model of a stochastic process is suggested which is formulated as an axiom, so that it cannot be either proved or refuted. Proceeding from this model, a formula for forecasting is found by means of constructing a chain of mathematical judgments. In the worst papers the solutions are simply given without even a clear formulation of the initial model. I cannot give here a complete list of the initial model–postulates because nobody has so far tried to classify and codify them. I shall confine myself to considering two well-formulated problems.

One of them is the well-known model of Box and Jenkins (1975). (For its brief description, see Nalimov, 1971.) The principal idea here is to pass from the study of the initial non-stationary process to the study of the sequence of processes created by the first, second, third, etc., differences and to investigate their mutual correlation. It is suggested that one should put down the model of the initial non-stationary process as follows:

$$x_{p+1} = (\gamma_{-l}\Delta^{l-1} + \ldots + \gamma_{-2}\Delta + \gamma_{-1} + \gamma_0 S + \ldots + \gamma_m S^{m+1})\alpha_p + \alpha_{p+1}$$

The letters Δ and S have the following meaning here:

$$\Delta\alpha_p = \alpha_p - \alpha_{p-1};\ \Delta^l\alpha_p = \Delta(\Delta^{l-1}\alpha_p);\ \Delta^0\alpha_p = \alpha_p;$$
$$S\alpha_p = \sum_{j=0}^{\infty} \alpha_{p-j};\ S^m\alpha_p = S(S^{m-1}\alpha_p);\ S^0\alpha_p = \alpha_p;$$
$$p = 0;\ \pm 1;\ \ldots; \qquad m \geqslant 1;\ l \geqslant 1;$$

and α_{p+1}; α_p; α_{p-1} are similarly distributed random values with mathematical expectation equal to zero. Having accepted their model, we obtain:

$$\Delta^{m+1}x_{p+1} = (\gamma_{-l}\Delta^{m+l} + \ldots + \gamma_{m-1}\Delta + \gamma_m)\alpha_p + \Delta^{m+1}\alpha_{p+1}$$

$$\Delta^{m+1}x_{p+1} = \sum_{j=0}^{l+m} d_j\alpha_{p-j} + \alpha_{p+1}$$

Hence it follows that all the serial correlation coefficients of order higher than $l + m + 1$ for the differences of the process $\Delta^{m+1}\alpha_p$ of order $m + 1$ should equal zero. If, for instance, all the serial correlation coefficients beginning with the fourth are equal to zero in the sequence created by the second difference, this means that the model with the three parameters γ_{-2}, γ_{-1}, γ_0 can be accepted. Other problems, some of them purely technical, arise here: to estimate these parameters, to correct them with respect to time (since the non-stationarity itself may prove non-stationary), and to evaluate the transfer function determining the inertia the system is subject to.

It is noteworthy that the initial model of a non-stationary stochastic process is formulated as an axiom. Its correctness cannot be tested directly by analyzing the experimental data. Nor can we give any physical interpretation of the model, but from it we receive a working formula allowing us to estimate, on the basis of the experimental results, the number of parameters γ whereby the model for forecasting will be determined. The initial model is an axiom constructed so as to use, in an intuitively obvious way, the correlation properties of the process, although the whole concept is in no way based on the well-developed correlation theory of stationary processes.

The next example is a model suggested by Legostayeva and Shiryayev (1971) for interpolating and extrapolating nonstationary stochastic processes. The latter is given by the model

$$\xi(t) = f(t) + \eta(t), \quad -\infty < t < \infty$$

where $\eta(t)$ is a white noise[5] process with mathematical expectation equal to zero, $f(t)$ is the trend of the process for which it is known that $f(t) \in F_n(M)$ and $F_n(M)$ is a class of real-valued functions $f(t)$ presented as

$$f(t) = a_0 + a_1 t + \ldots + a_n t^n + g(t)t^{n+1}$$

where the coefficients are $|a_i| < \infty$, $t = 0, 1, \ldots, n$ and the measured

[5] A random process with a constant spectral density. In other words, this is a noise with no frequency selectivity.

functions $g(t)$ are such that sup $|g(t)| \leqslant M$, M being an unknown constant. On the basis of the results of observations over the stochastic process, it is necessary to find the values of the parameters $a_0, a_1, \ldots, a_n$ in the best way, in a certain sense, by using a specifically determined minimax weight function for which a theorem is proved concerning necessary and sufficient minimax conditions on a certain class of functions. The model is given here in a polynomial form, and for this reason it is possible to make a comparison with the parameter estimates by the maximum likelihood method according to the usual pattern of regression analysis. In comparing maximal variances for the estimate of the parameter a_0, this method has an advantage of the factor 1.25; if the comparison is made in standard deviation having the same dimensions as the values measured, the gain will be only 12 percent; for the rest of the parameters, the gain will evidently be even less. It is important to note that the estimates turn out to be unbiased.

Here some questions naturally arise: in what way can a comparison of two different initial model axioms be made, and in what way are they to be compared with other models? Furthermore, will they yield results differing essentially from the extrapolation in a simple polynomial presentation of the results? The answers to these questions cannot be given in the general case simply for the reason that the models are formulated so as to prove incomparable. None of the models described above can be compared with models in which the forecasting is based on the moving average of an exponentially weighted means, even though all the models are built so as to yield an algorithm for the transformation of data which would give us processes of the "white noise" type. Here again, we are dealing with a mosaic structure of the model axioms.

Now let us consider an applied extremal problem solved outside the framework of an axiomatic–deductive construction. We have in mind a problem of adaptive optimization, namely, the problem of tracing an uncontrollable time trend in a technological process and of adapting it to changing conditions (Nalimov, 1971; Nalimov and Golikova, 1971). In order to solve the problem, a slight "rocking" of the technological process is suggested, i.e., running a permanent experiment directly at the plant, modifying the variables in a small interval in order not to change the technological regime too much. In accordance with this idea, the industrial process will yield not only the necessary production but also the information concerning the direction in which we should move in order to trace the shifting extremum. Having thus formulated the problem, we should worry not only about the experimental design, i.e., the optimal location of the experimental points, but about the whole strategy as well.

Several solutions of the problem have been suggested. One of them is the method of Evolutionary Operation, which consists in locating experimental points according to the scheme of a rotatable design (we spoke of this criterion above on p. 55) and repeating them many times to detect the signal on the background of large industrial interference. The strategy of movement turned out to be unformalized. After a cycle of experiments had been completed, the engineers gathered for a discussion and decided what to do next: to make the steepest ascent, to shift the center of the experiment, and to run a new series of experiments or, most likely, to include new variables in their consideration.

Another method is simplex designing, in which an algorithm for following the shifting extremum is already given. The algorithm consists of the following steps: the experiments are performed in $k + 1$ vertices of a regular k-dimensional simplex with k independent variables; then the vertex with the minimal yield of the process is reflected about the simplex verge opposite to it. The new simplex is then considered consisting of the new reflected vertex and k vertices of the initial simplex.

Then the same procedure is repeated and the minimal yield vertex is reflected again. Several more rules are included in the algorithm, allowing us to avoid "spiralling" around the point with the maximal error of the yield; we shall not consider these details of the algorithm. The attractiveness of the method is obvious: in order to make a decision at each step we use only one new experiment and k previous ones; the greater the number of independent variables k, the more efficient the method. The whole concept is developed outside the framework of an axiomatic–deductive construction. If the formulation of the algorithm is regarded as an axiom, no theorems can be inferred from it, for we cannot regard as a theorem the expression of the coordinates of the mirror reflection of a vertex. True, we can speak of the algorithm in mathematical language. We can say that, if the region of experimentation in the space of independent variables is given by a hypercube and the number of independent variables k satisfies the condition $(k + 1) \equiv 0 \pmod 4$, the experimental design used here is optimal in a broad sense: it is D-optimal, rotatable, and orthogonal. It is also possible to show that, for experiments being performed according to this design, the gradient of the linear approximation obtained will be directed from the center of the simplex to the verge where the reflection is performed. But these can hardly be called theorems, for they are all obvious properties of the given algorithm. The initial premises are not too clear either. The possibility of moving along such a simplex procedure is indubitable if we deal with a linear field; everything will also be in order if the linear field is cut with small ravines, if only the simplex scale is successfully chosen so as to jump over the ravines. But here our conclusions obviously lose their rigor.

Some curious difficulties have arisen in comparing the regular procedure of this kind with random ones. The simplex procedure has immediately been contrasted to the method of random search. In its simplest form the latter reduces to the following: an initial point x_k is chosen in a k-dimensional space; a straight line is drawn through it in a random direction; on this line, on both sides from x_k at the distance p_k two experiments are performed; the experiment with better results determines a new initial point x_{k+1} for a random construction of the second line, etc. Strictly speaking, a random search no longer includes the problem of experimental design; this is a procedure where only the strategy is given. The comparison of the random search method with the simplex procedure can be made only by way of simulating problems on computers. But even this approach requires that the criteria of comparison be chosen and that the conditions of simulating experiments be strictly stipulated. Meshalkin and Nguen (1966) demanded that p_k should equal p_r, the radius of the sphere circumscribed around the simplex. From the standpoint of the mathematician, such an approach proved quite logical; it resulted in constructing a precise mathematical system of judgments, and it appeared possible to prove a number of lemmas and theorems. On the part of the experimenter, however, such a requirement caused perplexity; in performing a random search, the researcher, even in the second experiment, crosses the boundaries of the cube limiting the experimentation region of the space of independent variables. The higher the dimension of the space of independent variables is, the less advantageous the conditions are under which the simplex procedure is placed: it will be performed in the sphere of a smaller radius than that of the random search procedure (see Nalimov, 1966).

In order to make the random search strategy comparable with the simplex procedure, the former should be modified in a special way. Here the very formulation of the problem becomes odd: an algorithm of an applied significance should be modified to become comparable with another one. Rastrigin (1966) gives an interesting collection of criteria for searching out an extremum. It is divided into local and global criteria. In local criteria, losses during searches are considered, i.e., "fast actions" at one step and the probability of an error—the probability of an erroneous step. In the non-local criteria, the number of trials is considered which is necessary for solving the problem set with a given "divergence" (precision) understood as the average deviation of the value found from the extremum in a given situation. Obviously, it does not demand too strong an imagination to increase the number of criteria for comparing two so difficult-to-compare strategies; using these criteria, we shall still obtain new results. Is there any sense in all these activities?

Thus we see that the concept of a simplex procedure does not possess a chain of syllogisms which, as previously mentioned, is at least an apparent feature of the construction of traditional mathematical judgments. Even if we agree to consider the coordinate values of the simplex-reflected vertices as theorems, the results will be fairly poor. The whole concept has a more profound sense which is intuitively clear but cannot be logically inferred from the initial postulate. The attempt made by mathematicians to find this sense by way of comparing the simplex procedure with other procedures was not a success. It is particularly important that it could not have been a success, since such a comparison requires the introduction of new axioms of comparison which determine the results of the comparison and which are in no logical way connected with the formulation of the initial axiom determining the strategy of the method. A variety of mutually uncoordinated axioms of comparison again create a mosaic structure. There is a temptation to call such mathematics vulgar, although it is real in applied problems.

Finally, I would like to answer one more question: Is what is now taking place in applied mathematics a situation similar to that in pure mathematics at the time, for instance, of Newton and Leibniz? Then the concept of mathematical structure did not yet exist. In any case, mathematicians had learned to differentiate before it was really understood what a function is (Box and Jenkins, 1975). It seems that in making such a comparison we have to point out an immense difference. Even at the seventeenth-century stage of their knowledge, mathematicians reached clear and unambiguous decisions although, as a rule, they could not formulate them as theorems. If we are allowed to speak from the standpoint of Platonic realism,[6] it will be possible to formulate a hypothesis that mathematicians acted as if they already surmised the existence of structures undiscovered at the time. In any case, this is how Bourbaki (1960, p. 188) describes the state of things in mathematics of the seventeenth century:

> One must take into account the fact that the way to modern analysis was not opened until Newton or Leibniz, turning their back on the past, undertook a provisional justification of the new methods. They were sustained not by the rigor of methods but by the fecundity of the results.

In applied mathematics or, to be more precise, in the applied mathematics that we are considering, there are a great number of results, but obviously no consistency among them.

[6] There are doctrines, shared by some mathematicians (evidently Gödel among others), according to which mathematicians do not create their structures but, like physicists, discover them.

Applied Mathematics as a Language: The Role of Sense Content Underlying the System of Signs

As stated at the beginning of this chapter, the "language of mathematics" is a frequently repeated juxtaposition, a specific linguistic cliché. Let us now subject this phrase to an analysis and attempt to reveal the meaning which it can have.

In many applied problems of a clearly non-mathematical character, the mathematical calculus is used simply as a language which allows us to obtain quickly logical consequences from the initial premises. This language is convenient for its compactness and precision. Since it is well known, there is no need to explain and justify the rules of inference each time anew. Finally, in the process of using this language, which is in a sense a universal language, associations arise with other problems that have been solved with the help of a similar series of judgments, and this gives additional cogency to the new constructions. Mathematics is used here simply as a language for the concise expression of a system of logical judgments. In this connection I would like to remind the reader of the widely known, but far from universally acknowledged, thesis of Frege and Russell that mathematics is merely a part of logic.

When using mathematics as a language, the researcher still does not act as a pure mathematician. He always takes into consideration what lies behind mathematical symbols in various concrete problems. And while the first serious difference between applied mathematics and pure mathematics consists in the absence of a system of judgments of unified logical structures rich in logical consequences in the former, the second difference, of no less importance, consists in the fact that in applied problems of the type in question it is necessary to observe fairly closely what lies behind the symbols. Several examples illustrating this statement are given below.

The first example is as follows. In studying the problems of measuring the growth of science (Nalimov and Mul'chenko, 1969), we formulated an informational model of the development of science. In this model publications are viewed as primary carriers of information. The following postulates were formulated which give the growth of publications in various situations:

$$\frac{dy}{dt} = ky$$

or

$$\frac{dy}{dt} = ky(b - y)$$

where y is the number of publications, k and b are certain constants, and t is time.

The first postulate states that the rate of growth of the number of publications should increase proportionally to their present number. This postulate can be accepted for a situation where there are no factors hampering the process of growth. The second postulate expresses the simplest mechanism of self-braking which begins to take effect only when the number of publications becomes comparable with the constant b. By integrating, we obtain in the first case an exponential function and in the second case an equation of an S-shaped logistic curve. These functions are further used to describe the really observable phenomena (function parameters are then, naturally, estimated, the adequacy hypothesis tested, etc.). The curves of growth given by the exponential function can be extrapolated into the future, which drives them to obviously absurd values. This indicates that the mechanism of growth must change. One can consider complicated situations when various countries and different branches of knowledge enter into the game. The observational results should then be presented by a sum of exponents, but this is not too convenient. When one expands the sum of exponents into a Taylor series and confines oneself to the first term, it is possible to limit oneself to presenting results by a varying exponent whose parameters remain constant only in a limited interval of time.

In short, on the basis of the above-mentioned, simply formulated postulates, we obtain rich logical consequences that allow us to discuss easily rather complicated situations. The validity of our judgments increases when we recall that analogous systems of considerations are used in biology to describe the processes of population growth, and in physics for deducing the law of light absorption or the law of radioactive decay. It is pleasant to realize that in all these cases we make use of the same logical structures by operating with the same universal language. But in such a way of reasoning, we always remember what lies behind the symbols and formulas composed of symbols. Imagine the following mental experiment: a set of publications and a portion of radioactive substance are delivered to the Moon. Both the growth of publications and the decay of radioactive substance follow an exponential function. But without any additional tests or reasoning, we shall say that the radioactive substance will continue to decay along the exponential curve while the publications will not grow. When solving the differential equation, we acted as pure mathematicians: we did not trouble about the meaning of the symbols. But when we interpret the resulting functions obtained, we already think about what lies behind the symbols; consequently, we do not think like pure mathematicians.

The second example is an error of interpretation of Zipf's law. Let us assume that there is a certain text with the overall number of words D_N constructed on the basis of a vocabulary containing N individual words. Arrange all N words according to the frequency of their occurrence in the text under consideration. Let the absolute frequency of the nth word (in the order of rarity) in the text be d_n. Then Zipf's law is put down as follows:

$$d_n = \frac{k}{n}$$

where k is a constant determined by the normalizing condition:

$$D_N = d_1 + k \ln N$$

Now let us assume that somebody wants to compute the value of D_{N+1} by using this expression and by substituting $N + 1$ for N under the logarithm sign. Can this be done? If the new $(N + 1)$th word of our vocabulary at the same time takes up the $(N + 1)$th place according to the frequency of its occurrence, this can certainly be done. But now imagine that in the vocabulary there appears a new word, e.g., "cosmonaut." It will not take up the last place in a series of words constructed according to their frequency of occurrence; a rearrangement of words will have to happen, and the parameter will no longer be constant, i.e., a re-normalization will take place. In this case we cannot compute the value D_{N+1} without knowing the new value of k. This specific restriction imposed on the normalizing expression is not expressed mathematically. The researcher should keep this in mind. If he intends to use the normalizing expression as an extrapolating formula, he should think of what places in the list will be taken up by new words. He should think of what is not expressed but only implied: this obviously does not correspond to the mode of thinking of a pure mathematician.

If one does not pay attention to the meaning behind the formulas, one can obtain quite improbable results. I once came across a publication where a normalizing expression analogous to the one mentioned above was used to study the system in development. Regarding N as a function of time, the author began differentiating the normalizing function with respect to time, assuming that the parameter k remains constant. On the basis of the ensuing results, he came to interesting conclusions. When his attention was drawn to the impossibility of doing this, the whole system of judgments collapsed, since there were no data available concerning the behavior of the derivative dk/dt. It would seem that no information about the system in development can be obtained from an expression which does not contain such information, but the author was eager to do so. The mathematization of knowledge often results not only in the use of mathematics but also in the abuse of it. It is sometimes possible for one to be struck by the thesis (which is, however, never formulated explicitly) that by using mathematics to describe observable phenomena it is possible to obtain information which is not present either in the observational results or in the postulates on which the initial models are constructed. I have come across publications where it is mathematically proven that a human being can have only seven levels of abstraction or it is proven that in any field of knowledge half of the publications covers this field of knowledge and the other half covers the adjacent ones. The question is whether it is possible to construct a system of postulates and definitions from which such conclusions could follow. More often than not, these conclusions are connected with the fact that mathematical expressions in applied problems are treated as in pure mathematics, without thinking too much of what lies behind the various expressions clad in mathematical symbols.

The third example is an approximating formula. Can one construct an approximating formula in applied mathematics guided only by the mutual arrangement of experimentally observed points and not bothering about the vaguely formulated subject matter which underlies these observations? I once came across a paper where the curve of growth of the number of scientists in the Soviet Union was approximated in a deliberately complicated manner. The author divided the curve into separate segments, and for each of them, he invented a specific mechanism described by different differential equations. The models thus obtained showed good agreement with the observed data, and, moreover, they were well coordinated. The author was so much carried away by his constructions that he even came to the point of thinking he had inferred his models not from a certain system of postulates but immediately from observational results! What is worth paying attention to here is that the unsmooth behavior of the growth curve is more reasonably explained not by the functioning of a specific complicated and frequently changing mechanism of growth, but rather by the arbitrary character of decision making by the administrative bodies financing the development of science and taking stock of the number of scientists. (The very definition of the term "scientists" and the system of taking stock of scientists change from time to time.) Then the breaks in the behavior of the curve can be described in terms of fluctuation. The decision concerning the choice of an approximating formula has to be made by taking into account considerations not formulated in the language of mathematics.

The fourth example concerns the application of the classical methods of mathematical physics. The heat conduction equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

can be solved for $-\infty < x < +\infty$, $-T < t \leqslant 0$, where T is a positive number. If we are given an initial condition of distribution of temperatures at the present moment

$$u(x, 0) = f(x)$$

then the solution describing the distribution of temperatures in the past can be expressed by the integral (John, 1955)

$$u(x, -t) = \int_{-\infty}^{+\infty} k(s, t) f(x + is) ds$$

where $$k(s, t) = (4\pi t)^{-1/2} e^{-(s^2/4t)}$$

The mathematician asks: How far back does it make sense to search for the distribution of temperatures in the study of space objects such as the moon, using this equation of heat transfer or its generalized form? The answer to the question should be sought with the help of some additional considerations which are again mathematically inexpressible. In solving the problem of determining the limits of a formula's applicability, we use information which it does not contain.

The last example concerns the use of probabilistic arguments in the field of applied research. Here it is fairly easy to formulate the problem in a deliberately nonsensical way. In the well-known British journal *Nature*, the question of the justification of statistical inference was recently discussed quite seriously. The following example was cited: four kings, Georges I, II, III, and IV of the Hanoverian House, died on one and the same day of the week, on Saturday. The probability of the random event is extremely small here: $(1/7)^4 = 1/2{,}500$. Would not the mathematician come to the conclusion that Saturday is a fatal day for Georges of the Hanoverian House? Of course not. Using some additional considerations he will reformulate the problem (for further detail, see Nalimov, 1971).

An interesting paradox has been formulated by Kendall (1966), a well-known English statistician. It concerns the experiment of tossing a coin. Not only the way the coin falls (heads or tails) is connected with this event but also the character of sound at the moment of the fall, the duration of the fall, and many other phenomena. The probability of the simultaneous occurrence of all these events is insignificantly small, but on the basis of all these considerations, the mathematician does not come to the conclusion that the coin will not fall. He takes into account a number of additional considerations and formulates the problem in another way. By the way, it therefore follows that we should not treat altogether seriously the statement that, in the process of random molecular combination, a live organism cannot appear (see Quastler, 1964). Even if it turns out that in a system of theoretical calculations the probability of a random emergence of life is equal to 10^{-255} or even less, it still seems sufficiently convincing only if the hypothesis as a whole does not arouse any objections proceeding from some other considerations which are much more general but hard to formalize.

Thus we see that, in the applied problems considered above, mathematics functions as a kind of language. In judgments made in this language, we consider relevant not only and not so much the grammar of this language as what we want to say about the subject matter on the basis of considerations stemming from our deeply *intuitive* ideas. Here it is appropriate to recall the school in the foundations of mathematics which is traditionally called intuitionism. It is connected with the names of Weyl (1927) and Heyting (1956). I shall comment only briefly on the complicated conceptions of this school, which is related not only to mathematics but also to the psychology of thinking. According to intuitionist mathematicians, logic is no more significant than a language whose persuasive power is determined by the intuitive clearness and immediate obviousness of every elementary step of reasoning. By now the majority of mathematicians have evidently abandoned such an attempt to establish the foundation of mathematics. I quote Bourbaki[7] (1960) on this point:

> The intuitionist school, the memory of which is doubtless destined to remain only as an historical curiosity, did at least render the service of forcing its adversaries, that is to say, the majority of mathematicians, to make their positions more precise and to be aware more clearly of both the logical and sentimental reasons for their confidence in mathematics. (p. 56)

Interest in these ideas is by no means exhausted since in the applied problems considered here mathematics functions as a language for which the cogency of judgments can be based in the same way in which the intuitionists once wanted to base the system of judgments in pure mathematics. Statements made in the mathematical language in applied problems should first and foremost be intuitively convincing; this is their basis. Here the distinction between pure and applied mathematics is especially precise.

[7] The statement of Bourbaki is probably too strongly worded. It should be remembered that many views of the intuitionists are shared by constructivistic mathematicians. Besides, a few mathematicians occupying themselves with the foundations mathematics still stick to the concept of intuitionism.

Language of Mathematics as a Metalanguage: Mathematical Structures as Grammar of This Metalanguage

We speak of a metalanguage when a hierarchical structure of language is considered. Our everyday language is a metalanguage to the "language" of objects surrounding us. In ordinary language, we use the names of objects and not the objects themselves. We speak of mathematics and its logical foundation in mathematical language. The subjects of mathematics are structures and logical deductions from them, expressed in the language of formulas. The subjects of metamathematics are statements about such formal systems. For instance, the statement "arithmetic is consistent" is a statement of metamathematics.

The language of mathematics used to describe applied problems functions as a metalanguage with respect to the language whereby the problems have been previously formulated and discussed. Sometimes statements made in the metalanguage acquire so general a character that this results in the creation of metatheories;[8] what is meant here is already a hierarchical structure of theories. The metatheory evaluates logical consistency of the theories which are hierarchically lower.

This happened, for instance, to mathematical statistics. Its language became a metalanguage with respect to that of various experimental sciences. In the language of mathematical statistics, statements are made about the judgments formulated in these object languages. These statements have acquired such a broad character that a metatheory has been created, the mathematical theory of experiment. Its principal ideas were formulated in detail in an earlier book (Nalimov, 1971). Here I shall briefly repeat the formulations given there: (i) the mathematical theory of experiment has allowed us to formalize precisely the decision-making process in the experimental testing of hypotheses; (ii) it has stipulated the randomization of experimental conditions in order to get rid of the biased estimates obtained in studying the complicated, so-called diffuse systems; (iii) it has formulated clear requirements for the algorithms of information reduction; (iv) it has formulated the concept of a sequential experiment; (v) it has formulated the concept of the optimal use of the space of independent variables.

In the statement of Kleene cited above (p. 52), it is said that the content of mathematics should be understood intuitively; with the help of intuition, we should understand the way the rules of formal mathematics

[8] The term "metatheory" appeared after the term "metaphysics," the latter being used for the first time by Andronicus of Rhodes. When classifying Aristotle's works, he introduced the term "metaphysics" in order to place the philosophical works by Aristotle on the first causes after the works on physics. The Greek word μετα has the meaning "after," behind."

function. In applied problems, mathematics itself functions as a metatheory, and for this reason it should also be intuitively grounded despite the apparently formal language in which it is expressed.

If mathematics plays the role of a language in applied problems, the mathematical structures of this language are naturally considered as the grammar of this language. The question can be asked whether it is necessary for the person whose use of the language will be purely pragmatic to know the grammar perfectly. The answer seems to be negative; in any case, it is possible to speak everyday language without knowing its grammar. I recall that in Russia during the first decade after the revolution of 1917, there existed rather a strange viewpoint according to which it was not necessary to teach grammar in the secondary schools. In fact, it was not taught, but the people graduating from school still had a full mastery of the language.

Above (see p. 64), I gave an example showing the way the language of differential calculus is used to discuss the problems of measuring the growth of science. Is it necessary that the participants in this discussion understand the foundations of calculus based on the concept of set theory? It would seem not: it suffices to have only a general understanding of the rules for differentiation and integration which would be quite similar to these existing in the time of Newton, Leibniz, and their closest followers.

I have already mentioned the fact that probability theory acquired the status of a modern mathematical subject only after Kolmogorov had given it an axiomatic structure. It turned out that probability theory could be constructed in the framework of the general theory of measure with a special assumption that the measure of the whole space should equal unity. (Probability can never exceed unity; this is the maximal probability of the certain event.) The probability theory formulated as a mathematical subject proved to be a part of a very general mathematical concept with a clear logical structure of an absolutely abstract character. But such an approach to the definition of probability appeared to be practically inconceivable for experimenters. It was the frequency theory of probability that had a strong effect upon them; according to it, probability is defined as the limit of the relative frequency of an event when the number of trials increases to infinity. From the viewpoint of the experimenter, this definition seems intuitively obvious though it is logically inconsistent. Kolmogorov (1956) writes that a definition of this kind ". . . would correspond to the definition of a point in geometry as the result of splitting a physical body an infinite number of times, halving its diameter every time." He noted further that a frequency theory of probability which involves such a passage to a limit is in reality a mathematical fiction since it is impossible to imagine an infinite sequence of trials where all the conditions would remain constant. True, he also notes that the solution of applied problems does not necessarily demand a formal definition of probability. It suffices here to speak of probability as a number around which relative frequencies are clustered under specifically formulated conditions, so that this tendency to cluster is manifested more and more accurately and precisely with the growth (up to a reasonable limit) of the number of tests. It is interesting that neither of the two definitions solves the paradoxes (see p. 68) that arise when one tries to apply probabilistic concepts to the description of real problems in a very formal way.

Note also that the grammar of the language of mathematics cannot always be used to construct a system of inferences for real problems. Two examples illustrate this.

In mathematical statistics, a theorem is proved stating that the regression coefficient estimates obtained in problems of a multidimensional regression analysis by the method of least squares prove to be unbiased and efficient in the class of all linear estimates. Generally speaking, this is true only if all independent variables and all the corresponding regression coefficients with the mathematical expectation different from zero are included in the consideration. But mathematicians never stipulate this condition, and they do not have to do so. The mathematician always deals only with the model which is expressed in mathematical symbols. He cannot take into account what is implied but not explicitly expressed. The experimenter's mode of thinking is different. Applying regression analysis, e.g., to describe a technical process in a plant, he is quite aware of the fact that far from all possible and really existing independent variables are included in the mathematical model. Many of them prove not to be included simply for the reason that it is practically impossible to measure them. In this case, the regression coefficient estimates will be biased. This bias can be so large that the results of regression analysis lose any sense. (An example illustrating this statement is analyzed in detail in Nalimov, 1971, p. 162.)

Another example is information reduction. This is a purely linguistic problem. Information contained in a long series of observations over a random variable must be expressed in a compact form. It was suggested that the grammar of these statements be built so that the parameter estimates whereby the statements are generated are unbiased, consistent, and efficient. For a long period of time, all the textbooks of mathematical statistics were full of such recommendations. But later the approach turned out to be too dogmatic. In real problems it is necessary to take into account one more property, namely, robustness, i.e., nonsensitivity to discrepancies from the initial premises concerning the distribution functions. We are more often than not dealing with a contaminated sample where observations taken from one universe mingle with those from another universe with different parameters. Estimates efficient with respect to uncontaminated samples can prove very poor *with respect to contaminated samples. We have to abandon the rules of* grammar inferred deductively and replace them with recommendations obtained as a result of simulating the problems on computers (for details, see Nalimov, 1971).

Variety of "Dialects" of the Metalanguage of Mathematics

It is well known that a single practical problem can often be expressed and discussed in a variety of mathematical "dialects." Sometimes it can be formulated at the level of deterministic ideas, the hypothetical mechanism governing the process being described with the help of differential equations. At another time the same problem may be discussed in probabilistic terms, and here again various dialects can be used: we speak in terms of classical mathematical statistics on one occasion and in terms of information theory on another. Let us assume, for instance, that we are discussing optimization of a certain technological process. We can try to give a strictly deterministic model of it. Then the problem of optimization will be reduced to an application of the calculus of variations to such new branches of study as the method of dynamic programming and the maximum principle of Pontryagin. If, however, we estimate the level of our knowledge of the mechanism of the process under consideration pessimistically enough, we shall have to confine ourselves to the use of the language of multidimensional regression analysis or of the language of the method of principal components and, perhaps, of that of factor analysis. If anyone finds the probabilistic language still very unpleasant, he may use Boolean algebra. In this case the intervals of variation of independent and dependent variables should be split into separate regions and coded in the binary number system. It is then possible to apply the method of minimizing Boolean functions in the algebra of logic. In this model the target function and predicates will be connected by the logical operators "and" and "or" (for a detailed description of a technological situation, see Shcheglov, 1972).

In a single seminar, one and the same problem can be discussed in a variety of dialects, whereas for ordinary language such a situation is very rare. An adequate translation from one mathematical dialect into another appears to be impossible, just as translation is, strictly speaking, impossible both for ordinary languages and for an abstract, completely formalized language.[9]

It is impossible to give a criterion which would allow us to prefer a certain mathematical dialect in describing a practical problem. Moreover, it is impossible even to suggest a criterion for testing the hypothesis that a certain dialect of the language of mathematics is acceptable for describing a certain situation. It might seem that the following statement could be chosen as such a criterion: the language is accepted for describing a practical problem if by means of it a mathematical model can be obtained which adequately describes the observed phenomenon. But here we may recall one of Russell's paradoxes (see Russell, 1961): assume that somebody regularly hires a taxi and constructs a graph marking the number of the day on the abscissa and the number of the taxi on the ordinate. If n observations are obtained, they can be represented by a polynomial of the degree $(n - 1)$. The curve corresponding to the polynomial will pass through all the points observed. The model will be adequate in some respects, though there do not remain any degrees of freedom to test its adequacy in the familiar statistical sense. However, it is inapplicable for predicting the number of the taxi that will be hired tomorrow.

The same experimental data could be presented as a stochastic process, in which case the problem of forecasting would acquire sense. The question of the choice of a model and consequently of the choice of a dialect cannot be solved by a mere test of the adequacy of the hypothesis. The same difficulty can arise in the problem of interpolation. I once saw a case where, under the experimental conditions, the experimenter could only receive experimental points situated on the left and on the right of a two-dimensional graph, the middle of the graph remaining empty. An approximating formula had to be found that would describe the behavior of the function in the region missing from the graph.

A mathematician immediately, and quite naturally, suggested approximating observational results by a higher-order polynomial. The multiextremal character of the graph of this function aroused the the indignation of the experimenters. This kind of conflict situation where the researcher–experimenter has, on the intuitive level, certain prior information of the mechanism of the process under study but cannot formulate it in a form acceptable to the mathematician is, in fact, very common.

The mathematization of knowledge which has begun recently leads to the appearance of many papers in which identical, or at least similar,

[9] It is interesting to note the following: in the abstract mathematical theory of context-free languages it is stated that the problem of finding the finite transformation reflecting the language generated by one context-free grammar to the language generated by another grammar is algorithmically unsolvable (see, e.g., Ginsburg, 1966).

situations are described by a variety of models formulated in different mathematical dialects. Broad application of mathematics only increases the Babelian difficulties in science. Will there appear a criterion that will restrain this process? At present it is hard to answer this question. Such a criterion could be the requirement that we consider legitimate only the use of those dialects of mathematics whose application leads to the creation of useful metatheories. An example of such a metatheory is the mathematical theory of experiment which came into being as a result of the broad application of probabilistic language to the description of experimental situations. But here another question emerges immediately: What is to be considered a useful metatheory? There may appear self-contained metatheories. In the process of creating a metatheory or a fragment of one, a researcher brought up on the tradition of pure mathematics may formulate postulates without caring much about their logical consequences. He may not be a bit worried about whether his logical constructions are realistic.

Polysemy of the Language of Mathematics

The development of the scientific era in which we are living began with a struggle against the polysemy of our language. The Cartesian school of philosophy demanded that scientific terms should be strictly defined. This demand undoubtedly had a favorable effect on the development of the exact and natural sciences. Yet, researchers did not apprehend it completely, and a strictly monosemantic language was not created. Quite recently, the English school of linguistic philosophy put forward the thesis that a natural language is always richer, as a result of its polymorphism,[10] than an artificial language with strictly defined terms (see Gellner, 1959). Earlier (Nalimov and Mul'chenko, 1972), we attempted to strengthen this statement on the basis of Gödel's theorem. From the theorem it follows that human thinking is richer than its deductive formulation. Communication between people takes place at a logical level where strictness is in some polite way broken by the polymorphism of the language, and Gödel's difficulty is thus overcome.

For a long time, the language of mathematics remained strictly monosemantic. It was used only to describe those well-organized systems which were dealt with in traditional physics. Lately, the language of mathematics has begun to be used for describing poorly organized, diffuse systems. In this process it has acquired certain features of polymor-

[10] What is meant here is the diversity of meanings ascribed to a sign within the limits of a language dialect. The term "polysemy" might be better, but Gellner (1959) used the term "polymorphism."

phism. The requirements placed on mathematical descriptions have become less strict. Whereas the description of real phenomena in mathematical language was earlier regarded as the expression of a law of nature, now it has become possible to speak of mathematical models. A single system under study can be described by a variety of mathematical models which may all simultaneously be legitimate. This question was considered in detail in Chapter 1 of *Teoriya Eksperimenta* (Nalimov, 1971). Further, we showed (Nalimov and Mul'chenko, 1972) that polymorphism can be observed within the limits of one model. This occurs in the problem of transforming variables in regression analysis where the parameters of transformation can be chosen arbitrarily from a broad range of possible values. In the problem of the spectral presentation of a stochastic process, the experimenter receives not one curve but a variety of curves of spectral density computed on the basis of different weighting functions – the so-called "spectral windows." The mathematician has no basis for choosing among these functions, which are constructed in such a way that an increase in precision of the estimate of the spectrum leads to an increasing bias.

This polymorphism of the language of applied mathematics increases its flexibility. The boundary between it and ordinary language becomes in a sense less obvious, and at the same time, a new difference from traditional mathematics appears. I could also speak of the unpleasant manifestations of the polysemy of mathematical language which arise in connection with solving problems of an applied character. Let us return to the fourth example analyzed above (pp. 67–68).

In order to estimate the temperature distribution in the past, we must know the initial conditions $u(x, 0) = f(x)$. In real problems we can deal only with an approximated sample estimate $\hat{f}(x) \to f(x)$. It turns out that small arbitrary changes in $f(x)$ and in a finite number of its derivatives may lead to great changes in $u(x, -t)$. The problem of temperature distribution for earlier values of time proves incorrect in the same way in which Hadamard formulated his criticism vis-à-vis the problem of Cauchy (the problem of Cauchy consists in finding a solution of a differential equation which satisfies the given initial conditions). Hadamard showed that the formulation of this problem is incorrect for elliptic equations because their solution does not depend continuously on the initial conditions. Here I shall not dwell on the question of incorrectly formulated problems; it has been considered in detail by other authors (e.g., Ivanov, 1963). The search for a correctly formulated problem is a struggle against the troublesome polysemy of mathematical language. Even if one has managed to formulate a problem correctly, it does not yet mean much. For instance, a correct numerical solution of the heat-transfer problem for the past (see John, 1955) does not remove the question of how far back it makes sense to compute it. Russell's paradox mentioned above concerning the prediction of the taxi number springs up in spite of the use of a correct (in this sense) problem formulation.

Mathematical Model as a Question to Nature Asked by a Researcher[11]

The application of mathematical language in research allowed us to formalize our ideas as to what a good experiment is. That gave birth to the mathematical theory of experimental design, which was discussed in an earlier section.

The principal idea of my earlier book (Nalimov, 1971) is that the proper design of experiments is possible only when a mathematical model of the process under investigation has been found; moreover, it is stated there that the possibility of planning and its efficiency depend entirely on how the model is stated. However, my book does not provide any information as to how the models are to be constructed. The reason for this is quite simple: *the construction of a mathematical model is an art, whereas the design of an experiment is mainly a procedure.* Obviously, it is much easier to discuss procedures.

Nevertheless, a few words can be said even on the subject of art. A *mathematical model*, in the sense used here, is a *question* put by the researcher to nature. What, then, are the semantics of a question. The subject of interrogative (erotetic) logic is being given much attention [see, for example, Hintikka (1972) and Kondakov (1971)].

Any question necessarily consists of two component parts: an assertive part which introduces some knowledge, thus making the question possible (this part can be regarded as a prerequisite of the question), and the interrogative part proper. The interrogative part can be neither true nor false; it can only be either relevant or irrelevant. The prerequisite of a question can be true, false, or inadequate for asking the given question, and this can be established only by introducing some other, external information not contained in the question proper. Let us consider two examples.

Example 1. A psychiatrist says to a boy patient, "An American rooster lays an egg on the territory of the USSR; to what country does the egg belong?" The boy answers with astonishment, "How can a rooster . . ."

[11] This section is a part of the booklet written together with T. I. Golikova, *Experimental Design Theory: the Achieved and the Expected*, and was published in the journal *Industrial Laboratory*, No. 10, 1977. The journal is translated into English in the United States, and I have borrowed the translated excerpt from it.

Such an answer satisfies the psychiatrist; the patient perceived the fallibility of the assertive part of the question.

Example 2. At philosophical conferences one is frequently asked the question: "Do you believe that absolute truth exists?" A prerequisite of this question is the implied assertion that there exists (or at least can exist) a language semantically rich enough to express absolute truth. (If such a language cannot exist in principle, what is the difference between the last assertion and the assertion that negates the existence of absolute truth?) However, if one recalls Gödel's theorem,[12] the implied prerequisite of the question seems rather doubtful. We thus see how a question gives rise to a prior question that must be answered before the main question becomes relevant.

Thus, a question can be neither false nor true and, consequently, is no assertion in the strict sense (Kondakov, 1971). Langer (1951), following Cohen (1929), says that a question is an ambiguous sentence whose determinant is the answer to it. Hence it is clear that a question can be irrelevant or even forbidden. The development of any culture is prescribed by a set of permissible and forbidden questions. In our culture, for example, forbidden questions are ones such as "Why does Ohm's law exist?" or "Wherefrom and when did Ohm's law appear?" (see Chapter 1) since any possible answer to these questions seems an absurdity.

Any science, if we speak in contemporary terms, begins with formulation of questions. Generally speaking, one can carry out observations and even conduct experiments without questions, but such activities can hardly be called scientific. Ethnographers tell us that among peoples with a pantheistic outlook one can meet observers who know all that can be seen about nature. However, they observe without questioning, and the reason they do not ask questions is that they have no theory for making the prerequisite of a question meaningful. Alchemy existed for nearly 2,000 years, and its goals resembled those of modern chemistry; in the process of its (rather slow) development, equipment used up to our time was devised, and a lot of discoveries were made. But alchemy was not a science. Experiments were conducted blindly with the single purpose of making gold; no questions were asked about nature, and in fact they could not be asked since there was no language for formulating theoretical concepts based on past experience. But even today, listening to reports or a defense of a thesis, one often feels inclined to ask: "What question is your paper answering?" More often than not the answer is: "Why should it answer a question? That has not been our aim. We have just made this and that . . ."

[12] Gnoseological problems generated by Gödel's theorem have been well treated by Nagel and Newman (1960).

A characteristic feature of contemporary science is thus still the fact that scientists aim at getting answers to precisely formulated questions based on previous knowledge. However, knowledge is always relative and changeable since science progresses dialectically, i.e., in a revolutionary manner (Kuhn, 1970*a*; see also pp. 9–14). Consequently, a more careful formulation would be that scientists ask questions about nature on the basis of contemporary prejudices.

Mathematics is a language in which questions can be asked in a surprisingly compact form with the aid of abstract symbolic notations. Let us assume that a researcher has formulated his problem in the form of a model

$$\eta = \alpha(\mathbf{x}\theta)$$

in which he wants experimentally to evaluate the vector of parameters θ. In such a formulation, the model written above is simply a well-asked question. Its prerequisite is a distinct separation between the dependent and independent variables responsible for the investigated process and the analytic description of the model proper; the interrogative part is the specification of the parameter vector which has to be numerically evaluated. In the case of insufficient a priori information, the prerequisite is weakened, and instead of a single model one can have several competitive models, or in place of a single small set of independent variables one has a multitude of variables from which one must select really significant ones by means of a screening experiment. A change in the prerequisite of the question causes an immediate change in its interrogative component.

Earlier (Nalimov, 1971), attention was drawn to the fact that even the most simple problem, such as weighing three objects with the aid of a balance, can be subject to planning provided a model is specified. Such a model must necessarily contain an assertive part, which in the case of weighing is a polynomial model without interaction members (the experimenter asserts on the basis of a priori information that no interaction takes place, and this is the prerequisite of the question). The investigator may have no a priori knowledge about the mechanism of the investigated phenomenon, but still the question may be asked on the basis of some information about the logical structure of some "blind" examination. As an example, one can use the model of staging a screening experiment in pharmacological isolation of therapeutically active or toxic preparations. [This is discussed in detail elsewhere (Nalimov and Golikova, 1976, Section 3 of Chapter 6); the reader can readily separate in the model its assertive and interrogative parts].

One can speak about a hierarchy of interrogative components associated with the mathematical model as a question. If we get an answer to the first interrogative component of the above model, i.e., the numerical estimates of the parameter vector θ are found, we immediately have a second, hierarchically higher, interrogative component: it is necessary to evaluate how the given model describes the problem (to test its adequacy, etc.); in this case the evaluation of parameters turns into the assertive component of the new question. If the model is found to be satisfactory, the information is included in the assertive part of the question, and a new interrogative component appears which otherwise can be formulated as follows: Where is the extremum? What is the response surface in the extremum region?

Any scientific hypothesis, especially when written in a mathematical form, can perhaps be regarded as a question. A probabilistic model of the semantics of everyday language using the Bayesian theorem has been proposed (Nalimov, 1974*b*, 1981).

$$p(\mu|y) = kp(\mu)p(y|\mu)$$

where $p(\mu)$ is the distribution function of the meaning of word μ given a priori; $p(y|\mu)$ is the likelihood function which defines the distribution of the semantic content of sentence y, provided our attention is drawn to the meaning of word μ; and $p(\mu|y)$ is the a posteriori probability that defines the distribution of the meaning of word μ in sentence y. This model will be discussed in detail in the subsequent chapters of this book. In a profound sense, the prerequisite of this model, if treated as a question, is the assertion that our language is discrete while our thinking is continuous (Nalimov, 1979). The interrogative component of the question is aimed at clarifying how on the basis of the written model one can explain the entire diversity of our verbal behavior.

It should certainly be possible to devise a meaningful classification of models by treating them as questions. But we are still not prepared for this. We shall restrict ourselves here to several brief remarks. One of the noteworthy classes of models is classification models. In a natural way they can be divided into models of logical classifications (for more detail, see Meyen and Shreider, 1976), such as the universal decimal classification used in libraries and models of numerical taxonomy (the method of principal components, cluster analysis, etc.). A characteristic feature of such models is that they describe the observed phenomena regardless of cause-and-effect relations. Numerical taxonomy models are gratifying in their "poverty." Their prerequisites are extremely limited: the observable variables cannot be divided into dependent and independent; the whole thing is limited to enumeration and defining a metric; a stopping rule is selected (in some models) for ceasing the classification procedure;

and the interrogative part is free from many claims and is limited to a search for the hierarchy of taxons in the given imperative metric.

It is interesting to note also that probabilistic models raise questions about the diffuse–behavioral descriptions of the universe (Gellner, 1959) regardless of cause-and-effect relations (For more details, see the next chapter). Differential equations appear any time a question is asked in a cause-and-effect formulation. Of special interest is the recent attempt to use digital computers to construct models with a weak assertive part and an interrogative part full of pretensions. As an example, we cite the grandiose American program for a comprehensive study of five ecosystems: the tundra, steppes, deserts, and leaf-bearing and coniferous forests. Comprehensive models have been constructed and partitioned into blocks containing up to a thousand parameters under very weak initial theoretical prerequisites. The models were to answer questions about the behavior of these ecosystems. It is still too early to speak about the outcome of this program. However, a critical analysis of this activity, based on a careful analysis of materials concerning three of the above-mentioned ecosystems, has been published recently (Mitchell et al., 1976).

It might, of course, be possible to formulate some criteria for good models. But once again we must limit ourselves to certain fragmentary remarks. Levins (1966) states that in models there is a trade-off between *generality, precision*, and *realism*. (Nothing similar is observed in the "laws of nature," and in this they differ from models describing diffuse systems.) Stressing of one of these three factors immediately weakens the others. We could add to this that, with a given assertive part and certain fixed experimental possibilities, the answer is the more definite the weaker the demands specified by the interrogative part of the model. Hence, it can so happen that in the description of chemical processes simple polynomial models can provide more than models with nonlinear parameters which pretend to give an adequate description of the mechanism of the process. (This very important problem will be raised again in Chapter 8.)

Specialists in experimental design seem now to be somewhat disillusioned. All was well when the topic was the design of so-called extremal experiments. All was clear: the model proposed by Box and Wilson in 1961 proved to be typical[13] of many situations, especially in technical sciences. Certain additional typical models, such as screening problem models, became known later. Effort had to be devoted not only to the selection and construction of the model (although the activity still was

[13] One possible reason that mathematical statistics is not very much appreciated in science is that many professional statisticians tend to reduce the entire diversity of real problems to certain typical models such as the models of the analysis of variance.

creative[14]) but mostly to the subsequent purely technical part of research activity. Many experimenters have learned how to do this by themselves without outside help. Specialists in experimental design should now become modelers, i.e., constructors of models. Herein probably lies the success both of their personal activities and of the field as a whole.

Is it possible to teach modeling, which is said to be more of an art than a science, and if so, what should be taught? To write a deeply intimate letter to one's friend is also an art, but nevertheless school children are taught writing even if it is known that not all will grasp the art. To educate model builders, one must teach them how to represent in a compact symbolic form assertions about the real world which are posed in a vague, uncertain manner. The model must clarify a specified problem or question. It must exhibit an economy of ends and means. It is an art.

Peculiarities of Teaching Applied Mathematics

The pragmatic meaning of the distinction between pure and applied mathematics first becomes obvious in the problems of teaching. If the logical structure of applied mathematics is different from that of pure mathematics, it should also be taught differently.

One should not think that in speaking of the peculiarities of applied mathematics I want to deny the role of mathematical structures or abstract mathematical constructions. But they seem to reflect a certain outlook rather than serve as a real instrument. Not everybody engaged in studying applied problems should possess this outlook or possess it to an equal extent. In any case, it would have been quite unrealistic to think that every experimental researcher could also become a mathematician. It is also unrealistic to believe that every mathematician will be able to solve applied problems using mathematics of a form foreign to him. Here again, I would like to raise the question of training scientists of an intermediate type (for details, see *On Teaching Mathematical Statistics to Researchers*, 1971). A scientist with a solid background in pure mathematics can prove to be quite helpless in applied mathematics. [After this section had been completed, my attention was drawn to a very interesting article by I. I. Blekhman, A. D. Myshkis, and Ya. G. Panovko (1976) devoted to the same problem.]

[14] Creative activity here means the selection of one out of several typical models, the selection of dependent and independent variables, the choice of the region of the independent-variable space in which the experiment is to be carried out, etc. It can be easily calculated that if the error in the selection of the span of variation of each independent variable is 15 percent, then, for example, in the estimation of second-order polynomial models for five factors, the efficiency of a D-test plan decreases by approximately 60 percent as referred to one parameter.

Concluding Remarks

I now return once more to the question of where the demarcation line lies between pure and applied mathematics. Naturally, there can be no precise boundary between them. We may think of a continuous scale of logical structures of various degrees of rigor. At one end is located pure mathematics in the sense of Bourbaki; at the other end, there are those of its applications which I consider related to the broadly developing mathematization of knowledge. The difference between pure and applied mathematics is most prominent when we contrast the extreme manifestations of what is essentially the same phenomenon. Theoretical mechanics and hydrodynamics get quite close to pure mathematics on this scale, and physics obviously occupies an intermediate position. Hutten (1956) made a very interesting statement concerning the logical structure of physics. He said that, in attempting to reconstruct any branch of knowledge, we must distinguish between three stages of formalization: mathematization, when mathematics is used merely as a language; axiomatization; *and the construction of interpretation rules. Further, he remarked that if* one looks at physics from this standpoint he will have to acknowledge that its formalization has been limited to the first stage, mathematization. Many attempts have been made to axiomatize physical theories, but only one of them, that of Carathéodory, an expert in mathematical analysis of thermodynamics, has gained general recognition. Even the axiomatization of mechanics by Newton has not been a success. As to the interpretation rules (translation from the mathematical language to the experimental one), they are, strictly speaking, absent from physics. A brilliant example is a report by Abel (1969), who presents a collection of diverse and uncoordinated statements of physicists and philosophers, students of the foundations of physics, concerning the interpretation (in the language of experimental physics) of the concept of probability waves (psi-waves). Abel begins his article with the question: "Can we be said to know something . . . which we have not been able to put into words?"

I believe that an opposition of the two extreme manifestations of one and the same phenomenon has its raison d'être at least in the fact that it serves to draw attention to that end of the logical scale where, under the guise of mathematizing knowledege, something is being done which is very far from what mathematics proper is.

Back to Contents
Next to Chapter 4