What Is a Random Sequence?

Sérgio B. Volchan

1. INTRODUCTION. What is randomness? Are there random events in nature? Are there laws of randomness?

These old and deep philosophical questions still stir controversy today. Some scholars have suggested that our difficulty in dealing with notions of randomness could be gauged by the comparatively late development of probability theory, which had a somewhat hampered development [**20**], [**21**]. Historians generally agree upon the year 1654 as a convenient landmark for the birth of mathematical probability. At that time, some reasonably well-articulated ideas on the subject were advanced in the famous correspondence of Pascal and Fermat regarding the division of stakes in certain games of chance. However, it was only in 1933 that a universally accepted axiomatization of the theory was proposed by A. N. Kolmogorov [**28**], with many contributions in between. That is, almost three hundred years after its beginnings, and a hundred years after Cauchy's work on the rigorization of analysis, probability theory finally reached maturity. It achieved the status of an autonomous discipline of pure mathematics, instead of being viewed as a mixed branch of applied mathematics and physics.

In contrast, the *uses* of notions of randomness are as old as civilization itself. It appeared in a variety of games of chance (coin-tossing, dice, etc.), as well as in divination, decision-making, insurance, and law. Many reasons for this discrepancy between theory and application have been put forward. One suggestion is that a full development of the theory, going beyond combinatorics, had to wait for the creation of the very sophisticated mathematical tools and concepts of set theory and measure theory. A more plausible reason could be that our cognitive (and even psychological) constitution, which might have evolved to look for patterns and trends even where there are none, is not well suited to grasp randomness.[1]

In support of that last idea, many psychological studies have shown that people (even experts) perform poorly when using intuition to deal with randomness [**2**]. One classical example[2] is the 'gambler's fallacy': the common (false) belief that, after a sequence of losses in a game of chance, there will follow a sequence of gains, and vice versa, in a kind of self-compensation.

What are the characteristics usually associated with randomness? A common idea is to identify randomness with *unpredictability*. This intuition originates in people's experience with games of chance. For example, a sequence of coin tosses looks very irregular, and no matter how many times we've tossed the coin, say a thousand times, no one seems to be able to predict the outcome of the next toss. That arguably explains the widespread use of randomizing devices, like coins, dice, and bones, to guarantee *fairness* in gambling[3] and decision-making.

However, one could question whether these are examples of "really" random phenomena. After all, actual coin-tossing (for example) is a purely mechanical process, governed therefore by Newton's laws of motion. Hence, its outcome is as predictable,

[1] The mathematician Emile Borel claimed the human mind is not able to simulate randomness [**34**].

[2] A more subtle one is the Monty Hall problem (see Snell and Vanderbei [**32**]).

[3] The development of the mathematical analysis of games of chance seems to have been motivated not only by the desire to devise winning strategies for the games but also by the desire to detect fraud in them [**4**].

in principle, as the motion of the planets, once the initial conditions are given.[4] The observed unpredictability results from a peculiar combination of circumstances [**24**], [**49**]. First, there is a kind of "instability" built into the system, of the kind usually associated with meteorological systems; i.e., it is a dynamical system displaying sensitive dependence on (some set of) initial conditions. That, coupled with *our* inability to know these conditions with infinite precision, results in unpredictability *in practice*, even though the process is totally lawful in the sense of classical mechanics. In other words, we have an instance of the phenomenon of "chaos."

It is reasonable to ask whether there are "intrinsic" (or ontological) notions of randomness. The usual suggestion is the notion of "lawlessness," also conceived of as "disorder," "irregularity," "structurelessness," or "patternlessness," that we'll discuss later. It certainly includes unpredictability in some sense. But one needs to be careful here. To begin with, one needs to distinguish between "local" irregularity versus "global" (or statistical) regularities observed in many chance phenomena [**43**], [**42**]. For example, although we cannot predict the outcome of *individual* coin tosses, it is an empirical fact that the proportion of heads (or tails) obtained after a *great number* of tosses seems to converge to or stabilize around 0.5. Besides, there is a recurrent sense of paradox lingering in the enterprise of looking for laws that govern lawlessness [**11**]: after all, this last property seems to mean exactly the absence of any subjugation to laws.

Concerning randomness in natural phenomena, it is not quite clear what one should look for. Conceivably, some quantum mechanical phenomenon, like radioactive decay [**23**], would be a good candidate to investigate. In this paper we won't discuss this very important topic. We will focus instead on the admittedly less ambitious but more manageable question of whether it is possible at least to obtain a *mathematically* rigorous (and reasonable) definition of randomness. That is, in the hope of clarifying the concept of chance, one tries to examine a mathematical model or idealization that might (or might not) capture some of the intuitive properties associated with randomness. In the process of refining our intuition and circumscribing our concepts, we might be able to arrive at some fundamental notions. With luck (no pun intended), these might in turn furnish some insight into the deeper problems mentioned. At the very least it could help one to discard some previous intuitions or to decide upon the need for yet another mathematical model.

The history of mathematics shows that this strategy is frequently fruitful. An example of this process was the clarification of the concept of 'curve'. Not only did it lead to the discovery of "pathological curves" (which are interesting mathematical objects in themselves, linked to fractals and Brownian motion) but also to the realization that smoothness is a reasonable requirement in the formalization of the intuitive notion of curve [**19**]. Another example, which is central to our discussion, was the clarification of the intuitive notion of computability (see the next section).

Of course, this is not an easy task. The proposed model or idealization should be simple, without also being totally trivial. One idea is to consider an abstraction of the coin-tossing experiment, the so-called Bernoulli trials. Representing the occurrence of heads by 0 and tails by 1, we associate a *binary string* to each possible outcome of a successive coin-tossing experiment. We then ask: When is a binary string random?

To appreciate the difficulties involved, let's examine the "paradox of randomness" [**14**]. It goes like this. Suppose you toss an honest coin repeatedly, say twenty-three times. Consider the following outcomes:

[4] The mathematician and former magician Persi Diaconis was able consistently to get ten consecutive heads in coin-tossing by carefully controlling the coin's initial velocity and angular momentum.

- 00000000000000000000000
- 01101010000010011110011
- 11011110011101011111011.

The first result is generally considered suspect, while the second and third "look" random. However, according to probability theory all three outcomes, and in fact all the 2^{23} possible outcomes, have the same probability of $1/2^{23}$. Why, then, do the last two outcomes *seem* random while the first does not?

It is conceivable that the ultimate reason for that perception "belongs to the domain of psychology" [**29**], to be found in the structure of our visual-cognitive apparatus. Such issues notwithstanding, the question is whether it is possible to distinguish random from nonrandom strings in a mathematically meaningful way. Note that our intuition cannot be trusted much in this task. It's enough to observe that the second string above consists of the first twenty-three digits of the binary expansion of $\sqrt{2} - 1$. So, although it "looks" random, in the sense of exhibiting no obvious pattern, its digits were obtained by a process (root extraction) that, by all reasonable standards, is *not* random. Note the overall similarity with the third string, obtained by coin-tossing.

For strings it is only possible to develop a notion of *degrees of randomness*, there being no sharp demarcation of the set of all strings into random and nonrandom ones [**7**]. In fact, once a certain binary string with m zeroes is considered random, there is no reason not to consider equally random the string obtained by adding (or subtracting) one more zero to it (or from it).

The situation becomes clearer if one considers instead the set of all *infinite* binary strings, or *sequences of bits*. Although in real life applications we are bound to encounter only *finite*, albeit very long, strings, it is nevertheless worth considering this further idealization. The idea of taking infinite objects as approximations to finite but very large ones is not new. For example, in equilibrium statistical mechanics, in order to have a sharp notion of a phase transition one has to work in the so-called thermodynamic limit, in which the number of particles tends to infinity (as does the volume, but in such a way that particle density remains constant).[5] The great advantage of working with sequences is that they are easier to handle mathematically. This curious and common state of affairs is probably a result of treating a completed infinity as one whole (though large) object, instead of having to keep track of a large (but finite) number of objects (which makes combinatorics such a difficult craft). In particular, it is possible to obtain a *sharp* result, that is, to write $\{0, 1\}^{\mathbb{N}} = \mathcal{R} \cup \mathcal{R}^c$, decomposing the set of sequences into random and nonrandom ones.

But now the question becomes: What does it mean to say that an *individual* infinite sequence of 0s and 1s is random? Historically, three main notions were proposed[6]:

- *stochasticness* or *frequence stability*, due to von Mises, Wald, and Church;
- *incompressibility* or *chaoticness*, due to Solomonoff, Kolmogorov, and Chaitin;
- *typicality*, due to Martin-Löf.

Interestingly, all these proposals ended up involving two notions apparently foreign to the subject of randomness: *algorithms* and *computability*. With hindsight, this is not totally surprising. In a sense to be clarified as we proceed, randomness will be closely associated with "noncomputability."

[5] As the late mathematical-physicist R. Dobrushin noted, infinity is a better approximation to Avogadro's number 6.0×10^{23} than to the number 100.

[6] In this paper we only examine these. For some recent developments, see Ambos-Spiess and Kucěra [**1**]. Comprehensive discussions can be found in the monographs [**7**], [**33**] or the reviews [**41**], [**36**], [**18**].

Let $\Sigma = \{0, 1\}$, and let Σ^* be the set of all strings (finite words) of 0s and 1s, including the empty string Λ. Call $\Sigma^{\mathbb{N}}$ the corresponding set of all binary sequences. A preliminary observation is that any reasonable notion of a random sequence makes sense only with respect to a *given* probability distribution on the set $\Sigma^{\mathbb{N}}$ [**41**]. In fact, a sequence with twice as many 0s as 1s would not be considered random if each digit occurs with probability $p = q = 1/2$ (honest coin), but it could be random if $p = 1/3$ and $q = 2/3$. Thus, although originally motivated by foundational issues, the following discussion *presupposes* the usual measure-theoretic probability theory.

In the following we deal with Bernoulli(p, q) probability measures on $\Sigma^{\mathbb{N}}$, where $p + q = 1$. These are product measures specified by their action on strings: to wit, the probability of the string $x_1 \dots x_k$ equals $p^m q^{k-m}$, where m is the number of 1s and $k - m$ the number of 0s in it. Also, we will sometimes identify a real number in $[0, 1]$ with a sequence in $\Sigma^{\mathbb{N}}$ through its binary expansion, $x = 0.x_1 x_2 \dots$.[7] Then the Bernoulli$(1/2, 1/2)$ product measure corresponds to the uniform measure on $[0, 1]$, that is, to Lebesgue measure λ.

2. THE LOGIC CONNECTION: ALGORITHMS AND COMPUTABILITY. If randomness is to be conceived of as lawlessness, then one has to respond to the query: What is a "law"? This is too broad a question, and we'll focus on the (necessarily narrow)[8] mathematical notion of 'law of formation', say, that generates the successive digits of a binary sequence. To clarify what that could mean, we make a short digression to recall some old controversies in the foundations of mathematics [**30**].

According to the "constructivist" school[9] mathematical objects exist insofar as we are able to construct them in some way. In particular, every infinite structure should be given by some method telling how to construct it. In contrast, from Hilbert's "formalist" point of view existence corresponds to absence of contradictions, i.e., to consistency in a formal axiomatic system.

In the context of his investigations on the nature of the real numbers, Borel (1909) suggested that we only have access to real numbers (in fact, to any mathematical object) that are specifiable or describable in a finite number of words (in some language). It was known that this notion led to the Richard-Berry paradox.[10] In fact, Borel used this paradox and a Cantor-like diagonal argument to conclude that the set of finitely describable reals, though the only ones accessible to us, are not *effectively enumerable*, meaning that we cannot decide whether or not a given finite description defines a real number [**47**]. Unfortunately, he didn't develop a rigorous notion of effectiveness. Like most mathematicians at the time, he used the *intuitive* notion of effective procedure, conceived of as some kind of finite, step-by-step recipe or prescription that, if duly followed, arrives at a desired result (e.g., solves a given problem). A rigorous notion of an effective procedure or algorithm (and of computability) appeared only in the 1930s, as the culmination of foundational investigations in mathematical logic, being one of the great achievements of modern mathematical science [**15**].

The need for a clarification of the concept of algorithm, a notion deeply entrenched in mathematical practice [**26**], was only gradually felt. Properly speaking, it is a *metamathematical* concept, and it came to the fore with the creation by David Hilbert of

[7] Except for $x = 1$, for which we take the expansion $.111\dots$, we choose the expansion ending in an infinite sequence of 0s whenever x is a dyadic rational.

[8] So as to avoid the discussion of the concept of "natural law" in sciences like physics.

[9] Which included Gauss, Kronecker, Lebesgue, Borel, Brouwer, and Weyl, to cite a few.

[10] A version of it goes like this: Define a natural number as "the least number that cannot be described in less than twenty words." Does this number exist? Any answer leads to a contradiction.

the new discipline of *metamathematics*.[11] In his famous list of twenty-three problems proposed at the Second International Congress of Mathematics in Paris (1900), problem number ten, though not mentioning the word algorithm, asked for a "procedure" that would decide in a finite number of steps whether a given Diophantine equation[12] has an integer solution. Almost three decades later, in 1928, Hilbert and Ackermann formulated the *Entscheidungsproblem* (or Decision Problem) for first order logic, "the principal problem of mathematical logic."[13] If someone suspected that such problems had a negative solution (which, by the way, was not the case for Hilbert) then, in order to prove it so, one would need to be able to survey the class of allowed effective procedures. Only then would it be meaningful to say that a given (class of) problems has no effective solution.

Around 1936, due to the efforts of such logicians as Gödel, Church, Kleene, Post, Markov, and Turing, many apparently different candidates for "the" adequate notion of algorithmic or effective procedures were available [**22**]. Probably the simplest was the one proposed by Turing in 1936, which we describe next.

Turing's model of computation is an idealization based on his analysis of the steps performed by a human calculator (a 'computor'). It consists of an infinite one-dimensional tape (i.e., there are no memory limitations), equally divided into cells, and of a control head with a cursor capable of moving along the tape. Suppose that each cell can contain only the symbols 0, 1, or □ (blank). The set of tape symbols is $S = \{0, 1, \Box\}$, while the set of input symbols is $\Sigma = \{0, 1\}$.

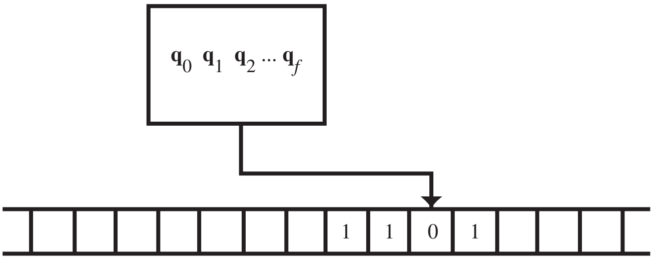

Figure 1. A Turing machine.

The control head acts through the read/write cursor that scans one cell at a time. At each given discrete instant, the control head can be in any one of a finite number of *internal states* belonging to a set $Q = \{q_0, q_1, \dots, q_f\}$, among which there are the special *initial* state q_0 and the *final* (or terminating) state q_f. Depending on the symbol being scanned and the state the control head is in, the cursor then writes a tape-symbol on the cell, moves one cell either to the right (R) or left (L), after which the control jumps to another state.

Hence, at each time, the machine's operation is completely determined by the current scanned symbol and the current state of the control head. That is, each *step* consists in the execution of a quintuple $(q, s; s', q', m)$, in the following sense. If q in Q is the current state and s in S is the symbol scanned, then s' in S is then printed on the cell, the control jumps to the new state q' in Q, the cursor moving to the next cell to

[11] The application of mathematical reasoning and methods to formal objects like formulas, axiom systems, and proofs, which themselves become the target of mathematical investigation.

[12] An equation of general form $P(z_1, \dots, z_n) = 0$, where P is a polynomial in n variables with integer coefficients.

[13] It asks for a general procedure that decides, in a finite number of steps, whether a formula of first order logic is or is not a theorem.

the right or left, signified by m in $\{L, R\}$. The head is now scanning a new cell and is in another state, so it consults the corresponding instruction and the cycle begins anew.

So we can define a *Turing machine* (TM) as a list of quintuples, that is, in terms of its *instructions*. More formally, a Turing machine is a function with domain in $Q \times S$ and range in $Q \times S \times \{0, 1\}^*$. An important consequence is that there are only *countably many* Turing machines. In fact, one codifies the instructions by an alphabet (say, binary). As the set of instructions for each machine is always finite, one can then order the machines by the increasing size of the instruction set (number of symbols in its codification) and, for each size, by lexicographic order. In this way we get an enumeration $TM_1, TM_2, \ldots$ of all Turing machines.

A *computation* by a Turing machine consists of the following. Initially, an input x from $\{0, 1\}^*$ is written on the tape. To the left and right of it, all cells are blank. The cursor is positioned at the leftmost symbol of x, and the control state is set at the initial state q_0. From there the machine follows the set of instructions. If the machine eventually reaches the final state q_f, the computation halts and the output is the string $TM(x)$ in $\{0, 1\}^*$ left on the tape (the computation "converges"). Otherwise, the machine goes on forever (the computation "diverges"). Therefore, each Turing machine defines a *partial* function from the set of strings $\{0, 1\}^*$ to itself.[14] In other words, a computation is nothing more nor less than *symbol processing*.

Now each string in $\{0, 1\}^*$ is a binary representation of a positive integer through an encoding function $e : \{0, 1\}^* \to \mathbb{N}$. A partial function $f : \mathbb{N} \to \mathbb{N}$ is said to be *Turing computable* if there is a Turing machine TM such that, for every n in the domain of f, there is an input w in $\{0, 1\}^*$ with $n = e(w)$ for which the machine eventually stops and such that the output $TM(w)$ satisfies $f(n) = e(TM(w))$.[15] From the countability of the collection of all Turing machines, it follows immediately that the set of (partial) Turing computable functions is a countable subset of the uncountable set of all partial functions from $\mathbb{N}$ to $\mathbb{N}$. In this sense, very few functions are computable.

Another important result that emerged from Turing's paper was the proof that there are (infinitely many) *universal Turing machines*, i.e., machines that can *simulate* the operation of any other Turing machine. The input of such a universal machine U consists of an adequate encoding of the instruction set of the machine TM that we want to simulate, followed by its input w. The output of U is then $TM(w)$ [**35**].

Strictly speaking, we have assumed here that the input to a Turing machine consists of the *data* and the *program*, suitably codified into one block of bits. Alternatively, one could think of the data and program coming separately. Such an approach entails modifying the machine by introducing a data (or work) tape and a program tape. In this case the computer defines a partial binary function $\phi : \{0, 1\}^* \times \{0, 1\}^* \to \{0, 1\}^*$. Moreover, as will become clear later, it is convenient to consider a restricted class of Turing machines called *prefix-free* machines.

A language[16] $\mathcal{L}$ is *prefix-free* if and only if no string of $\mathcal{L}$ is a prefix of another string of $\mathcal{L}$. So, for example, a binary prefix-free language cannot have both strings 10111 and 101110101 as words. A prefix-free machine is a Turing machine such that, whenever $\phi(p, q)$ is defined (the computation with program p and input q converges) and the string p is a prefix of string p' with $p \neq p'$, then $\phi(p', q)$ is not defined. In

[14] In case the function is defined for *all* strings in $\{0, 1\}^*$, then the function is said to be *total*.

[15] Also, a subset A of $\mathbb{N}$ is said to be *recursive* if its characteristic function is Turing computable. This means that there is an algorithm to decide whether or not an element belongs to A. On the other hand, if we only require that A could be effectively counted, then it is said to be a *recursively* or *effectively enumerable* (r.e.) set. More formally, A is recursively enumerable when it is the range of a computable function.

[16] I.e., a collection of strings (words) over a finite alphabet. For example, any subset of $\{0, 1\}^*$ is a language over $\{0, 1\}$.

other words, it defines a computable function $\phi : \{0,1\}^* \times \{0,1\}^* \to \{0,1\}^*$ such that, for all q in $\{0,1\}^*$, the function $\phi_q : \{0,1\}^* \to \{0,1\}^*$ given by $\phi_q(p) = \phi(p,q)$ has a prefix-free domain. To satisfy this requirement, it suffices that the machine have a finite program tape and a read-only cursor able to move exclusively to the right (see Figure 2). The usual properties of the previous machines are preserved. In particular, the existence of a universal prefix-free machine is assured.

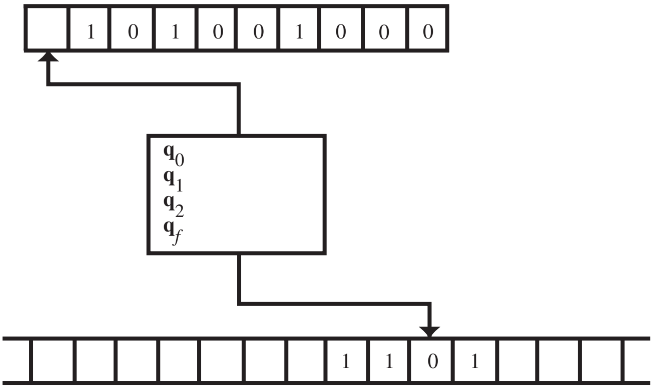

Figure 2. A prefix-free Turing machine.

As mentioned, there were alternative suggestions put forward as the adequate formalization of the intuitive notion of computability: Church's lambda-calculus, Kleene's general recursive functions, Post's automata, and Markov algorithms. Very soon, however, it was proved that all those apparently distinct notions were in fact equivalent, that is, they defined the same set of functions: namely, the *computable* or *partial recursive* functions [**40**].

This, and other factors, led to the daring suggestion that the "right" mathematically precise concept of computability had been found. This is the content of the famous

Church-Turing Thesis. The class of intuitively computable functions coincides with the class of Turing computable functions.

Note that this is not really a "thesis" awaiting a proof. It is more like a rigorous definition proposed for the intuitive (nonrigorous) notion of computability. So, when people started to refer to a method or procedure being *effective*, *algorithmic*, or *mechanical*,[17] it was intended to mean that it can be implemented by a Turing machine.

Using his concept of effective procedures, Church (1936) was able to give a negative answer to Hilbert's *Entscheidungsproblem*. Shortly afterwards, Turing used a diagonal argument to prove the undecidability of the Halting Problem and, by reducing the Decision Problem to it, was also able to prove its undecidability.

Now, we have earlier made the suggestion that random sequences would be those sequences that are "lawless." If by that expression we mean that they should present no nontrivial regularities or patterns *whatsoever*, then *no* sequence would be random. In fact, a theorem of van der Waerden [**8**] asserts that in *every* binary sequence one of the two symbols must occur in arithmetical progressions of every length. Therefore, such a broad concept of "law" doesn't work.

[17] The use of the adjective "mechanical" can be misleading. It intends to convey the idea of a routine process that can be carried out without the need of ingenuity, but not necessarily in the sense of physical mechanics.

Suppose that we restrict the notion of law to that of a "rule of formation" that gives the consecutive digits of the sequence. Given the discussion above, we are led by the Church-Turing thesis to the idea of an effective rule, as defined by a computable function. Hence, we are naturally led to conceive of lawlessness as the *absence of computable regularities*. We can now see, by using a cardinality argument, that this proposal is inadequate as well. For, to get the set of lawless sequences we would need to discard from $\Sigma^{\mathbb{N}}$ all the lawful ones. The latter, however, constitute at most a denumerable set of sequences. The resulting set of candidate random sequences is too large. In fact, consider the set of sequences such that $x_{2n} = x_{2n+1}$ for all $n \geq 1$. Such sequences are too "locally" ordered to be random, but because there are uncountably many of them, some would have been included in the set of random sequences.[18]

3. RANDOMNESS AS STOCHASTICNESS.

The sixth problem in Hilbert's famous list (1900) is the following [**13**]:

> The investigations on the foundations of geometry suggest the problem: To treat in the same manner, by means of axioms, those physical sciences in which mathematics plays an important part; first of all, the theory of probability and mechanics.

Note that Hilbert considered probability theory to be a branch of physics. Although a commonly held view at the time, this is somewhat surprising coming from a formalist. In fact, the whole point of an axiomatic formulation is to highlight the formal aspects of a mathematical theory, irrespective of its initial motivations and/or further interpretations and applications [**6**].

In 1919 the physicist Richard von Mises proposed to develop the theory of probability as part of physics [**45**]. In a very influential work[19] he suggested that the theory should be based on the notion of random sequences, which he called "collectives" (*Kollectivs*, in German).[20] The basic insight was the global statistical regularity observed in random experiments like coin-tossing, namely, *frequency stability*.

Definition 3.1. An infinite binary sequence $x = x_1 x_2 \ldots$ is random if it is a *collective*; i.e., if it has the following two properties:

I. Let $f_n = \sharp\{m \leq n : x_m = 1\}$ be the number of 1s among the first n terms in the sequence. Then

$$\lim_{n\to\infty} \frac{f_n}{n} = p$$

exists and $0 < p < 1$.

II. If $\Phi : \{0,1\}^* \longrightarrow \{0,1\}$ is an *admissible* partial function (i.e., a *rule for the selection* of a subsequence of x such that x_n is chosen precisely when $\Phi(x_1 x_2 \ldots x_{n-1}) = 1$), then the subsequence $x_{n_1} x_{n_2} \ldots$ so obtained has Property I for the same p.

[18] Alternatively, not every lawless sequence would satisfy the Law of Large Numbers, which is taken to be a necessary property of random sequences (see Section 3). This law holds except for a set of Lebesgue measure zero, which can be uncountably large.

[19] Kolmogorov himself recognized von Mises's influence.

[20] There is an anecdote about Banach who, when asked by state authorities about the relevance of his work, replied that he was working on collectives; thereafter he wasn't bothered anymore. I thank Professor Paul Schweitzer for telling me of this episode.

Property I is known as the *Law of Large Numbers*,[21] which in measure-theoretic probability theory is a theorem, holding for almost all sequences x. Property II, which is the requirement that frequency stability be preserved under the operation of extracting infinite subsequences, eliminates such trivially nonrandom sequences as 01010101010101 . . . . It has an interpretation in terms of gambling, the source of many of von Mises's arguments. In that context it is called the *Law of Excluded Gambling Strategy*: a gambler betting in fixed amounts cannot make more profit in the long run by betting according to a "system" than by betting at random. This is certainly an intuitive requirement. Imagine a game in which you bet on the successive bits of an apparently random sequence supplied by a casino. Now, if that sequence is the binary representation of the number π and you are able to recognize it as such, then by predicting in advance the next digit, you would have a gambling "system."

Many criticisms were directed at von Mises's proposals. Not only were his arguments, based on gambling-house notions, considered inexact or at best semimathematical, but also the central notion of "admissible" selections was not clarified at all. Surely, many examples of admissible selection rules could be given: select x_n for which n is prime (or for which n is given by some other arithmetic law) or choose the x_ns that immediately follow the occurrence of 001 (in which case the choice will depend on the elements of the sequence), etc. However, if *arbitrary* selection rules (i.e., arbitrary subsequences) are allowed, then collectives don't even exist. This is known as "Kamke's argument." In fact, given a sequence $\{n_k\}_{k\geq 1}$ with $n_1 < n_2 < \dots$, consider a selection from x of the subsequence $x_{n_1}x_{n_2}\dots$. If all subsequences are allowed, there will be one such that $x_{n_k} = 1$ for all k and another with $x_{n_k} = 0$. Therefore, no sequence x could be a collective.

Kamke's objection didn't disturb von Mises, it would seem, because it used the unlimited (nonconstructive) concept of existence of mathematical objects of set-theoretic mathematics. In particular, in the argument above no "rule" for finding the subsequence is mentioned. We see here a need to clarify the notion of rules and the concomitant tension of the constructivist versus nonconstructivist viewpoints.

The next natural move was to restrict in suitable fasion the set of admissible selections. In 1937, Abraham Wald showed that, if the set $\mathcal{S}$ of admissible place selections is *countable*, then collectives do exist [**27**]. More precisely, let

$$C(\mathcal{S}, p) = \left\{ x \in \Sigma^{\mathbb{N}} : \forall \Phi \in \mathcal{S}, \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (\Phi x)_k = p \right\},$$

where $0 < p < 1$, be the set of collectives with respect to $\mathcal{S}$.

Theorem 3.2 (Wald). *For any countable $\mathcal{S}$ and any p in* $(0, 1)$, $\sharp\, C(\mathcal{S}, p) = 2^{\aleph_0}$*; that is, $C(\mathcal{S}, p)$ has the cardinality of the continuum.*

This result still left entirely open the question of which class $\mathcal{S}$ to choose. In 1940, the logician Alonzo Church proposed that, in order to isolate precisely those sequences that are "intuitively" random, the set of admissible place selections should consist of the *computable* or *partial recursive* functions. That is, only "effectively calculable" selections should be admitted. Thus a central notion of the Theory of Algorithms (or Recursive Function Theory or Computability Theory) entered the scene.

As a first attempt at the clarification of the concept of an individual random sequence, the Mises-Wald-Church viewpoint had some desirable traits. First of all, col-

[21] In a weaker version, it was first proven by Jakob Bernoulli, appearing in his posthumously published book *Ars Conjectandi* (1713).

lectives are abundant, constituting a set of measure one in $\Sigma^{\mathbb{N}}$. In addition, no collective can be generated by an algorithm. For, if it could be, then one could construct computable selection rules Φ_0 and Φ_1 as follows: for all $n \geq 1$,

$$\begin{cases} \Phi_0(x_1 x_2 \dots x_{n-1}) = 1 & \text{if } x_n = 0, \\ \Phi_1(x_1 x_2 \dots x_{n-1}) = 1 & \text{if } x_n = 1. \end{cases}$$

Furthermore, every random sequence is *Borel-normal* in base 2, meaning that every block of bits of size k in the sequence x appears in it with the "right" asymptotic frequency of $1/2^k$. This property, a case of the Law of Large Numbers, was discovered by Borel in 1909. Moreover, he proved that, relative to Lebesgue measure, almost every real number x in [0, 1] is *absolutely* normal; i.e., x satisfies the frequency condition for blocks of digits in any base representation. Borel normality is clearly a desirable property for any reasonable notion of randomness. In fact, it was initially proposed as *the* defining characteristic of a random real number. Unfortunately, not every normal number is random. For example, Champernowne proved in 1934 that the number

$$.0123456789101112131415\dots$$

is Borel-normal in base 10 [**37**], though clearly computable. The same is true of the Copeland-Erdös (1946) number $.23571113171923\dots$, obtained by concatenation of the prime numbers. Curiously, aside from these somewhat contrived examples, and in spite of the fact that most (in the sense of Lebesgue measure) real numbers in [0, 1] are normal, it is not known whether such fundamental mathematical constants as π, e, $\sqrt{2}$, or $\log 2$ are normal.[22]

A definitive blow against the idea of randomness as stochasticness was struck in 1939 by Jean Ville in his detailed analysis of the notion of collectives [**44**]. He showed that collectives are not "random enough" by proving that there are collectives satisfying

$$\frac{f_n}{n} \geq \frac{1}{2}$$

for all n, i.e., showing a preference for 1s over 0s (though still having limiting relative frequency equal to 1/2). Moreover, according to Levy's Law of the Iterated Logarithm, the set of sequences that exhibit this behavior has Lebesgue measure zero. In other words, the "collectives" don't satisfy all the *laws of randomness* of probability theory, understood to mean the *laws holding with probability one*, which renders the Mises-Wald-Church notion of randomness unsatisfactory.

4. RANDOMNESS AS INCOMPRESSIBILITY. The source of the "paradox of randomness" mentioned in the introduction is that we don't expect a regular outcome from a random experiment. This seems to be the intuition behind the famous "argument from design" frequently used against Darwin's theory of evolution: how could an intricately designed structure such as the human eye have evolved by pure chance? It's inconceivably improbable and therefore a "cause" must be responsible for its occurrence.[23]

[22] We refer to their fractional parts, which are numbers in [0, 1]. For some fascinating ideas in this direction, see Bailey and Crandall [**3**].

[23] Of course, the cause is nothing but Darwin's natural selection, which is essentially a *nonrandom* process, as forcefully described by the biologist R. Dawkins [**16**].

A possible abstract version of this argument, based on an original intuition due to Laplace, goes as follows. Suppose the object of interest is a binary string w (i.e., assume that all objects of interest can be codified as such strings) of length $|w| = n$. The question is to decide whether this string appeared by pure chance (i.e., by coin-tossing) or was "designed." The probability of its being generated by chance is 2^{-n}. Now, suppose it was generated by some simple "mechanism" (or "cause") that itself can be codified (described) by a string with m bits, with *m much smaller* than n. This would mean w is so regular that it can be (implicitly) described by the much smaller string representing its generating "cause." Therefore, it is 2^{n-m} more likely that w was generated by some cause than at random [**25**].

The idea of randomness as incompressibility was proposed independently (and almost simultaneously) by Ray Solomonoff, Andrei Kolmogorov, and Gregory Chaitin. The intuition is that a string is "irregular" or "patternless" if it cannot be "described" more efficiently than by giving the whole string itself. This is the notion of program-size algorithmic (or descriptive) *complexity*.[24] From its vantage point a string is random if no program of size substantially smaller than the string itself can generate or describe it.

For example, some (base 10) numbers, even large ones like 1,000,000,000,000, have short representations; a case in point is the number just cited, which can be expressed as 10^{12}. But it is difficult to describe economically, say, the number 5,172,893,164,583, except by writing it down digit for digit. The situation is exemplified even more dramatically by very long strings. Take, for instance, the highly "ordered" binary string $111\dots1$ consisting of 10,000 copies of the digit 1. It can be described by the program "print 1, 10,000 times." Compared to the sequence itself, which is 10,000 bits long, its description needs a little more than $\log_2 10{,}000 \approx 14$ bits. On the other hand, to describe an arbitrary "disordered" 10,000-bit-long string, we will most probably need to print the string itself, which serves as its own (very long) description. As observed in [**39**], randomness in this sense is due either to "extreme disorder" or to an "exaggerated order": the string is so uncharacteristically "complicated" that its precise description cannot be shorter than itself.

To formalize these notions, consider w in $\{0, 1\}^*$, and let U be a universal Turing machine. Let $U(p)$ be the output of machine U when fed an input p from $\{0, 1\}^*$, and let $|p|$ be the length (in bits) of the word p. Then we record:

Definition 4.1 (Kolmogorov-Chaitin). The *descriptive* or *algorithmic complexity* $K_U(w)$ of a word w with respect to the machine U is given by

$$K_U(w) = \begin{cases} \infty & \text{if there is no } p \text{ such that } U(p) = w, \\ \min\{|p| : U(p) = w\} & \text{otherwise.} \end{cases}$$

In other words, $K_U(w)$ is the size of the smallest input program p that, when fed to the Turing machine U, outputs ("prints") w and stops. Alternatively, it is the length of the shortest binary program p that "describes" or "codifies" the "object" w. This definition is universal or machine-independent in the following sense.

Theorem 4.2 (Invariance Theorem). *If U is a universal Turing machine, then for any universal Turing machine $\tilde{U}$ it is true that*

$$K_U(w) \le K_{\tilde{U}}(w) + c_{\tilde{U}}$$

for all w in $\{0, 1\}^*$, where $c_{\tilde{U}}$ is a constant independent of w.

[24] To be distinguished from resource-based or computational complexity.

Proof. Let $\tilde{U}(q) = w$ and let $s_{\tilde{U}}$ be the program that simulates the machine $\tilde{U}$ in the machine U: $U(s_{\tilde{U}}q) = \tilde{U}(q) = w$. Then $p = s_{\tilde{U}}q$ has length $|p| = |s_{\tilde{U}}| + |q|$, and letting $c_{\tilde{U}} = |s_{\tilde{U}}|$, we have

$$K_U(w) = \min_{\{p:U(p)=w\}} |p| \leq \min_{\{q:\tilde{U}(q)=w\}} (|q| + c_{\tilde{U}}) = K_{\tilde{U}}(w) + c_{\tilde{U}}. \qquad \blacksquare$$

One says that $K_U(w)$ is the "optimal" complexity[25] of w in the sense that $|K_{U_1}(w) - K_{U_2}(w)| < c = c(U_1, U_2)$ for all w in $\{0, 1\}^*$ and any pair of universal Turing machines U_1 and U_2. One can then fix once and for all a universal Turing machine U and write $K(w) = K_U(w)$.

The following result shows that there are few words with low complexity.

Proposition 4.3. $\sharp\{w \in \Sigma^* : K(w) < k\} < 2^k$.

Proof. One can list all computer programs of length less than k, ordered by increasing size:

$$\Lambda, 0, 1, 00, 01, 10, 11, \ \ldots\ , 111\ldots 11,$$

yielding a total of $1 + 2 + 4 + \cdots + 2^{k-1} = 2^k - 1$ programs. As each program generates at most one output for each input, we obtain the stated result.

■

Kolmogorov proposed to call an infinite sequence random if it has initial segments (or prefixes) of high complexity. Such sequences are said to be *incompressible*.

Definition 4.4. A sequence x in $\Sigma^{\mathbb{N}}$ is *incompressible* (or *chaotic*) when there is a constant c such that

$$K(x(n)) \geq n - c$$

for all n, where $x(n) = x_1 x_2 \ldots x_n$.

Unfortunately, the Swedish mathematician Per Martin-Löf showed that no such sequence exists.

Theorem 4.5 (Martin-Löf). *If $f : \mathbb{N} \to \mathbb{N}$ is a computable function such that*

$$\sum_{n=1}^{\infty} 2^{-f(n)} = \infty,$$

then for any binary sequence $x = x_1 x_2 \ldots$ it is the case that

$$K(x(n)) < n - f(n)$$

for infinitely many values of n.

In particular, the theorem holds for $f(n) = \log_2 n$. Therefore, every binary sequence drops infinitely often below $n - \log_2 n$, that is, far below its own length. This deadlock

[25] Sometimes also called "entropy," adding to the multitude of different notions bearing that name.

led Martin-Löf to formulate his own notion of randomness, which we will discuss in the next section.

On the other hand, Chaitin, Levin, Schnorr, and others were able to show that the incompressibility idea can be made consistent by suitably restricting the class of algorithms (or Turing machines). Chaitin's concept of *prefix-algorithms* is probably the simplest. The idea is reminiscent of the notion of prefix-free codes or languages used in information theory. As we saw in Section 2, a prefix-algorithm is a partial computable function ϕ_q, for fixed q, whose domain is prefix-free. Thus if the corresponding Turing machine halts for inputs p and p', then neither input is an initial segment of the other. There is a corresponding universal prefix-free Turing machine and basically all notions developed earlier carry over to the treatment of such machines.

In contrast to ordinary algorithms, prefix-algorithms are *self-delimiting*. That is, suppose the string p is a "description" of the string w by a Turing machine. The following might well happen without the prefix proviso. The machine, when presented with the input p, would first scan it from tip to tip in order to obtain its length $|p| = n$ and only then begin the bit-by-bit computation on p. In this case the complexity of w might well amount to $n + \log_2 n$ instead of n. This cannot happen for prefix-algorithms. By construction the program tape is read only to the right, and the machine stops at the last bit of p.

This motivates the replacement of the previous notion of complexity with the following one. If U is a universal prefix algorithm, then the *(Chaitin)-complexity* $C_U(w)$ of a string w relative to U is given by

$$C_U(w) = \min\{|p| : U(p, \Lambda) = w\}.$$

By the Invariance Theorem, we can fix a universal machine U and declare $C(w) = C_U(w)$ to be the complexity of w. Then one replaces the previous definition by:

Definition 4.6. A sequence x in $\Sigma^{\mathbb{N}}$ is *incompressible* (or *chaotic*) when there is a constant c such that

$$C(x(n)) \geq n - c$$

for all n, where $x(n) = x_1 x_2 \ldots x_n$.

The great technical advantage of working with prefix algorithms comes from a result in information theory called *Kraft's inequality* (1949) (see Rozenberg and Salomaa [**38**, p. 164]). It asserts that every prefix-free language $\mathcal{L}$ over $\{0, 1\}$ satisfies

$$\sum_{w \in \mathcal{L}} 2^{-|w|} \leq 1.$$

A generalization of this inequality known as the *Kraft-Chaitin inequality* [**7**], [**9**] is a crucial ingredient in the proof of the equivalence of the notion of incompressible sequences and another, conceptually very different notion, that of *typical* sequences. This is the topic of the next section.

5. RANDOMNESS AS TYPICALITY. Intuitively, we consider something to be "typical" when it is unexceptional or ordinary. If the objects in question are binary sequences, we could say that typical sequences are "featureless." Also, we would like the set of typical sequences to be much "bigger" than the set of nontypical ones, in some sense. But in exactly what sense? In the context of measure theory, which offers a generalization of the notions of area and volume, a set is considered "small" or unexceptional when it has measure zero.

One of the earliest applications of this notion was in celestial mechanics, in connection with the problem of the stability of the solar system. There, the idea was to prove stability by showing that the set of initial conditions of the equations of motion that led to some catastrophic event (e.g., planetary collisions or dispersion to infinity) would be a set of measure zero. In other words, the set of "problematic" points is negligible and has no physical significance.[26] Of course, there is some arbitrariness to this notion: what is considered negligible in one context might nonetheless be important in another one.[27]

Let us return, however, to binary sequences. In the essay cited earlier, Ville observed that the inadequacy of the Mises-Wald-Church notion of stochastic sequences was that it is based on a *single* "law of randomness," namely, the Law of Large Numbers. He suggested that a truly random sequence should satisfy all such laws. More precisely, a sequence is typical according to Ville if it satisfies all properties that occur with probability one in $\Sigma^{\mathbb{N}}$; i.e., if $\{\Sigma_\alpha : \alpha \in I\}$ is the collection of all sets of probability one, then x would be typical if:

$$x \in \bigcap_\alpha \Sigma_\alpha.$$

Formulated so simply, this cannot work. For example, let λ be the Bernoulli $(1/2, 1/2)$ distribution. Then $\lambda(\{x\}) = 0$ for all x or, equivalently, $\lambda(\Sigma^{\mathbb{N}} - \{x\}) = 1$. Therefore, $\cap_\alpha \Sigma_\alpha = \emptyset$ and there would be no typical sequences! To save the idea, one needs to restrict in some fashion the collection of sets of probability one. This is a familiar situation and once again the notion of algorithms or effective procedures comes to the rescue.

In 1966 Martin-Löf (then working as a postdoc in Moscow) advanced the notion of sets of *effective* measure one. Recall first the notion of a *null set* in $\Sigma^{\mathbb{N}}$: it is a set that can be covered by certain elementary sets such that the cover has measure as small as we want. The basic building blocks of the cover are the *cylinder sets*, $\Gamma_w = \{x \in \Sigma^{\mathbb{N}} : x = wy\}$, where w belongs to $\{0, 1\}^*$. These are the sets of all sequences with a given prefix w. Note that in terms of the real numbers in $[0, 1]$, each cylinder set is a dyadic interval $(0.w, 0.w + 2^{-|w|}]$.

Let μ be a probability measure on $\Sigma^{\mathbb{N}}$. We say that a subset N of $\Sigma^{\mathbb{N}}$ is μ-**null** if and only if, for each rational number $\epsilon > 0$, there is a sequence of words $w_0, w_1, \ldots$ in $\{0, 1\}^*$ such that

(i) $N \subset \bigcup_{k \geq 1} \Gamma_{w_k}$

and

(ii) $\sum_{k \geq 1} \mu(\Gamma_{w_k}) < \epsilon$.

A set N is said to be *effectively μ-null* provided there exists an algorithm (i.e., a Turing machine) that, for each rational $\epsilon > 0$ and nonnegative integer k, computes w_k for which the foregoing conditions (i) and (ii) are satisfied. A set of effective μ-measure one is defined by complementation. Note that in the case of the Bernoulli$(1/2, 1/2)$-measure λ, we have $\lambda(\Gamma_w) = 2^{-|w|}$.

[26] More recently, these ideas have been also used in the foundations of statistical mechanics.

[27] In the previous example, it is clear that planetary collisions are very important events in astronomy and the planetary sciences.

A probability measure μ is said to be *computable* when, for each positive rational number ϵ and each w in $\{0, 1\}^*$, there exists a Turing machine computable function F taking $(\epsilon, w) \to F(\epsilon, w)$ such that

$$|F(\epsilon, w) - \mu(\Gamma_w)| \leq \epsilon.$$

For example, the Bernoulli$(1/2, 1/2)$-measure is computable, because 2^{-n} is rational.

With these notions, Martin-Löf obtained the following result.

Theorem 5.1 (Martin-Löf). *Let μ be a computable probability measure. The intersection of sets of effective μ-measure one is nonempty and is a set of effective μ-measure one.*

This shows that the identification of the set of random sequences with the set of typical sequences is consistent: a sequence is *Martin-Löf-random* when it belongs to all sets of effective measure one (for some computable measure).

Martin-Löf's ideas can be interpreted in terms of the concept of an *effective statistical sequential test* for randomness. It consists of a recursively enumerable sequence of dyadic intervals $\{I_m^n\}_{n\geq 1}$ such that, for each fixed m, $\mu(I_m^n) < 2^{-m} = \epsilon$. To apply the test on a sequence x means to choose a confidence level m (or ϵ) and check whether or not x belongs to I_m^n for some $n \geq 1$. In the affirmative case, x is rejected (it fails the test), being considered nonrandom *at the level m*. On the other hand, if x is not rejected from a certain level on, then it passes (or succeeds in) the test. Therefore, each test eliminates a certain regularity property that is considered incompatible with the sequence being random. Hence, a sequence is said to be Martin-Löf-random if it passes *all* the effective sequential tests for randomness. The above theorem amounts to the statement that there is a *universal* (or *maximal*) sequential test that, if passed, defines a sequence as being random.

Corollary 5.2. *A computable sequence x in $\Sigma^{\mathbb{N}}$ is μ-typical if and only if $\mu(\{x\}) > 0$.*

In particular, Lebesgue-almost-every x in $[0, 1]$ is nontypical. Although, from this point of view, most real numbers are nontypical, it is impossible to construct a single concrete example of such a number by *algorithmic* means (i.e., they are noncomputable).

The crowning achievement of these investigations is the following impressive result, connecting two apparently very different notions of randomness (for a proof, see Li and Vitányi [**33**] or Calude [**7**]).

Theorem 5.3 (Levin-Schnorr-Chaitin). *A binary sequence is typical with respect to the* Bernoulli$(1/2, 1/2)$ *measure if and only if it is chaotic with respect to that distribution.*

The set $\mathcal{R}$ of Martin-Löf-random real numbers in $[0, 1]$ is, by construction, a set of Lebesgue-measure one. Each of its elements x is noncomputable, that is, there is no algorithm that generates the binary digits of x. However, one can *define*, in the style of classical mathematics, particular examples of such numbers. The most famous is the so-called *Chaitin's Ω-number*:

$$\Omega = \sum_{\{p\in\Sigma^*:U(p,\Lambda)<\infty\}} 2^{-|p|},$$

for a fixed universal prefix-free Turing machine U. The sum is taken over all inputs p in Σ^* for which the computation converges (U halts).

Note that $\Omega > 0$, for U halts for *some* input p. By Kraft's inequality, $\Omega \leq 1$; since U does not halt for *all* inputs, we have $\Omega < 1$. Therefore $0 < \Omega < 1$. The number Ω is called the *halting probability* with respect to U: Ω is the probability that the machine U stops when fed an input p chosen "at random" (i.e., by tossing an honest coin).

It can be shown that Ω has certain curious features [**38**]. First of all, it is an incompressible number, hence (Martin-Löf)-random.[28] In spite of its being noncomputable, Ω can be *estimated*, and it is known that $0.00106502 < \Omega < 0.217643$. Also, if the prefix $\Omega(n)$ of size n is known, then we can decide all halting problems codifiable in less than n bits. It follows that Ω has the following property. Let β be a formula in some axiomatic mathematical system A, and let $TM(A, \beta)$ and $TM(A, \neg\beta)$ be Turing machines that check whether β or $\neg\beta$ is a theorem (merely by verifying all possible proofs in the system). If a big enough initial prefix $\Omega(n)$ is known, one can decide whether β is provable, not provable, or independent in A!

How big a prefix is needed depends on the sizes of F and β, that is, on how compactly expressible they are. A reasonable estimate (for humanly interesting cases) reckons that some 10,000 digits would suffice. This would encompass such classical conundrums as Goldbach's conjecture and Riemann's hypothesis. The catch is that, even if $\Omega(n)$ were known, it would be computationally useless: the computation time $t(n)$ to find from $\Omega(n)$ all the halting programs of size less than n increases faster than any computable function.

6. CONCLUSION.

The notion of Martin-Löf-random sequences is mathematically consistent and, unexpectedly, coincides with an apparently very different notion of randomness, namely, that of incompressibility.[29] Coupled with the fact that no serious flaw, analogous to the ones that surfaced in the theory of the collectives, has yet been found, this state of affairs makes a strong argument supporting the concept of Martin-Löf-random sequences as the best candidate for the mathematical definition of randomness.

As observed in [**17**], this parallels the story leading to the proposal of the Church-Turing thesis. By the same token, the main theoretical objections are the ones inherited through the use of the Church-Turing thesis.[30] For example, isn't there an overemphasis on the concept of computability? After all, as remembered in [**40**, p. 316], "most of mathematics is about noncomputable objects and even noncountable ones." Is the concept of randomness, founded in the concept of absence of *computable* regularities, the only adequate and consistent one? In which directions, if any, should one look for alternatives?

Even more problematic are some allegedly deep connections to randomness in the *physical* world, which are advanced on the basis of the given mathematical definition. What is not quite clear is the soundness of such questions as: "Is the universe recursive/computable?" After all, computability is about algorithmic procedures that in turn refer to a class of methods used by humans for solving certain mathematical problems, methods that may appear in connection with physical models. Frequently a distinction isn't clearly drawn between the mathematical formalism used in a physical theory and the referents the theory is supposed to investigate.[31] It seems to presuppose the idea that the "universe" can somehow be identified with a kind of big universal Turing ma-

[28] And according to Borel, would not be a real number at all.

[29] For other equivalences, see Calude [**8**].

[30] See Copeland [**12**] for a nice discussion of the thesis and some misunderstandings of it.

[31] It is analogous to saying that quantum mechanics is Hilbert space analysis instead of a theory about atoms and molecules. Or, that classical mechanics is dynamical systems theory instead of a theory of interacting point particles.

chine putting out bits for us to decode. Besides being strongly anthropomorphic, this is an extreme simplification of nature and is at odds with the pictures presented to us by the natural sciences. Of course, this is not to say that investigation of the computability properties of, say, the solution of certain partial differential equations describing physical processes is unimportant. The question is whether or not our description of them as computable bears any relevance to the natural processes themselves.

Finally, what about applications? This is probably where the theory has the least to offer, although that was not the main purpose of the investigation to begin with (however, see Li and Vitányi [**33**]). Consider the quest for long tables of random numbers. These are in great demand in statistics (in conjunction with the problem of random sampling), in computer simulations (for instance, in applications of the Monte Carlo method), and in cryptography. It would be desirable to generate such numbers quickly, preferably in a reproducible way, without a simultaneous overburdening of memory resources. But these are requirements that cannot be satisfied by a "truly" random sequence. In fact, such sequences are noncomputable, so cannot be efficiently stored. Nor can one generate high complexity binary strings from a small random "seed." Moreover, the complexity function $C(x)$ is itself noncomputable [**7**], [**33**], therefore one cannot algorithmically decide whether a certain string is random. One settles for a more pragmatic principle: "if it acts randomly, it is random." That is, one resorts to *pseudorandom bit generators* [**31**], which are completely computable methods to generate large "random-looking" bits. Although predictable in principle, one tries to design them so that they display sufficient (pseudo)randomness for applications.

ACKNOWLEDGEMENTS. Work partially supported by FINEP (Pronex project). The author is thankful to the anonymous referee for many valuable criticisms and suggestions that greatly improved the article. It is also a pleasure to thank Professor Paul Schweitzer for his careful review of an earlier draft.

REFERENCES

1. K. Ambos-Spies and A. Kučera, Randomness in computability, in *Computability Theory: Current Trends and Open Problems* (P. Cholak et al., eds.), Contemporary Mathematics, American Mathematical Society, Providence, 2000.
2. H. Bar-Hillel and W. A. Wagenaar, The perception of randomness, *Advances in Applied Mathematics* **12** (1991) 428–454.
3. D. H. Bailey and R. C. Crandall, On the random character of fundamental constant expansions, http://www.perfsci.com/free/techpapers/freepapers.html, May 2000.
4. D. Bellhouse, The role of roguery in the history of probability, *Statistical Science* **8** (3) (1993) 410–420.
5. D. J. Bennett, *Randomness*, Harvard University Press, 1998.
6. M. Bunge, Three faces and three masks of probability, *Probability in the Sciences*, E. Agazzi, ed., Kluwer Academic Publishers, Dordrecht, 1988.
7. C. S. Calude, *Information and Randomness*, Springer-Verlag, New York, 1994.
8. C. S. Calude, Who is afraid of randomness? http://www.cs.auckland.ac.nz/staff-cgi-bin/mjd/secondcgi.pl, Sept. 2000.
9. C. S. Calude and C. Grozea, The Kraft-Chaitin Inequality revisited, http://www.cs.auckland.ac.nz/staff-cgi-bin/mjd/secondcgi.pl, April 1996.
10. G. Chaitin, A theory of program size formally identical to information theory, *J. ACM* **22** (1975) 329–340.
11. A. Compagner, Definitions of randomness, *Amer. J. Phys.* **59** (8) August (1991) 700–705.
12. B. J. Copeland, The Church-Turing Thesis, *Stanford Encyclopedia of Philosophy*, http://www.plato.stanford.edu/, January 1997.
13. L. Corry, David Hilbert and the axiomatization of physics, *Arch. Hist. Exact Sci.* **51** (1997) 83–198.
14. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., New York, 1991.
15. M. Davies, *The Universal Computer: The Road from Leibniz to Turing*, W. W. Norton & Company, New York, 2000.
16. R. Dawkins, *The Blind Watchmaker*, W. W. Norton & Company, New York, 1986.

17. J.-P. Delahaye, *Information, Complexité et Hasard*, Éditions Hermès, Paris, 1999.
18. C. Dellacherie, Nombres au hazard. De Borel à Martin-Loef, *Gazette des Math.*, Soc. Math. France, **11** (1978) 23–58.
19. S. Fefferman, Mathematical intuition vs. mathematical monsters, *Synthese* **125** (2000) 317–332.
20. I. Grattan-Guinness, *The Rainbow of Mathematics*, W. W. Norton & Company, New York, 1997.
21. I. Hacking, *The Emergence of Probability*, Cambridge University Press, 1975.
22. R. Herken, ed., *The Universal Turing Machine: a Half-Century Survey*, 2nd ed., Springer-Verlag, New York, 1995.
23. H. Inoue, H. Kumahora, Y. Yoshizawa, M. Ichimura, and O. Mitayake, Random numbers generated by a physical device, *Appl. Statist.* **32** (2) (1983) 115–120.
24. J. Keller, The probability of heads, *Amer. Math. Monthly* **93** (1986), 191–197.
25. W. Kirchherr, M. Li, and P. Vitanyi, The miraculous universal distribution, *Math. Intell.* **19** (4) (1997) 7–15.
26. D. E. Knuth, Algoritmic thinking and mathematical thinking, *Amer. Math. Monthly* **92** (1985) 170–181.
27. D. E. Knuth, *The Art of Computer Programming*, vol. 2, 3rd ed., Addison-Wesley, Boston, 1998.
28. A. N. Kolmogorov, *Foundations of the Theory of Probability*, 2nd ed., Chelsea Pub., New York, 1956.
29. A. N. Kolmogorov and V. A. Uspenskii, Algorithms and randomness, *Theory Probab. Appl.* **32** (3) (1987) 389–412.
30. S. Körner. *The Philosophy of Mathematics*, Hutchinson University Library, London, 1968.
31. J. Lagarias, Pseudorandom numbers, *Statistical Science* **8** (1) (1993) 31–39.
32. J. Laurie Snell and R. Vanderbei, Three Bewitching Paradoxes, in *Topics in Contemporary Probability and Its Applications*, CRC Press, Boca Raton, FL, 1995, pp. 355–370.
33. M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, 2nd ed., Graduate Texts in Computer Science, Springer-Verlag, New York, 1997.
34. P. Martin-Löf, The literature on von Mises' kollectives revisited, *Theoria* **XXXV** (1969) 12–37.
35. M. Minsky, *Computation: Finite and Infinite Machines*, Prentice-Hall, 1967.
36. A. A. Muchnik, A. L. Semenoff, and V. A. Uspensky, Mathematical metaphysics of randomness, *Theoretical Computer Science* **201** (1998) 263–317.
37. I. Niven, *Irrational Numbers*, The Carus Mathematical Monographs, No. 11, John Wiley & Sons, New York, 1967.
38. G. Rozenberg and A. Salomaa, *Cornerstones of Undecidability*, Prentice Hall, Upper Saddle River, NJ, 1994.
39. G. Sacks, Teoria della ricorsività, *Rend. Sem. Mat. Univ. Pol. Torino* **55** (1) (1997) 1–17.
40. R. Soare, Computability and Recursion, *Bulletin of Symbolic Logic* **2** (3) (1996) 284–321.
41. V. A. Uspenskii, A. L. Semenov, and A. Kh. Shen, Can an individual sequence of zeros and ones be random?, *Russian Math. Surveys* **45** (1) (1990) 105–162.
42. M. van Lambalgen, Von Mises' definition of randomness reconsidered, *The Journal of Symbolic Logic* **52** (3) (1987) 725–755.
43. M. van Lambalgen, Randomness and infinity, http://www.illc.uva.nl/Publications/reportlist.php, January 1995.
44. J. Ville, *Étude Critique de la Notion de Collectif*, Gauthier-Villars, Paris, 1939.
45. R. von Mises, *Probability, Statistics and Truth*, reprint, Dover, Mineola, NY, 1981.
46. R. von Mises and J. L. Doob, Discussions of papers in probability theory, *Annals of Mathematical Statistics* **12** (2) (1941) 215–217.
47. J. von Plato, *Creating Modern Probability*, Cambridge University Press, 1998.
48. P. Wagner, *La Machine en Logique*, PUF, Paris, 1998.
49. Y. Zeng-yuan and Z. Bin, On the sensitive dynamical system and the transition from the apparently deterministic process to the completely random process, *Applied Mathematics and Mechanics* **6** (3) (1985) 193–211.

SÉRGIO B. VOLCHAN earned an M.S. in physics from the Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Brazil. He then turned to mathematics, receiving his Ph.D. from the Courant Institute of Mathematical Sciences in 1995. After spending two years as a researcher at the Instituto de Matemática Pura e Aplicada (IMPA), he moved to the mathematics department of PUC-Rio, where he is assistant professor. His research interests are probability theory and its applications to mathematical physics.
Deptartamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro, Rua Marquês de São Vicente 225, Gávea, 22453-900 Rio de Janeiro, Brasil
volchan@mat.puc-rio.br