THE JOURNAL OF SYMBOLIC LOGIC
Volume 57, Number 2, June 1992

A SURVEY/EXPOSITORY PAPER

KOLMOGOROV AND MATHEMATICAL LOGIC

VLADIMIR A. USPENSKY[1]

INTRODUCTION

There are human beings whose intellectual power exceeds that of ordinary men. In my life, in my personal experience, there were three such men, and one of them was Andrei Nikolaevich Kolmogorov. I was lucky enough to be his immediate pupil. He invited me to be his pupil at the third year of my being student at the Moscow University. This talk is my tribute, my homage to my great teacher.

Andrei Nikolaevich Kolmogorov was born on April 25, 1903. He graduated from Moscow University in 1925, finished his post-graduate education at the same University in 1929, and since then without any interruption worked at Moscow University till his death on October 20, 1987, at the age $84\frac{1}{2}$.

Kolmogorov was not only one of the greatest mathematicians of the twentieth century. By the width of his scientific interests and results he reminds one of the titans of the Renaissance. Indeed, he made prominent contributions to various fields from the theory of shooting to the theory of versification, from hydrodynamics to set theory. In this talk I should like to expound his contributions to mathematical logic.

Here the term "mathematical logic" is understood in a broad sense. In this sense it, like Gallia in Caesarian times, is divided into three parts:

(1) mathematical logic in the strict sense, i.e. the theory of formalized languages including deduction theory,[2]

(2) the foundations of mathematics, and

(3) the theory of algorithms.

Received April 16, 1991.

[1]This paper was presented on July 28, 1989, as an invited address at Logic Colloquium '89, in Berlin. I am deeply indebted to the Chair of the Program Committee, Prof. Dr. Ernst-Jochen Thiele of the Technische Universität Berlin, who invited me, arranged for the typing of the manuscript, and then submitted it to this JOURNAL.

[2]As Alonzo Church writes in his celebrated monograph [Ch 56, §07], "The subject of formal logic, when treated by the method of setting up a formalized language, is called *symbolic logic*, or *mathematical logic* or *logistic*."

0022-4812/92/5702-0001/$03.80

Kolmogorov made a first-rate contribution in each of these topics.

Let me mention here that it was Kolmogorov who in spring of 1972 gave the first course in mathematical logic at the Moscow University that was required for all mathematics students (Kolmogorov's lectures were attended also by some eminent mathematicians, including Andrei Markov, then head of Mathematical Logic Department of the University). Kolmogorov is co-author (with A.G. Dragalin, now in Hungary) of two university textbooks on mathematical logic:

Introduction to Mathematical Logic, 1982 [ANK & D 82]

and

Mathematical Logic: Additional Chapters, 1984 [ANK & D 84],

both in Russian. At the time of his death Kolmogorov was the head of the Mathematical Logic Department at Moscow University as well as the head of the Scientific Committee on Mathematical Logic of the Soviet Academy of Sciences.

But certainly these engagements, however important, are not the most important. The most important are his services to scientific knowledge, his influence on the forming of the contemporary field of mathematical logic.

Here is the list of Kolmogorov's major achievements in the field of the mathematical logic and its applications:

1. The very first mathematical study of intuitionistic logic; the very first axiom system for this logic, anticipating Heyting's formalization.
2. Not only a Brouwer-style critique of classical (i.e. traditional) mathematics but also a positive study of the intuitionistic validity of that mathematics.
3. Formulation of intuitionistic logic as a logic not of assertions but of problems.
4. A result on translatability of classical mathematics into intuitionistic mathematics, the first example of the so-called "embedding operation".
5. A definition of the most general concept of an algorithm.
6. Founding the theory of numberings (also called numerations, or enumerations).
7. Founding the theory of complexity for finite objects (properly speaking, for constructive objects).
8. Founding the algorithmic theory of information.
9. Forming the basis for the algorithmic theory of probability.

This talk will consist of three parts:

Part I: Investigation of Intuitionistic Logic.
Part II: The Pure Theory of Algorithms.
Part III: Applications of Algorithm Theory to Probability Theory and to Information Theory.

Part I.
Investigation of Intuitionistic Logic

On September 30, 1925, at the age of twenty two, Kolmogorov finished the paper *On the principle "tertium non datur"* (or *On the law of the excluded middle*) [ANK 25]. Not only was it Kolmogorov's first paper in mathematical logic, it was also the very first paper in his country, Russia, which contains mathematical results in this field of mathematics (earlier papers deal with elementary Boolean treat-

ment of propositions only, and contained rather algebraic results if any). This youthful paper demonstrates some main features of Kolmogorov's scientific talent: the problems considered possess philosophical profundity, and pure mathematical results naturally arise as the answers to epistemologically oriented questions.

Indeed, in his paper of 1925 Kolmogorov, starting with Brouwer's criticism of illegitimate use of the law of the excluded middle (alias tertium non datur), and even, as we shall see, strengthening this criticism, posed two sound questions. First, why has the illegitimacy of this use often gone unnoticed and, second, why, specifically, has this illegitimate use not yet led to contradictions? In his paper Kolmogorov has answered both these questions. First, answers Kolmogorov, the use of the dubious excluded middle principle is often unnoticed because this use is quite justified if the final conclusion is of finitary character. And this answer of a twenty-two-year-old youth daringly challenges a remark by Brouwer who thinks that finitary conclusions based on a transfinite use of the law of the excluded middle must be considered unreliable (see [ANK 25, V, §1]). Second, answers Kolmogorov, this use never leads to a contradiction because if it does lead to one, then some contradiction can be constructed without this use.

Kolmogorov begins by writing down a well-known system of six axioms of propositional logic (or the logic of judgements, as Kolmogorov calls it) introduced by Hilbert in 1922 [Hi 22, p. 153]—four axioms of implication (AI)

Ax.1. $A \to (B \to A)$,

Ax.2. $[A \to (A \to B)] \to (A \to B)$,

Ax.3. $[A \to (B \to C)] \to [B \to (A \to C)]$,

Ax.4. $(B \to C) \to [(A \to B) \to (A \to C)]$,

and two axioms of negation (AN)

Ax.5. $A \to (\neg A \to B)$,

Ax.6. $(A \to B) \to [(\neg A \to B) \to B]$.

Here and below the rules of inference are modus ponens and substitution. We call the calculus produced by Axioms (AI) + (AN) the *Hilbert calculus* and denote it by $\mathfrak{H}$. (Strictly speaking, $\mathfrak{H}$, as Kolmogorov defines it, is $\mathfrak{B}$ + (PDN) i.e. the axiom system $\mathfrak{B}$ (which will be given below) plus the principle of double negation (PDN) $\neg\neg A \to A$; but the calculi $\mathfrak{B}$ + (PDN) and (AI) + (AN) are equivalent, as is proved in [ANK 25, II, §§7–8].)

The axioms of implications are intuitionistically appropriate. As for (AN), the axiom Ax.6 is a version of the principle of excluded middle and so is invalid intuitionstically. What is new is that Kolmogorov brings the axiom Ax.5 into question and finally rejects it because it "does not have and cannot have any intuitive foundation since it asserts something about the consequences of something impossible: we have to accept B if the true judgement A is regarded as false" (II, §4). He explains also why this Axiom 5 was not criticized by Brouwer: the point is that this axiom "is used only in a symbolic presentation of the logic of judgments; therefore it is not affected by Brouwer's critique" (I, §6). So Kolmogorov rejects both Hilbert's axioms of negation and introduces his own new axiom of negation

(K) $(A \to B) \to [(A \to \neg B) \to \neg A]$,

which he calls the *principle of contradiction.* The system of five axioms, the four axioms of implication (AI) and the axiom (K), Kolmogorov calls "the system $\mathfrak{B}$".

Thus we have the historically first attempt to formalize intuitionistic logic by proposing an axiom system.

"The system $\mathfrak{B}$ is nowadays known as the minimal calculus and differs from Heyting's system in that the latter contains in addition to axioms (AI) and (K) also the axiom $A \to (\neg A \to B)$" [HW 67]. Strictly speaking, Heyting's calculus of 1930 as well as the minimal calculus (as the latter was defined by Johansson in [Jo 37]) deals with implication and negation, and also with conjunction, disjunction and equivalence. But if a formula contains implications and negations only it is all the same, as far as its formal provability is concerned, to appeal to the full calculus (with all five connectives) or to its implicational-negational fragment only. More exactly: If a formula containing negation and implication only is provable in the full Heyting system, then it is also deducible from only those axioms which contain no connectives except negation and implication. This effect is discussed in [Ch 56, §26] (see also the supplement "Errata" on p. 377 in the second printing of [Ch 56]) and in [Pl 88, §5]. If a formula containing negation and implication only is provable in Johansson's full minimal calculus, then it is also provable in the system $\mathfrak{B}$. This effect is demonstrated in [Pl 88, §6]. There is no similar effect for the classical propositional logic and the axiom system (AI): this (AI) is not classically complete for implication alone. The question of classical completeness of (AI) for implication alone is raised in [ANK 25, II, §2]. The following example gives a negative answer to this question: the so-called Pierce's law $[(A \to B) \to A] \to A$ is a classical tautology but not provable intuitionistically (because $[(A \to \neg A) \to A] \to A$ is equivalent to the principle of double negation and hence is not provable in $\mathfrak{B}$). "It is now a familiar result that it is necessary to add Pierce's law, $[(A \to B) \to A] \to A$, to render the positive implicational calculus classically complete." [HW 67].

The said enables us, following Hao Wang, to call Kolmogorov's system $\mathfrak{B}$ by the same name "minimal calculus" as the full calculus of Johansson. And this terminology is in some concord with the terminology of Church, who called the full minimal calculus with all five connectives the "minimal propositional calculus of Kolmogoroff and Johansson" [Ch 56, §26]. Church writes in his celebrated monograph: "Kolmogoroff considers primarily not the full minimal calculus but the part of it obtained by suppressing the three primitive connectives, conjunction, disjunction, equivalence, and the axioms containing them [...] Addition of the three axioms for disjunction [...] is mentioned by Kolmogoroff in a footnote, but the full minimal calculus and the name ("minimal calculus" or "Minimalkalkül") first occur in Johansson's paper" [Ch 56, n. 210].

In the final section of his paper Kolmogorov brings into consideration also axioms of the first order predicate calculus. So it may be said that in this paper he introduced the first formalization not only for intuitionistic propositional logic but for intuitionistic predicate logic too; and the latter formalization is nothing but the minimal predicate calculus [Ch 56, Exercise 38.10] (up to the restriction to two primitive connectives), which calculus, consequently, is also due to Kolmogorov.

But the formalization of intuitionistic logic and the invention of the minimal calculus are subordinate topics of Kolmogorov's paper. Let us recall its main goals: to explain why illegitimate use of the excluded middle principle goes unnoticed and does not lead to a contradiction.

In the first place, why is it unnoticeable—under the presupposition, of course, that we deal with finitary judgments only? (The questionableness of application of that dubious principle to transfinite judgments is fully recognized and discussed in detail in Kolmogorov's paper.) Because, answers Kolmogorov, all the laws of classical propositional logic, including the law of the excluded middle, can be freely and legitimately used when applied to finitary judgments—and, more generally, to all judgments that satisfy the principle of double negation (all finitary, as well as all negative judgments are among the latter). More precisely, Kolmogorov (III, §4) demonstrates the following fact:

Let $\phi(X_1, X_2, \ldots)$ be a propositional formula with connectives $\neg$ and $\rightarrow$ only and with variables $X_1, X_2, \ldots$, and let ϕ be provable in the classical calculus $\mathfrak{H}$; then

$$\neg\neg A_1 \rightarrow A_1, \neg\neg A_2 \rightarrow A_2, \ldots \vdash \phi(A_1, A_2, \ldots),$$

where $\vdash$ indicates inference in the minimal calculus $\mathfrak{B}$, letters $A_1, A_2, \ldots$ being treated as constants, not variables (that means no substitution instead of A_i is allowed).

The precise boundary of the domain in which the application of classical propositional logic is justified intuitionistically has just been found: this domain coincides with the domain in which the principle of double negation is applicable. Here it is appropriate to stop and to think this brilliant epistemological result over. This result remains valid if ϕ contains together with negations and implications also conjunctions; however it does not hold when ϕ is allowed to contain disjunctions [Pl 88, §8].

REMARK. If one permits ϕ to contain all propositional connectives and understands $\vdash$ as deducibility in the full minimal calculus of Johansson, then, to secure validity of this new version of Kolmogorov's theorem, the "diagnostic formula" $\neg\neg A \rightarrow A$ should be replaced by the formula $(\neg\neg A \rightarrow A) \,\&\, (A \vee \neg A)$. And if one changes the "minimal" meaning of $\vdash$ to the intuitionistic one, i.e. understands $\vdash$ as deducibility in the intuitionistic calculus, then one should take $A \vee \neg A$ as the "diagnostic" formula. In a more general case let us consider some propositional (subclassical) calculus and let $\vdash$ denote inference in this calculus. Now one can seek a "diagnostic" formula $\delta(X)$ such that, for any $\phi(X_1, X_2, \ldots)$, if ϕ is classically provable then $\delta(A_1), \delta(A_2), \ldots \vdash \phi(A_1, A_2, \ldots)$, where $A_1, \ldots, A_n$ are arbitrary constant propositions. It is not obligatory for that δ to exist at all. But, as we just have seen, for the intuitionistic calculus δ is $X \vee \neg X$, for the Kolmogorov minimal calculus δ is $\neg\neg X \rightarrow X$, and for the Johansson minimal calculus δ is $(\neg\neg X \rightarrow X) \,\&\, (X \vee \neg X)$. And in a still more general predicate case a set of "diagnostic" δ_n should be considered, each δ_n having the form $\delta_n(W^n)$, where W^n is an n-ary predicate variable (see [Us Pl 91]).

Kolmogorov's theorem just formulated yields the following obvious corollary.

COROLLARY 1. *In any classical tautology—in which the connectives occurring are only implication and negation—let every variable be replaced by its negation; then the resulting formula is a theorem of the minimal calculus $\mathfrak{B}$.*

And this Corollary 1 trivially implies Corollary 2, whose formulation is obtained from that of Corollary 1 by changing "negation" to "double negation". None of the corollaries are mentioned in [ANK 25], but, formulating Corollary 2, Church indicates that this result is "substantially that of Kolmogoroff" [Ch 56, Ex.26.20].

Now, in the second place, why does one never obtain a contradiction while using the law of the excluded middle? Kolmogorov answers this question by demonstrating the following effect. Had one obtained a contradiction by using the law of the excluded middle, then one could obtain another contradiction without use of this law. I believe this was chronologically the first theorem on relative consistency of a calculus, i.e. consistency of a calculus under the presumption that some other calculus is consistent. Nowadays the very method by which this relative consistency has been achieved seems quite natural and almost obvious—it is the so-called embedding of one calculus, which consistency is to be checked, into another, which is presupposed to be consistent. But it was Kolgomgorov who invented this embedding method in the paper under discussion. This method manifests itself in a so-called embedding operation. Let L_1 and L_2 be two formal languages. A mapping of the set of formulas of L_1 to the set of formulas of L_2 is called, in the broadest sense, an *embedding operation* if it preserves some specific deduction features of formulas as, for example, the mapping $^\circ$:

$$(\vdash_{L_1} \mathbb{A}) \Leftrightarrow (\vdash_{L_2} \mathbb{A}^\circ),$$

or something like this. So there can be various species of embedding operations. Kolmogorov did not use the term "embedding operation", but presented chronologically the first instance of such an operation. According to Kolmogorov we consider a mapping *, which satisfies for any $\mathbb{A}$ and $\mathbb{U}_i$ the following two conditions:

$$(\mathbb{U}_1, \ldots, \mathbb{U}_k \vdash_{L_1} \mathbb{A}) \Rightarrow (\mathbb{U}_1^*, \ldots, \mathbb{U}_k^* \vdash_{L_2} \mathbb{A}^*),$$

$$(\mathbb{U}_1, \ldots, \mathbb{U}_k \vdash_{L_1} \neg\mathbb{A}) \Rightarrow (\mathbb{U}_1^*, \ldots, \mathbb{U}_k^* \vdash_{L_2} \neg\mathbb{A}^*).$$

We take the first condition from [ANK 25, IV, §3, footnote 18]. Any such mapping will be called an *embedding operation* (in a narrow sense). Now if an embedding operation exists for languages L_1 and L_2, then L_1 is obviously consistent relative to L_2 in the following sense: if a contradiction can be deduced in L_1 from some $\mathbb{U} = (\mathbb{U}_1, \ldots, \mathbb{U}_k)$, then necessarily a contradiction can be deduced in L_2 from the corresponding $\mathbb{U}^* = (\mathbb{U}_1^*, \ldots, \mathbb{U}_k^*)$. If the rules of inference in L_2 and $\mathbb{U}^*$ are beyond doubt, then L_1 with the axioms $\mathbb{U}$ is consistent. This is the main pattern of demonstrating relative consistency, and this pattern was invented by Kolmogorov. But he not only created this pattern, he proposed a specific embedding operation—now called the *Kolmogorov embedding operation*—for the case where L_1 and L_2 are respectively classical and intuitionistic mathematics. The Kolmogorov operation puts a double negation (which he denotes by "n") before each subformula. For example, if S is $\neg(A \to \forall x B)$, where A and B are atomic, then S^* is $n\neg n(nA \to n\forall x nB)$. So Kolmogorov obtains the consistency of classical mathematics up to intuitionistic mathematics, provided the axioms $\mathbb{U}$ of classical mathematics are such that their asterisk-translations $\mathbb{U}^*$ are intuitionistically true. His operation not only gives a metamathematical result of translatability of two very fundamental theories into each other, but simultaneously suggests an intuitionistically appropriate meaning for classical judgments: to receive such a meaning it suffices to interpret each proposition not as the assertion of itself as such (as it is interpreted in classical mathematics) but as the assertion of its double negation.

Two comments on Kolmogorov's demonstration of the consistency of classical mathematics.

Comment 1. When demonstrating that his operation is an embedding operation, Kolmogorov confined himself to the case when the formulas under consideration are built up by implications and negations only and the rules of inference are propositional rules. This restriction is not essential, and in Exercise 38.12 of Church's monograph the reader is invited to establish a similar result for the full first order predicate calculus (and, as Church writes, "this result is due in substance to Kolmogoroff" [Ch 56, n. 357]).

Comment 2. Kolmogorov's method leans upon the presumption that for every axiom A of mathematics the corresponding A^* is intuitionistically true. Let us listen to Hao Wang [HW 67]: "From this it would seem to follow that all classical mathematics is intuitionistically consistent (V, §1). As we know, however, this conclusion, even today, has not yet been firmly established so far as classical analysis and set theory are concerned. On the other hand, it seems not unreasonable to assert that Kolmogorov did foresee that the system of classical number theory is translatable into intuitionistic theory and therefore is intuitionistically consistent. In fact, it is not hard to work out his general indications and verify such a conclusion."

The paper [AMK 25] was published in Russian. Its English translation appeared only in [Hei 67]. It seems Church was the first Westerner who appreciated Kolmogorov's ideas. There are no references to [ANK 25] in [Jo 37] nor in [Gö 33]. In [Gö 33] Gödel demonstrated the relative consistency of the classical arithmetic using, in essence, Kolmogorov's method of embedding a classical theory into an intuitionistic one; in that paper Gödel proposed his own version of an embedding operation.

In 1984 Kolmogorov himself expressed the task of his paper of 1925 in the following words: "The paper [ANK 25] was intended to be an introductory part of a larger work. The constructions of models of different branches of classical mathematics within the frames of intuitionistic mathematics had to substantiate their consistency (while the consistency of intuitionistic mathematics was regarded as a consequence of its intuitive persuasiveness). To substantiate the consistency of classical propositional logic such a way was, of course, superfluous; however, it was believed that the method could be applied to substantiate the consistency of classical arithmetics (cf. Gödel's paper of 1933)" [ANK 85a].

As already mentioned, in [ANK 25] Kolmogorov proposed an intuitionistic meaning for the formulae of classical mathematics. And in the paper *Zur Deutung der intuitionistischen Logik* [ANK 32], that was finished in Göttingen on January 15, 1931, he, on the contrary, breathed into the formulae of intuitionistic propositional calculus a meaning which does not depend on any philosophical premises of intuitionism. Specifically, Kolmogorov proposed to treat every such formula not as an assertion but as a problem.

The concept of a problem is one of the fundamental concepts of mathematics, and it may be that Kolmogorov was the first who took it as the subject of technical discourse. According to Kolmogorov, a *problem* is a demand to find some entity satisfying a given condition; this entity may be of different kinds, say a set of numbers as in the problem of solving an equation, or an inference as in the

problem of proving a theorem. Any entity that satisfies the condition is called a *solution* of the problem. And to find such an entity is, by definition, to *solve* the problem. So Kolmogorov proposed that together with the logic of assertions, which in fact is traditional classical logic, there is a logic of problems. Formulae of the latter coincide in their form with ordinary propositional formulae, but the meaning is quite different. Propositional letters now denote not assertions but problems, $a \wedge b$ means the problem "solve both problems a and b"; $a \vee b$ means "solve at least one of the problems a and b"; $a \supset b$ means "assuming a solution of a is given, solve b" or, what means the same, "*reduce* b to a"; and $\neg a$ means "assuming a solution of a is given, obtain a contradiction" (it is interesting that Kolmogorov regards $\neg a$ as a more strong demand than to demonstrate the unsolvability of a).

So every propositional formula $p(a, b, c, \ldots)$ is a problem provided $a, b, c, \ldots$ are problems. If $x, y, z, \ldots$ are variables, $\vdash p(x, y, z, \ldots)$ is, by Kolmogorov's definition, the problem "find a general method (nowadays we say "algorithm") for solving $p(a, b, c, \ldots)$ whatever problems $a, b, c, \ldots$ are". It turns out that $\vdash \alpha$ is a solvable problem provided α is provable in Heyting's calculus.

In Hao Wang's words, "it seems not unreasonable to assert that Kolmogorov did foresee" the subsequent Kleene-Nelson semantics of realizability. As Kolmogorov himself indicated in the last footnote in his paper ("a comment added in proof"), his interpretation is close to Heyting's ideas of [Hey 31] (see also [Hey 34] and the letters of Kolmogorov to Heyting recently published [ANK 88]). But, as Heyting admitted [Hey 34, *Intuitionismus*, §5, no. 2], Kolmogorov goes further and suggests meaning independent of intuitionistic premises. It is appropriate to mention also Kolmogorov's following contributions to the field:

1. Comprehension of Heyting's calculus as a problem logic. Kolmogorov makes the clear distinction, which is absent in Heyting's work, between a statement ("Aussage") and a problem ("Aufgabe") and, correspondingly, between the logic of statements (of judgments, of propositions) and the logic of problems.

2. Understanding $a \supset b$ as the problem of reducibility of the problem b to the problem a. Heyting [Hey 34, *Intuitionsmus*, §5, no. 2] views the meaning of $a \supset b$ in a construction leading from any *proof* of a to a *proof* of b. That every two problems a and b generate the problem of reducibility of b to a is an essential step in logical thinking. This step is necessary to bring into life, say, Post's theory of reducibility of decision problems and so the general idea of relative computability. I believe Kolmogorov was the first who made this step.

The paper [ANK 32] consists of two sections. So far I have been discussing the first section, and in this section its author acts as a classical mathematician. But in the second section the author acts as an intuitionist. And as such he expressed the following view: for a universal statement it is, generally speaking, senseless to consider its negation as a definite statement. Thus the law of the excluded middle is trivially applicable to every statement whose negation makes sense. So the very subject of intuitionistic logic as a logic of statements evaporates. And from the intuitionistic point of view, logic can exist only as problem logic.

In 1984 Kolmogorov described the task of his paper as follows: "The paper [ANK 32] was written in the hope that in time the logic of solution of problems will become a permanent part of a course in logic. It was supposed to create a

united logical technique dealing with objects of two types—with statements and with problems." [ANK 85a]. Of course, Kolmogorov did not claim that he had fulfilled this task in his paper. But the direct and indirect influence of his ideas, especially in his country, is beyond any doubt. His study of problem logic was carried on by his students (Yu. T. Medvedev and others) and students of his students (D. P. Skvortsov and others). About this topic, see [Us Pl 85, IV].

Part II.
The Pure Theory of Algorithms

"The distinction between "the constructive" and "the nonconstructive" occupies a high place in all contemporary mathematical thinking" wrote Kolmogorov in one of his prefaces [ANK 54]. And the first section of [ANK 70] is titled "The growing role of finite mathematics". The interest in the distinction just mentioned and in finitary mathematics that stimulated his early work in logic remained with him all during his life, and this interest, as we shall see, brought him in the evening of his life to creating a surprising alliance between constructiveness and randomness. Naturally, he cannot pass by the very notion of algorithm, and he wrote the article "Algorithm" for the second edition of the ***Greater***[3] ***Soviet Encyclopaedia*** [ANK 50].

The history of development of the notion of algorithm is still awaiting its historian. I believe that Émil Borel was the first who mentioned this general notion in print, back in 1912. In 1936 celebrated papers of Post, Church, Turing, and Kleene appeared. But I would like to stress that these celebrated papers, though proceeding from the most general intuitive notion of algorithm, did not set out to investigate this very notion in all its generality; they had another task—to formalize the notion of a computable function and to present some restricted class of such algorithms that simultaneously can be described with mathematical rigour and are powerful enough to ensure a computation of any function which is computable in an intuitive (informal) sense. (In [Us Se 81, Part I, §2] and [Us Se 87, no. 1.2.0] such a class is called *representative*.) I believe Kolmogorov was the first one who tried to apprehend the concept of an algorithmic process in all its generality and to offer a mathematical definition of this concept. When I was a student at Moscow University there was a rumor about Kolmogorov's lecture in which he made an attempt to define the concept of an algorithmical process in terms of operating with boxes insertable one into another.

At the end of 1951 Kolmogorov gave me a sheet of paper which I still have today. The sheet contained a general definition of the notion of an algorithm created by him just then. That definition was elaborated, as Kolmogorov instructed me, in my graduation work in the spring of 1952, then was expounded in March of 1953 in a lecture of Kolmogorov's to the Moscow Mathematical Society [ANK 53], and, finally, was published in 1958 by Kolmogorov and me [ANK & U 58].

[3] Literally "Big", but usually called "Greater" in Western citations.

Of course, many of the features of Kolmogorov's general approach to algorithms seem obvious now; but let us not forget that Kolmogorov was the first who explicitly formulated these features.

So what is Kolmogorov's approach? It can be divided into two stages. The first stage is a philosophical scheme, and the second is a mathematical realization of the scheme.

The scheme, as it is exposed in [ANK 53] and [ANK & U 58], includes the following ideas. Suppose an algorithm Γ is given. Then:

1. At any moment there is some state of the algorithmic process. For instance, in a computer it is something like the assignment of zeros and ones in computer cells. Speaking on Turing machines, Kleene [Kl 52, §67] uses the term "situation" (and Turing himself, "complete configuration"), and in this case it consists of three components: the state of the tape, the mark of the head, and the location of the head on the tape.

2. For some states S the next state S^*, the successor of S, is defined.

3. The algorithmic process, beginning with some initial state (the input), proceeds by moving from any state S to the state S^*.

4. The process goes until it reaches a state S whose successor S^* is undefined, and then the so-called "resultless stop" occurs, or a state S for which there comes some stopping signal, and then we have a "resultful stop", this S being in this case the final result (the "solution", as Kolmogorov called it). So Γ is defined, i.e. comes to a result, only for those initial states beginning with which one obtains a resultful stop.

5. There is a partially defined mapping Ω_Γ from the set of all conceivable states to the same set such that $\Omega_\Gamma(S) = S^*$; the existence of such a mapping secures for the algorithm Γ what is called "the property of determiniativeness". So the process has *iterative* character and consists in iterating the mapping Ω_Γ (that mapping was called by Kolmogorov the "operator of immediate transformation").

6. The operator Ω_Γ is a *restricted local action* (see [Us Se 87, no.1.0.3]). This means that for any conceivable S there is an "essential part" of it, called *active zone*, and

(6.1) the size of this zone is bounded for the given algorithm Γ;

(6.2) the value $\Omega_\Gamma(S)$ depends not on the argument S in the whole but only on its active zone; and

(6.3) the operator $\Omega_\Gamma(S)$ transforms only the active zone of S and does not touch the rest of S; so it can do nothing but replace the active zone by something else, this something depending only on the zone replaced.

7. The stopping signal, regarded as a function of a state, also depends on the active zone, but on nothing else.

8. The uniform boundedness of the active zone implies that for Γ there can be only a finite set of nonisomorphic zones, and consequently Ω_Γ (as well as the stopping signal) can be described in finitary terms; and this secures the finitary character of the algorithm Γ.

Now let us proceed to a mathematical model of these ideas. We must explain first what process states are, or should be. Of course, they are constructive objects. But what are constructive objects?

As Kolmogorov believed, each state of every algorithmic process (or, what is the same, every constructive object) is an entity of the following structure. This entity consists of elements and connections; the total number of them is finite. Each connection has a fixed number of elements connected. Each element belongs to some type; each connection also belongs to some type. For every given algorithm the total number of element types and the total number of connection types are bounded.

This is all quite natural and rather obvious. So a constructive object (and any algorithm process state is such an object) can be regarded as a finite first-order structure of a finite signature, the signature being fixed for every particular algorithm.

But there is a noteworthy restriction on this structure. Kolmogorov demanded not only that each connection should have a fixed number of elements involved, but that each element can be involved only in a restricted set of connections (or relation instances in the first-order structure); the upper bound of number of connections involving any given element is fixed for every algorithm.

This restriction is essential. Indeed, let us consider an algorithm using directed graphs as process states. Kolmogorov's restriction not only forbids us to use graphs with unbounded fan-out (that is, the set of edges coming out from any node) but also (and this is crucial!) forbids us to use graphs with unbounded fan-in: the number of edges entering a node must be bounded for any fixed algorithm. We shall return to this point when discussing the so-called Schönhage machines.

It is not hard to find a geometrical presentation for Kolmogorov's idea of a constructive object. Such a presentation takes the form of a nondirected graph whose nodes (but not edges) are marked by some labels; those labels correspond not only to the element types and to the connection types but also, for any connection, to the positions of all elements incident to this connection, and, for any element, to the positions of all connections incident to this element. Then the following property of the graph necessarily holds: for any node, all nodes immediately (i.e. by only one edge) connected with it bear different labels. We require that the graph have exactly one distinguished node, called the *active node*.[4]

Now you have what is called a *Kolmogorov complex*: a finite undirected connected graph with labelled nodes and with the two conditions just formulated: different labels for the immediate neighbors of any node and the existence of exactly one active node.

Hence we can talk in graph terms only. First of all, let us stress that the variety of feasible labels is bounded for any fixed algorithm.

And now, about the active zone. It is, by Kolmogorov's definition, the neighborhood of the active node of some radius, this radius being fixed for any particular algorithm Γ. The immediate transformation Ω_Γ replaces the active zone by

[4]The active node together with distinction of labels secures for any graph of the kind described the existence of some intrinsic coordinate system such that each node possesses its own coordinate by which it can be identified.

some new subgraph fully determined by this zone. We omit here the technical details explaining how the substitution is actually performed. For those details see [Us Se 81, part I, §2], [Us Se 87, no.1.0.7] or [Us Se 87a, V].

In computer science they usually speak about *Kolmogorov machines*. Computer scientists have less interest than logicians in the foundational problem of whether the general concept of an algorithm can be strictly and appropriately defined. They treat Kolmogorov's formulation as a description of some computing device whose storage (which they often call "tape") can change its topology. The advantage of this device from a computer scientist's point of view is that it gives a good measure of time complexity and allows one to prove nice theorems about this measure. For this purpose it is more convenient to use Leonid Levin's version of a Kolmogorov machine, which was published by Yuri Gurevich a few years ago in his essay opening "The Logic in Computer Science Column" in BEATCS [Gu 88]. In that version all graphs are directed but symmetric, and nodes have no labels but edges are colored in such a way that edges coming out from any node have different colors. And the immediate transformation of S into S^* can be very easily described in Levin's terms. But Kolmogorov and we, following him, are here interested mainly in foundational questions.

Certainly every function computable by a Kolmogorov machine is partial recursive in an appropriate sense and vice versa; so one obtains a version of the Church-Post-Turing-Kleene thesis. But here one deals with a stronger effect. Since Kolmogorov's formulation claims to be an adequate definition of the most general concept of an algorithm, "every computation, performing only one restricted local action at a time, can be viewed as (not only simulated by, but actually being) the computation of an appropriate KU[5] machine (in the more general form). In a sense, this is stronger than Turing's thesis" [Gu 88, p. 174]. From this thesis it follows, in particular, for an appropriate understanding of real-time computation, that any function real-time computable by some computing device is also real-time computable by a suitable Kolmogorov machine, provided the device performs a restricted local action at a time. The latter condition rules out such devices as, say, random access machines. But what is "to perform a restricted local action at a time"? It must be said that the answer is not so clear.

To illustrate the difficulties let us slightly generalize the genuine Kolmogorov machine and introduce the *directed* Kolmogorov machine. To avoid confusion the previous genuine Kolmogorov machine will be called the *undirected* Kolmogorov machine. The only difference between undirected and directed machines is that the former deal with undirected graphs and the latter deal with directed ones. In the case of directed graphs there is the following requirement now: for any node A all the nodes which are the ends of directed edges coming out from A have different labels. (For Levin's presentation the requirement that the graph be symmetric is discarded.) The variety of edges coming out from a node—the fan-out— remains bounded in the process of computing by a fixed directed machine, but the variety of nodes entering a node—the fan-in—can grow unboundedly. So it is obvious

[5]KU means Kolmogorov-Uspensky; that is what these machines are called in [Gu 88].

that a computation of such a machine cannot be viewed as a computation of any undirected Kolmogorov machine. Thus we seemingly have a contradiction with Kolmogorov's thesis just formulated in the quotation from Gurevich's paper. But is it fair to view a single step of a directed machine as a restricted local action taking into account that in the performing such an action the fan-in can exceed any previously given bound? Perhaps it will be more correct to call actions of this kind *semi-restricted* or *semi-local* (as in [Us Se 87, no.1.2.1] and [Us Se 87a, VI]).

All that we have just said can be applied to the so-called Schönhage machine. In 1970 Schönhage published a description of his storage modification machines, the main property of which was their ability to change the topology of their storage [Schö 70]. As Schönhage himself admitted later [Schö 80], he knew nothing about the earlier invention by Kolmogorov. Though Schönhage used different terms, the Schönhage machines can be described as some kind of directed Kolmogorov machines.[6] It seems that Schönhage machines are more powerful than the genuine (that is, undirected) Kolmogorov machines: probably they can compute some functions essentially faster. At any case they are not in conformity with the genuine formulation of Kolmogorov. But perhaps they are not in conformity with human intuition, either—here I mean the intuition of realizability in physical time and physical space of an algorithmic process reduced to fairly elementary steps. As we have seen, the question is rooted in the problem of what is the admissible variety of constructive objects that can serve as states of a single algorithmic process.

Kolmogorov never lost his interest in algorithms till the end of his life. In the autumn of 1952 I became his postgraduate student, and one of the first instructions he gave me was to translate from German the book ***Rekursive Funktionen*** by Rózsa Péter [Pé 51]—apparently the first book ever published on that subject. He edited the publication of Russian translation (1954) and wrote a substantial preface to it [ANK 54]. I believe that in this preface there appears for the first time the idea of using a nonextendable function (i.e. a partial recursive function which cannot be extended to the total recursive function) as the technical source to get a recursively enumerable but nonrecursive set. And this was just one of many valuable suggestions that his students received from him. So, independently of Kleene he made the important observation that if in a formal language a set of valid formulas and a set of invalid formulas are recursively nonseparable, then there is no semantically complete logistic system for this language. It is precisely on this observation the papers [Us 53] and [Us 53a] are based.

In the autumn of 1953 Kolmogorov invited me to run with him a seminar on algorithm theory—more accurately, on the theory of recursive functions and recursively enumerable sets. The name of the seminar was "Recursive Arithmetic", and it was held once a week through the academic year 1953/54. Though not so crowded, this seminar, perhaps the first in the USSR on this topic, counted for something in the algorithmic education of Moscow logicians. At that seminar a

[6]That is why in [Sli 81] Schönhage machines are called Kolmogorov-Schönhage machines (while undirected Kolmogorov machines are called Kolmogorov-Uspensky machines).

set-theoretic presentation of the theory of computable functions and enumerable sets resembling N. N. Luzin's geometric presentation of descriptive set theory was tested and polished. I remember S. I. Adyan and Yu. T. Medvedev among the regular participants in that seminar. And among the results of the seminar I would like to mention the characterization of hyperimmune sets by means of nonexistence of any majorizing recursive function, which is adopted as a definition of those sets in [Ro 67, §9.5]. The problem of what nonmajorizable sets are was presented by Kolmogorov, in the autumn of 1953, to the seminar's participants, and then was answered simultaneously and independently by Medvedev and me. It turned out that those sets are exactly the hyperimmune ones in the original sense of Post and Dekker (see [Us 57, §6] and [Ro 67, §9.5, Theorem XV]).

At that seminar the theory of numerations, or of numberings, sprang up. The basic tasks and basic notions of this theory were explicitly suggested by Kolmogorov to the seminar's participants in February of 1954. The source was the study of Kleene's notation system for ordinal numbers; so this theory of Kleene's constituted the chief historical reason for the theory of numberings. In his brief speech Kolmogorov introduced the notion of a *numbering* (or of a numeration, or an enumeration[7]) as a mapping of a set of natural numbers onto a set of arbitrary elements (the "numbered set"); then Kleene's ordinal notation system turned out to be a numbering. In the same talk Kolmogorov suggested that we study various numbering as such in an abstract way, and introduced the principal notion of the future theory: the notion of reducibility of numberings, as well as the very term "reducible".

Considering two arbitrary numberings α and β of the same set, Kolmogorov distinguished the case where there is an algorithm which for any element of the numbered set and for any α-number of that element produces some β-number of the same element; and Kolmogorov marked this situation by the wording "α is *reducible* to β". These ideas by Kolmogorov were presented for the first time in [Us 55]. Nowadays the theory of numberings is a rather flourishing branch of algorithm theory (see, for example, [Er 77]) with profound applications to computer science, since any program system makes a numbering of the set of computable functions (see a numbering definition of the notion of a "mode", or "method" of programming in [Us 56] or in [Us 56a]; essentially that terminology was used in [ANK 65, §3]).

In the sixties (more precisely, in the years 1963–64) Kolmogorov came to creating his celebrated complexity theory and wrote his first paper on this topic [ANK 65]. There are two kinds of complexity—complexity of processes (time complexity, space complexity, and so on) and complexity of objects. The latter was the subject of Kolmogorov's studies.

Kolmogorov's ideas on complexity (and this was a rather ordinary fate of Kolmogorov's ideas) were published incompletely and with a great delay: the paper [ANK 70], though prepared in 1970 as a draft of Kolmogorov's talk at the

[7]I prefer the word "numbering" as in "Gödel numbering".

International Congress of Mathematicians at Nice, was published only in 1983. The main means of dissemination of Kolmogorov's ideas on complexity were his discussions with his students (at his seminars and privately). In several aspects Kolmogorov's views were summed up in his talk (jointly with the author of the present essay) at the First World Congress of the Bernoulli Society at Tashkent, 1986 [ANK & U 86; ANK & U 87].

That things differ not only in size but also in complexity is obvious. But Kolmogorov was the first who tried to measure the complexity of a thing numerically, by a number. His idea is very simple and reduces the notion of complexity of an object to the notion of size, though not of the object itself but of some description of that object. So there are *objects* and there are *descriptions*. An object is complex if all its descriptions are large, and it is simple if at least one of its description is small. Hence it is natural to define the *complexity* of an object as the minimal size of a description of this object.

There are, of course, many *modes of description*, and each mode leads to its own particular complexity measure. So one needs to choose a good mode of description to obtain a good complexity measure. The less the value of complexity for all the objects, the better is the description mode. But there is no mode that gives the least possible complexity for every object. So it is necessary to invent a reasonable method of comparison of description modes, i.e. to invent a reasonable ordering on the variety of modes. Such a method does exist: let us say that one mode is *no worse* than a second mode if the complexity related to the first mode is less than the complexity related to the second, up to an additive constant.

Then there is a so-called *optimal* mode of description which is the best with respect to the ordering just defined; in fact, there are many such optimal modes. The complexity, related to any optimal mode, of an object is "the true complexity" of the object. So there are many "true complexities" of the same object, but any two of them differ less than by a constant, with the constant depending not on the object but only on two corresponding optimal modes.

As to what a mode of description is, it is a function S from the set $\mathbb{N}$ of natural numbers, these numbers serving as descriptions, to a set D whose elements serve as objects. (So this S is a numbering of D; see above. Here is a connection between two theories, discovered by Kolmogorov—numbering theory and complexity theory.) If one restricts oneself to computable function S, then one can prove the existence of optimal modes.

In this presentation we have followed [ANK 70, §3]. To avoid terms like "the complexity related to an optimal mode of description" or "the true complexity", let us use, as a synonym, the term *entropy*. And, following [Us Se 81, Part I, §17], [Us Se 87, §1.17], [ANK & U 86, §1.4], or [ANK & U 87, no.1.4], let us present now a more explicit picture.

The objects under consideration are constructive objects. The descriptions are constructive objects, too. So we can encode objects and descriptions by finite binary words and then identify the encoded entities with their codes. Thus we transfer from objects and descriptions of arbitrary nature to binary words. The set of all (finite) binary words is denoted by Ξ. Any *mode of description* is a set $E \subset \Xi \times \Xi$; if $\langle p, y \rangle \in E$, then p is a *description* of y in the mode E.

The *complexity* $K_E(y)$ of y with respect to E is defined as

$$K_E(y) = \min\{l(p) \mid \langle p, y\rangle \in E\},$$

where $l(p)$ is the length of the word p; as usual, $K_E(y) = \infty$ if there is no p such that $\langle p, y\rangle \in E$.

Let E and F be two modes of description. We say that *the mode E is no worse than F* if for some constant C not depending on y and for every $y \in \Xi$

$$K_E(y) \leq K_F(y) + C.$$

Finally, let $\mathfrak{A}$ be a class of modes. A mode $A \in \mathfrak{A}$ is called *optimal* with respect to $\mathfrak{A}$ if it is no worse than any $F \in \mathfrak{A}$. The complexity related to any optimal mode of description is called *entropy*.

From now on (this is important!) only enumerable (i.e. recursively enumerable) modes of description will be considered.

We say a mode of description E *preserves equality* if

$$\langle p_1, y_1\rangle \in E \,\&\, \langle p_2, y_2\rangle \in E \,\&\, (p_1 = p_2) \Rightarrow (y_1 = y_2).$$

Two words u and v are said to be *comparable* if either u is an initial segment of v or v is an initial segment of u; when u and v are comparable we write $u \gamma v$. We say a mode E *preserves comparability* if

$$\langle p_1, y_1\rangle \in E \,\&\, \langle p_2, y_2\rangle \in E \,\&\, (p_1 \gamma p_2) \Rightarrow (y_1 \gamma y_2).$$

Originally Kolmogorov proved the existence of optimal modes in the class $\mathfrak{K}$ of all enumerable modes that preserve equality (see [ANK 65, §3] and [ANK 70, §3]). But Kolmogorov's theorem remains valid for several other important classes. For instance, there are optimal modes in the class $\mathfrak{L}$ of all enumerable modes that preserve comparability. This was proved by Kolmogorov's student Leonid Levin [Le 73], though in a slightly different system of notions and terms.

For any mode of description that is optimal with respect to $\mathfrak{K}$ the complexity related to this mode is called the *simple Kolmogorov entropy*; this notion (but not this wording) was introduced by Kolmogorov in [AK 65, §3] and [ANK 70, §3]. For any mode of description that is optimal with respect to $\mathfrak{L}$ the complexity related to this mode is called *monotone entropy*; it was introduced by Levin [Le 73].

In presenting the simple Kolmogorov entropy we have almost literally followed its presentation in [ANK 70, §3]. But this entropy first appeared in [ANK 65, §3] by the formula

$$K_A(y) = K_A(y \mid x),$$

where $K_A(y \mid x)$ is "the conditional complexity of y given x" with respect to A, and A is an optimal mode of conditional description or, closer to [ANK 65], an "optimal method of programming". The definitions are easy:

1. A *mode of conditional description* is an enumerable set $G \subset \Xi \times \Xi \times \Xi$ such that

$$\langle p, x, y_1\rangle \in G \,\&\, \langle p, x, y_2\rangle \in G \Rightarrow (y_1 = y_2).$$

If $\langle p, x, y\rangle \in G$, we say p is a *conditional description* of y given x. (In [ANK 65] Kolmogorov used not an enumerable set G but a computable function $y = \phi(p, x)$; he called ϕ "a programming method" and p "a program".)

2. The *conditional complexity* of y given x with respect to G is

$$K_G(y \mid x) = \min\{l(p) \mid \langle p, x, y\rangle \in G\}.$$

3. A mode A of conditional description is called *optimal* if for any mode ϕ

$$K_A(y \mid x) \leq K_\phi(y \mid x) + C_\phi,$$

where the constant C_ϕ does not depend on x and y.

The main theorem of [ANK 65, §3] states that there are optimal modes of conditional description. Nowadays for any optimal mode of conditional description the conditional complexity related to this mode is called the *conditional Kolmogorov entropy*. It is easy to see that for any optimal conditional A the formula $K_A(y) = K_A(y \mid 1)$ gives an optimal unconditional simple Kolmogorov entropy (and every such entropy can be obtained by means of that formula provided the conditional A is appropriately chosen).

This suggests considering two remaining classes of unconditional modes of description—the class $\mathfrak{D}$ and the class $\mathfrak{P}$. The class $\mathfrak{D}$ consists of all enumerable modes $E \subset \Xi \times \Xi$ that satisfy the condition

$$\langle p_1, y_1\rangle \in E \,\&\, \langle p_2, y_2\rangle \in E \,\&\, (p_1 = p_2) \Rightarrow (y_1 \gamma y_2),$$

and $\mathfrak{R}$ consists of all enumerable modes $E \subset \Xi \times \Xi$ that satisfy the condition

$$\langle p_1, y_1\rangle \in E \,\&\, \langle p_2, y_2\rangle \in E \,\&\, (p_1 \gamma p_2) \Rightarrow (y_1 = y_2).$$

For each of these two classes the Kolmogorov-style theorem stating the existence of optimal modes is valid. In such a way one gets the *decision entropy* [Zv Le 70], which is, by definition, the entropy related to an optimal mode in $\mathfrak{D}$, and the *prefix entropy* [Le 76], which is, by definition, related to an optimal mode in $\mathfrak{P}$.

A useful and detailed comparative presentation of the various versions of entropy has been published by V′yugin [V′yu 81].

Nowadays the theory of complexity for constructive (finite) objects is recognized as one of the major achievements of the theory of algorithms and as an integral part of this theory (see, for example, [Ma 77]).

To form a due appreciation of Kolmogorov's role in founding complexity theory let us try to trace the main aspects of his definition of entropy:

1. Kolmogorov's definition suggested the first means to measure the complexity of an object numerically.

2. This definition serves as the basis for the algorithmic theory of information (see Part II of this essay).

3. It opens a way to other versions of complexity measure which were introduced later by his student Levin and turned out to be essential for the algorithmic theory of probability.

4. It has many other applications. Various applications of the Kolmogorov theory of complexity and a vast bibliography are presented in [Li Vi 88].

To conclude this part, let us note that Kolmogorov founded not only the theory of complexity but also a scientific school in this field. A number of prominent specialists belong to that school (among them P. Martin-Löf, Ya. M. Barzdin', L. A. Levin, A. K. Zvonkin, P. Gács, A. L. Semenov, V. V. V'yugin, A. Kh. Shen', N. K. Vereshchagin, E. A. Asarin, V. V. Vovk, and An. A. Muchnik).

Part III.
Applications of the Theory of Algorithms to Probability Theory and Information Theory

Kolmogorov did not personally work on applications of algorithm theory to computer science. However, he took an interest in these applications and expressed the opinion that mathematical logic (including the theory of algorithms) and computer science must develop in close contact. He was even ready to rename the Department of Mathematical Logic, which he chaired, the Department of Mathematical Logic and Informatics. But personally he was concerned with building bridges from the abstract theory of algorithms not to computer science but to two other applied branches of mathematics—probability theory and information theory.

It is well known that Kolmogorov is one of the founding fathers of contemporary probability theory. As was usual for him, he was interested first of all in the essence of the things. He wanted to know the very nature of probability and randomness. And in one of his last talks (at the fourth USSR-Japan symposium, Tbilisi, 1982) he said: "There emerges the problem of finding the reasons for the applicability of the mathematical theory of probability to the phenomena of the real world. (...) Since randomness is defined as absence of regularity, we should primarily specify the concept of regularity. The natural means of such a specification is the theory of algorithms and recursive functions; the first attempt at its application in probability theory was that made by Church [Ch 40]" [ANK 82]. And nineteen years earlier he wrote: "I have already expressed the view (...) that the basis for the applicability of the results of the mathematical theory of probability to real "random phenomena" must depend on some form of the *frequency concept of probability*, the unavoidable nature of which has been established by von Mises in a spirited manner" [ANK 63].

Von Mises reduced randomness to stability of frequencies. From the logical point of view, there are at least two reasons why his approach is important. First, this approach permits one to recognize as random or nonrandom an individual sequence of symbols (only infinite sequences are considered)—an effect impossible in the classical theory of probabilities. Second, it gives us an opportunity to use substantially the notion of an algorithm. That opportunity was taken by Church and Kolmogorov. (A survey on various definitions of randomness for a single sequence and on their algorithmic versions is presented in [La 87].)

To briefly summarize von Mises' ideas let us confine ourselves to the two-letter alphabet of symbols $\{0, 1\}$ and let us consider an infinite sequence of zeros and ones. Does it make sense to assert that the sequence is random with respect to the

Bernoulli distribution with p and q as probabilities of zero and of one? Von Mises believed that the sequence must be recognized as random if every legitimate subsequence of it has the property of stability of frequencies. We say that a binary sequence possesses the *property of frequency stability* with limit p if $\lim(V_n/n) = p$, $n \to \infty$, where V_n is the number of zeros in the initial segment of length n. By von Mises' definition, a binary sequence is random (with respect to p and q) if every suitably, or admissibly, chosen subsequence of it has frequency stability with limit p.

The main problem here is what is the suitable, or admissible, way to choose a subsequence. Examples of suitably and nonsuitably chosen subsequences are obvious. It is admissible to compose a subsequence by picking exactly those terms of the original sequence whose subscripts are prime numbers [Mi 28, p. 32], or a subsequence consisting of exactly those terms which succeed the terms whose values are ones [Mi 28, p. 36]. It is not admissible to compose a subsequence by picking exactly those terms which themselves are equal to one. But the general idea of admissibility, or suitability, still remains rather vague. Though most likely Kolmogorov's solution of the problem is not a final one, it has not yet been surpassed and gives the most advanced definition of a von Mises random sequence known today.

Let us consider the subject in detail. To begin with, recall that the problem is to define what subsequence of a given sequence is admissible (suitable, legitimate) in the sense that the frequency stability in each such subsequence secures the randomness of the entire sequence. So we have to formulate the notion of a *selection rule*, which allows us to form a subsequence by making a legitimate selection of terms of the entire sequence. As we have already said, the first attempt to present a precise definition of an admissible selection rule dates back to 1940 and is due to Alonzo Church [Ch 40]. To comprehend Kolmogorov's definition it is proper to begin with Church's.

Church required the existence of an algorithm that could determine whether or not to select the term a_s depending on the values of the preceding terms a_0, $a_1, \ldots, a_{s-1}$ of the considered sequence. Thus the domain of this algorithm is the set Ξ of all binary words, and the range is the two-element set $\{\text{Yes}, \text{No}\}$. The algorithm operates as follows. Let the sequence being considered be $a_0 a_1 a_2 \cdots$, and suppose that its initial segment $a_0 a_1 \cdots a_{s-1}$ has been admitted as the input to the algorithm. Then if "Yes" results at the output, the term a_s must be chosen for inclusion in the subsequence being generated; but if "No" is the output, then a_s is passed over. The computable function $\varphi\colon \Xi \to \{\text{Yes}, \text{No}\}$, defined by this algorithm, is uniquely determined by the set $\{X \mid \varphi(X) = \text{Yes}\}$. Therefore, in order to specify a Church admissible selection rule, it suffices to designate some decidable set $D \subset \Xi$ and then to put $\varphi(x) =$ "Yes" if $x \in D$ and $\varphi(x) =$ "No" if $x \in \Xi \setminus D$.

Church's precise formulation is perhaps the closest to von Mises' original concept. It does not allow one to form a subsequence by selecting those terms whose successors are equal to one, and it is unclear whether von Mises would permit such a subsequence. The matter is that von Mises considered alternation of sequence members as a process going on in real physical time, like flipping a coin. It is obvious, however, that the particular subsequence just mentioned also must possess

the property of frequency stability provided the entire sequence is random. But this observation still does not lead to a direct contradiction between Church's definition and the traditional idea of randomness.

Such a contradiction was obtained later when it was discovered that Church random sequences can violate some laws of probability theory (see, for example, [Ja 70] or [Kn 69, no.35, ex. 31]). From a logician's point of view there is, however, a more remarkable contradiction found by Loveland [Lo 66]. Loveland indicated a Church random sequence that does not remain Church random after a computable permutation of its terms. This effect obviously contradicts the human intuition of randomness. Thus one finds an important algorithmic property of randomness: a random sequence must remain random after any computable permutation. It is essential that any new sequence $a_{\gamma_0}, a_{\gamma_1}, \ldots$ formed from $a_0, a_1, \ldots$ by means of some Church admissible rule be a *strict* subsequence. The word "strict" means that the terms of the subsequence should proceed one after another in the same order as in the entire sequence, i.e.

$$\gamma_0 < \gamma_1 < \gamma_2 < \cdots. \tag{1}$$

It is restriction (1) that makes Loveland's example possible.

So it was necessary to create a new and more general definition of an admissible selection rule. Such a definition was proposed by Kolmogorov in 1963 in Remark 2 of his article [ANK 63]. We shall present this definition in the words of Kolmogorov's talk [ANK 82]:

"According to [ANK 63], the rule of selection is given by means of algorithm (or, if you like, by Turing machine). Selection of the next member of the sequence takes place in the following way. The input information consists of the finite sequence of the members $n_1, n_2, \ldots, n_k$ and values $x_{n_1}, x_{n_2}, \ldots, x_{n_k}$ of the members of the original sequence. The output of the algorithm is, firstly, the number n_{k+1} of the next scanned element $x_{n_{k+1}}$ (this number must coincide with none of those $n_1, \ldots, n_k$; as to the order of the numbers $n_1, \ldots, n_k, n_{k+1}$ no restrictions are put to it); secondly, the indication whether $x_{n_{k+1}}$ is selected only to be scanned or the algorithm decided to include $x_{n_{k+1}}$ into the sequence selected.

"On the next step of the algorithm's work its input consists already of a longer sequence of numbers $n_1, \ldots, n_{k+1}$ and values $x_{n_1}, \ldots, x_{n_{k+1}}$; the algorithm naturally starts its work from the empty set.

"Expansion, as compared to [Ch 40], consists in the fact that the order of members in the selected subsequence should not obligatory coincide with their order in the original sequence."

Thus the main feature of Kolmogorov's definition is that the requirement (1) has been discarded and the terms of the sequence are allowed to proceed in a new order.

To avoid any vagueness, let us note that the algorithm occurring in the above quotation from [ANK 82] need not converge totally, for all instances of input information; if the algorithm diverges, i.e. has no output, the selection procedure stops and the subsequence turns out to be finite; in this case there is no frequency stability requirement for this subsequence (the same absence of the frequency stability requirement holds in the case when a Church admissible selection rule

gives a finite subsequence). The sequence is recognized as random or, more exactly, *Kolmogorov stochastic* if any subsequence obtained from it by a Kolmogorov admissible selection rule has the property of frequency stability provided the subsequence is infinite. A more formalized definition of a Kolmogorov admissible selection rule and hence of a Kolmogorov stochastic sequence can be found in [Us Se 87, no.2.6.1], [ANK & U 86, §1.6] or [ANK & U 87, no.1.6].

Kolmogorov's definition, being the most general up to now, eliminates Loveland's undesirable effect. But problems remain. Here is one of them. Is it reasonable to require that each sequence which is obtained from a random sequence by means of an admissible (in some sense) selection rule be itself random in the same sense? In any case, it is still unknown whether any subsequence formed by applying a Kolmogorov admissible selection rule to a Kolmogorov stochastic sequence is itself a Kolmogorov stochastic sequence. Plausibly it could turn out that this is not so because the class of Kolmogorov stochastic sequences turns out to be too broad: using the method of [La 87], Shen′ recently [Shen′ 88] constructed an example of such a Kolmogorov stochastic sequence that is not random in the sense of being chaotic or being typical; see below. A general discussion about what the selection rules should be is presented in [Us Se 87, no.2.6.2]; see also [Shen′ 82].

And now we move to the main contribution of Kolmogorov to the algorithmic theory of probability. That is the concept of chaoticness. Though Kolmogorov personally was not fully successful in the ultimate formalization of this concept, he suggested the basic ideas for such a formalization. In 1971–72 Kolmogorov's design was fulfilled by his student Leonid Levin, and then independently by Claus Peter Schnorr.

To make our presentation more clear and to avoid any confusion of concepts and terms, let us distinguish between the general intuitive notion of *random* sequence and the no less intuitive notions of a *stochastic* sequence and a *chaotic* sequence. The term "stochastic" was proposed by Kolmogorov for Mises random sequences: to be *stochastic* means that the sequence and all admissibly selected subsequences of it possess the property of frequency stability. The term "chaotic" was approved by Kolmogorov; being *chaotic* means that the sequence demonstrates an absence of regularity, that there is no simple law governing the alternation of its terms. And it is generally accepted that every random sequence is both stochastic and chaotic. So the sequences that are random in von Mises', Church's and Kolmogorov's sense should be called respectively "stochastic", "Church stochastic" and "Kolmogorov stochastic".

Let us proceed to chaoticness. This concept, though not in its final details, was suggested by Kolmogorov to his students in the sixties. I believe these ideas were presented also in Kolmogorov's talk at the International Congress of Mathematicians at Nice in 1970—as far as I can judge by reading the draft [ANK 70] of that talk.

In §6 of [ANK 70] Kolmogorov declares an intention to call a binary sequence $x = (x_1, x_2, \ldots, x_n, \ldots)$ chaotic[8] with respect to $p = q = \frac{1}{2}$ if its initial segments

[8]Kolmogorov used the word "random".

$x^n = (x_1, \ldots x_n)$ have the complexity

$$K(x^n) \geq n - C, \tag{2}$$

where C is a constant depending on x but not on n. Simultaneously, there and then, Kolmogorov cites a theorem by Martin-Löf [ML 66] according to which there is no sequence that satisfies (2) provided K is the simple Kolmogorov entropy. Nevertheless, as it clearly follows from [ANK 70], Kolmogorov does not abandon his intention and his hope.

Martin-Löf's negative effect of apparent nonexistence of chaotic sequences was rooted in some specific features of the class of modes of description that was adopted by Kolmogorov in order to define the complexity measure. If one changes this class $\mathfrak{R}$ to Levin's class $\mathfrak{L}$ (or to Levin's $\mathfrak{P}$; see Part II of this essay), then Martin-Löf effect will be eliminated: there do exist sequences x satisfying (2) provided K is monotone entropy or prefix entropy. And in these cases one obtains a proper definition of chaoticness, as Kolmogorov predicted. In other words, the monotone entropy and prefix entropy used by Levin turned out to be good enough to meet the Kolmogorov scheme for defining chaoticness.

But what are "proper" and "good enough"? These expressions have the following strict sense. For the monotone entropy KM the inequality

$$KM(x^n) \geq n - C \tag{3}$$

with C not depending on n gives the *definition of a chaotic sequence* (as a sequence that satisfies (3)) which is equivalent to the earlier definition of a random sequence given by Martin-Löf, also a student of Kolmogorov. Martin-Löf's definition [ML 66a] was to some extent induced by Kolmogorov's attempt to define the concept of a random finite string by means of tests; problems of the definition of randomness for finite strings, however, are beyond this essay; see on these problems, for example, [Shen' 83], [ANK & U 86, Chapter II], or [ANK & U 87, no. 2].

In terminology introduced in [ANK & U 86] and [ANK & U 87], Martin-Löf has defined the concept of a *typical* sequence. (Speaking informally, the property of being typical is the property of belonging to any reasonable majority). So one can say that the chaotic sequences defined in monotone entropy terms by Levin coincide with the typical sequences defined by Martin-Löf. This coincidence was established by Levin in 1971 (see [Le 73]). If in (3) one changes the monotone entropy KM to the prefix entropy KP, one obtains the same class of chaotic sequences [V'yu 81, Corollary 3.2]. And the same class will be obtained if one changes KM to the so-called *process complexity*, introduced by C. P. Schnorr [Schn 73]; and Schnorr [Schn 73] has demonstrated the coincidence of chaotic sequences generated by his process complexity with Martin-Löf typical sequences. (In [Schn 77] Schnorr discarded the process complexity and adopted a version of the definition of monotone entropy.)

The coincidence of the class of all chaotic sequences (and one obtains the same class starting from either Levin's monotone entropy, or Levin's prefix entropy, or Schnorr's process complexity) with the class of all typical sequences, proved by Martin-Löf, enables us to think that we have finally got the proper class of "truly random" sequences. And this is a justification of Kolmogorov's formula (2). Let

us remark that any chaotic sequence is Kolmogorov stochastic, but not vice versa because of Shen"s example [Shen′ 88, Theorem 1].

Up to now we have spoken only of randomness with respect to the uniform Bernoulli distribution where $p = q = \frac{1}{2}$. Generalization to arbitrary p and q (provided they are computable reals) is obvious. And a further generalization is also possible, to arbitrary computable probability distributions; the equivalence of being chaotic to being typical remains valid.

Though the simple Kolmogorov entropy turned out to be somewhat inadequate to define randomness, this notion much facilitated the development of the algorithmic theory of information. Indeed, the *conditional Kolmogorov entropy* (which is nothing but a conditional version of the simple Kolmogorov entropy) served as the cornerstone in founding this theory. And this founding was performed by Kolmogorov [ANK 65], [ANK 68] (see also the paper [Zv Le 70], thoroughly edited by Kolmogorov; the definition and the theorems in §5 of that paper belong to Kolmogorov).

According to Kolmogorov, the basic concepts of information theory—either Shannon's traditional probabilistic theory or Kolmogorov's new algorithmic one—are

(A) the amount of information $I(x:y)$, contained in an object x, about an object y,

(B) the amount of information $H(y\,|\,x)$ necessary to describe an object y provided an object x is already given, and

(C) the (unconditional) entropy $H(y)$ of an object y.

Of these three quantities the second is initial, and the other two are defined by the equalities

$$H(y) = H(y\,|\,e), \tag{4}$$

$$I(x:y) + H(y\,|\,x) = H(y), \tag{5}$$

where e is an "a priori known object".

In the traditional probabilistic theory of information both objects x and y are random variables; in the algorithmic theory of information they are constructive objects. The immediate tasks of the algorithmic version of the theory are:

1°. to check whether basic formulae holding in the probabilistic case remain valid in the algorithmic case, and

2°. to establish relations between the two versions, algorithmic and probabilistic, of information theory quantities, especially entropy.

As to the first task, let us recall that in the probabilistic case, which is explored well enough, the following formulae are valid:

(P1) $H(y\,|\,y) = 0$,

(P2) $I(x:y) \geq 0$,

(P3) $I(x:y) = I(y:x)$,

(P4) $H(\langle x, y\rangle) = H(x) + H(y\,|\,x)$.

In the algorithmic theory of information one takes $K(y\,|\,x)$ as $H(y\,|\,x)$. Here $K(y\,|\,x)$ is the conditional Kolmogorov entropy (see Part II, above). Hence $H(w)$ turns out to be the simple Kolmogorov entropy of w. As this entropy is defined up to a bounded summand, "all statements of the algorithmic theory of information

in their general formulations are true only up to bounded terms" [ANK 68, no.3]. So it is not unexpected that in the algorithmic version the formulae (P1) and (P2) change to inequalities

(A1) $0 \leq H(y \mid y) \leq C$,

(A2) $I(x:y) \geq -C$,

where C is a positive constant not depending on x and y. More surprising is the appearance of logarithmic terms in the algorithmic analogues of (P3) and (P4):

(A3) $I(x:y) = I(y:x) + O(\log_2 H(\langle x, y\rangle))$,

(A4) $H(\langle x, y\rangle) = H(x) + H(y \mid x) + O(\log_2 H(\langle x, y\rangle))$;

this is a theorem of Kolmogorov (and independently Levin; see [Zv Le 70, Theorem 5.2]).

Now we look at relations between the algorithmic and the probabilistic versions of the theory of information. As both entropies, Kolmogorov's and Shannon's, related to the number of digits that are needed to encode a message, it was expected that those two quantities are close to one another. And they really are. This fact is reflected in two important theorems established, directly or indirectly, by Kolmogorov. (Versions of these theorems are formulated in [Ba 77]; see also [Shen' 87a].) Those theorems of Kolmogorov connect the two so-called specific entropies, the Kolmogorov and the Shannon, of a long word.

THEOREM 1. *Let A be a binary word of length n and let p and q be the frequencies of zeros and ones in A. Let $K(A)$ be the simple Kolmogorov entropy of A, and let H be the Shannon entropy of the random variable taking its two values with probabilities p and q: $H = -p\log_2 p - q\log_2 q$. This H can be treated as the specific (i.e. taken per letter) Shannon entropy of A. The specific Kolmogorov entropy is, respectively, $K(A)/n$. Then*

$$K(A)/n \leq H + C\log_2 n, \tag{6}$$

where C does not depend on n [Zv Le 70, Theorem 5.1].

THEOREM 2. *Consider the Bernoulli probability distribution with the probabilities p and q on the set of all infinite binary sequences. Let $H = -p\log_2 p - q\log_2 q$ be the specific Shannon entropy. The specific Kolmogorov entropy of a given sequence w is, by definition, $\lim_{n\to\infty} K(w^n)/n$, where w^n is the initial segment of w having the length n. Almost certainly the specific Kolmogorov entropy does exist and is equal to H.*

This fact can be easily inferred from formula (5.18) of [Zv Le 70], which is part of the proof of Kolmogorov's Theorem 5.3 of [Zv Le 70].

And about 1985 Kolmogorov wrote in [ANK 87a, III]:

"The algorithmic approach is based on using the theory of algorithms to define the notion of entropy, or complexity, of a finite object and the notion of the information in a finite object about another one. An intuitive difference between "simple" and "complex" objects was felt, apparently, for a long time. On the way to a formalization of this difference, an obvious difficulty arises: what has a simple description in one language, may have no simple description in another language; and it is not clear which mode of description should be chosen. The main discovery, due to R. Solomonoff and myself, is the following: with the aid of the theory of algorithms it is possible to restrict this arbitrary choice and to define the complexity almost invariantly (the replacement of one mode of description by another

leads merely to the adding of a bounded summand).[9] Among my papers the most complete exposition of this idea is contained in [ANK 70] (see also A. Kh. Shen"s comment on [ANK 65; ANK 68; ANK 70])".

The comment of Shen', mentioned by Kolmogorov, is [Shen' 87a].

References

[ANK 25] A. N. Kolmogorov, *On the principle "tertium non datur"*, ***Mathematicheskiĭ Sbornik***, vol. 32 (1924/25), pp. 646–667; English translation, *On the principle of excluded middle*, in [Hei 67], pp. 416–437.

[ANK 32] ———, *Zur Deutung der intuitionistischen Logik*, ***Mathematische Zeitschrift***, vol. 35 (1932), pp. 58–65.

[ANK 50] ———, *Algorithm*, ***Greater Soviet encyclopedia***, 2nd ed., Vol. 2, Gosudarstvennoe Nauchnoe Izdatel'stvo "Bol'shaya Sovetskaya Èntsiklopediya", Moscow, 1950, p. 65. (Russian)

[ANK 53] ———, *On the notion of algorithm*, ***Uspekhi Matematicheskikh Nauk***, vol. 8 (1953), no. 4 (56), pp. 175–176 (Russian); reprinted in [ANK 87], Russian, p. 24.

[ANK 54] ———, *Translation editor's preface*, in the Russian translation of [Pé 51], IL, Moscow, 1954, pp. 3–10. (Russian)

[ANK 63] ———, *On tables of random numbers*, ***Sankhyā: The Indian Journal of Statistics, Series A***, vol. 25 (1963), pp. 369–376.

[ANK 65] ———, *Three approaches to the definition of the concept of the "amount of information"*, ***Problemy Peredachi Informatsii***, vol. 1 (1965), no. 1, pp. 3–11; English translations: (a) ***Problems of Information Transmission***, vol. 1 (1965), no. 1, pp. 1–7; (b) ***International Journal of Computer Mathematics***, vol. 2 (1968), pp. 157–168; (c) ***Selected Translations in Mathematical Statistics and Probability***, vol. 7, American Mathematical Society, Providence, Rhode Island, 1968, pp. 293–302.

[ANK 68] ———, *Logical basis for information theory and probability theory*, ***IEEE Transactions on Information Theory***, vol. IT-14 (1968), pp. 662–664; Russian version, ***Problemy Peredachi Informatsii***, vol. 5 (1969), no. 3, pp. 3–7.

[ANK 68a] ———, *Some theorems on algorithmic entropy and the algorithmic quantity of information*, ***Uspekhi Matematicheskikh Nauk***, vol. 23 (1968), no. 2 (140), p. 201. (Russian)

[ANK 70] ———, *Combinatorial foundations of information theory and the calculus of probabilities*, ***Uspekhi Matematicheskikh Nauk***, vol. 38 (1983), no. 4 (232), pp. 27–36; English translation, ***Russian Mathematical Surveys***, vol. 38 (1983), no. 4, pp. 29–40. [This paper was written in 1970.]

[ANK 82] ———, *On logical foundations of probability theory*, ***Probability theory and mathematical statistics (proceedings of the fourth USSR-Japan symposium held at Tbilisi, 1982)***, Lecture Notes in Mathematics, vol. 1021, Springer-Verlag, Berlin, 1983, pp. 1–5.

[ANK 85] ———, ***Selected works***. Vol. I: ***Mathematics and mechanics***, "Nauka", Moscow, 1985; English translation, Kluwer, Dordrecht, 1991.

[ANK 85a] ———, *About my papers on intuitionistic logic*, in [ANK 85], p. 393 (Russian), 451–452 (English).

[ANK 87] ———, *Selected works*. Vol. III: ***Theory of information and theory of algorithms***, "Nauka", Moscow, 1987; English translation, Kluwer, Dordrecht (to appear).

[ANK 87a] ———, *About my papers on information theory and some of its applications*, in [ANK 87], Russian, pp. 251–253.

[ANK 88] ———, *Letters of A. N. Kolmogorov to A. Heyting*, ***Uspekhi Matematicheskikh Nauk***, vol. 43 (1988), no. 6 (264), pp. 75–77; English translation, ***Russian Mathematical Surveys***, vol. 43 (1988), no. 6, pp. 89–93.

[9] As far as I'm aware, the first publication in print that contained a reconstruction of the theory of information on an algorithmic base was a paper by R. Solomonoff published in 1964. I came to a similar conception in 1963–64, not knowing Solomonoff's papers, and published my first paper [ANK 65] on this topic at the beginning of 1965. [Kolmogorov's footnote, from [ANK 87a].]

[ANK & D 82] A. N. KOLMOGOROV and A. G. DRAGALIN, ***Introduction to mathematical logic***, Moskovskiĭ Gosudarstvennyĭ Universitet, Moscow, 1982. (Russian)

[ANK & D 84] ———, ***Mathematical logic: supplementary chapters***, Moskovskiĭ Gosudarstvennyĭ Universitet, Moscow, 1984. (Russian)

[ANK & U 58] A. N. KOLMOGOROV and V. A. USPENSKY, *On the definition of an algorithm*, ***Uspekhi Matematicheskikh Nauk***, vol. 13 (1958), no. 4 (82), pp. 3–28; English translation, ***American Mathematical Society Translations***, ser. 2, vol. 29 (1963), pp. 217–245.

[ANK & U 86] ———, *Algorithms and randomness*, ***Proceedings of the first world congress of the Bernoulli Society (Tashkent, 1986)***. Vol. 1: ***Probability theory and applications***, VNU Science Press, Utrecht, 1987, pp. 3–53.

[ANK & U 87] ———, *Algorithms and randomness*, ***Teoriya Veroyatnosteĭ i eë Primeneniya***, vol. 32 (1987), pp. 425–455; English translation, ***Theory of Probability and Its Applications***, vol. 32 (1987), pp. 389–412.[10]

[BA 77] YA. M. BARZDIN', *Algorithmic information theory*, ***Mathematical encyclopedia***. Vol. 1, Izdatel'stvo "Sovetskaya Èntsiklopediya", Moscow, 1977, cols. 219–222; English translation, Kluwer, Dordrecht, 1988, vol. 1, pp. 140–142.

[Ch 40] A. CHURCH, *On the concept of a random sequence*, ***Bulletin of the American Mathematical Society***, vol. 46 (1940), pp. 130–135.

[Ch 56] ———, ***Introduction to mathematical logic***. Vol. 1, Princeton University Press, Princeton, New Jersey, 1960.

[Er 77] YU. L. ERSHOV, ***The theory of numberings***, "Nauka", Moscow, 1977; German translation of an earlier version, Parts I, II, III, ***Zeitschrift für Mathematische Logik und Grundlagen der Mathematik***, vol. 19 (1973), pp. 289–388; vol. 21 (1975), pp. 473–584; vol. 23 (1977), pp. 289–371.

[Gö 33] K. GÖDEL, *Zur intuitionistischen Arithmetik und Zahlentheorie*, ***Ergebnisse eines Mathematischen Kolloquiums***, vol. 4 (1931/32), pp. 34–38 (1933).

[Gu 88] YU. GUREVICH, *The logic and computer science column*, ***Bulletin of the European Association for Theoretical Computer Science***, No. 35 (June 1988), pp. 71–82.

[Hei 67] J. VAN HEIJENOORT (editor), ***From Frege to Gödel: a source-book in mathematical logic, 1879–1931***, Harvard University Press, Cambridge, Massachusetts, 1967.

[Hey 30] A. HEYTING, *Die formalen Regeln der intuitionistischen Logik*, ***Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-Mathematische Klasse***, 1930, pp. 42–56.

[Hey 31] ———, *Die intuitionische Grundlegung der Mathematik*, ***Erkenntnis***, vol. 2 (1931/32), pp. 106–115.

[Hey 34] ———, ***Mathematische Grundlagenforschung. Intuitionismus. Beweistheorie***, Springer-Verlag, Berlin, 1934.

[Hi 22] D. HILBERT, *Die logischen Grundlagen der Mathematik*, ***Mathematische Annalen***, vol. 88 (1923), pp. 151–165.

[HW 67] HAO WANG, Preface to the English translation of [ANK 25], in [Hei 67], pp. 414–416.

[Ja 70] K. JACOBS, *Turing-Maschinen und zufällige* 0-1-*Folgen*, ***Selecta mathematica*** (K. Jacobs, editor), Vol. II, Springer-Verlag, Berlin, 1970, pp. 141–167.

[Jo 37] I. JOHANSSON, *Der Minimalkalkül, ein reduzierter intuitionistischer Formalismus*, ***Compositio Mathematica***, vol. 4 (1937), pp. 119–136.

[Kl 52] S. C. KLEENE, ***Introduction to metamathematics***, Van Nostrand, Princeton, New Jersey, 1952.

[Kn 69] D. E. KNUTH, ***The art of computer programming***. Vol 2: ***Seminumerical algorithms***, Addison-Wesley, Reading, Massachusetts, 1969.

[La 87] M. VAN LAMBALGEN, *Von Mises' definition of random sequences reconsidered*, this JOURNAL, vol. 52 (1987), pp. 725–755.

[Le 73] L. A. LEVIN, *On the notion of a random sequence*, ***Doklady Akademii Nauk SSSR***, vol. 212 (1973), pp. 548–550; English translation, ***Soviet Mathematics Doklady***, vol. 14 (1973), pp. 1413–1416.

[10]There are two regrettable mistakes in the English version. On p. 394, line 2 from the bottom, and p. 395, lines 1 and 3 from the top, the word "countable" should be replaced by "enumerable (i.e. recursively enumerable)". Also, in the top line on p. 395, the word "in" should be deleted.

[Le 76] ———, *Various measures of complexity for finite objects (axiomatic description)*, ***Doklady Akademii Nauk SSSR***, vol. 227 (1976), pp. 804–807; English translation, ***Soviet Mathematics Doklady***, vol. 17 (1976), pp. 522–526.

[Li Vi 88] M. Li and P. M. B. Vitányi, *Kolmogorov complexity and its applications*, ***Handbook of theoretical computer science***. Vol. A: ***Algorithms and complexity***, (J. van Leeuwen, editor), Elsevier, Amsterdam, and MIT Press, Cambridge, Massachusetts, 1990, pp. 187–254.

[Lo 66] D. Loveland, *A new interpretation of the von Mises' concept of random sequence*, ***Zeitschrift für Mathematische Logik und Grundlagen der Mathematik***, vol. 12 (1966), pp. 279–294.

[Ma 77] Yu. I. Manin, *Kolmogorov complexity*, Chapter VI, §9 in Yu. I. Manin, ***A course in mathematical logic***, Springer-Verlag, Berlin, 1977, pp. 225–230.

[Mi 28] R. von Mises, ***Wahrscheinlichkeit, Statistik und Wahrheit***, J. Springer, Wien, 1928.

[ML 66] P. Martin-Löf, *On the concept of a random sequence*, ***Teoriya Veroyatnosteĭ i eë Primeneniya***, vol. 11 (1966), pp. 198–200; English translation, ***Theory of Probability and Its Applications***, vol. 11 (1966), pp. 177–179.

[ML 66a] ———, *The definition of random sequences*, ***Information and Control***, vol. 9 (1966), pp. 602–619.

[Pé 51] R. Péter, ***Rekursive Funktionen***, Akademischer Verlag, Budapest, 1951.

[Pl 88] V. E. Plisko, *The Kolmogorov calculus as a fragment of minimal calculus*, ***Uspekhi Matematicheskikh Nauk***, vol. 43 (1988), no. 6 (264), pp. 79–91; English translation, ***Russian Mathematical Surveys***, vol. 43 (1988), no. 6, pp. 95–110.

[Ro 67] H. Rogers, Jr., ***Theory of recursive functions and effective computability***, McGraw-Hill, New York, 1967; 2nd ed., MIT Press, Cambridge, Massachusetts, 1987.

[Schn 73] C. P. Schnorr, *Process complexity and effective random tests*, ***Journal of Computer and System Sciences***, vol. 7 (1973), pp. 376–388.

[Schn 77] ———, *A survey of the theory of random sequences*, ***Basic problems in methodology and linguistics: proceedings of the fifth international congress of logic, methodology and philosophy of science, part III*** (R. E. Butts and J. Hintikka, editors), Reidel, Dordrecht, 1977, pp. 193–211.

[Schö 70] A. Schönhage, *Universelle Turing Speicherung*, ***Automatentheorie und formale Sprachen (Tagung, Oberwolfach, 1969***; J. Dörr and G. Hotz, editors), Bibliographisches Institut, Mannheim, 1970, pp. 369–383.

[Schö 80] ———, *Storage modification machines*, ***SIAM Journal on Computing***, vol. 9 (1980), pp. 490–508.

[Shen′ 82] A. Kh. Shen′, *Frequency approach to defining the concept of a random sequence*, ***Semiotika i Informatika***, vol. 18, VINITI, Moscow, 1982, pp. 14–42. (Russian)

[Shen′ 83] ———, *The concept of Kolmogorov* (α, β)*-stochasticity and its properties*, ***Doklady Akademii Nauk SSSR***, vol. 271 (1983), pp. 1337–1340; English translation, ***Soviet Mathematics Doklady***, vol. 28 (1983), pp. 295–299.

[Shen′ 87] ———, *Tables of random numbers*, in [ANK 87], Russian, pp. 270–274.

[Shen′ 87a] ———, *Algorithmic theory of information*, in [ANK 87], Russian pp. 257–261.

[Shen′ 88] ———, *On relations between different algorithmic definitions of randomness*, ***Doklady Akademii Nauk SSSR***, vol. 302 (1988), pp. 548–552; English translation, ***Soviet Mathematics Doklady***, vol. 38 (1989), pp. 316–319.

[Sli 81] A. O. Slisenko, *Complexity problems in computational theory*, ***Uspekhi Matematicheskikh Nauk***, vol. 36 (1981), no. 6 (222), pp. 21–103; English translation, ***Russian Mathematical Surveys***, vol. 36 (1981), no. 6, pp. 23–125.

[So 64] R. Solomonoff, *A formal theory of inductive inference. Part I*, ***Information and Control***, vol. 7 (1964), pp. 1–22.

[Us 53] V. A. Uspensky, *Gödel's theorem and the theory of algorithms*, ***Uspekhi Matematicheskikh Nauk***, vol. 8 (1953), no. 4 (56), pp. 176–178. (Russian)

[Us 53a] ———, *Gödel's theorem and the theory of algorithms*, ***Doklady Akademii Nauk SSSR***, vol. 91 (1953), pp. 737–740; English translation, ***American Mathematical Society Translations***, ser. 2, vol. 23 (1963), pp. 103–107.

[Us 55] ———, *Systems of enumerable sets and their numberings*, ***Doklady Akademii Nauk SSSR***, vol. 105 (1955), pp. 1155–1158. (Russian)

[Us 56] ———, *Computable operations and the notion of a program*, ***Uspekhi Matematicheskikh Nauk***, vol. 11 (1956), no. 4 (70), pp. 172–176. (Russian)

[Us 56a] ———, *The notion of a program and computable operators*, ***Proceedings of the third all-union mathematical congress***, Vol. 1: ***Sectional reports***, Izdatel'stvo Akademii Nauk SSSR, Moscow, 1956, p. 186. (Russian)

[Us 57] ———, *Some notes on recursively enumerable sets*, ***Zeitschrift für Mathematische Logik und Grundlagen der Mathematik***, vol. 3 (1957), pp. 157–170; English translation, ***American Mathematical Society Translations***, ser. 2, vol. 23 (1963), pp. 89–101.

[Us Pl 85] V. A. USPENSKY and V. E. PLISKO, *Intuitionistic logic*, in [ANK 85], pp. 394–404 (Russian), 452–466 (English).

[Us Pl 91] ———, *Diagnostic propositional formulas*, ***Vestnik Moskovskogo Universiteta, Seriya I: Matematika, Mekhanika***, vol. 1991, no. 3, pp. 7–12; English translation, ***Moscow University Mathematics Bulletin***, vol. 46 (1991), no. 3 (to appear).

[Us Se 81] V. A. USPENSKY and A. L. SEMENOV, *What are the gains of the theory of algorithms: basic developments connected with the concept of algorithm and with its application in mathematics*, ***Algorithms in modern mathematics and computer science (Urgench, 1979)***, Lecture Notes in Computer Science, vol. 122, Springer-Verlag, Berlin, 1981, pp. 100–234.

[Us Se 87] ———, ***Theory of algorithms: main discoveries and applications***, "Nauka", Moscow, 1987. (Russian)

[Us Se 87a] ———, *Algorithms, or machine, of Kolmogorov*, in [ANK 87], Russian pp. 279–289.

[V'yu 81] V. V. V'YUGIN, *Algorithmic entropy (complexity) of finite objects and its application to defining randomness and amount of information*, ***Semiotika i Informatika***, vol. 16, VINITI, Moscow, 1981, pp. 14–43. (Russian)

[Zv Le 70] A. K. ZVONKIN and L. A. LEVIN, *Complexity of finite objects and the algorithm-theoretic foundations of the notions of information and randomness*, ***Uspekhi Matematicheskikh Nauk***, vol. 25 (1970), no. 6 (156), pp. 85–127; English translation, ***Russian Mathematical Surveys***, vol. 25 (1970), no. 6, pp. 83–124.

DEPARTMENT OF MATHEMATICAL LOGIC
FACULTY OF MECHANICS AND MATHEMATICS
MOSCOW UNIVERSITY
119899 MOSCOW V-234, USSR