CDMTCS
Research
Report
Series

A Glimpse into Algorithmic Information Theory

Cristian S. Calude
Department of Computer Science
University of Auckland

CDMTCS-093
February 1999

Centre for Discrete Mathematics and
Theoretical Computer Science

A Glimpse into Algorithmic Information Theory

Cristian S. Calude
Department of Computer Science
University of Auckland
Private Bag 92019 Auckland
New Zealand
Email: cristian@cs.auckland.ac.nz

Abstract

This paper is a subjective, short overview of algorithmic information theory. We critically discuss various equivalent algorithmical models of randomness motivating a "randomness hypothesis". Finally some recent results on computably enumerable random reals are reviewed.

1 Randomness: An Informal Discussion

In which we discuss some difficulties arising in defining randomness.

Suppose that one is watching a simple pendulum swing back and forth, recording 0 if it swings clockwise at a given instant and 1 if it swings counterclockwise. Suppose further that after some time the record looks as follows:

$$10101010101010101010101010101010.$$

At this point one would like to deduce a "theory" from the experiment.[1] The "theory" should account for the data presently available and make "predictions" about future observations. How should one proceed? It is obvious that there are many "theories" that one could write-down for the given data, for example:

$$10101010101010101010101010101010000000000000000000000000000000\ldots$$
$$10101010101010101010101010101010111111111111111111111111111111\ldots$$
$$10101010101010101010101010101010001001001001001001001001001001\ldots$$
$$10101010101010101010101010101010101010101010101010101010101010\ldots$$
$$10101010101010101010101010101010000111000111000111000111000111\ldots$$
$$10101010101010101010101010101010101011101110010110100011000111\ldots$$

Consistently with the requirements formulated above, each "theory" starts with the experimental data (which is finite) and continues with "predictions" about the system future (which is potentially infinite). The results of the experiment display a simple pattern, always 10's, so probably the best prediction is that the system will continue to produce 10's forever. Is there any rational, objective way of deciding among various possible "theories" that does not rely only upon intuition? Occam's razor states that the "best" theory is the "simplest" theory: "Nunquam ponenda est pluralitas sine necesitate". What is a "simple theory"? For Solomonoff [35] the "simplest theory" is the one

[1] People have studied pendulums for centuries and apparently everything is known about them: they are the "epitome of regularity". In fact, pendulums are unpredictable in different ways as chaos theory has showed on various occasions, e.g., in Hall [24].

with the shortest length, i.e., the one printed by a shortest length computer program. First we have to produce the experimental data; then, we have to "guess" a continuation. For example, the following program will account for the first sequence:

```
PRINT  10101010101010101010101010101010, PRINT  0.
```

Can we do it better? Given the regularity of the record a considerable shorter program can be written to produce it, namely the program "`PRINT 10 16 times`". This program can be used to print out the first five "theories" above:

```
PRINT 10 16 times, PRINT 0
PRINT 10 16 times, PRINT 1
PRINT 10 16 times, PRINT 001
PRINT 10 forever
PRINT 10 16 times, PRINT 000111
```

The program "`PRINT 10 16 times`" can be generalised to a "law" expressed by the program "`PRINT 10 X times`". Note that the length of printouts predicted by this program grows much faster than the length of the program itself. Can we write a simple program to print the last "theory"? In this case the continuation of the experiment does not follow an obvious pattern ...or maybe there is no such pattern. To understand this "theory" we will follow Chaitin[2] [11] who engaged in defining and studying the "complexity" of finite binary strings. A typical question motivating this approach is: Are the first one million digits of the binary expansion of the number π less complex than a string produced by flipping a fair coin one million times?[3] Chaitin defined the complexity of a finite binary string as the size of the smallest program which calculates it. If the string can be compressed into a very short program then one might conclude that the string has a pattern, that it follows a law, that it is simple; if the string cannot be compressed at all, then it is maximally complex or random. Let's test this idea on some examples. Suppose that someone claims to have tossed a fair coin 64 times and the result is:

$$x = 0101010101010101010101010101010101010101010101010101010101010101$$

Then, the experiment is repeated and the result is:

$$y = 1010101010010010101001011011000110001100111100110011111000011011$$

Almost everyone would be surprised or suspicious to see x, but the string y probably would be acceptable. Why? Here is a typical flawed explanation: the probability of x is extraordinarily small, i.e., 2^{-64}, so it is unreasonable to believe that x has been actually produced by a real experiment. However, from a probabilistic point of view there is nothing special about x: all of the 2^{64} possible strings of length 64 have equal probabilities of appearance, 2^{-64}. The difference between x and y *is not* probabilistic, but *structural*: while x is ordered, there is no apparent pattern in y.[4] Laplace [27], pp.16-17, was, in a sense, aware of the above phenomenon:

> *In the game of heads and tails, if head comes up a hundred times in a row then this appears to us extraordinary, because after dividing the nearly infinite number of combinations that can arise in a hundred throws into regular sequences, or those in which we observe a rule that is easy to grasp, and into irregular sequences, the latter are incomparably more numerous.*

[2] A that time a undergraduate at City University of New York; see [18]. Kolmogorov [28] developed similar ideas.

[3] Record 1 for heads and 0 for tails.

[4] Of course, we may argue that the presence of the pattern 01 in x has no significance at all because a) the number of tosses is relatively small, b) finding patterns and meanings is just a human subjective predilection.

Instead of computing probabilities of specific strings let's instead discuss about the "typicality" of some strings with respect to some particular stochastic processes. For the process of flipping a fair coin, incompressible strings are typical, and highly compressible strings are atypical. And, because the number of highly compressible strings (of a given length) is *small*, the occurrence of such string is *extraordinary*: our surprise regarding x is justified.

Of course, the above explanation is informal, and a lot need to be done to turn it into a rigorous theory. Before presenting some technical details let indulge ourselves in a simple counting analysis. A string of length n will be said to be c–incompressible if its compressed length is greater than or equal to $n-c$. For example, the 16–incompressible strings of length 64 are exactly the strings that can be compressed to a length of 48 or larger. Note that every $n+1$–incompressible string is n–incompressible, so every 5–incompressible string is 4–incompressible. As the number of strings of length n is 2^n, it turns out that at least half of all the strings of every length are 1–incompressible, at least $\frac{3}{4}$ are 2–incompressible, at least $\frac{7}{8}$ are 3–incompressible, so on. In general, at least $1-\frac{1}{2^c}$ of all strings of length n are c–incompressible. For example, about 99.9% of all strings of length 64 cannot be compressed by more than 16% and about 99.99999998% of these strings cannot be compressed by more than 50%.

2 Algorithmic Models for Randomness

In which different proposals for defining "algorithmically random sequences" are discussed and a "randomness hypothesis" is formulated.

The discussion in the first section suggests that:

1. *a "random" sequence should be typical,* i.e., it should belong to any "reasonable" majority;

2. *a "random" sequence should be chaotic,* i.e., no simple law should be capable to produce the terms of the sequence.

To address typicality let's isolate the set of all sequences having "all verifiable" properties that from the point of view of classical probability theory are satisfied with "probability one" with respect to the unbiased discrete probability. Let us denote by Σ the binary alphabet $\{0,1\}$ and by Σ^* the set of all binary strings. The unbiased discrete probability on Σ which gives equal preference to 0 and 1 (both appear with probability 1/2) induces the product measure μ on the set of all Borel subsets of the set of all binary sequences Σ^ω. If $x = x_1x_2\ldots x_n$ is a string of length n, then the *cylinder* induced by x, $x\Sigma^\omega$, i.e., the set of all sequences starting with $x_1x_2\ldots x_n$, will have the probability $\mu(x\Sigma^\omega) = 2^{-n}$. This number can be interpreted as "the probability that a sequence $\mathbf{y} = y_1y_2\ldots y_n\ldots$ has the first element $y_1 = x_1$, the second element $y_2 = x_2, \ldots$, the nth element $y_n = x_n$". Independence means that the probability of an event of the form $y_i = x_i$ does not depend upon the probability of the event $y_j = x_j$. Note that every open set, i.e., a union of cylinders, is μ measurable. Finally, $S \subset \Sigma^\omega$ is a *null set* in case for every real $\varepsilon > 0$ there exists an open set which contains S and has measure less than ε. For instance, every enumerable subset of Σ^ω is a null set.

A property P of sequences is said to be *true almost everywhere (in the sense of μ)* in case the set of sequences not having the property P is a null set. The main example of such a property is the famous *Law of Large Numbers* discovered by Borel: *For every sequence $x = x_1x_2\ldots x_m\ldots$ and natural number $n \geq 1$, the limit of $S_n(\mathbf{x})/n$, when n tends to ∞, exists almost everywhere in the sense of μ and has the value 1/2;* (here

$S_n(\mathbf{x}) = x_1 + x_2 + \cdots + x_n$). In other words, *there exists a null set $S \subset \Sigma^\omega$ such that for every $\mathbf{x} \notin S$, we have $S_n(\mathbf{x})/n = 1/2$.*

It is clear that a sequence satisfying a property false almost everywhere with respect to μ is very "particular". Accordingly, it is tempting to say that

> *a sequence $\mathbf{x}$ is "random" if it satisfies every property true almost everywhere with respect to μ.*

Unfortunately, we may define, for every sequence $\mathbf{x}$, the property $P_{\mathbf{x}}$:

> *a sequence $\mathbf{y}$ satisfies $P_{\mathbf{x}}$ if and only if for every $n \geq 1$ there exists a natural $m \geq n$ such that $x_m \neq y_m$.*

Every $P_{\mathbf{x}}$ is an asymptotic property which is true almost everywhere with respect to μ and $\mathbf{x}$ does not have property $P_{\mathbf{x}}$. Accordingly, *no sequence can verify all properties true almost everywhere with respect to μ.* The above definition is vacuous!

The above analysis may suggest that there is no truly lawless sequence. Indeed, a "universal" non-trivial property shared by all sequences was discovered by van der Waerden:

> *In every binary sequence at least one of the two symbols must occur in arithmetical progressions of every length.*

Looking at the proof of van der Waerden's result (and of a few similar ones) we notice that they are all *non-constructive.* To be more precise, there is no algorithm which will tell in a finite amount of time which alternative is true: 0 occurs in arithmetical progressions of every length or 1 occurs in arithmetical progressions of every length.

As a consequence we have to consider not *all* asymptotic properties true almost everywhere with respect to μ, but only a *countable set* of such properties. So, the important question becomes: *Which properties should be considered?* Clearly, the "larger" the chosen class of properties is, the "more random" will be the sequences satisfying those properties. A constructive selection seems to be suggested by both statistical practice and philosophical intuition. One such definition, proposed by Martin-Löf [29], is based on randomness tests. We fix a standard computable bijective function $\langle , \rangle$ defined on $\mathbf{N} \times \Sigma^*$ with values in Σ^*; here $\mathbf{N}$ is the set of non-negative integers. For a set $A \subseteq \Sigma^*$ let $A_k = \{x \in \Sigma^* \mid \langle k, x \rangle \in A\}$. A *Martin-Löf test* is a computably enumerable (c.e.) set $A \subset \Sigma^*$ such that $\mu(A_i\Sigma^\omega) \leq 2^{-i}$, for all natural i. The set $\bigcap_{i \geq 0}(A_i\Sigma^\omega)$ is the set of all sequences which do not pass the randomness test A. We now define:

> *a sequence $\mathbf{x}$ is Martin-Löf random if for every Martin-Löf test A,*
> $\mathbf{x} \notin \bigcap_{i \geq 0}(A_i\Sigma^\omega)$.

Martin-Löf [29] proved the existence of a *universal Martin-Löf test,* a test W with the property that for every Martin-Löf test A there is a constant c such that $A_n \subseteq W_{n+c}$, for all n. So, Martin-Löf 's definition can be rephrased as:

> *a sequence $\mathbf{x}$ is Martin-Löf random if and only if $\mathbf{x}$ passes a universal Martin-Löf test.*

It is almost immediate that typicality is guaranteed as for each Martin-Löf test A, the set $\bigcap_{i \geq 0}(A_i\Sigma^\omega)$ is constructively null:

Theorem 1 *Constructively, with probability one (in the sense of μ), every sequence is Martin-Löf-random.*

So, with probability one every binary sequence is random—the set of random sequences is *large.* However, from a topological point of view[5], *the set of Martin-Löf random sequences is constructively a first Baire category set*, i.e., it is *small*, cf. Calude, Chiţescu [5]. This is a natural example where the popular analogy between measure-theoretic small sets (null sets) and topological small sets (sets of first Baire category) fails to work even in a constructive way.

Solovay [36] has proposed another measure-theoretic definition of randomness which aims to capture typicality:

> *A sequence x is Solovay random if for every c.e. set $A \subset \Sigma^*$ such that $\sum_{i \geq 1} \mu(A_i \Sigma^\omega) < \infty$, there exists a natural N such that for all $i > N$, $x \notin A_i \Sigma^\omega$.*

To address the second property associated with randomness, i.e., "chaoticity", we follow Chaitin's complexity-theoretic approach. To this aim we employ self-delimiting Turing machines M: such a machine reads the program from right to left only, never going back, so its accepted programs (i.e., its domain) form a prefix-free set. If, after finitely many steps, M halts with the program tape head scanning the last bit of the input x, then the computation is a success, and we write $M(x) < \infty$; the output of the computation is the string $M(x) \in \Sigma^*$. Otherwise, the computation is a failure, we write $M(x) = \infty$, and there is no output. The program set

$$PROG_M = \{x \in \Sigma^* \mid M(x) < \infty\}$$

is a c.e. *instantaneous code (or prefix-free set)*, i.e., no program leading to a halting computation can be the prefix of another such program. Conversely, *every prefix-free c.e. set of words is the domain of some self-delimiting Turing machine.*

Let M be a self-delimiting Turing machine. The *program-size complexity* of the string $x \in \Sigma^*$ (relative to M) is

$$H_M(x) = \min\{|y| \mid y \in \Sigma^*,\ M(y) = x\},$$

where $\min \emptyset = \infty$. It was shown by Chaitin [13] (see Calude [2]) that there is a self-delimiting Turing machine U that is *universal*, in the sense that, for every self-delimiting Turing machine M, there is a constant c_M (depending upon M) with the following property: if $M(x) < \infty$, then there is an $x' \in \Sigma^*$ such that $U(x') = M(x)$ and $|x'| \leq |x| + c_M$.[6] Clearly, every universal self-delimiting machine produces every string. Every two universal self-delimiting machines U and V induce roughly the same program-size complexity, i.e., $H_U(x) = H_V(x) + O(1)$. We denote by x^* the *canonical program* of x, i.e., $x^* = \min\{y \in \Sigma^* \mid U(y) = x\}$, where the minimum is taken according to the quasi-lexicographical order $0, 1, 00, 01, 10, 11, 000, \ldots$ Clearly, the program-size complexity of a string x is exactly the length of its canonical program.[7]

The probability that a program generated by a coin tossing run on M will produce the output x is the measure of the cylinder induced by the set of strings y which output x on M:

$$P_M(x) = \sum_{\{M(y)=x\}} 2^{-|y|}.$$

[5] Σ comes equipped with the discrete topology and Σ^ω is endowed with the product topology.

[6] This "optimality" holds true for space complexity, but it seems to fail to be true for time complexity: the simulation time is much bigger than the original computational time.

[7] By definition, x^* is the most compact way (for U) to store x: the computation $U(x^*) = x$ produces x by freeing $|x| - |x^*|$ bits of memory. What is the least thermodynamic cost of generating a string x from the canonical program x^*? Zurek [37] has proven that *the computation $U(x^*) = x$ can be achieved reversibly, with no cost in terms of entropy increase.* Let's note that a reversible computation, i.e., a computation which can be undone, can be performed only by using computer memory to keep track of the exact logical path from input to output (see more in [6] and [3]): thermodynamic irreversibility is inevitable only in the presence of logically irreversible operations.

In the case $M = U$ we simply write $PROG = PROG_U$[8], $H(x) = H_U(x)$, $P(x) = P_U(x)$. The relation between program-size complexity (H) and algorithmic probability (P), known as the *Coding Theorem* (Chaitin [13], Gács [23], Calude [2]) is given by the following formula:

$$H(x) = -\log_2 P(x) + \mathrm{O}(1).$$

The program-size complexity is only up to a fixed constant different from the entropy;[9] the probability of computing x is concentrated in the canonical program x^* computing x. [10]

We are now in a position to give two complexity-theoretic definitions of random sequences.

> *An infinite sequence* $\mathbf{x}$ *is Schnorr random if there is a constant* c *such that* $H(\mathbf{x}(n)) > n - c$, *for every integer* $n > 0$ (Chaitin [13]).

> *An infinite sequence* $\mathbf{x}$ *is Chaitin random if* $\lim_{n\to\infty} H(\mathbf{x}(n)) - n = \infty$ (Chaitin [13]).

Finally, we present Hertling and Weihrauch topological approach to define randomness initiated in [25, 26]. A *randomness space* is a triple (X, B, μ), where X is a topological space, $B : \mathbf{N} \to 2^X$ is a total numbering of a subbase of the topology of X, and μ is a measure defined on the σ-algebra generated by the topology of X.[11] Let $(W_n)_n$ be a sequence of open subsets of X; a sequence $(V_n)_n$ of open subsets of X is called W–*computable* if there is a c.e. set $A \subseteq \mathbf{N}$ such that $V_n = \bigcup_{\pi(n,i)\in A} W_i$ for all $n \in \mathbf{N}$.[12] Next we define $W_i' = W'(i) = \bigcap_{j\in D_{(1+i)}} W_j$, for all $i \in \mathbf{N}$, where we have used the bijection defined on $D : \mathbf{N} \to \{E \mid E \subseteq \mathbf{N} \text{ is finite}\}$ by $D^{-1}(E) = \sum_{i\in E} 2^i$. Note that if B is a numbering of a subbase of a topology, then B' is a numbering of a base of the same topology. A *randomness test on* X is a B'–computable sequence $(W_n)_n$ of open sets with $\mu(W_n) \leq 2^{-n}$, for all $n \in \mathbf{N}$.

> *An element* $x \in X$ *is called random if* $x \notin \bigcap_{n\in\mathbf{N}} W_n$, *for every randomness test* $(W_n)_n$ *on* X.

Consider now the topological space Σ^ω and the numbering B of a subbase (in fact a base) of the topology is given by $B_i = \nu(i)\Sigma^\omega = \{p \in \Sigma^\omega \mid \nu(i) \text{ is a prefix of } p\}$, where $\nu : \mathbf{N} \to \Sigma^*$ is the length–lexicographical bijection between $\mathbf{N}$ and the set Σ^*. The last definition reads:

[8] Sometimes $PROG$ is denoted by K.

[9] Let's illustrate the Coding Theorem with an example from Cover and Thomas [21]. We look at a monkey trying to "type" the entire work of Shakespeare, of say $1,000,000$ bits long. If the monkey types "at random" on a dumb typewriter, the probability that the result is Shakespeare's work is $2^{-1,000,000}$; if the monkey sits in front of a computer terminal, then the algorithmic probability that it types the same text is $2^{-H(\text{Shakespeare})} \asymp 2^{-250,000}$, an event with an extremely small chance to happen, but still more likely than the first event. The use of the typewriter reproduces exactly the input produced by the typing while a computer "runs" the input and produces an output. Consequently, a random input to a computer is much more likely to produce an "interesting" output than a "random" input to a typewriter. Is this a way to create "sense" out of "nonsense"?

[10] We are now in a position to come back to Occam's Razor principle by calculating the probability of seeing a 1 next after having obtained n 1's in the experiment described in the first section. This conditional probability is the ratio of the probability of all sequences with an initial string 1^{n+1} (i.e., the cylinder $1^{n+1}\Sigma^\omega$) to the probability of all sequences having an initial string 1^n. The canonical programs carry most of the probability, hence we can approximate the probability that the next bit is a 1 with the probability of the program "`PRINT 1's forever`". Consequently, $\Sigma_y P(1^{n+1} y \Sigma^\omega) \approx P(111\ldots 11\ldots) = c > 0$. It is more difficult to estimate the probability that next bit is 0 which is $1/(cn+1)$, cf. Cover and Thomas [21], p. 169.

[11] Recall that a subbase of a topology is a set β of open sets such that the sets $\bigcap_{W\in E} W$, for finite, nonempty sets $E \subseteq \beta$ form a basis of the topology.

[12] $\pi(n,i)$ is a computable bijection, for exmple, $\pi(n,i) = (n+i)(n+i+1)/2 + i$.

A sequence is Hertling–Weihrauch random if it is random in the space (Σ^ω, B, μ).

All the above approaches lead to the same class of sequences:

Theorem 2 *Let* $\mathbf{x} \in \Sigma^\omega$. *The following statements are equivalent:*

1. *The sequence* $\mathbf{x}$ *is Martin-Löf random.*
2. *The sequence* $\mathbf{x}$ *is Chaitin random.*
3. *The sequence* $\mathbf{x}$ *is Schnorr random.*
4. *The sequence* $\mathbf{x}$ *is Solovay random.*
5. *The sequence* $\mathbf{x}$ *is Hertling–Weihrauch random.*

In what follows we will simply call "random" a sequence satisfying one of the above equivalent conditions. Theorem 2 motivates the following "randomness hypothesis":

> *A sequence is "algorithmically random" if it satisfies one of the equivalent conditions in Theorem 2.*

Various arguments supporting this hypothesis, e.g., *random sequences are Borel absolutely normal*[13] are analysed in Calude [2]. Here is another argument due to Fouché [22]: *if* $X \subseteq \Sigma^\omega$ *is a* Σ^0_1 *set and has measure one, then it contains at least one random sequence.* In particular, if X is Π^0_1 set which contains some random sequence, then it has non-zero measure. So, if a Π^0_1 event reflected in some random sequence, then the event must be probabilistically significant.

3 Information: Quantity vs Structure

In which we contrast the properties of three reals aiming to encode the same amount of information as the halting problem.

We shall consider three reals Ω, Ξ, and Υ. The first real is the halting probability of U to stop on a program whose bits have been obtained by tossing a fair coin, Chaitin's Ω number:[14]

$$\Omega = \sum_{x \in \Sigma^*} P_U(x) = \sum_{x \in PROG} 2^{-|x|}.$$

The second one is the halting problem characteristic real: put *string*(i) the ith binary string in quasi-lexicographical and define

$$\Xi = \sum_{string(i) \in PROG} 2^{-i}.$$

Finally, let $time(string(i))$ be the running time of the computation $U(string(i))$[15], and define the third real

[13] Every string appears in a random sequence with the probability 2^{-n}, where n is the length of the string.

[14] For more about Ω see Bennett [1], Calude, Salomaa [10], Calude, Meyerstein [9], Rozenberg, Salomaa [31].

[15] $time(string(i))$ is a positive integer in case $string(i) \in PROG$, and $time(string(i)) = \infty$, in the opposite case.

$$\Upsilon = \sum_i 2^{-i}/time(string(i)). \tag{1}$$

All three numbers can be computed from $PROG$. Can we do it better, i.e., compute some of these reals without using $PROG$? The answer is negative for Ω (cf. Chaitin [13]):

Theorem 3 *Ω is random.*

Proof. [16] Let f be a computable one-to-one function which enumerates $PROG$, the domain of U. Let $\omega_k = \sum_{j=0}^{k} 2^{-|f(j)|}$. Clearly, (ω_k) is an increasing sequence of rationals converging to Ω. Consider the binary expansion of $\Omega = 0.\Omega_0\Omega_1\cdots$.

We define a self-delimiting Turing machine M as follows: on input $x \in \Sigma^*$ compute $y = U(x)$ and the smallest number (if such a number exists) t with $\omega_t \geq 0.y$. Let $M(x)$ be the first (in quasi-lexicographical order) string not belonging to the set $\{U(f(0)), U(f(1)), \ldots, U(f(t))\}$ if both y and t exist, and $M(x) = \infty$ if $U(x) = \infty$ or t does not exist.

If $x \in PROG_M$ and x' is a word with $U(x) = U(x')$, then $M(x) = M(x')$. Applying this to an arbitrary $x \in PROG_M$ and the canonical program $x' = (U(x))^*$ of $U(x)$ yields

$$H_M(M(x)) \leq |x'| = H(U(x))\,. \tag{2}$$

Furthermore, by the universality of U, there is a constant $c > 0$ with

$$H(M(x)) \leq H_M(M(x)) + c \tag{3}$$

for all $x \in PROG_M$. Now, fix n and assume that x is a word with $U(x) = \Omega_0\Omega_1\cdots\Omega_{n-1}$. Then $M(x) < \infty$. Let t be the smallest non-negative integer (computed in the second step of M) with $\omega_t \geq 0.\Omega_0\Omega_1\cdots\Omega_{n-1}$. We have

$$0.\Omega_0\Omega_1\cdots\Omega_{n-1} \leq \omega_t < \omega_t + \sum_{s=t+1}^{\infty} 2^{-|f(s)|} = \Omega \leq 0.\Omega_0\Omega_1\cdots\Omega_{n-1} + 2^{-n}\,.$$

Hence, $\sum_{s=t+1}^{\infty} 2^{-|f(s)|} \leq 2^{-n}$. This implies $|f(s)| \geq n$, for every $s \geq t+1$. From the construction of M we conclude that $H_U(M(x)) \geq n$. Using (3) and (2) we obtain

$$n \leq H(M(x)) \leq H_M(M(x)) + c \leq H(U(x)) + c = H(\Omega_0\Omega_1\cdots\Omega_{n-1}) + c\,,$$

which proves that the sequence $\Omega_0\Omega_1\cdots$ is random. ■

What about Ξ? It is not difficult to see that Ξ *is not computable* (otherwise we could use it to test whether $string(i)$ is in $PROG$, contradicting the undecidability of the halting problem): again we need $PROG_U$ to compute Ξ. Is Ξ random? The answer is negative as the following Diophantine argument (due to Chaitin [16]) shows. Using a classical arithmetization we can obtain effectively a Diophantine equation[17] $E(n, x_1, \ldots, x_m) = 0$ with the property that for every natural n, $\Omega_n = 1$ if and only if the equation $E(n, x_1, \ldots, x_m) = 0$ has infinitely many solutions in $x_1, \ldots, x_m$. A similar representation can be obtained for Ξ in case we replace the question "the equation $E(n, x_1, \ldots, x_m) = 0$ has infinitely many solutions" with a weaker question, namely, "the equation $E'(n, x_1, \ldots, x_m) = 0$ has finitely many solutions", for a suitable equation

[16] The proof is taken from [7].

[17] An exponential Diophantine equation is an equation $E(a_1, \ldots, a_n, x_1, \ldots, x_m) = 0$ built from non-negative integers $a_1, \ldots, a_n$ by using only the operations of addition, multiplication and exponentiation.The most famous example of exponential Diophantine equation is Fermat's equation $x^n + y^n = z^n$, proven to admit no integer solution greater than 3.

E'. Indeed, the question whether an arbitrarily given exponential Diophantine equation $E'(n, x_1, \ldots, x_m) = 0$ has finitely many solutions is undecidable, but the answers to such questions have no maximal information content because they are not independent. To see this let's consider such an equation $E'(n, x_1, \ldots, x_m) = 0$, with $m = 0, 1, 2, \ldots, t-1$, and consider the binary string $b = b_0 b_1 b_2 \ldots b_{t-1}$ where $b_i$is 0 if the answer corresponding to $m = i$ is negative and 1 in the opposite case. The quantity of information encoded by b is about $\log_2 t$ bits, much less than the length of b which is t. The reason is simple: knowing *how many* equations have a positive answer gives enough information to find exactly *which* equations do have a positive answer. Consequently, Ξ *is neither computable nor random.*

The computations of both Ξ and Ω require extra information, for example, the one given by $PROG$. If we know Ξ we can compute Ω, and conversely, if we know Ω we can compute Ξ; so, in an informal sense, Ξ encodes the same quantity of information as Ω. However, their structures are quite different. What about Υ? Obviously, knowing Ω or Ξ is enough to compute Υ, because $string(i) \in PROG$ if and only if $time(string(i)) < \infty$. Is the converse implication true? The answer is negative:

Theorem 4 *The real Υ is computable.*

Proof. We construct an algorithm computing, for every positive integer n, the nth digit of Υ. The idea is simple: only the terms $2^{-i}/time(string(i))$ for which $time(string(i)) = \infty$ do produce perturbations in (1) because at every finite step of the computation they appear to be non-zero when, in fact, they are zero! The solution is to run all non-stopping programs $string(i)$ enough time such that their cumulated contribution is too small to affect the nth digit of Υ. ■

All three numbers, Ω, Ξ, and Υ are c.e., i.e., they are all limits of increasing, computable sequences of rationals[18]; however, Ω is random (so, non-computable), Ξ is non-computable, but non-random, while Υ is even computable.

Let's have a closer analysis of the above numbers. We start by defining the concept of "random real". A real $\alpha \in (0,1)$ is random in case its binary expansion is a random sequence. Actually, the choice of base is irrelevant, cf. Theorem 6.111 in Calude [2]. This can be seen also by defining directly random reals using Hertling-Weihrauch's approach. To this aim we consider the set of reals $\mathbf{R}$ with the usual Lebesgue measure μ and B the numbering of a base of the real line topology defined by $B_{\pi(i,j)} = \{x \in \mathbf{R} \mid |x - \nu_D(i)| < 2^{-j}\}$,[19] where $\nu_D(<k, l, m>) = (k-l)2^{-m}$ is defined on the set of dyadic reals $D = \{x \in \mathbf{R} \mid x = (i-j)2^{-k}$, for some $i, j, k\}$. For the unit interval we work with the restriction of the Lebesgue measure and $B_i \cap [0,1]$.

The first question is: are there random c.e. reals? Chaitin's Ω number is a first example. Solovay[36] has introduced the notion of Ω–like real and proved that every such number is c.e. and random. To define Ω–like reals we need the relation of domination. Following Solovay we say that the real α *dominates* the real β (we write $\alpha \geq_{dom} \beta$) if there are a partial computable function f from rationals to rationals, and a constant $c > 0$ with the property that if p is a rational number less than α, then $f(p)$ is (defined and) less than β, and satisfies the inequality $c(\alpha - p) \geq \beta - f(p)$. Informally, $\alpha \geq_{dom} \beta$ if there is an effective way to get a good approximation to β from below from any good approximation to α from below. For c.e. reals domination can also be expressed as follows: a c.e. real α dominates a c.e. real β if and only if there are two computable, non-decreasing sequences (a_i) and (b_i) of rationals and a constant $c > 0$ such that $\lim_n a_n = \alpha$, $\lim_n b_n = \beta$, and $c(\alpha - a_n) \geq \beta - b_n$, for all n.

[18] Equivalently, a real is c.e. if the set of rationals less than it is c.e.

[19] Recall that π is a computable bijection.

A real α is Ω-like if it is the limit of a universal computable, increasing sequence of rationals[20] (Solovay [36]).

Theorem 5 *Let $0 < \alpha < 1$. The following conditions are equivalent:*

1. *The real α is a Chaitin Ω real, i.e., for some universal self-delimiting Turing machine U, $\alpha = \Omega_U$.*
2. *The real α is Ω–like.*
3. *The real α is c.e. and dominates all c.e. reals.*
4. *There exists a universal computable, increasing sequence of rationals converging to α.*
5. *Every computable, increasing sequence of rationals with limit α is universal.*
6. *The real α is c.e. and random.*

The implication $1 \Longrightarrow 2$ was proved by Solovay [36]; the equivalence $2 \Longleftrightarrow 3$ can be found in Chaitin [15]; the implication $2 \Longrightarrow 1$ and equivalences $2 \Longleftrightarrow 4 \Longleftrightarrow 5$ are proved in Calude, Hertling, Khoussainov, Wang [7]; finally, the implication $6 \Rightarrow 2$ was proved by Slaman [32]. Recently, Slaman [33] proved that the measure of any section of a universal Martin-Löf test W, $\mu(W_n\Sigma^\omega)$, is Ω–like.

Let's now compare the information of an Ω-like c.e. real Ω to that of a non-Ω-like c.e. real α. Clearly, either Ω contains more information or at least its information is structured in a more useful way (because we can find a good approximation from below to any c.e. real from a good approximation from below to Ω). Sometimes we need to compute not just an arbitrary approximation (say, of precision 2^{-n}) from below to a c.e. real, instead, but a special approximation, namely the first n digits of its binary expansion. Is the information in Ω organized in such a way as to guarantee that for any c.e. real α there exists a total computable function $g : \mathbf{N} \to \mathbf{N}$ (depending upon α) such that from the first $g(n)$ digits of Ω we can actually compute the first n digits of α? The answer is negative if one demands that the computation is to be done by a total computable function, but it is positive if we instead work with a partial computable function.

For a set $A \subset \Sigma^*$ we denote by χ_A the characteristic function of A. For example, $\Xi = \chi_{PROG}$. We say that A is *Turing reducible* to B (we write $A \leq_T B$) if there is a B–oracle Turing machine φ^B such that $\varphi^B(x) = \chi_A(x)$. We say that A is *weak truth-table reducible* to B (we write $A \leq_{wtt} B$) if $A \leq_T B$ via a Turing reduction which on input x only queries strings of length less than $g(x)$, where $g : \Sigma^* \to \mathbf{N}$ is a fixed computable function. Finally, A is *truth-table reducible* to B (we write $A \leq_{tt} B$) if there is a computable sequence of Boolean functions $\{F_x\}_{x\in\Sigma^*}$, $F_x : \Sigma^{r_x+1} \to \Sigma$, such that for all x, we have $\chi_A(x) = F_x(\chi_B(0)\chi_B(1)\cdots\chi_B(r_x))$. Note that in contrast with tt-reductions, a wtt-reduction may diverge. If $r \in \{\text{T, tt, wtt}\}$, then $A =_r B$ in case $A \leq_r B$ and $B \leq_r A$. A c.e. set A is tt(wtt)-c*omplete* if $PROG \leq_{tt} A$ ($PROG \leq_{wtt} A$). See Odifreddi [30] or Soare [34] for more details.

For every infinite sequence $\mathbf{x} \in \Sigma^\omega$, let $A_{\mathbf{x}} = \{v \in \Sigma^* \mid 0.v \leq 0.\mathbf{x}\}$ and $A^{\#}_{\mathbf{x}} = \{string(n) \mid x_n = 1\}$. Then, if $0.\mathbf{x}$ is a c.e. real, then $A_{\mathbf{x}}$ is a c.e. set which is Turing equivalent to $A^{\#}_{\mathbf{x}}$; however, $A^{\#}_{\mathbf{x}}$ is not necessarily c.e. Now, $0.\mathbf{x} \leq_{tt} 0.\mathbf{y}$ in case $A^{\#}_{\mathbf{x}} \leq_{tt}$

[20] A computable, increasing, and converging sequence (a_i) of rationals is called *universal* if for every computable, increasing and converging sequence (b_i) of rationals there exists a number $c > 0$ such that $c(\alpha - a_n) \geq \beta - b_n$ for all n, where $\alpha = \lim_n a_n$ and $\beta = \lim_n b_n$.

$A^{\#}_{\mathbf{y}}$. It is easy to see that $0.\mathbf{x} \leq_{tt} 0.\mathbf{y}$ if and only if there are two computable functions $g : \mathbf{N} \to \mathbf{N}$ and $F : \Sigma^* \to \Sigma^*$ such that $\mathbf{x}(n) = F(\mathbf{y}(g(n)))$, for all n.

The following result summarizes the reducibility relations between Ω, Ξ, Υ and $PROG$:

Theorem 6 *The following statements are true:*

1. $\Xi \not\geq_{dom} \Omega$,
2. $A_\Omega =_T A_{\chi_K} =_T \Xi$,
3. *For every c.e. real* α, $\alpha \leq_{tt} \Xi$,
4. $\Xi \not\leq_{tt} \Omega$, *so* Ω *is not tt-complete,*
5. *for any c.e. real* $0.\mathbf{x}$ *there exist a computable function* g *and a partial computable function* F *with* $\mathbf{x}(n) = F(\Omega(g(n)))$ *for all* n, *so* Ω *is wtt-complete.*

So, the $\leq_{tt}$–preorder has a maximum among the c.e. reals, but this maximum is not Ω, as no random c.e. real is maximal; Ω is maximal for the $\leq_{tt}$–preorder. The first four statements come from Calude, Hertling, Khoussainov, Wang [7]; the last statement was proven in Calude, Nies [8].

A final question: Is the compatibility between randomness and computable enumerability jeopardizing the randomness hypothesis (randomness should exclude constructivity)? The answer is negative: the compatibility is a consequence of the fact that there is no sequence passing any test of randomness—as discussed in section 2. Note that the c.e. of Ω is of very little help in computing the digits of $\Omega_0, \Omega_1, \ldots$ as this set is immune (cf. Calude, Chiţescu [4]), that is no infinite subset of it is c.e.

Acknowledgement 7 *I wish to express my gratitude to Dr. Philip Polkinghorne, surgical specialist, who restored my sight. This work was partially supported by the grant AURC A18/XXXXX/62090/F3414056, 1996.*

References

[1] C. H. Bennett, M. Gardner. The random number omega bids fair to hold the mysteries of the universe, *Scientific American* 241 (1979), 20–34.

[2] C. S. Calude. *Information and Randomness. An Algorithmic Perspective*, Springer-Verlag, Berlin, 1994.

[3] C. S. Calude, J. L. Casti. Parallel thinking, *Nature* 392 (1998), 549–551.

[4] C. Calude, I. Chiţescu. Qualitative properties of P. Martin-Löf random sequences, *Boll. Unione Mat. Ital.* VII, Ser. B3, 240 (1989), 229-240.

[5] C. Calude, I. Chiţescu. Random sequences: some topological and measure-theoretical properties, *An. Univ. Bucureşti, Mat.-Inf.* 2 (1988), 27–32.

[6] C. S. Calude, J. Casti, M. Dinneen (eds.). *Unconventional Models of Computation*, Springer-Verlag, Singapore, 1998.

[7] C. S. Calude, P. Hertling, B. Khoussainov, and Y. Wang. Recursively enumerable reals and Chaitin Ω numbers, in: M. Morvan et al. (eds.), *Proceedings of the 15th Symposium on Theoretical Aspects of Computer Science (Paris)*, Springer–Verlag, Berlin, 1998, 596–606.

[8] C. S. Calude, A. Nies. Chaitin Ω numbers and strong reducibilities, *J. UCS* 11 (1997), 1161–1166.

[9] C. S. Calude, F. W. Meyerstein. Is the universe lawful?, *Chaos, Solitons & Fractals*, (to appear in 1999).

[10] C. Calude, A. Salomaa. Algorithmically coding the universe, in G. Rozenberg, A. Salomaa (eds.). *Developments in Language Theory*, World Scientific, Singapore, 1994, 472–492.

[11] G. J. Chaitin. On the length of programs for computing finite binary sequences, *J. Assoc. Comput. Mach.* 13 (1966), 547–569. (Reprinted in: [17], 219-244.)

[12] G. J. Chaitin. On the length of programs for computing finite binary sequences: statistical considerations, *J. Assoc. Comput. Mach.* 16 (1969), 145-159. (Reprinted in: [17], 245–260.)

[13] G. J. Chaitin. A theory of program size formally identical to information theory, *J. Assoc. Comput. Mach.* 22 (1975), 329–340. (Reprinted in: [17], 197–223.)

[14] G. J. Chaitin. Information-theoretic characterizations of recursive infinite strings, *Theoret. Comput. Sci.* 2 (1976), 45–48. (Reprinted in: [17], 203–206.)

[15] G. J. Chaitin. Algorithmic information theory, *IBM J. Res. Develop.* 21 (1977), 350-359, 496. (Reprinted in: [17], 44–58.)

[16] G. J. Chaitin. *Algorithmic Information Theory*, Cambridge University Press, Cambridge, 1987. (third printing 1990)

[17] G. J. Chaitin. *Information, Randomness and Incompleteness, Papers on Algorithmic Information Theory*, World Scientific, Singapore, 1987. (2nd ed., 1990)

[18] G. J. Chaitin. *Information-Theoretic Incompleteness*, World Scientific, Singapore, 1992.

[19] G. J. Chaitin. *The Limits of Mathematics*, Springer-Verlag, Singapore, 1997.

[20] G. J. Chaitin. *The Unknowable*, manuscript, 31 December 1998, 121 pp.

[21] T. M. Cover, J. A. Thomas. *Elements of Information Theory*, Wiley, New York, 1991.

[22] W. L. Fouché. Descriptive complexity and reflective properties of combinatorial configurations, *J. London. Math. Soc.* 54 (1996), 199-208.

[23] P. Gács. On the symmetry of algorithmic information, *Soviet Math. Dokl.* 15(1974), 1477-1480; correction, Ibidem 15(1974), 1480.

[24] N. Hall (ed.). *The New Scientist Guide to Chaos*, Penguin, London, 1991.

[25] P. Hertling, K. Weihrauch. Randomness spaces, in: K. G. Larsen, S. Skyum, and G. Winskel (eds.). *Automata, Languages and Programming, Proceedings of the 25th International Colloquium, ICALP'98* (Aalborg, Denmark), Springer-Verlag, Berlin, 1998, 796–807.

[26] P. Hertling and K. Weihrauch. Randomness spaces, *Research Report 079, CDMTCS*, Auckland, January 1998.

[27] P. S. Laplace. *A Philosophical Essay on Probability Theories*, Dover, New York, 1951.

[28] A. N. Kolmogorov. Three approaches for defining the concept of "information quantity", *Problems Inform. Transmission* 1 (1965), 3–11.

[29] P. Martin-Löf. The definition of random sequences, *Inform. and Control* 9 (1966), 602–619.

[30] P. Odifreddi. *Classical Recursion Theory*, North-Holland, Amsterdam, 1989.

[31] G. Rozenberg, A. Salomaa. *Cornerstones of Undecidability*, Prentice-Hall, Englewood Cliffs, 1994.

[32] T. A. Slaman. *Random Implies Ω-Like*, manuscript, 14 December 1998, 2 pp.

[33] T. A. Slaman. *Recursive Enumerability and Randomness,* manuscript, 11 January 1999, 5 pp.

[34] R. I. Soare. *Recursively Enumerable Sets and Degrees*, Springer-Verlag, Berlin, 1987.

[35] R. J. Solomonoff . A formal theory of inductive inference, Part 1 and Part 2, *Inform. and Control* 7 (1964), 1–22 and 224–254.

[36] R. Solovay. *Draft of a paper (or series of papers) on Chaitin's work ... done for the most part during the period of Sept. – Dec. 1974*, unpublished manuscript, IBM Thomas J. Watson Research Center, Yorktown Heights, New York, May 1975, 215 pp.

[37] W. H. Zurek. Thermodynamic cost of computation, algorithmic complexity and the information metric, *Nature* 341 (1989), 119–124.