On the history of martingales in the study of randomness

Laurent Bienvenu, Glenn Shafer, and Alexander Shen[1]

Abstract

Martingales played an important role in the study of randomness in the twentieth century. Jean Ville invented martingales in the 1930s in order to improve Richard von Mises' concept of a collective, and Claus-Peter Schnorr made martingales algorithmic in the 1970s in order to advance the study of algorithmic randomness.

Keywords: martingale, collective, complexity, randomness, semimeasure

1 Introduction

Jean Ville introduced martingales into mathematical probability in the 1930s in order to improve Richard von Mises' concept of a random sequence, or collective. When the study of random sequences was revived by Andrei Kolmogorov and others in the 1960s, martingales again found their place.

In its broadest outlines, the story we tell here is about the different approaches to the definition of an individual random sequence. Richard von Mises proposed to define this notion in terms of limiting frequency and selection rules. Then Ville showed that martingales (capital processes for gambling strategies) do the job more completely than selection rules (with respect to classical probability theory).

The contributions of von Mises, Ville, and Abraham Wald in the 1930s (and the notion of an individual random object in general) were neglected in the 1940s and 1950s, because measure proved a more expeditious way of modernizing classical probability. Probability theory's predictions are events to which it gives measure (probability) near or equal to one, and whose failure therefore has measure near or equal to zero – and why say more? Misbehaving frequencies and unbounded martingales are merely examples of sets of measure zero.

[1]Laurent Bienvenu (Laurent.Bienvenu@lif.univ-mrs.fr) is a von Humboldt Fellow at the Ruprecht-Karls-Universität Heidelberg. Glenn Shafer (gshafer@rutgers.edu) is a professor in the Rutgers Business School and in the Department of Computer Science, Royal Holloway, University of London. Alexander Shen (alexander.shen@lif.univ-mrs.fr) is a researcher at LIF (Laboratoire d'Informatique Fondamentale, CNRS & University Aix-Marseille; supported in part by ANR grant NAFIT-08-EMER-008-01).

In the 1960s, tools from the theory of computation permitted the revival of the study of randomness. Kolmogorov (and later Gregory Chaitin) proposed to define random objects as objects of maximal complexity. Per Martin-Löf showed that the notion of measure zero can also be made algorithmic. His work on algorithmic measure zero inspired Schnorr's work on algorithmic martingales. The relations between the definitions of randomness that use complexity, effective measure and martingales were established in the 1970s by Schnorr, Levin and others. These results now form the basis of algorithmic randomness theory.

We begin the article by reviewing the contributions of von Mises, Wald, and Ville. Von Mises first introduced collectives in 1919. In Section 2, we recall the concept and von Mises' motivation for introducing it. In Section 3, we review how Wald, writing in the 1930s, clarified the concept and demonstrated its consistency. In Section 4, we review how Ville, in the thesis and book he published in 1939, defined a stronger concept based on martingales.

After pausing, in Section 5, to consider how collectives fell out of fashion in the 1950s, we describe developments in the 1960s and 1970s. In Section 6, we review the invention of the concept of algorithmic complexity and describe work by Ray Solomonoff, Kolmogorov, and Chaitin in the 1960s. In Section 7, we explain how Martin-Löf came to define randomness for infinite sequences in the mid 1960s. In Section 8, we review Schnorr's introduction of algorithmic martingales around 1970. In Section 9, we discuss semimeasures, introduced by Levin in a paper with Zvonkin in 1970, and their relation to martingales. Finally, in Section 10, we discuss how Schnorr and Levin related randomness to complexity using monotone complexity (discovered by Schnorr and Levin) and prefix complexity (introduced by Levin and rediscovered and made popular by Chaitin).

In a brief epilogue, Section 11, we say a few words about subsequent developments in algorithmic complexity and martingales, particularly those related to von Mises' original project of providing a foundation for probability and its applications.

In addition to published sources, we have drawn on interviews with Peter Gács, Leonid Levin, and Per Martin-Löf, and on discussions at a meeting at Dagstuhl in late January and early February 2006. Marcus Hutter recorded historical talks at Dagstuhl by Christian Calude, Claus-Peter Schnorr, and Paul Vitányi and posted them at `http://www.hutter1.net/dagstuhl`. We have also profited from discussions with Leonid Bassalygo, Vladimir Uspensky, Vladimir Vovk, Vladimir Vyugin, and others. In an appendix, we reproduce several documents on which we have drawn: a letter from Andrei Kolmogorov to Maurice Fréchet, abstracts of talks by Kolmogorov at the Moscow Mathematical Society, and letters from Levin to Kolmogorov.

A preliminary version of this article [4], by Laurent Bienvenu and Alexander Shen, contains additional information about the history of algorithmic information theory; see also [75].

2 Richard von Mises' collectives

In a celebrated article published in 1919 [53], Richard von Mises (1883–1953) raised a question that was widely discussed during the following twenty years: how can we give a mathematical account of the notion of an individual random sequence?

The problem of deciding whether a particular sequence is random was hardly novel

in 1919. Should we disbelieve the fairness of a lottery if we repeatedly observe that the winning numbers are always even? If we see letters on a table arranged to spell ROMA or CONSTANTINOPOLITANENSIBUS, can we rule out the arrangement having happened by chance? Aren't these orderings as likely as any others? Such questions were debated by d'Alembert, Condorcet, and Laplace in the eighteenth century [15], and they were taken up in nearly every treatise on probability in the nineteenth century.

But von Mises, who was a philosopher as well as an applied mathematician [76], had a new idea. He believed that random sequences should be considered the subject matter of mathematical probability, just as lines and planes are the subject matter of geometry. The axioms of probability theory should therefore be statements about the properties of random sequences, or collectives (Kollektiv in German). In order to construct a simple mathematical theory, we should take these random sequences to be infinite, von Mises thought, just as Euclid and Hilbert took lines and planes to be infinite.

Von Mises formulated two axioms for collectives. For simplicity, we state them for a collective with two values, e.g., a sequence of heads and tails obtained by coin tossing:

I. There exists a limiting frequency: if s_N is the number of heads among the first N coin tosses, the ratio s_N/N converges to some real number p as N tends to infinity.

II. This limiting frequency is stable: if we select a subsequence according to some selection rule, then the resulting subsequence (if infinite) has the same limiting frequency.

A selection rule is a mathematical rule that decides whether a term is selected or not using only the values of the preceding terms but not the value of the term in question. For example, a selection rule may select terms whose numbers are prime, or terms that immediately follow heads in the sequence, but not the terms that are heads themselves.

Axiom I made sense to mathematicians as a result of work by Émile Borel in 1909 [5]. Borel had shown that convergence of s_N/N to a limit can be expected with probability one in the case of independent trials. This limit is, of course, the probability for heads. Axiom II is also persuasive. Suppose someone tells you that flipping a coin produced the sequence

$$10101010101010101010101010101\ldots$$

where 1 (heads) and 0 (tails) alternate. Would you believe this? Probably not. The limiting frequency of 1s in this sequence exists and is equal to 1/2. But the sequence is too regular. This is where axiom II comes in: if one selects from this sequence the bits in even positions, one gets the sequence

$$11111111111111111111111111111\ldots$$

in which the frequency of 1s is different (1 instead of 1/2). We can win for sure by betting only on the trials in this subsequence. By ruling this out, von Mises explained, the second axiom expresses a classical principle, the principle that excludes systems for beating the odds.

According to von Mises, probability theory is about the properties of collectives and about operations that transform collectives into other collectives. He used the following example: take a collective (a sequence of 1s and 0s) and cut it into 3-bit groups. Then replace each group by an individual bit according to majority rule. Probability theory has to find the limiting frequency of the resulting sequence if the limiting frequency of the original one is known.

When he launched the concept of a collective, von Mises was already prominent because of his work in mechanics; he was director of an institute of applied mathematics at Berlin starting in 1920. But he devoted a good deal of his subsequent career to promoting collectives. His book on the topic, *Wahrscheinlichkeit, Statistik und Wahrheit*, first appeared in 1928 [54] and subsequently saw multiple editions in German and in English. In 1931 he published a textbook on probability and statistics based on collectives [55]. After fleeing from Berlin to Turkey in 1933, he emigrated to the United States in 1939, where he became a professor at Harvard. His later publications on probability include a debate with the United States mathematician Joseph Doob in September 1940 at a meeting of the Institute of Mathematical Statistics, published in 1941 [57, 58], and a posthumous book edited by his widow, *Mathematical Theory of Probability and Statistics* [59].

Von Mises realized that he had not demonstrated the logical consistency of his axioms or the existence of sequences satisfying them. But he managed to gain sufficient attention for his ideas that others undertook these tasks. Among them were the United States mathematician Arthur Copeland (1898–1970) and the German philosophers Hans Reichenbach (1891–1953) and Karl Popper (1902–1994). Reichenbach was a colleague of von Mises in Berlin and also emigrated to Turkey and then to the United States. Popper was Viennese; he emigrated to New Zealand in 1937 and then to England in 1949. The three authors, Copeland in 1928 [16], Reichenbach in 1932 [65], and Popper in 1935 [64], made suggestions that turned to out to be equivalent to each other and closely related to the concept of a normal number, already developed in Borel's 1909 article. Their suggestions boiled down to requiring von Mises' axiom II only for selection rules that select just the trials for which the r preceding trials, for a specified r, match a specified string of 1s and 0s of length r. It is easy to give an algorithm for constructing sequences whose limiting frequencies are not affected by such selections, and for this very reason, von Mises did not consider this solution satisfactory. In von Mises' eyes, a sequence that can be predicted could not be considered random.

For fuller reviews of the work stimulated by von Mises during the 1920s and 1930s, see Martin-Löf [48] and Chapter 6 of von Plato [63]. These authors discuss in particular the work of Erhard Tornier, who proposed replacing von Mises' single random sequence with an infinite matrix consisting of many sequences that might result from an infinite sequence of trials. William Feller collaborated with Tornier.

3 Abraham Wald's clarification

It was Abraham Wald, the star of Karl Menger's mathematical seminar in Vienna, who reformulated the second axiom in a way that satisfied von Mises.

Karl Menger (1902–1985), son of the Austrian economist Carl Menger, was loosely associated with Moritz Schlick's seminar on the philosophy of science, which became known in retrospect as the Vienna circle. After earning his doctorate in mathematics in 1924, Menger worked for two years with L. E. J. Brouwer in Amsterdam before returning to Vienna, where he eventually obtained a post in the university and organized his own seminar on mathematics. For eight years, from 1928–29 through 1935–36, the seminar's proceedings were published as a journal, the *Ergebnisse eines Mathematischen Kolloquiums*.[2]

[2] In 1998, Springer reprinted the proceedings, along with several introductory articles in English, in a

Prominent contributors to the seminar included Nachman Aronszajn, Kurt Gödel, Marston Morse, John von Neumann, Albert Tarski, and Norbert Wiener. The most important contributor was Menger's most brilliant student, Abraham Wald (1902–1950). Wald was the same age as Menger but was a latecomer to the university. He had been born in Transylvania, where his father was an orthodox Jewish baker, and the family had come to Vienna after the Romanian annexation of the region during World War I. Unable to study at the Vienna gymnasium, he passed the examination for entrance to the university after attending an engineering school and being tutored by his brother in mathematics [60, 90]. But by the time he completed his doctorate in 1931, he was contributing to almost every topic in the seminar. Because his religion barred him from a university post of his own, he continued to contribute to the seminar while earning a living in Oskar Morgenstern's economics institute, until the worsening political situation forced Menger to end the seminar in 1936.

Collectives came into Menger's seminar by way of Schlick's, where Menger had heard Karl Popper present his ideas on collectives. Menger asked Popper to come to his own seminar to give a more precise mathematical explanation [52]. Popper did so on February 6, 1935, and Wald immediately responded with a proposal of his own.

A selection rule, in the case of a sequence of 1s and 0s, can be thought of as a function s from $\{0,1\}^*$ to $\{0,1\}$, where $\{0,1\}^*$ is the set of all finite strings of 1s and 0s. Applying s to an infinite sequence $\omega_1\omega_2\ldots$ means that we select all terms ω_i such that $s(\omega_1\omega_2\ldots\omega_{i-1}) = 1$; the selected terms are listed in the same order as in the initial sequence. For those who accept a nonconstructive concept of mathematical existence, there obviously exist s that change the limiting frequency of 1s in a particular $\omega_1\omega_2\ldots$. There are many s, for example, that select just the terms for which $\omega_i = 0$, thus changing the limiting frequency to 0, and others that select just the terms for which $\omega_i = 1$, thus changing the limiting frequency to 1. But Menger and his seminar were attuned to Brouwer's constructivism and to the developments of the day in logic, and so when Wald thought about functions from $\{0,1\}^*$ to $\{0,1\}$, he did not necessarily consider all functions that exist in a nonconstructive sense. He thought it might be appropriate instead to consider only functions that can be constructed in some particular system of arithmetic. There will not be so many of these functions, because a logical system can have only a countable number of symbols, and from these we can construct only a countable number of formulas. The question, therefore, is whether there exist sequences that have an invariant limiting frequency with respect to the countable set of selection rules that can be constructed in a given system, and if so, in what sense these sequences can themselves be constructed.

Wald's first publication on the topic was a note in French in the *Comptes rendus* in Paris [84], submitted by Émile Borel for the session of January 20, 1936. In this note, Wald asserts without proof the existence of collectives when the number of selection rules is countable. Like von Mises, he considers more than the binary case, but with respect to the binary case, his assertion says that for any countable family of selection rules and for any $p \in (0,1)$ there exist a continuum of sequences that satisfy axioms I and II with limiting frequency p.

This result by itself (the *Comptes rendus* note says nothing about constructivity) was

single volume [19].

hardly surprising even at the time. Classical probability theory, modernized by Borel in 1909 [5], Maurice Fréchet in 1915 [23], and Andrei Kolmogorov in 1933 [29], had taught mathematicians that the disjunction of a countable number of events of probability zero itself has probability zero. It is obvious (and was proven rigorously by Doob in 1936 [20]) that if you apply a selection rule s to a sequence of independent random variables $\omega_1\omega_2\ldots$, each equal to 1 with probability p and to 0 with probability $1-p$, then you obtain a subsequence that has the same distribution (let us call it the Bernoulli distribution with parameter p). Borel had shown that the probability is zero that the frequency of 1s in a realization of the Bernoulli distribution with parameter p fails to converge to p. So the probability is also zero that any of the countable number of subsequences obtained from a countable number of selection rules fails to converge to p. The complement of this event has probability one and therefore has the cardinality of the continuum.

In a footnote, Wald says that he will give proofs in the seventh volume of Menger's *Ergebnisse*, the volume for 1934–35. But the article containing these proofs appeared instead in the eighth and final volume, for 1935–36, which did not appear in print until 1937 [85]. Written in German, this article also gives the explanation, missing from the *Comptes rendus* note, that an insistence on constructivity justifies the consideration of countable systems of selection rules. The article uses what we now consider an informal concept of constructivity: a sequence $a_1a_2\ldots$ was considered to be constructively (or effectively) defined if for each a_i there is a procedure for determining the value of a_i in a finite number of steps, but it was not clear what was meant by a procedure, whether the procedure might depend on i, etc. Wald shows that for a countable system of constructively defined selection rules, there exists a constructively defined collective (Theorem V, p. 49). He does this by constructing the collective recursively from the selection rules. In modern terminology, he uses the selection rules as oracles.

Let us explain Wald's recursion in the simple case where we consider only a finite system of selection rules, say a set S consisting of n selection rules, and we want to construct a collective ω consisting of 1s and 0s with limiting frequency 1/2. Suppose we have constructed $\omega_1\ldots\omega_{i-1}$ and now want to specify ω_i. Let S_i be the subset of S consisting of the rules in S that will include the ith entry of ω in the subsequence they select when applied to ω:

$$S_i = \{s \in S \mid s(\omega_1\ldots\omega_{i-1}) = 1\}.$$

Because we have already constructed $\omega_1\ldots\omega_{i-1}$, we have determined S_i and also the preceding S_j (those for $1 \le j < i$). Let k_i be the number of the preceding S_j that are equal to S_i, and set

$$\omega_i = \begin{cases} 1 & \text{if } k_i \text{ is even,} \\ 0 & \text{if } k_i \text{ is odd.} \end{cases}$$

(In particular, $\omega_1 = 1$, because $k_1 = 0$; there are no j satisfying $1 \le j < 1$.) If we fix a subset A of S and select from ω the subsequence consisting of the ω_i for which $S_i = A$, we get the alternating sequence $101010\ldots$. By considering the 2^n different subsets A of S, we partition ω into 2^n subsequences, all equal to $101010\ldots$. Each of these has limiting frequency 1/2, and so ω does as well. If we apply a selection rule $s \in S$ to ω, we pick up the entries in half these 2^n subsequences, those corresponding to the subsets of S that contain s, and the limiting frequency will still be 1/2.

The construction for countably many selection rules builds on this simple picture: we add new rules one by one at intervals so great that the boundary effects cannot affect the limiting frequency.

Wald considers not only collectives with entries drawn from $\{0, 1\}$, but also collectives drawn from any finite set (Theorem I, p. 45). He also considers collectives with entries drawn from an infinite set M (Theorems II–IV, pp. 45–47). He finds, as Copeland had found using his more restrictive concept of a selection rule, that the theory works if M is countable or if one considers only a restricted class of events, e.g., those that are Peano-Jordan measurable. Wald's Theorems V–VI (p. 49) observe that the resulting collectives can be effectively constructed.

In October 1937, Wald presented his results in a celebrated colloquium on probability at Geneva. This colloquium, chaired by Maurice Fréchet, brought together most of the world's leading probabilists for the last time before the war. In addition to Wald and Fréchet, attendees included Harald Cramér, Wolfgang Doeblin, William Feller, Bruno de Finetti, Werner Heisenberg, Eberhard Hopf, Bohuslav Hostinsky, Paul Lévy, Jerzy Neyman, George Polya, and Hugo Steinhaus. The session on foundations was remembered for its lively discussion of collectives, which were criticized by Feller, Fréchet, Lévy, and others. The second installment of the proceedings, published in 1938, included articles on collectives by Wald [86] and von Mises [56]. Wald, still writing in German, stated the theorems he had proven in his *Ergebnisse* article and refuted some of the criticisms. Von Mises, who had not been at the colloquium but wrote in French, embraced Wald's analysis fully, seeing it as a vindication of his confidence that his axioms were logically consistent.

Wald and von Mises both took a practical tone. They considered probability an applied field. A mathematical theory of probability can involve idealization, such as the consideration of infinite sequences instead of long finite ones, but the test of its adequacy should be whether it covers important applications. In any particular application only finitely many selection rules can be relevant. Wald pointed to a logical system of arithmetic permitting the formulation of only countably many selection rules not because he imagined using so many, but to make the point that no one could conceivably need a collective to do more.

Wald's introduction of constructivity into the discussion of collectives coincided with a debate among logicians concerning how this notion should be made precise. The debate was motivated by David Hilbert's question of whether there exists a procedure for separating mathematical truths from falsehoods, and it was eventually settled by a consensus around *Church's thesis*, the thesis that effective calculability should be identified with a precise concept that had been given different but equivalent definitions by Alonzo Church and his students, Gödel, and Alan Turing (see, e.g., [17]). In 1940 [14], Church suggested using this new precise concept of effective calculability, now usually called simply *computability*, to define collectives. Under Church's definition, a sequence of 1s and 0s is a collective with probability p if the limiting frequency of 1s is p in the sequence and in any subsequence selected by a computable selection rule. With this definition, as Church explained, the existence of collectives can be proven nonconstructively, following Wald or using Doob's result. But a collective cannot be constructed, because the set of all computable selection rules, while countable, is not effectively enumerable. (It is a subset of a set that can be effectively enumerated, but it cannot be effectively enumerated itself.)

4 Jean Ville's martingales

Jean André Ville (1910-1989) was a participant in Menger's seminar when Karl Popper and Abraham Wald gave their talks on collectives in February 1935. The most brilliant of the first students to earn the degree in probability that Fréchet introduced at the University of Paris in 1931, Ville had been awarded scholarships to study in Berlin in 1933–34 and in Vienna in 1934–35. Fréchet had sent Ville to Berlin to get started on a doctoral thesis in analysis, but Ville was more interested by the new mathematics and new applications he encountered in Menger's seminar, and he was particularly fascinated by collectives.

As a student in France, Ville had learned a way of thinking about the application of probability theory that was quite different from that of von Mises. According to Cournot's principle,[3] which was popular among French probabilists when Ville was a student, probability theory makes contact with the empirical world only by making predictions with probability near or equal to one. The law of large numbers is one such prediction: the frequency of 1s in a sequence of tosses of a fair coin will converge to 1/2. The law of the iterated logarithm is another: the frequency will oscillate around 1/2, converging at a certain specified rate. From this point of view, von Mises was too exclusively focused on the convergence of frequencies. What about the other predictions probability theory makes with probability one? Will collectives in von Mises' sense also satisfy them? Not necessarily, Ville concluded. There are some probability-one predictions that we cannot guarantee a collective to have through our choice of the system of selection rules. Or to put the point positively, there are properties with measure zero that will be possessed by some collective no matter what system of selection rules we adopt.

Ville first made his point in the *Comptes rendus* in July 1936 [81], in a concise note without examples or proofs that considered only the familiar case of collectives consisting of 1s and 0s with limiting frequency 1/2. Under the Bernoulli measure, the sequences that are not collectives with respect to a given countable system of selection rules have measure zero. But, Ville asserted, not every property of measure zero can be ruled out in this way. He further asserted that this shortcoming of collectives can be corrected by replacing the system of selection rules by a martingale, i.e., a betting strategy.[4] Ville considered strategies satisfying the following conditions:

> *Ville's conditions.* (1) You start with unit capital. (2) At every trial, you bet only a fraction α of your current capital, where $0 \leq \alpha \leq 1$, on 1 or on 0, so that your capital will remain nonnegative no matter how the trial comes out.

It is easy to show that the resulting capital will remain bounded with probability one. So there exist a continuum of sequences for which it remains bounded; we may call these collectives with respect to the betting strategy. Ville asserted without proof that for any property E to which the Bernoulli measure assigns measure zero, there exists a strategy

[3] For a history of Cournot's principle and examples of statements embracing it by Jacques Hadamard, Paul Lévy, Maurice Fréchet, and Émile Borel, see [72]. The principle was named after Cournot by Fréchet around 1950.

[4] For centuries the word *martingale* has referred to the strategy for betting that doubles one's bet after every loss. See Roger Mansuy's article in this issue of the *Electronic Journal for History of Probability and Statistics*.

satisfying his conditions for which the capital is unbounded if E happens. Thus we can rule out any property of measure zero for our collectives by properly choosing the strategy.

For those not steeped in the philosophy of the French probabilists, or for whom probability could only mean frequency, Ville's results may not have seemed well motivated. William Feller, in a two-sentence review in *Zentralblatt* (Zbl 0014.16802), summarized what Ville claimed to have proven while making it clear that he could not figure out why Ville should want to prove it.

Ville's ideas received a fuller hearing the following year, when Fréchet presented them to the colloquium at Geneva as part of a wide-ranging argument against collectives and in favor of the axiomatic approach perfected by Andrei Kolmogorov. In Fréchet's contribution to the colloquium's proceedings [24], published in 1938, we see for the first time in print an example of a property of measure zero that cannot be ruled out by a system of selection rules. Probability theory tells us that the frequency of 1s should oscillate above and below 1/2 as it converges to 1/2. But a collective need not have this property. Its frequency can instead approach the limit from above, for example. It is instructive to point out (although Fréchet did not) that this happens in the construction by Wald that we reviewed in Section 3. The sequence constructed there is the result of interleaving many copies of the sequence $101010\ldots$, and because the frequency of 1s in any prefix (finite initial segment) of each copy is always greater than or equal to 1/2, this must also be true for the whole sequence. This shows that no matter what countable system of selection rules we adopt, there will be a collective in which the frequency converges to 1/2 from above. We cannot force the frequency to oscillate above and below 1/2 as required by the law of the iterated logarithm by a clever choice of the selection rules.

Wald stood his ground. At Geneva, he protested that those who had criticized the theory of collectives for excluding some sequences were now criticizing it because it did not exclude enough sequences ([24], p. 35). In his contribution to the proceedings [86], he questioned whether every asymptotic property should be accorded the same significance as the convergence of frequencies.[5] Then, conceding that strengthening the concept of a collective so as to guarantee other asymptotic properties is of some interest, he proposed a way to do this while preserving von Mises' emphasis on frequencies. Call a selection rule *singular* if the sequences of 1s and 0s for which it produces infinite subsequences have total measure zero, he proposed, and call a collective ω with respect to a countable system S of selection rules *strong* if no singular selection rule in S produces an infinite subsequence when applied to ω. There exists a continuum of strong collectives for any countable system of selection rules.[6] For every property A of probability zero, there is a singular selection rule that produces infinite subsequences when applied to sequences in A; so by adding this selection rule to S, we can guarantee that every strong collective avoids the property A. And we can do this for countably many A.

Fréchet expressed his admiration for Wald's ingenuity but objected that the new concepts weakened the simplicity that made von Mises' picture attractive. We might add that they threaten to push frequencies out of the picture. Why not make all the

[5] An asymptotic property of $\omega_1\omega_2\ldots$ is one that does not depend on any finite prefix. In 1933 [29], Kolmogorov had shown that the probability of an asymptotic property is either zero or one.

[6] In the case of the singular rules, the sequence must be outside the set of probability zero on which the rule produces an infinite subsequence; in the case of the nonsingular rules, it must be outside the set of probability zero on which the rule produces a subsequence that does not converge to 1/2.

selection rules singular, and why not combine all the asymptotic properties we want, including the frequency properties, into one property A, to be enforced by means of just one singular selection rule? It takes only one more step to get us to Ville's picture: Define the singular selection rule using a strategy whose capital process is unbounded on A. For example, include the next bit ω_i every time the capital hits a new high. To the best of our knowledge, neither Wald nor anyone else ever promoted the concept of a strong collective further.[7] Wald was simply marshalling every argument he could think of against Fréchet's equally broad offensive.

In March 1938, Hitler annexed Austria. Wald fled from Vienna to Transylvania and then immigrated to the United States in the summer of 1938; most of his family perished in the Holocaust. One of his first publications in the United States was a very positive review, in 1939 [87], of von Mises' *Probability, Statistics, and Truth*, the English version of the second edition of *Wahrscheinlichkeit, Statistik, und Wahrheit*. The review only obliquely touched on his own contribution and made no reference to Ville's. Wald's initial employment in the United States was with the Cowles Commission, which had already offered him a position in 1937, but he quickly moved to Columbia University. In 1946, he became head of a newly created Department of Mathematical Statistics at Columbia. By the time of his death in 1950, in an airplane accident in India, he was widely regarded as the leading voice in mathematical statistics in the world. In an obituary ([90], p. 13), his colleague Jacob Wolfowitz ascribed to him "an unusual aversion to all forms of controversy".

Von Mises, like Wald, was unconvinced by Fréchet's arguments. He accepted Ville's theorem that there exist asymptotic properties that have probability zero under the theory of denumerable probabilities (this was Borel's name for the extension of classical probability theory to infinite sequences of trials) and that are satisfied by some collectives, no matter what system of selection rules is adopted. But he saw no problem with this – no reason to modify the theory of collectives to avoid it ([56], p. 66).

As for Ville's proposal to substitute a martingale for a system of selection rules, it is not clear that anyone understood Fréchet's explanation of it. Von Mises admitted that he did not understand Ville's theory ([56], p. 66). Wald had not mentioned Ville's work in his response to Fréchet, and he seems never to have mentioned it subsequently. De Finetti, in his summary of the colloquium, incorrectly stated Ville's definition of a collective relative to a martingale ([18], p. 22). Decades later, in 1964, Lévy wrote to Fréchet that he had never quite understood Ville's definition of a martingale, and that Michel Loève and Aleksandr Khinchin had told him that they had never understood it either ([3], p. 292).

Ville might have been better served to speak for himself. But the work on martingales was his thesis. French practice did not permit him to publish his proofs until the thesis was accepted, and this was delayed by Fréchet's insistence that he add enough analysis to make it respectable. He did this during the academic year 1937–38, using the concept of a martingale to prove new results for stochastic processes in discrete time and trying to extend these results to continuous time in the framework being developed by Doob. Borel and Fréchet finally allowed Ville to defend his thesis only in March 1939, on the eve of World War II. Borel then published it in his series of monographs on probability

[7] However, we may retrospectively note that when we consider *computable* singular selection rules, we get exactly the class of Martin-Löf random sequences, see Section 7 below.

[82]. This was a prestigious publication, at least in France, and the book was widely distributed, though apparently not widely read.

As Fréchet had explained at Geneva, but too cryptically, Ville found it convenient to work not with the strategies he initially called martingales but with the capital processes they determine. A strategy tells us how to bet on ω_n after seeing $x = \omega_1 \dots \omega_{n-1}$. In the usual case of 1s and 0s, this means that it tells us, as a function of x, whether to bet on $\omega_n = 1$ or $\omega_n = 0$ and how much to bet. The strategy together with the initial capital determines, for every finite string x of 1s and 0s, the capital we will have after observing x, say $m(x)$. The condition that the bets be at even odds dictates that

$$m(x) = \frac{m(x0) + m(x1)}{2}. \tag{1}$$

Any function m satisfying (1) for every finite string x is a capital process arising from a strategy and from some initial capital, and uniquely determines that strategy and initial capital. Because of this one-to-one correspondence, and because capital processes play the most direct role in his theory, Ville transferred the name martingale from the strategies to the capital processes. He called any function on finite strings satisfying (1) a martingale.

Ville was particularly interested in nonnegative martingales – martingales that start with a positive initial capital, say $m(\Box) = 1$, where $\Box$ is the empty string, and satisfy $m(x) \geq 0$ for every finite string x. These conditions are equivalent to what we called Ville's conditions above; your capital remains nonnegative for sure if and only if you never bet more than you have.

Each of the selection rules considered by Wald and von Mises excluded a property of measure zero. Wald considered countably many selection rules, and the union of a countable number of sets of measure zero still has measure zero. So Wald could exclude certain properties of measure zero. Ville could do better: he could exclude any property of measure zero, and he could do it with a single nonnegative martingale; he did not need a countable system of them. To see Ville's picture clearly, we need to understand two points:

1. If $m_1, m_2, \dots$ are nonnegative martingales starting with 1, then the weighted sum $\sum_i \alpha_i m_i$, where the α_i are positive real numbers adding to 1, is also a nonnegative martingale starting with 1. It is obtained by dividing our initial capital among the strategies that produce the m_i: we assign initial capital α_i to the strategy that makes, at each trial, α_i times the bet made by the strategy that produces m_i when you start with 1. The sum $\sum_i \alpha_i m_i$ is unbounded if and only if one of the m_i is unbounded; it therefore excludes the union of the sets of measure zero excluded inidividually by the m_i.

2. If a nonnegative martingale m is unbounded on an event E, then there is another martingale that tends to infinity on E. This is because we can stop the strategy at an arbitrarily large value for m, and by taking a weighted sum of stopped versions of m for increasingly large values ($1/\alpha_i$, for example), we obtain a martingale that tends to infinity on E. So instead of saying that a sequence is a collective with respect to a nonnegative martingale m if m is bounded on the sequence, we can say it is a collective with respect to m if m does not tend to infinity on the sequence.

Ville's claim that for any event E of measure zero there exists a nonnegative martingale that is unbounded on E is not difficult to prove. One begins with the observation that for every set E of sequences of 1s and 0s of length N that contains a fraction ϵ or less of such sequences, there is a nonnegative martingale that multiplies its initial capital by $1/\epsilon$ on E.

One of von Mises' arguments for his second axiom was that it prevents a gambler from making money by selecting trials on which to bet. Ville argued that this "principle of the excluded gambling system" should apply equally to a strategy that can vary the amount bet, and so his martingale theory of collectives is a natural strengthening of von Mises' and Wald's theory. But whereas Ville's 1936 note in the *Comptes rendus* had positioned his theory as a new and better theory of collectives, his thesis and book were positioned, as their title said, as a critique of collectives. He probably had no choice; he had to accept the view of his mentors that probability should be seen as an application of functional analysis and measure theory. To the extent that it is independently axiomatized, it should start with an axiomatic system like Kolmogorov's or like Borel's, which differed from Kolmogorov's only in that conditional probability was taken as primitive and related to unconditional probability by the axiom $P(A\&B) = P(A)P(B|A)$ ([82], p. 10).

In order to make the thesis a book, Ville added two philosophical chapters, one at the beginning and one at the end. But the mathematical exposition in the middle remained a thesis rather than a more mature exposition. A whole chapter is devoted to an elaborate notation for working with sequences of 1s and 0s, and another is devoted to Popper and Reichenbach. The simple explanation we have given concerning how to construct a collective that approaches 1/2 from above is obscured by the apparatus, to the extent that some recent readers have resorted to working out their own constructions [42].

The thesis and book were reviewed in half a dozen mathematical journals. Two of the reviews, de Finetti's review of the thesis in *Zentralblatt* (Zbl 0021.14505) and Doob's review of the book in the *Bulletin of the American Mathematical Society* (45(11):824, 1939), mentioned how martingales could replace systems of selection rules in the definition of collectives. The others gave the impression that Ville was merely reviewing the literature on collectives.

It was only through Doob that Ville's work on martingales contributed to mathematical probability in the second half of the twentieth century. Giving Ville full credit for inventing the concept of a martingale, Doob developed the study of martingales within measure-theoretic probability, where they have become increasingly central. (See Paul-André Meyer's article on the history of stochastic processes from 1950 through the 1980s in this issue of the *Electronic Journal for History of Probability and Statistics*.)

5 The status quo of the 1950s

The 1937 colloquium at Geneva is sometimes seen as a watershed. A substantial mathematical literature had been devoted to collectives during the 1920s and 1930s, but the Geneva colloquium showed that most probabilists favored working in the measure-theoretic framework of Kolmogorov's axioms. By the 1950s, almost all mathematical work in probability and statistics was in Kolmogorov's framework, and little mathematical attention was being paid to collectives. For most working mathematicians, there was no need to justify the notion of a probability measure by means of an additional layer: it

was much simpler to consider the measure as a primary object, not something generated by an underlying collective.

Von Mises' collectives did remain a topic of discussion among philosophers and philosophically minded mathematicians and statisticians, at least in the West.[8] Most people, including most philosophers and mathematicians, intuitively identified probability with frequency, and the theory of collectives was the simplest way to make that identification into a theory. The notion of irregularity embodied in von Mises' second axiom was sometimes influential, moreover, even when collectives were not mentioned; see for example R. A. Fisher's comments about relevant subsets in his 1956 book on statistical inference ([22], pp. 34–35).

Even among philosophers, however, Ville's concept of a collective based on martingales seems to have completely disappeared by the 1950s. Church's 1940 article [14], often regarded as the last word on collectives, had made no mention of Ville's work. The French logician and philosopher Jean Cavaillès wrote about Ville's ideas in 1940 [7], but his example was not followed by philosophers writing in English. (Cavaillès became a leader in the resistance to the German occupation of France and was shot by the Gestapo in 1944.)

Whereas von Mises energetically promoted his theory for decades, Ville, as we have seen, was already diffident about collectives based on martingales in his 1939 thesis, and he then went on to other things. Mobilized in the fall of 1939 along with all other French reservists, Ville spent a year at the front and then a year in a German prison camp before returning to France in June 1941. During the remainder of the war, he worked mainly on statistical problems, returning to martingales only briefly, when he tried to use them to study Brownian motion but then realized that the results he was obtaining had already been found by different methods by Lévy. After the war, Ville worked on Shannon information, signal theory, and mathematical economics.

The degree to which Ville's collectives based on martingales had been forgotten in the 1950s can be measured by the ill informed praise for his thesis when he was appointed to a chair in economic mathematics in the Paris Faculty of Sciences in 1959. His fellow mathematician Luc Gauthier, in the report on Ville's work that justified the vote to appoint him, recalled that the thesis had earned the highest praise from Fréchet and Borel. The foundations of measure theory were far from clarified at the time, Gauthier added, and Ville's thesis had strongly contributed to their being put in order.[9]

6 The invention of the algorithmic definition of randomness in the 1960s

The study of random sequences revived in the 1960s, when it became clear that new ideas from mathematical logic and programming could be used to characterize the complexity of sequences. The complexity of a sequence or other finite object can be defined as the length of the shortest program that generates it (this is *description* complexity, as opposed to *computation* complexity, since we ignore the time and other resources needed), and

[8] Concerning criticism of von Mises by Soviet philosophers, see Siegmund-Schultze [77].

[9] In French: "La thèse de Monsieur Jean VILLE, intitulée Étude critique de la notion de Collectif, est une étude sur les fondements du calcul des probabilités, qui a eu les plus vifs éloges de Monsieur FRECHET et de Monsieur BOREL. Il est de fait que les assises de la théorie de la mesure étaient loin d'être clarifiées à l'époque où Jean VILLE a fait sa thèse, et que cette dernière a fortement contribué à sa mise au point." (Archives Nationales, Fontainebleau, Cote 19840325, art. 542.)

the most complex objects can be considered random.

The idea of measuring the complexity of a message by the length of its shortest encoding, as well as the idea of calling the most complex messages the most random, had become familiar in the 1940s and 1950s to students of Shannon's information theory. Shannon considered only very specific encodings, but mathematical logicians found reasons for studying compressibility more abstractly. As A. A. Markov explained in 1964, one reason came from the quantitative analysis of undecidability:

> Undecidable algorithmic problems were discovered in many fields, including the theory of algorithms, mathematical logic, algebra, analysis, topology and mathematical linguistics. Their essential property is their generality: we look for an algorithm that can be applied to every object from some infinite class and always gives a correct answer. This general formulation makes the question not very practical. A practical requirement is that the algorithm work for every object from some finite, though probably very large, class. On the other hand, the algorithm itself should be practical.... An algorithm is an instruction, and it is natural to require that this instruction not be too long, since we need to invent it.... So an algorithmic problem could be unsolvable in a practical sense even if we restrict inputs to a finite set. ([44], p. 161)

The key step in defining algorithmic complexity was the realization and demonstration that there exist decompression algorithms that are universal and provide (in an asymptotic sense that we will review) shortest possible descriptions for finite objects. The shortest description of an object with respect to such a universal algorithm is the object's algorithmic complexity (or Kolmogorov complexity, as we now say). Once this definition is established, it makes sense to take the second step and say that objects with maximal complexity (i.e., longest descriptions) among the objects of some class are random in this class.

Kolmogorov took these two steps in a celebrated article published in 1965 [31]. In this section, we review what is known about how Kolmogorov came to these ideas. We also discuss two other authors who arrived independently at similar ideas at around the same time: Ray Solomonoff and Gregory Chaitin.

6.1 Kolmogorov

Milestones for the evolution of Kolmogorov's thinking about algorithmic complexity and randomness in the early 1960s are provided by the titles of talks that he gave at the Moscow Mathematical Society:

1. *Редукция данных с сохранением информации* (Data reduction that conserves information), March 22, 1961.

2. *Что такое "информация"*? (What is information?), April 4, 1961.

3. *О таблицах случайных чисел* (On tables of random numbers), October 24, 1962. This talk probably corresponds to the article Kolmogorov published in *Sankhyā* in 1963 [30].

4. *Мера сложности конечных двоичных последовательностей* (A complexity measure for finite binary strings), April 24, 1963.

5. *Вычислимые функции и основания теории информации и теории вероятностей* (Computable functions and the foundations of information theory and probability theory), November 19, 1963.

6. *Асимптотика сложности конечных отрезков бесконечной последовательности* (Asymptotic behavior of the complexities of finite prefixes of an infinite sequence), December 15, 1964. The title suggest that this talk might have discussed Martin-Löf's results, but Martin-Löf remembers discussing them with Kolmogorov only the following spring (see Section 7).

Three later talks about algorithmic complexity, given from 1968 to 1974, have short published abstracts, which are translated in Appendix B.

In his obituary for Kolmogorov written in 1988 [62], K. R. Parthasarathy recalled that Kolomogorov had traveled by sea to India in the spring of 1962 to work at the Indian Statistical Institute and receive an honorary degree from the University of Calcutta. When he arrived in Calcutta, he told the students at the institute about his work, while on the ship, "on tables of random numbers, and the measurement of randomness of a sequence of numbers using ideas borrowed from mathematical logic." This may refer to the work that Kolmogorov published in *Sankhyā* in 1963 [30]. The third talk in the list above, on October 24, 1962, would have been given after he returned to Moscow from India.

In the *Sankhyā* article, Kolmogorov does not yet adopt the idea that maximally complex sequences are random. Instead, he offers a finitary version of von Mises' picture, in which random sequences are those whose frequencies are not changed by the simplest selection rules. In the article, Kolmogorov writes as follows:

> 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. However, for a long time I had the following views:[10]
>
> (1) The frequency concept based on the notion of *limiting frequency* as the number of trials increases to infinity, does not contribute anything to substantiate the applicability of the results of probability theory to real practical problems where we have always to deal with a finite number of trials.
>
> (2) The frequency concept applied to a large but finite number of trials does not admit a rigorous formal exposition within the framework of pure mathematics.
>
> Accordingly I have sometimes put forward the frequency concept which involves the conscious use of certain not rigorously formal ideas about 'practical reliability', 'approximate stability of the frequency in a long series of trials', without the precise definition of the series which are 'sufficiently large'...
>
> I still maintain the first of the two theses mentioned above. As regards the second, however, I have come to realize that the concept of random distribution of a property in a large finite population can have a strict formal

[10] This is corroborated by a letter Kolmogorov wrote to Fréchet in 1939 (Appendix A.)

> mathematical exposition. In fact, we can show that in sufficiently large populations the distribution of the property may be such that the frequency of its occurrence will be almost the same for all sufficiently large sub-populations, when the *law of choosing these is sufficiently simple*. Such a conception in its full development requires the introduction of a measure of the complexity of the algorithm. I propose to discuss this question in another article. In the present article, however, I shall use the fact that there cannot be a very large number of simple algorithms.

Whereas von Mises considered an infinite binary sequence random if the frequency of 1s has a limit and the selection rules we consider do not change this limit, Kolmogorov now considered a finite binary sequence random if the simplest selection rules do not change the frequency of 1s very much. Whereas Wald had relied on the constructible selection rules being countable, Kolmogorov relied on simple rules being few in number. His formalization of the idea of a selection rule also differed from von Mises; for example, it allowed the decision whether to include a particular term to depend on later as well as earlier terms. He did not, however, consider anything like a martingale for testing randomness. We have no evidence that he ever took notice of Ville's work.

The article was received by *Sankhyā* in April 1963. Kolmogorov's hint that he will write another article showing how to measure the complexity of an algorithm suggests that he may have already worked out the difficulties in defining algorithmic complexity when he submitted the article. This is also suggested by the title of the talk he gave at the Moscow Mathematical Society on April 24, 1963. We can be confident, in any case, that he had the definition by the autumn of 1964, because we have Per Martin-Löf's testimony that he learned about it then from Leonid Bassalygo [51]. Bassalygo confirms this (in a private communication to Alexander Shen); he recalls a walk with Kolmogorov in the early spring or late autumn in which Kolmogorov tried to explain the definition, which he found difficult to grasp.

Bassalygo was not the only person to have difficulty understanding Kolmogorov's definition of algorithmic complexity. The problem lies in sorting out and keeping in mind the sense in which the measurement of complexity is invariant when we change from one universal algorithm to another. If we write $K_{\mathcal{A}}(x)$ for the shortest description of a finite string x by a universal algorithm $\mathcal{A}$ and $K_{\mathcal{B}}(x)$ for the shortest description by a second algorithm $\mathcal{B}$, then the universality of A implies that there exists a constant c such that

$$K_{\mathcal{A}}(x) \leq K_{\mathcal{B}}(x) + c$$

for all x, no matter how long. Because the constant c might be very large, this inequality has only an asymptotic significance: it says that A does at least nearly as well as B for very complex x, those for which $K_{\mathcal{A}}(x)$ and $K_{\mathcal{B}}(x)$ are both so large that c is negligible in comparison. If we compare $\mathcal{A}$ to yet another algorithm $\mathcal{C}$ instead of $\mathcal{B}$, the constant c may change. So when we choose $\mathcal{A}$ as our standard for measuring complexity – i.e., set $K(x)$ equal to $K_{\mathcal{A}}(x)$ and call it the algorithmic complexity of x,[11] we must keep in mind that this algorithmic complexity $K(x)$ is meaningful only up to an arbitrary constant that is independent of x. Because of this arbitrary constant, the number $K(x)$ does not have any meaning or use for a particular string x. But we can use the function K to

[11] Many authors now use $C(x)$ instead of $K(x)$.

make asymptotic statements about the complexity of strings as they are made longer and longer. These subtleties and limitations have served as a brake on interest in algorithmic complexity. Some people are confused by the definition; others find it too asymptotic for their taste.

Kolmogorov was the first to publish a precise statement of the definition of algorithmic complexity and a proof of the existence of universal algorithms. In the 1965 article in which he first did this [31], he contrasted this new way of measuring information to the familiar idea of Shannon information or entropy. The proposal to consider maximally complex objects random appears only in a single sentence at the end of the article.

There are now many tutorials that provide further explanations concerning the definition of Kolmogorov complexity and the existence of universal algorithms. See, e.g., [41, 74].

6.2 Solomonoff

Kolmogorov's invention of algorithmic complexity was anticipated by Ray Solomonoff (born 1926). Solomonoff issued technical reports explaining the idea in 1960 and 1962, before Kolmogorov had arrived at it, and he also anticipated Kolmogorov in publication, with articles in *Information and Control* in 1964 [78, 79].

Solomonoff was interested in inductive inference. He proposed to formalize Occam's razor by basing predictions on the simplest law that fits the data – i.e., the simplest program that generates it. He proved the invariance of the length of this program, which is the same as proving the universality of Kolmogorov's measure of complexity. He also defined a universal prior distribution for prediction by averaging all possible laws, giving smaller weights to laws with longer programs required to describe them, and he conditioned this universal prior on what has been observed so far to make predictions.

The shortcoming of this early work, which helps explain its lack of influence, is its lack of rigor. Solomonoff did not do mathematics with the rigor that might be expected for so abstract a topic. He acknowledged this in the reports and articles themselves. A proof of invariance can be extracted from Solomonoff's article [78], but what is being proven is not clearly stated and the reasoning is introduced with an apology: "an outline of the heuristic reasoning behind this statement will give clues as to the meanings of the terms used and the degree of validity to be expected of the statement itself." Elsewhere in the article, he writes, "If Eq. (1) is found to be meaningless, inconsistent or somehow gives results that are intuitively unreasonable, then Eq. (1) should be modified in ways that do not destroy the validity of the methods used in Sections 4.1 to 4.3." Kolmogorov's student Leonid Levin remembers that when Kolmogorov instructed him to read and cite Solomonoff, he was frustrated by this aspect of the work and soon gave up.

Kolmogorov made a point of acknowledging Solomonoff's priority in publication after he learned about it. In [32] he wrote: "As far as I know, the first paper published on the idea of revising information theory so as to satisfy the above conditions [dealing with individual objects, not random variables] was the article of Solomonov [78]. I came to similar conclusions, before becoming aware of Solomonoff's work, in 1963–1964, and published my first arcticle on the subject [31] in early 1965". Unlike Kolmogorov, Solomonoff had not used the concept of algorithmic complexity to define randomness; Solomonoff was interested instead in induction.

Solomonoff's 1964 articles also contain other ideas that were developed much later. In

Section 3.2 (in the first of the two articles), for example, Solomonoff gives a simple formula for predictions in terms of conditional a priori probability, using monotonic machines much before Levin and Schnorr. In 1978, Solomonoff formally proved that this formula works for all computable probability distributions [80].

6.3 Chaitin

Gregory Chaitin was born in the United States in 1947, into a family from Argentina. He recalls that in an essay he wrote as he entered the Bronx High School of Science in 1962, he suggested that a finite binary string is random if it cannot be compressed into a program shorter than itself [13]. He entered City College in 1964, and after his first year there, in the summer of 1965, he wrote "a single paper that is of a size of a small book" [13]. A condensed version was published in two parts in the *Journal of the ACM*. In the first part, published in 1966 [9], he defines the complexity of a binary string in terms of the size of a Turing machine; in the second, submitted in November 1965 but published only in 1969 [10], he defines complexity more generally, in the same way as Kolmogorov did in his 1965 article.

Chaitin and his family returned to Buenos Aires in 1966, and he joined IBM Argentina as a programmer in 1967. His work on algorithmic complexity made a jump forward when he visited IBM's Watson Laboratory in New York for a few months in 1974. He joined this laboratory full-time in 1975 and spent the period from 1976 to 1985 concentrating on IBM's RISC (Reduced Instruction Set Computer) project. He resumed his work on algorithmic information theory in 1985 and has continued it since. Since 2000, he has been a visiting professor in the Computer Science Department at the University of Auckland in New Zealand.

We will discuss some of Chaitin's work in the 1970s in Section 10. His most famous discovery, which we will not discuss in this article, is probably his proof of Gödel's incompleteness theorem based on the Berry paradox [11].

7 Per Martin-Löf's definition of randomness

The Swedish mathematician Per Martin-Löf (born 1942) went to Moscow to study with Kolmogorov during 1964–65, after learning Russian during his military service. In a recent interview with Alexander Shen [51], he explained that he had not previously worked on randomness and did not immediately do so when he arrived. Kolmogorov first gave him a problem in discriminant analysis, which he solved but considered insufficiently challenging. In late autumn 1964, however, Leonid Bassalygo told him about Kolmogorov's new ideas about complexity and randomness, which he found very exciting. He set about learning about recursive function theory and soon obtained interesting results about unavoidable oscillations in complexity in the prefixes of infinite binary sequences, which he discovered when trying to make the complexity of these prefixes as large as possible.

In March 1965, in a train to the Caucasus, Martin-Löf told Kolmogorov about two theorems he had proven on these oscillations. Kolmogorov was so interested that he asked Martin-Löf to present his results as a sequel to a lecture that Kolmogorov gave in Tbilisi, on their way back to Moscow in late March. Martin-Löf wrote two papers in Russian on the oscillations; the second appeared in 1966 [45]; the first was incorporated into an article that appeared in English in 1971 [50].

Kolmogorov had been interested in finite sequences, but in order to get away from the finitary theory's annoying constants, Martin-Löf investigated instead the question of how to define randomness for an infinite binary sequence. Martin-Löf's first thought was that an infinite binary sequence $\omega_1\omega_2\ldots$ might be considered random if the complexity of a prefix $\omega_1\ldots\omega_n$ is always maximal up to a constant, i.e.,

$$K(\omega_1\ldots\omega_n) = n + O(1). \tag{2}$$

(This means that there exists a constant c such that $n - c \leq K(\omega_1\ldots\omega_n) \leq n + c$ for all n.) But there are no sequences with this property, Martin-Löf discovered, because of the unavoidable oscillations in complexity.

By the time he left Moscow in July 1965, Martin-Löf was on his way to a definition of randomness for infinite sequences using an approach that mixed logic with measure theory: effectively null sets. In his interview with Alexander Shen [51], Martin-Löf recalls that although he was not familiar with the work of Wald, Church, and Ville, he had absorbed from his reading of Borel the idea that a random sequence should avoid properties with probability zero, or null sets (see, for example, [6]). It is impossible to avoid all null sets; any single sequence itself has probability zero. But it is possible to avoid countably many null sets, and Martin-Löf realized that only countably many can be effectively constructed.

Whereas Wald had constructed null sets by way of selection rules, and Ville had constructed them by way of martingales, Martin-Löf considered how null sets are defined in measure theory. Consider as usual the simple case of the Bernoulli measure with $p = 1/2$. Ever since Borel's 1909 article, mathematicians had understood that this measure is the same as Lebesgue measure on the interval $[0, 1]$ when each real number in $[0, 1]$ is identified with the sequence of 1s and 0s formed by its dyadic expansion. Measure theory says that a subset A of $[0, 1]$ is null (has measure zero or probability zero) if for every $\varepsilon > 0$ there exists a sequence of intervals covering A whose total measure is at most ε. Martin-Löf called A *effectively null* if there exists an algorithm that takes any positive rational ε as input and generates a sequence of intervals that cover A and have total measure at most ε. It is obvious that the union of all effectively null sets is a null set, since there are only countably many algorithms. Sequences that do not belong to any effectively null set therefore exist and form a set with measure one. These are the sequences Martin-Löf considered random. Now they are called *Martin-Löf random* sequences.

Martin-Löf also proved that the union of all effectively null sets is effectively null – in other words, there exists a largest effectively null set. This maximal set consists of all nonrandom sequences. A set A is effectively null if and only if A is a subset of this maximal effectively null set, i.e., A does not contain any random sequence.

Martin-Löf arrived at his definition and results while back in Sweden during the academic year 1965–66. He published them in 1966, in an article that was received by the journal on April 1, 1966 [46]. Later in April, he gave four lectures on his results at the University of Erlangen-Nürnberg, and notes from his lectures [47], in German, were widely distributed, making his and Kolmogorov's work on complexity and randomness relatively well known in Germany.

In his first Erlangen lecture, Martin-Löf contrasted the foundations for probability proposed by von Mises and Kolmogorov. Von Mises, he explained, wanted to base proba-

bility on the concept of a collective, whereas Kolmogorov had proposed to begin with the axioms for probability and base applications on two ideas: that frequency approximates probability when an experiment is repeated, and that an event of very small probability can be expected not to happen on a single trial (Cournot's principle). He cited Ville's book, the Geneva colloquium, and other contributions to the literature on collectives and declared that Ville's counterexample, in which the convergence to 1/2 is from above, had brought discussion of von Mises' Axiom II to an end for the time being.

In his 1966 article and in his Erlangen lectures, Martin-Löf begins how own contribution with the concept of a universal test for the randomness of finite sequences. This is a reformulation of Kolmogorov's definition of randomness for finite sequences by means of a universal algorithm, but Martin-Löf found it could be adapted more readily to infinite sequences. He showed that there exists a universal sequential test for the randomness of infinite sequences, and that this way of defining randomness for infinite sequences is equivalent to the definition in terms of the maximal effectively null set.

Martin-Löf never had an opportunity to discuss his results with Kolmogorov, but they were mentioned in a detailed survey article [91], published in 1970 by Leonid Levin and Alexander Zvonkin, two of Kolmogorov's students, on Kolmogorov's suggestion; Kolmogorov carefully reviewed this article and suggested many corrections. In addition to Martin-Löf's results, the article covered other results about complexity and randomness obtained by the Kolmogorov school in Moscow.

Martin-Löf later studied the earlier literature on random sequences in more detail and published a review of it in 1969 in English in the Swedish philosophical journal *Theoria* [48]. This was the first survey in the English language of the work by von Mises, Wald, and Ville, and others that we mentioned in Sections 2, 3, and 4 above, and in some respects it rescued Ville from obscurity. Whereas the influence of Ville's martingales in measure-theoretic probability was by way of Doob, its influence in algorithmic randomness seems to have been by way of Martin-Löf.

8 Claus-Peter Schnorr's computable martingales

Claus-Peter Schnorr (born 1943), who was looking for new research topics after earning a doctoral degree for work in mathematical logic at Saarbrücken in 1967, encountered algorithmic randomness through the notes from Marin-Löf's Erlangen lectures. Building on Martin-Löf's results, Schnorr brought martingales back into the story. His work on algorithmic martingales during the late 1960s culminated, in 1970, in his habilitation and in a series of lectures that appeared as a book in 1971 [67]. (See also [66, 68, 69].)

According to Schnorr's talk at Dagstuhl [70], he never read Ville's book, having learned about the notion of a martingale indirectly. Schnorr's book is the first publication in which martingales were used in connection with algorithmic randomness.

Schnorr studied *computable* and *lower semicomputable* martingales. A function f (arguments are finite strings of 1s and 0s, values are reals) is called computable if there is an algorithm that computes the values of f with any given precision: given x and positive rational ε, the algorithm computes some rational ε-approximation to $f(x)$. A function is lower semicomputable if there is an algorithm that, given x, generates a sequence of rational numbers that approach $f(x)$ from below. It is easy to see that f is computable if and only if both f and $-f$ are lower semicomputable.

Schnorr characterized Martin-Löf randomness in terms of martingales as follows: an infinite binary sequence is Martin-Löf random if and only if no lower semicomputable nonnegative martingale wins against it (by becoming unbounded). (The initial capital can be noncomputable in this setting.) He also brought the notion of a *supermartingale*, introduced into measure-theoretic probability by Doob in the 1950s, into the theory of algorithmic randomness. A function m on finite strings is a supermartingale if it satisfies the supermartingale inequality,

$$m(x) \geq \frac{m(x0) + m(x1)}{2}.$$

This can be the capital process of a gambler who is allowed to throw money away at each trial. Schnorr proved that lower semicomputable supermartingales characterize Martin-Löf randomness in the same way as lower semicomputable martingales do.

But Schnorr was dissatisfied with this formulation. He proved that there exists a sequence that wins against all computable martingales but is not Martin-Löf random, and he considered computability more appropriate as a condition on martingales than semicomputability. Why should we generate approximations from below but not above? He concluded that semicomputable martingales (or supermartingales) are too broad a class, and that the corresponding class of sequences, the Martin-Löf random sequences, is too narrow.

Trying to find a definition of randomness that better matched his intuition, Schnorr considered a smaller class of effectively null sets, now sometimes called *Schnorr null*. For an effectively null set A there exists an algorithm that given $\varepsilon > 0$ generates a sequence of intervals that cover A and have total measure *at most* ε. For a Schnorr null set, this total measure should *equal* ε. This may sound a bit artificial, but it is equivalent to asking for a computably converging series of lengths of covering intervals. The sequences that are outside all Schnorr null sets he called random ("zufällig" in German; we now call them *Schnorr random*). Schnorr proved that this class of sequences is indeed larger than the class of Martin-Löf random sequences. He also proved that a sequence is Schnorr random if and only if no computable martingale computably wins on it; this means that there exists a nondecreasing unbounded computable function $h(n)$ such that the player's capital after n steps is greater than $h(n)$ for infinitely many n.

Schorr also considered a natural intermediate requirement: no computable martingale wins (computably or not) on a sequence, i.e., all computable martingales are bounded on its prefixes. Schnorr proved that this class (now its members are sometimes called *computably random* sequences) is broader than the class of Martin-Löf random sequences; much later Wang [88] showed that it is still smaller than the class of all Schnorr random sequences.

Schnorr's work during this period also contained many other ideas that endured and were developed further much later. For example, he considers how fast a player's capital increases during the game. If a sequence violates the strong law of large numbers, there exists a computable martingale that wins exponentially fast against it, but the violation of more delicate laws may involve slower growth in the player's capital. In the past ten years, the growth of martingales has been connected to notions of effective dimension [43].

One of Schnorr's goals was to develop concepts of pseudorandomness. An object with a short description can be called pseudorandom if the time needed to decompress

the description is unreasonably large. So Schnorr considered complexity with bounded resources in his book. He later worked in computational cryptography, where more recent and more practical theories of pseudorandomness are used [28].

9 Leonid Levin's semimeasures

Semimeasures, which are closely related to supermartingales, were introduced in the 1970 article by Levin and Zvonkin [91].

Let Σ be the set of all finite and infinite binary sequences, and let Σ_x be the set of all extensions (finite and infinite) of a binary string x. Then $\Sigma_x = \Sigma_{x0} \cup \Sigma_{x1} \cup \{x\}$. A *semimeasure* is a measure on Σ. It is convenient to specify a semimeasure in terms of the value it assigns to Σ_x for each x, say $q(x)$. A nonnegative real-valued function q on finite strings defines a semimeasure if and only if

$$q(x) \geq q(x0) + q(x1) \tag{3}$$

for every finite string x. We usually assume also that $q(\Box) = 1$ (this says that the measure assigns the value 1 to the whole set Σ; it is a probability measure). The difference between the two sides of the inequality (3) is the measure of the finite string x. A semimeasure is said to be *lower semicomputable* if the function $x \mapsto q(x)$ is lower semicomputable.

As Levin showed in the article with Zvonkin, lower semicomputable semimeasures are output distributions of randomized algorithms. Consider a black box that has a random bit generator inside and, being started, produces a string of 1s and 0s bit by bit (pausing between each bit for an unpredictable amount of time). This machine can produce both finite (if no bits appear after some moment) and infinite sequences and therefore determines a probability distribution on Σ. This distribution is a lower semicomputable semimeasure and every lower semicomputable semimeasure (that equals 1 on the entire set Σ) can be obtained in this way.

What is the connection between semimeasures and supermartingales? As Ville had explained in 1939 ([82], pp. 88–89), a nonnegative martingale m is a ratio of two probability measures. To see what this means, write $p(x)$ for the probability the Bernoulli measure with parameter $1/2$ assigns to x being a prefix of the infinite binary sequence. Then $p(x) = (1/2)^n$, where n is the length of x. Because $p(x0) = p(x1) = (1/2)p(x)$, Equation (1) tells us that

$$m(x)p(x) = m(x0)p(x0) + m(x1)p(x1). \tag{4}$$

If m is nonnegative and starts at 1, this implies that $m(x)p(x)$ can be interpreted as the value assigned to Σ_x by a probability measure. Writing $q(x)$ for $m(x)p(x)$, we have $m(x) = q(x)/p(x)$. Every nonnegative martingale $m(x)$ starting at 1 can be represented in this way, and every such ratio is a nonnegative martingale starting at 1. This generalizes to supermartingales and semimeasures. If q is a semimeasure and p is a probability measure, then the ratio $q(x)/p(x)$ is a nonnegative supermartingale starting at 1, and every nonnegative supermartingale starting at 1 can be obtained in this way. Lower semicomputable semimeasures correspond to lower semicomputable supermartingales.

The article with Zvonkin also included Levin's proof of the existence of a maximal lower semicomputable semimeasure, called the *universal semimeasure* or *a priori probability on a binary tree*. This is a lower semicomputable semimeasure r such that for any

other lower semicomputable semimeasure q there exists a constant c such that

$$r(x) \geq \frac{q(x)}{c}$$

for any finite string x.

Semimeasures can be used to define supermartingales with respect to any measure, not only uniform Bernoulli measure. Ville had already shown that the representation of a martingale as a ratio of measures generalizes to the case where p is any measure on $\{0,1\}^\infty$: a martingale with respect to p is the ratio of some measure q to p. A supermartingale with respect to an arbitrary measure p is similarly the ratio of a semimeasure q to p. This implies that for any measure p there exists a maximal lower semicomputable p-supermartingale: it is the ratio of the universal semimeasure r (perhaps conditioned on p in a certain sense) to p. This connects maximal p-supermartingales for different p: when we switch from semimeasures to supermartingales, one object (the universal semimeasure) is transformed into a family of seemingly different objects (maximal lower semicomputable supermartingales with respect to different measures).

Zvonkin and Levin's 1970 article [91] had the ingredients needed to provide a criterion of randomness in terms of semimeasures: a sequence ω is Martin-Löf random with respect to a computable measure p if and only if the ratio $r(x)/p(x)$ is bounded for prefixes x of ω, where $r(x)$ is the universal semimeasure. (This statement is a reformulation of Schnorr's characterization of Martin-Löf randomness in terms of lower semicomputable supermartingales.) However, Levin discovered this result only later (see Levin's letters to Kolmogorov in Appendix C).

10 Characterizing Martin-Löf randomness using complexity

The goal of characterizing the randomness of an infinite sequence in terms of the complexity of its prefixes was finally achieved in the 1970s by Schnorr and Levin. To do this (and this itself was a very important development), they modified the definition of algorithmic complexity. Schnorr and Levin introduced monotone complexity, and Levin and Chaitin introduced prefix complexity.

The history of these discoveries is complicated, because different people, working independently, sometimes used slightly different definitions, and sometimes the results remained unpublished for several years or were published without proofs in a short and sometimes cryptic form. We begin this section with some biographical information about Levin, which explains in part why this happened with some of his results.

10.1 Leonid Levin in the Soviet Union

In a recent interview [40], Leonid Levin recalls that he was thinking about the length of the shortest arithmetic predicate that is provable for a single value of its parameter when he was a student in a high school for gifted children in Kiev in 1963–64. He realized that he did not know how to make this definition invariant – i.e., how to make the complexity independent of the specific formalization of arithmetic. The following year, 1964–65, he was studying in a boarding school for gifted children in Moscow, founded by Kolmogorov, and he posed his question to A. Sossinsky, a teacher there. Sossinsky asked Kolmogorov about the question, and Kolmogorov replied that he had answered it in a forthcoming article.

In January 1966, Levin entered Moscow State University, becoming a first-year undergraduate in the middle of the academic year. This was unusual, but it was permitted for students at Kolmogorov's school that year, because the Soviet Union was changing from an 11-year to a 10-year curriculum. Early during his study at the university, he obtained a result on the symmetry of information, which he hoped to use to convince Kolmogorov to be his adviser. But Kolmogorov was always busy, and the appointment to talk with him was postponed several times from February to August 1967. Finally, when Levin called him again, Kolmogorov agreed to see him and mentioned that he would tell him something he had just discovered – that information is symmetric. Levin was surprised: "But, Andrei Nikolaevich, this is exactly what I wanted to tell you." — "But do you know that the symmetry is only up to logarithmic terms?" — "Yes." — "And you can give a specific example?" — "Yes." The results they had discovered independently were published without proof by Kolmogorov in 1968 [32], and the proofs were published in the 1970 article by Zvonkin and Levin [91]. Levin continued to work with Kolmogorov during his undergraduate years, but because Kolmogorov did not officially belong to the Mathematical Logic Division of the Mathematics Department, where Levin was enrolled, V. A. Uspensky, who had been Kolmogorov's student in the 1950s, served as Levin's official advisor.

The typical track for a future mathematician in the Mathematics Department of Moscow State University at that time was 5 years of undergraduate studies plus 3 years of graduate school. Then the student was supposed to defend a thesis, becoming a "kandidat" (кандидат физико-математических наук), which is roughly equivalent to having a doctoral degree in the United States. To enter graduate school after finishing 5 years of undergraduate studies, one needed a good academic record and a recommendation from the local communist party and komsomol. Komsomol (**ком**мунистический **сою**з **мол**одёжи, Communist Union of Young People) was almost obligatory for those from 14 to 28 years of age. Most university students were members, although there were some exceptions and the requirement was never formalized as a law.

Being Jewish, already a handicap at that time, and also a nonconformist, Levin created a lot of trouble for the local university authorities as an undergraduate. He became an elected local komsomol leader but did not follow the instructions given by his Communist Party supervisors. Noisy and arrogant, as he later described himself ([73], p. 152), he got away with his behavior because the local authorities did not want to take disciplinary actions that would show higher-ups they were having difficulties, but this tolerance faded after the Prague Spring of 1968, and when Levin finished his undergraduate studies in 1970, his misbehavior was mentioned in his graduation letter of recommendation. Not surprisingly, he was not admitted to the graduate school. But with the help of the university rector, I. G. Petrovsky, Kolmogorov managed to secure a job for him in the university's statistical laboratory, which Kolmogorov headed.

An individual could defend a "kandidat" thesis without having been enrolled in a graduate program. So Levin prepared a thesis, consisting of results he had published in the 1970 article with Zvonkin, along with a few others. It was clearly impossible to defend it in Moscow, but a defense finally took place in Novosibirsk in Siberia in 1971. Very untypically, it was unsuccessful. Though all the reviews were positive, the jury not only rejected the thesis, but they included reference to Levin's "unclear political position" in their report. It effectively barred him from defending any thesis in the Soviet Union.

Levin realized that he might be soon barred from publishing in Soviet journals, and many important results, including the definition of prefix complexity, remain unpublished at the time. So he rushed a number of articles into print from 1973 to 1977. These articles were short and cryptic, containing many claims without proofs and many ideas that were understood only much later.

Some of Levin's results also appeared in a paper published in 1974 by Peter Gács. While working in Hungary, Gács had read Kolmogorov's 1965 article, Martin-Löf's lecture notes from Erlangen, and Zvonkin and Levin's 1970 article, and he had begun corresponding with Levin. He spent the 1972–73 academic year in Moscow working with Levin.

Levin was eventually given permission to leave the Soviet Union. As he recalls, the KGB made it known to him through Kolmogorov that they thought emigration was his best option. (Kolmogorov did not say whether he agreed with their advice.) In 1978, Levin immigrated to the United States, where he became well known for work in a number of areas of theoretical computer science, including one-way functions, holographic proofs, and for discovering (independently from Cook and Karp) the phenomenon of NP-completeness (the article [36] appeared while he was still in Russia).

10.2 Monotone complexity: Levin and Schnorr

By 1971–72, Levin and Schnorr had both realized, independently, that the oscillations in complexity that had stood in the way of Martin-Löf's goal of characterizing randomness by requiring maximal complexity for all prefixes can be eliminated if the algorithms or machines used to define complexity are required to be monotone in some sense.

We see the idea of monotone complexity already in Appendix C's Letter II from Levin to Kolmogorov, written in January 1971 or earlier. There Levin explains that an algorithm A is monotone if whenever $A(x)$ is defined and y is a prefix of x, $A(y)$ is also defined and is a prefix of $A(x)$. Let us define monotone complexity as the minimal length of a program (for an optimal monotone algorithm) that produces x. Levin formulates the following criterion: a sequence is Martin-Löf random with respect to a computable measure p if and only if the monotone complexity of its prefixes equals $-\log_2 p(x)+O(1)$. For the uniform Bernoulli measure this means that $\omega_1\omega_2\ldots$ is random if and only if the monotone complexity of $\omega_1\ldots\omega_n$ equals $n+O(1)$. Note that the monotone complexity of any string of length n is at most $n+O(1)$, and this criterion characterizes random sequences as sequences whose prefixes have maximal possible complexity.

Schnorr advocated a version of monotone complexity, which he called *process complexity*, in May 1972 at the Fourth ACM Symposium on the Theory of Computing (STOC), in Denver [68]. In the proceedings, he proved that a sequence is Martin-Löf random if and only if its n-bit prefix has monotone complexity $n+O(1)$. This seems to be first time this result appears in print, but as Schnorr pointed out, the basic properties of monotone algorithms had already been studied by himself [67] and by Zvonkin and Levin [91].

In an article that appeared in 1973 [35], Levin proved essentially the same result using a slightly different version of monotone complexity, which Schnorr adopted in a subsequent article [69]. Levin also noted that the same proof works for *a priori complexity* – i.e., minus the binary logarithm of the universal semimeasure on the binary tree. This statement is equivalent to Schnorr's characterization of randomness in terms of semicomputable supermartingales.

10.3 Prefix complexity

Prefix complexity can be defined in different ways. First, the prefix complexity of a natural number i can be defined as $-\log_2 m_i$ where m_i is the maximal lower semicomputable converging series of non-negative reals. (A series $\sum_i a_i$ is lower semicomputable if the function $i \mapsto a_i$ is lower semicomputable, i.e., for every i one can effectively generate approximations to a_i from below.) The prefix complexity of binary strings is then defined using some computable bijection between strings and natural numbers. (Of course, we need to prove that there exists a maximal converging lower semicomputable series; it can be done in the same way as for universal semimeasures on the binary tree.)

Another definition explains the name used: the *prefix complexity* of a string x is the length of the shortest program p, considered as a bit string, that produces x, assuming that the programming language used has the following "prefix" (self-delimiting) property: if some program p produces some output, any extension of it produces the same output.

Levin and Gács were the first to publish a definition of prefix complexity. They did so in Russian in 1974. Levin's 1974 article [37] appeared in English translation in 1976, and Gács' 1974 article, which attributed the idea to Levin, appeared in English translation in 1975 [25] (see [26]). The two authors' articles state, without proof, the equivalence of the two definitions mentioned above. Levin's article refers for details to an unpublished paper of his and to Gács' article. The unpublished paper appeared only in 1976 [38].

The prefix complexity defined as $-\log_2 m_i$ (but not the other definition) appeared also in Levin's unpublished 1971 thesis. In the 1970 article [91] there is a footnote suggesting consideration of the a priori probability of the string 0^n1 (n zeros followed by one); this quantity coincides with m_i. But this idea is not developed further in the article.

Chaitin independently worked out similar ideas during his work at the Watson Laboratory in 1974, and his resulting article, which appeared in 1975 [12], contained similar definitions and a proof that prefix complexity and minus the logarithm of the maximal converging series are equal – the first published proof of this result. This article by Chaitin is also the first publication to state that prefix complexity characterizes Martin-Löf randomness: a sequence $\omega_1\omega_2\ldots$ is Martin-Löf random with respect to the uniform Bernoulli measure if and only if the prefix complexity of $\omega_1\ldots\omega_n$ is at least $n - O(1)$. (For prefix compleixity the upper bound $n+O(1)$ is no longer valid, but the lower bound still provides a randomness criterion.) In the article [12], Chaitin attributes this result to Schnorr: Chaitin suggested the requirement "prefix complexity of $\omega_1\ldots\omega_n$ is at least $n-O(1)$" as the definition of randomness (now this is often called "Chaitin randomness") and Schnorr, acting as a referee of the paper, informed Chaitin about the equivalence. In his talk at Dagstuhl [70], Schnorr says, "I knew the first paper of Chaitin that has been published one year later after the Kolmogorov 1965 paper, but the next important paper made Chaitin one of the basic investigators of complexity. This was a paper on self-delimiting or prefix-free descriptions, and this was published in 1975 in the *Journal of the ACM*. In fact I was a referee of this paper, and I think Chaitin knew this because I've sent my personal comments and suggestions to him, and he used them."

Chaitin's definition of prefix complexity was slightly different from Levin's: whereas Levin required that extensions of a program p that produces x should produce x, too, Chaitin required that such extensions always produce nothing. Both restrictions reflect (in different ways) the intuitive idea of a self-delimiting program, which allows the ma-

chine to find out the program has ended without the use of an end-marker. The differences are not important; the two definitions lead to the same quantity up a $O(1)$ term and so are equivalent.

The possibility of switching back and forth between two definitions of prefix complexity (in terms of a series and self-delimiting programs) is an important technical advantage. Another advantage of prefix complexity over complexity as originally defined (*plain* complexity) is that it allows an improvement in the result on symmetry of information originally discovered by Kolmogorov and Levin. We can relate the complexity of a pair to the conditional complexities with an $O(1)$ error term instead of the the logarithmic error term obtained by Kolmogorov and Levin. This was discovered independently by Levin and Chaitin; the first proofs were published in Gács' 1974 article [25] and Chaitin's 1975 article [12].

11 Epilogue

The mathematical theory of randomness and algorithmic information theory have continued to develop since the seminal works of the 1960s and 1970s. In recent decades, they have benefited from advanced techniques of recursion theory and have been applied to other areas of mathematics. Recent books include Christian Calude's *Information and Randomness. An Algorithmic Perspective* [8] and Ming Li and Paul Vitányi's *An Introduction to Kolmogorov Complexity and Its Applications* [41], both of which have copious historical notes. Many interesting recent results can be found in books by Andre Nies [61] and by Rod Downey and Denis Hirschfeldt [21].

Most of the work on algorithmic randomness since the 1970s has been concerned with infinite sequences. But Kolmogorov was always more interested in finite random objects, because only finite objects can be relevant to our experience. Some of his ideas for using the theory of complexity in probability modeling were extended by his student Evgeny Asarin [1, 2].

Martingales, which can have a finite or infinite horizon, have also recently been considered as a foundation for probabilistic reasoning independently of the classical axioms [71]. Instead of forbidding a nonnegative martingale to diverge to infinity in an infinite number of trials, one forbids it to multiply its initial capital by a large factor in a finite number of trials. Predictions are made and theorems proven by constructing martingales. Tests are conducted by checking whether martingales do multiply their initial capital handsomely. The picture that emerges is a little different from classical probability theory, because the logic does not depend on there being enough bets to define probability distributions.

Appendix

A Letter from Kolmogorov to Fréchet

The Fréchet papers in the archives of the Academy of Sciences in Paris include a letter in French to Fréchet, in which Kolmogorov elaborates briefly on his philosophy of probability. This translation is published with permission from the Academy.

Moscow 6, Staropimenovsky per. 8, flat 5
3 August 1939

Dear Mr. Fréchet,

I thank you sincerely for sending the proceedings of the Geneva Colloquium, which arrived during my absence from Moscow in July.

The conclusions you express on pp. 51–54 are in full agreement with what I said in the introduction to my book:

> In the pertinent mathematical circles it has been common for some time to construct probability theory in accordance with this general point of view. But a complete presentation of the whole system, free from superfluous complications, has been missing...

You are also right to attribute to me (on p. 42) the opinion that the formal axiomatization should be accompanied by an analysis of its real meaning. Such an analysis is given, perhaps too briefly, in the section "The relation to the world of experience" in my book. Here I insist on the view, expressed by Mr. von Mises himself (Wahrscheinlickeitsrechnung 1931, pp. 21–26), that "collectives" are finite (though very large) in real practice.

One can therefore imagine three theories:

A A theory based on the notions of "very large" finite "collectives", "approximate" stability of frequencies, etc. This theory uses ideas that cannot be defined in a purely formal (i.e., mathematical) manner, but it is the only one to reflect experience truthfully.

B A theory based on infinite collectives and limits of frequencies. After Mr. Wald's work we know that this theory can be developed in a purely formal way without contradictions. But in this case its relation to experience cannot have any different nature than for any other axiomatic theory. So in agreement with Mr. von Mises, we should regard theory B as a certain "mathematical idealization" of theory A.

C An axiomatic theory of the sort proposed in my book. Its practical value can be deduced directly from the "approximate" theory A without appealing to theory B. This is the procedure that seems simplest to me.

Yours cordially,
A. Kolmogoroff

B Abstracts of Kolmogorov's talks

Abstracts of some of the talks at the meetings of Moscow Mathematical Society were published in the journal "Успехи математических наук" (Uspekhi matematicheckikh nauk). Here we reproduce translations of the abstracts for three talks by Kolmogorov, in 1967, 1971, and 1974, on algorithmic information theory. The translations are by Leonid Levin; we have edited them slightly.

B.1 A. N. Kolmogorov, "Several theorems about algorithmic entropy and algorithmic amount of information". The talk was on October 31, 1967; the abstract appeared in Volume 23, no. 2, March-April 1968.

The algorithmic approach to the foundations of information theory and probability theory was not developed far for several years after its appearance, because some questions raised at the very start remained unanswered. Now the situation has changed somewhat. In particular, it is ascertained that the decomposition of entropy $H(x,y) \sim H(x) + H(y|x)$ and the formula $J(x|y) \sim J(y|x)$ hold for the algorithmic concept only with accuracy $O([\log H(x,y)])$ (Levin, Kolmogorov).

The fundamental difference between the algorithmic definition of a Bernoulli sequence (a simplest collective) and the definition of Mises-Church, stated earlier, is concretized in the form of a theorem: there exist Bernoulli (in the sense of Mises-Church) sequences $x = (x_1, x_2, ...)$ with density of ones $p = \frac{1}{2}$, with initial segments of entropy ("complexity") $H(x^n) = H(x_1, x_2, ..., x_n) = O(\log n)$ (Kolmogorov).

For understanding of the talk an intuitive, not formal, familiarity with the concept of a computable function suffices.

B.2 A. N. Kolmogorov, "Complexity of specifying and complexity of constructing mathematical objects". The talk was on November 23, 1971; the abstract appeared in Volume 27, no. 2, 1972.

1. Organizing machine computations requires dealing with evaluation of (a) complexity of programs, (b) size of memory used, (c) duration of computation. The talk describes a group of works that consider similar concepts in a more abstract manner.

2. It was noticed in 1964–1965 that the minimal length $K(x)$ of the binary representation of a program specifying the construction of an object x can be defined invariantly up to an additive constant (Solomonoff, A. N. Kolmogorov). This permitted using the concept of *definition complexity* $K(x)$ of constructive mathematical objects as the basis for a new approach to the foundations of information theory (A. N. Kolmogorov, Levin) and probability theory (A. N. Kolmogorov, Martin-Löf, Schnorr, Levin).

3. Such characteristics as "required memory volume," or "required duration of work" are harder to free of technical peculiarities of special machine types. But some results may already be extracted from the axiomatic "machine-independent" theory of a broad class of similar characteristics (Blum, 1967). Let $\Pi(p)$ be a characteristic of "construction complexity" of the object $x = A(p)$ by a program p, and let $\Lambda(p)$ be the length of the program p. The formula $K^n\Pi(x) = \inf(\Lambda(p) : x = A(p), \Pi(p) = n)$ defines the "n-complexity of definition" of object x (when the condition is unsatisfiable, the inf is considered infinite).

4. Barzdin's Theorem on the complexity $K(M_\alpha)$ of prefixes M_α of an enumerable set of natural numbers (1968) and results of Barzdin, Kanovich, and Petri on corresponding complexities $K^n\Pi(M_\alpha)$, are of general mathematical interest, as they shed some new light on the role of extending previously used formalizations in the development of mathematics. The survey of the state of this circle of problems was given in the form free from any cumbersome technical apparatus.

B.3 A. N. Kolmogorov, "Complexity of algorithms and objective definition of randomness". The talk was on April 16, 1974; the abstract appeared in Volume 29, no. 4 (155), 1974.

To each constructive object corresponds a function $\Phi_x(k)$ of a natural number k – the log of minimal cardinality of x-containing sets that allow definitions of complexity at most k. If the element x itself allows a simple definition, then the function Φ drops to 1 even for small k. Lacking such a definition, the element is "random" in a negative sense. But it is positively "probabilistically random" only when the function Φ, having taken the value Φ_0 at a relatively small $k = k_0$, then changes approximately as $\Phi(k) = \Phi_0 - (k - k_0)$.

C Levin's letters to Kolmogorov

These letters are not dated but were written after the submission of [91] in August 1970 and before Kolmogorov left (in January 1971) for an oceanographic expedition on the ship Dmitry Mendeleev. Copies (the typescript for the first two letters and the handwritten manuscript for the third one) provided by Leonid Levin and translated by Alexander Shen.

C.1 Letter I

Dear Andrei Nikolaevich! A few days ago I obtained a result I like a lot. Maybe it could be useful to you if you work on these topics while traveling on the ship.

This result gives a formulation for the foundations of probability theory different from Martin-Löf. I think it is closer to your initial idea about the relation between complexity and randomness and is much clearer from the philosophical point of view (as, e.g., [Yu. T.] Medvedev says).

Martin-Löf considered (for an arbitrary computable measure P) an algorithm that studies a given sequence and finds more and more deviation from the P-randomness hypothesis. Such an algorithm should be P-consistent, i.e., find deviations of size m only for sequences in a set that has measure at most 2^{-m}. It is evident that a number m produced by such an algorithm on input string x should be between 0 and $-\log_2 P(x)$. Let us consider the complementary value $(\log_2 P(x)) - m$ and call it the "complementary test" (the consistency requirement can be easily reformulated for complementary tests).

Theorem. *The logarithm of a priori probability [on the binary tree]* $-\log_2 R(x)$ *is a P-consistent complementary test for every measure P and has the usual algorithmic properties.*

Let me remind you that by a priori probability I mean the universal semicomputable measure introduced in our article with Zvonkin. [See [91].] It is shown there that it [minus its logarithm] is numerically close to complexity.

Let us consider a specific computable measure P. Compared to the universal Martin-Löf test f (specific to a given measure P) our test is not optimal up to an additive constant, but is asymptotically optimal. Namely, if the universal Martin-Löf test finds a deviation m, our test finds a deviation at least $m - 2\log_2 m - c$. Therefore, the class of random infinite binary sequences remains the same.

Now look how nicely it fits the philosophy. We say that a hypothesis "x appeared randomly according to measure P" can be rejected with certainty m if the measure P

is much less consistent with the appearence of x than a priori probability (this means simply that $P(x) < R(x)/2^m$. This gives a law of probability theory that is violated with probability at most 2^{-m}. Its violation can be established effectively since R is semi-computable [enumerable from below]. But if this law holds, all other laws of probability theory [i.e., all Martin-Löf tests] hold, too. The drawback is that it gives a bit smaller value of randomness deficiency (only $m - 2\log_2 m - c$ instead of m), but this is a price for the universality (arbitrary probability distribution). The connection with complexity is provided because $-\log_2 R(x)$ almost coincides with the complexity of x. Now this connection does not depend on the measure.

It is worth noting that the universal semicomputable measure has many interesting applications besides the above mentioned. You know its application to the analysis of randomized algorithms. Also it is often useful in proofs (e.g., in the proof of J. T. Schwartz' hypothesis regarding the complexity of almost all trajectories of dynamic systems). Once I used this measure to construct a definition of intuitionistic validity. All this show that it is a rather natural quantity.

L.

C.2 Letter II

Dear Andrei Nikolaevich!

I would like to show that plain complexity does not work if we want to provide an *exact* definition of randomness, even *for a finite case*. For the uniform distribution on strings of fixed length n the randomness deficiency is defined as n minus the complexity. For a non-uniform distribution length is replaced by minus the logarithm of the probability.

It turns out that even for a distribution on a finite set the randomness deficiency could be high on a set of large measure.

Example. Let

$$P(x) = \begin{cases} 2^{-(l(x)+100)} & \text{if } l(x) \le 2^{100} \\ 0 & \text{if } l(x) > 2^{100}. \end{cases}$$

Then $|\log_2 P(x)| - K(x)$ exceeds 100 *for all* strings x.

A similar example can be constructed for strings of some fixed length (by adding zero prefixes). The violation could be of logarithmic order.

Let me show you how to sharpen the definition of complexity to get an exact result (both for finite and infinite sequences).

Definitions. Let A be a monotone algorithm, i.e., for every x and every y that is a prefix of x, if $A(x)$ is defined, then $A(y)$ is defined too and $A(y)$ is a prefix of $A(x)$. Let us define

$$KM_A(x) = \begin{cases} \min\ l(p)\colon x \text{ is a prefix of } A(p) & \\ \infty & \text{if there is no such } p \end{cases}$$

The complexity with respect to an optimal algorithm is denoted by $KM(x)$.

Let $P(x)$ be a computable distribution on the Cantor space Ω, i.e., $P(x)$ is the measure of the set Γ_x of all infinite extensions of x.

Theorem 1.

$$KM(x) \le |\log_2 P(x)| + O(1);$$

Theorem 2.

$$KM((\omega)_n) = |\log_2 P((\omega)_n)| + O(1)$$

for P-almost all ω; here $(\omega)_n$ stands for n-bit prefix of ω. Moreover, the probability that the randomness deficiency exceeds m for some prefix is bounded by 2^{-m}.

Theorem 3. *The sequences ω such that*

$$KM((\omega)_n) = |\log_2 P((\omega)_n)| + O(1);$$

satisfy all laws of probability theory (all Martin-Löf tests).

Let me use this occasion to tell you the results from my talk in the laboratory [of statistical methods in Moscow State University]: why one can omit non-computable tests (i.e., tests not definable without a strong language).

For this we need to improve the definition of complexity once more. The plain complexity $K(x)$ has the following property:

Remark. Let A_i be an effectively given sequence of algorithms such that

$$K_{A_{i+1}}(x) \le K_{A_i(x)}$$

for all i and x. Then there exists an algorithm A_0 such that

$$K_{A_0}(x) = 1 + \min_i K_{A_i}(x).$$

Unfortunately, it seems that $KM(x)$ does not have this property. This can be corrected easily. Let A_i be an effective sequence of monotone algorithms with finite domain (provided as tables) such that

$$KM_{A_{i+1}}(x) \le KM_{A_i(x)}$$

for all i and x. Let us define then

$$\overline{KM}_{A_i}(x) = \min_i KM_{A_i}(x).$$

Among all sequences A_i there exists an optimal one, and the compexity with respect to this optimal sequence is denoted by $\overline{KM}(x)$. This complexity coincides with the logarithm of a universal semicomputable semimeasure [=a priori probability on the binary tree].

Theorem 4. $\overline{KM}(x)$ *is a minimal semicomputable [from above] function that makes Theorem* 2 *true.*

Therefore no further improvements of $\overline{KM}$ are possible.

Now consider the language [=set] of all functions computable with a fixed noncomputable sequence [oracle] α. Assume that α is complicated enough, so this set contains the characteristic function of a universal enumerable set [$\mathbf{0}'$].

We can define then a relativized [языковую in the Russian original] complexity $\overline{KM}_\alpha(x)$ replacing algorithms by algorithms with oracle α, i.e., functions from this language.

Definition. A sequence ω is called *normal* if

$$\overline{KM}((\omega)_n) = \overline{KM}_\alpha((\omega)_n) + O(1).$$

For a finite sequence ω_n we define the "normality deficiency" as

$$\overline{KM}(\omega_n) - \overline{KM}_\alpha(\omega_n).$$

Theorem 5. *A sequence obtained by an algorithm from a normal sequence is normal itself.*

Theorem 6. *Let P be a probability distribution that is defined (in a natural encoding) by a normal sequence. Then P-almost every sequence is normal.*

This theorem exhibits a law of probability theory that says that a random process cannot produce a non-normal sequence unless the probability distribution itself is not normal. This is a much more general law than standard laws of probability theory since it does not depend on the distribution. Moreover, Theorem 5 shows that this law is not restricted to probability theory and can be considered as a univeral law of nature:

Thesis. Every sequence that appears in reality (finite or infinite) has normality deficiency that does not exceed the complexity of the description (in a natural language) of how it is physically produced, or location etc.

It turns out that this normality law (that can be regarded as not confined to probability theory) and the law corresponding to the universal computable test together imply any law of probability theory (not necessary computable) that can be described in the language. Namely, the following result holds:

Theorem 7. *Let P be a computable probability distribution. If a sequence ω is normal and passes the universal computable P-test, then ω passes any test defined in our language (i.e., every test computable with oracle α).*[12]

Let us give one more interesting result that shows that all normal sequences have similar structure.

Theorem 8. *Every normal sequence can be obtained by an algorithm from a sequence that is random with respect to the uniform distribution.*

C.3 Letter III. This letter has no salutation. Levin recalls that he often gave notes like this to Kolmogorov, who rarely had much time to hear lengthy explanations and preferred something written in any case.

We use a sequence α that provides a "dense" coding of a universal [recursively] enumerable set. For example, let α be the binary representation of [here the text "the sum of the a priori probabilities of all natural numbers" is crossed out and replaced by the following:] the real number

$$\sum_{p \in A} \frac{1}{p \cdot \log^2 p}$$

where A is the domain of the optimal algorithm.

A binary string p is a "good" code for x if the optimal algorithm converts the pair $(p, K(x))$ into a list of strings that contains x, and the logarithm of the cardinality of this list does not exceed $K(x) + 3 \log K(x) - l(p)$. (The existence of such a code means that x is "random" when $n \geq l(p)$.)

We say that a binary string p is a canonical code for x if every prefix of p either is a "good" code for x or is a prefix of α, and $l(p) = K(x) + 2 \log K(x)$.

[12] In a footnote in the letter, Levin adds, "Note that for every set of measure 0 there exists a test (not necessary computable) that rejects all its elements."

Theorem 1. *Every x (with finitely many exceptions) has a canonical code p, and p and x can be effectively transformed into each other if $K(x)$ is given.*

Therefore, the "non-randomness" in x can appear only due to some very special information (a prefix of α) contained in x. I cannot imagine how such an x can be observed in (extracted from) the real world since α is not computable. And the task "to study the prefixes of a specific sequence α" seems to be very special.

References

[1] Asarin, Evgeny A., Some properties of Kolmogorov δ-random finite sequences, *Theory of Probability and its Applications* 32:507–508, 1987.

[2] Asarin, Evgeny A., On some properties of finite objects random in the algorithmic sense, *Soviet Mathematics Doklady* 36:109–112, 1988.

[3] Barbut, Marc, Bernard Locker, and Laruent Mazliak, *Paul Lévy – Maurice Fréchet: 50 ans de correspondance en 107 lettres*, Hermann, 2004.

[4] Bienvenu, Laurent, and Shen, Alexander. *Algorithmic information theory and martingales*, `arxiv:0906.2614`.

[5] Borel, Émile, Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico de Palermo*, 27:247–270, 1909.

[6] Borel, Émile, *Probabilité et certitude*, Presses Univ. de France, Paris, 1950.

[7] Cavaillès, Jean, Du collectif au pari. *Revue de métaphysique et de morale* 47:139–163, 1940.

[8] Calude, Cristian S., *Information and Randomness. An Algorithmic Perspective.* Springer-Verlag, 1994. Second edition, 2002.

[9] Chaitin, Gregory J., On the length of programs for computing finite binary sequences, *Journal of the ACM* 13:547–569, 1966.

[10] Chaitin Gregory J., On the length of programs for computing finite binary sequences: statistical considerations, *Journal of the ACM*, 16:145–159, 1969.

[11] Chaitin, Gregory J., Computational complexity and Gödel's incompleteness theorem, *ACM SIGACT News* 9: 11–12, April 1971.

[12] Chaitin, Gregory J., A theory of program size formally identical to information theory, *Journal of the ACM* 22:329–340, 1975. Received April 1974; revised December 1974.

[13] Chaitin, Gregory J., Algorithmic information theory. Some recollections. 25 May 2007, `http://www.cs.auckland.ac.nz/~chaitin/60.html`

[14] Church, Alonzo, On the concept of a random sequence. *Bull. Amer. Math. Soc.* 46(2):130–135, 1940.

[15] Bru, Bernard, and Pierre Crépel (eds), *Condorcet: Arithmétique politique; Textes rares ou inédits (1767–1789).* Institut National d'Études Démographiques, Paris, 1994.

[16] Copeland Sr., Arthur H., Admissible numbers in the theory of probability, *American Journal of Mathematics* 50:535–552, 1928.

[17] Davis, Martin (ed), *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, Raven Press, New York, 1965.

[18] de Finetti, Bruno, *Compte rendu critique du colloque de Genève sur la théorie des probabilités*, Hermann, Paris, 1939. This is the eighth installment (number 766) of [89].

[19] Dierker, E., and K. Siegmund (eds), *Karl Menger: Ergebnisse eines Mathematischen Kolloquiums*. Contributions by J.W. Dawson jr., R. Engelking, W. Hildenbrand, foreword by G. Debreu, and afterword by F. Alt, Springer, 1998.

[20] Doob, Joseph, Note on probability. *Annals of Mathematics* 37(2):363–367, 1936.

[21] Downey, Rod G., and Hirschfeldt, Denis, *Algorithmic Randomness and Complexity*, Springer, to appear in 2010. ISBN: 978-0387955674.

[22] Fisher, R. A., *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinburgh, 1956.

[23] Fréchet, Maurice, Définition de l'intégrale sur un ensemble abstrait. *Comptes rendus* 160: 839–840, 1915.

[24] Fréchet, Maurice, Exposé et discussion de quelques recherches récentes sur les fondements du calcul des probabilités. Pp. 23–55 of the second installment (No. 735) of [89], 1938.

[25] Gács, Peter, On the symmetry of algorithmic information, *Soviet Math. Doklady* 15(5):1477–1480, 1974. Original: Гач, Петер, О симметрии алгоритмической информации. *Доклады Академии наук СССР* 218(6):1265–1267, 1974. Submitted April 9, 1974.

[26] Gács, Peter, Review of *Algorithmic Information Theory* by Gregory J. Chaitin, *The Journal of Symbolic Logic* 54(2):624–637, June 1989.

[27] Gács, Peter, J. Tromp, and Paul Vitányi, Algorithmic statistics. *IEEE Transactions on Information Theory* 47(6):2443–2463, 2001.

[28] Goldreich, Oded, *An Introduction to Cryptography*. Vol. 1. Cambridge University Press, 2001.

[29] Kolmogorov, Andrei N., *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, 1933.

[30] Kolmogorov, Andrei N., On tables of random numbers, *Sankhyā: The Indian Journal of Statistics*, Series A, 25(4):369–376, 1963.

[31] Kolmogorov, Andrei N., Three approaches to the quantitative definition of information, *Problems of Information Transmission* 1:1-7, 1965. Original: Колмогоров, Андрей Николаевич, Три подхода к определению понятия "количество информации", *Проблемы передачи информации*, 1(1):3–11, 1965.

[32] Kolmogorov Andrei N., Logical basis for information theory and probability theory, *IEEE Trans. Inform. Theory* 14:662–664, 1968. Footnote: "Manuscript received December 13, 1967. The work is based on an invited lecture given at the International Symposium on Information Theory, San Remo, Italy, September, 1967. Translation courtesy of AFOSR, USAF. Edited by A.V.Balakrishnan." Russian version: Колмогоров, Андрей Николаевич, К логическим основам теории информации и теории вероятностей, *Проблемы передачи информации* 5(3):3–7, 1969.

[33] Kolmogorov, Andrei N., Combinatorial foundations of information theory and the calculus of probabilities. *Russian Mathematical Surveys*, 38:29–40, 1983.

[34] Levin, Leonid A., Some syntactic theorems on the calculus of finite problems of Ju. T. Medvedev, *Soviet Math. Doklady* 10(2):288–290, 1969. Original: *Доклады Академии наук СССР* 185(1):32–33, 1969.

[35] Levin, Leonid A., On the notion of a random sequence, *Soviet Math. Doklady* 14(5):(1413–1416), 1973. Original: Левин, Леонид Анатольевич, О понятии случайной последовательности, *Доклады Академии наук СССР* 212(3):548–550, 1973. Submitted July 1, 1972.

[36] Levin, Leonid A., Universal sequential search problems, *Problems of Information Transmission*, 9(3):265–266, 1973. Original: Левин, Леонид Анатольевич, Универсальные задачи перебора. *Проблемы передачи информации*, 9(3):115–116, 1973. Submitted June 7, 1972.

[37] Levin L.A., Laws of information conservation (nongrowth) and aspects of the foundation of probability theory, *Problems of Information Transmission* 10:206–210, 1974. Original: Левин, Леонид Анатольевич, Законы сохранения (невозрастания) информации и вопросы обоснования теории вероятности. *Проблемы передачи информации* 10(3):30–35, 1974. Submitted Jan. 9, 1974, sent to the printer August 1974.

[38] Levin, Leonid A., Various measures of complexity for finite objects (axiomatic description), *Soviet Math. Doklady* 17(2):522–526, 1976. Original: Левин, Леонид Анатольевич, О различных мерах сложности конечных объектов (аксиоматическое описание). *Доклады Академии наук СССР* 227(4):804–807, 1976. Submitted: June 7, 1975.

[39] Levin, Leonid A., Randomness conservation inequalities: Information and independence in mathematical theories. *Information and Control* 61:15–37, 1984.

[40] Levin, Leonid A., Interview with Alexander Shen (June 2008, unpublished).

[41] Li, Ming, and Paul Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, Springer, New York, 3rd edition, 2008.

[42] Lieb, Elliot H., Daniel Osherson, and Scott Weinstein, *Elementary Proof of a Theorem of Jean Ville*, see `arxiv:cs/0607054v1` at `arxiv.org`.

[43] Lutz, Jack H., Dimension in Complexity Classes, *SIAM Journal on Computing* 32:1236–1259 (2003).

[44] Марков, Андрей Андреевич, О нормальных алгорифмах, связанных с вычислением булевых функций. *Известия Академии наук СССР, серия математическая* 31:161–208, 1967.

[45] Martin-Löf, Per (Мартин-Лёф П.) О понятии случайной последовательности. (On the notion of a random sequence.) *Теория вероятностей и её применения* 11(1):198–200, 1966.

[46] Martin-Löf, Per, The definition of random sequences, *Information and Control* 9(6):602–619, 1966.

[47] Martin-Löf, Per, *Algorithmen und zufällige Folgen. Vier Vorträge von Per Martin-Löf (Stockholm) gehalten am Mathematischen Institut der Universität Erlangen-Nürnberg.* Erlangen, 1966. This document, dated 16 April 1966, consists of notes taken by K. Jacobs and W. Müller from lectures by Martin-Löf at Erlangen on April 5, 6, 14, and 15. There are copies in several university libraries in Germany and the United States.

[48] Martin-Löf, Per, The literature on von Mises' Kollektivs revisited, *Theoria*, 35(1):12–37, 1969.

[49] Martin-Löf, Per, Notes on constructive mathematics. Stockholm, Almqvist & Wiksell, 1970.

[50] Martin-Löf, Per, Complexity oscillations in infinite binary sequences, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 19:225–230, 1971.

[51] Martin-Löf, Per, private communication to Alexander Shen, 2008.

[52] Menger, Karl, The formative years of Abraham Wald and his work in geometry, *Annals of Mathematical Statistics* 23:14–20, 1952.

[53] von Mises, Richard, Grundlagen der Wahrscheinlichkeitsrechnung, *Mathematische Zeitschrift* 5(191):52–99, 1919.

[54] von Mises, Richard, *Wahrscheinlichkeit, Statistik und Wahrheit*, Vienna: Springer-Verlag, 1928.

[55] von Mises, Richard, *Wahrscheinlichkeitsrechnung und ihre Anwendungen in der Statistik und der theoretischen Physik*: Leipzig and Vienna: F. Deuticke, 1931.

[56] von Mises, Richard, Quelques remarques sur les fondements du calcul des probabilités, Pp. 57–66 of *Les fondements du calcul des probabilités*, the second installment (number 735) of [89], 1938.

[57] von Mises, Richard, On the foundations of probability and statistics, *Annals of Mathematical Statistics* 12:191–205, 1941.

[58] von Mises, Richard, and J. L. Doob, Discussion of papers on probability theory. *Annals of Mathematical Statistics* 12:215–217, 1941.

[59] von Mises, Richard, *Mathematical Theory of Probability and Statistics*. (Edited and complemented by Hilda Geiringer.) New York and London: Academic Press, 1964.

[60] Morgenstern, Oskar, Abraham Wald, 1902–1950. *Econometrica* 19(4):361–367, 1951.

[61] Nies, Andre, *Computability and Randomness* (Oxford Logic Guides). Oxford University Press, 2009. ISBN: 978019923076.

[62] Parthasarathy, K. R., Obituary of Andrei Kolmogorov, *Journal of Applied Probability* 25(1)1:445–450, March 1988.

[63] von Plato, Jan, *Creating Modern Probability*, Cambridge University Press, 1994.

[64] , Popper, Karl R., *Logik der Forschung: Zur Erkenntnistheorie der modernen Naturwissenschaft*, Vienna, Springer, 1935. An English translation, *The Logic of Scientific Discovery*, with extensive new appendices, was published by Hutchinson, London, in 1959.

[65] Reichenbach, Hans, Axiomatik der Wahrscheinlichkeitsrechnung, *Mathematische Zeitschrift* 34:568–619, 1932.

[66] Schnorr, Claus-Peter, A unified approach to the definition of random sequences, *Mathematical Systems Theory* 5(3):246–258 (1971).

[67] Schnorr, Claus-Peter, *Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Begründung der Wahrscheinlichkeitstheorie*, Springer, 1971.

[68] Schnorr, Claus-Peter, The process complexity and effective random tests, *Proceedings of the fourth annual ACM symposium of Theory of Computing, Denver, Colorado, May 1–3, 1972.* Journal version: Process complexity and effective random tests, *Journal of Computer and System Sciences* 7:376–388, 1973.

[69] Schnorr, Claus-Peter, A survey of the theory of random sequences, In: Butts R.E., Hintikka J., eds., *Basic problems in methodology and linguistics*, Dordrecht, D. Reidel, pp. 193–211, 1977.

[70] Schnorr, Claus-Peter, A talk during Dagstuhl seminar 06051, 29 January – 3 February 2006, `http://www.hutter1.net/dagstuhl/schnorr.mp3`.

[71] Shafer, Glenn, and Vladimir Vovk. *Probability and Finance: It's Only a Game!* Wiley, New York, 2001.

[72] Shafer, Glenn, and Vladimir Vovk, The sources of Kolmogorov's *Grundbegriffe*. *Statistical Science* 21(1):70–98, 2006. A longer version is at www.probabilityandfinance.com as Working Paper 4.

[73] Shasha, Dennis, and Cathy Lazere, *Out of their Minds: The Lives and Discoveries of 15 Great Computer Scientists*, Springer, New York, 1998.

[74] Shen, Alexander, *Algorithmic Information theory and Kolmogorov complexity*, Uppsala Universitet, Technical Report 2000-034, `http://www.it.uu.se/research/publications/reports/2000-034`.

[75] Shen, Alexander. *Algorithmic information theory and foundations of probability*, `arxiv:0906.4411`.

[76] Siegmund-Schultze, Reinhard, A non-conformist longing for unity in the fractures of modernity: Towards a scientific biography of Richard von Mises (1883–1953). *Science in Context* 17(3):333–370, 2004.

[77] Siegmund-Schultze, Reinhard, Mathematicians forced to philosophize: An introduction to Khinchin's paper on von Mises' theory of probability. *Science in Context* 17(3):373–390, 2004. A translation of Khinchin's paper into English, by Oscar Sheynin, follows on pp. 391–422.

[78] Solomonoff, Ray J., A formal theory of inductive inference, Part I, *Information and Control* 7(1):1–22, March 1964.

[79] Solomonoff, Ray J., A formal theory of inductive inference, Part II, *Information and Control* 7(2):224–254, June 1964.

[80] Solomonoff, Ray J., Complexity-based induction systems: Comparisons and convergence theorems. *IEEE Transactions on Information Theory* IT-24(4):422–432, July 1978.

[81] Ville, Jean-André, Sur la notion de collectif. *Comptes rendus* 203:26–27, 6 July 1936.

[82] Ville, Jean, *Étude critique de la notion de collectif*, Gauthier-Villars, Paris, 1939.

[83] Vovk, Vladimir, and Glenn Shafer, Kolmogorov's contributions to the foundations of probability. *Problems of Information Transmission* 39:21–31. 2003.

[84] Wald, Abraham, Sur la notion de collectif dans le calcul des probabilités. *Comptes rendus* 202:180–183, 20 January 1936.

[85] Wald, Abraham, Die Wiederspruchsfreiheit des Kollektivbegriffes der Wahrscheinlichkeitsrechnung, *Ergebnisse eines Mathematischen Kolloquiums* 8:38–72, 1937. Reprinted in [19]

[86] Wald, Abraham, Die Widerspruchfreiheit des Kollectivbegriffes, Pp. 79–99 of *Les fondements du calcul des probabilités*, the second installment (number 735) of [89], 1938.

[87] Wald, Abraham, Review of *Probability, Statistics and Truth*, by Richard von Mises. *Journal of the American Statistical Association* 34(207):591–592, 1939.

[88] Wang, Yongge, A separation of two randomness concepts, *Information Processing Letters*, 69(3):115–118, 1999.

[89] Wavre, Rolin, ed. *Colloque consacré à la théorie des probabilités*, Paris, Hermann, 1938–1939. Proceedings of a conference held in Geneva in October 1937, published as eight installments in Hermann's *Actualités Scientifiques et Industrielles*. The first seven appeared in 1938 as numbers 734 through 740; the eighth, de Finetti's summary of the colloquium, appeared in 1939 as number 766 [18].

[90] Wolfowitz, Jacob, Abraham Wald, 1902–1950, *Annals of Mathematical Statistics* 23(1):1–13, 1952.

[91] Zvonkin, A. K., and L. A. Levin, The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Math. Surveys*, 25(6):83–124, 1970. Original: Звонкин, Александр Калманович; Левин, Леонид Анатольевич, Сложность конечных объектов и обоснование понятий информации и случайности с помощью теории алгоритмов. *Успехи математических наук* 25(6):85–127, 1970.