

INCOMPLETENESS THEOREMS FOR RANDOM REALS

Advances in Applied Mathematics 8
(1987), pp. 119–146

G. J. Chaitin

*IBM Thomas J. Watson Research Center, P.O. Box 218,
Yorktown Heights, New York 10598*

Abstract

We obtain some dramatic results using statistical mechanics–thermodynamics kinds of arguments concerning randomness, chaos, unpredictability, and uncertainty in mathematics. We construct an equation involving only whole numbers and addition, multiplication, and exponentiation, with the property that if one varies a parameter and asks whether the number of solutions is finite or infinite, the answer to this question is indistinguishable from the result of independent tosses of a fair coin. This yields a number of powerful Gödel incompleteness-type results concerning the limitations of the axiomatic method, in which entropy–information measures are used. © 1987 Academic Press, Inc.

1. Introduction

It is now half a century since Turing published his remarkable paper *On Computable Numbers, with an Application to the Entscheidungsproblem* (Turing [15]). In that paper Turing constructs a universal Turing machine that can simulate any other Turing machine. He also uses Cantor's method to diagonalize over the countable set of computable real numbers and construct an uncomputable real, from which he deduces the unsolvability of the halting problem and as a corollary a form of Gödel's incompleteness theorem. This paper has penetrated into our thinking to such a point that it is now regarded as obvious, a fate which is suffered by only the most basic conceptual contributions. Speaking as a mathematician, I cannot help noting with pride that the idea of a general purpose electronic digital computer was invented in order to cast light on a fundamental question regarding the foundations of mathematics, years before such objects were actually constructed. Of course, this is an enormous simplification of the complex genesis of the computer, to which many contributed, but there is as much truth in this remark as there is in many other historical "facts."

In another paper [5], I used ideas from algorithmic information theory to construct a diophantine equation whose solutions are in a sense random. In the present paper I shall try to give a relatively self-contained exposition of this result via another route, starting from Turing's original construction of an uncomputable real number.

Following Turing, consider an enumeration r_1, r_2, r_3, \dots of all computable real numbers between zero and one. We may suppose that r_k is the real number, if any, computed by the k th computer program. Let $.d_{k1}d_{k2}d_{k3}\dots$ be the successive digits in the decimal expansion of r_k . Following Cantor, consider the diagonal of the array of r_k ,

$$\begin{aligned} r_1 &= .d_{11}d_{12}d_{13}\dots \\ r_2 &= .d_{21}d_{22}d_{23}\dots \\ r_3 &= .d_{31}d_{32}d_{33}\dots \end{aligned}$$

This gives us a new real number with decimal expansion $.d_{11}d_{22}d_{33}\dots$. Now change each of these digits, avoiding the digits zero and nine. The result is an uncomputable real number, because its first digit is

different from the first digit of the first computable real, its second digit is different from the second digit of the second computable real, etc. It is necessary to avoid zero and nine, because real numbers with different digit sequences can be equal to each other if one of them ends with an infinite sequence of zeros and the other ends with an infinite sequence of nines, for example, $.3999999\dots = .4000000\dots$.

Having constructed an uncomputable real number by diagonalizing over the computable reals, Turing points out that it follows that the halting problem is unsolvable. In particular, there can be no way of deciding if the k th computer program ever outputs a k th digit. Because if there were, one could actually calculate the successive digits of the uncomputable real number defined above, which is impossible. Turing also notes that a version of Gödel's incompleteness theorem is an immediate corollary, because if there cannot be an algorithm for deciding if the k th computer program ever outputs a k th digit, there also cannot be a formal axiomatic system which would always enable one to prove which of these possibilities is the case, for in principle one could run through all possible proofs to decide. Using the powerful techniques which were developed in order to solve Hilbert's tenth problem (see Davis *et al.* [7] and Jones and Matijasevič [11]), it is possible to encode the unsolvability of the halting problem as a statement about an exponential diophantine equation. An exponential diophantine equation is one of the form

$$P(x_1, \dots, x_m) = P'(x_1, \dots, x_m),$$

where the variables x_1, \dots, x_m range over natural numbers and P and P' are functions built up from these variables and natural number constants by the operations of addition, multiplication, and exponentiation. The result of this encoding is an exponential diophantine equation $P = P'$ in $m + 1$ variables n, x_1, \dots, x_m with the property that

$$P(n, x_1, \dots, x_m) = P'(n, x_1, \dots, x_m)$$

has a solution in natural numbers x_1, \dots, x_m if and only if the n th computer program ever outputs an n th digit. It follows that there can be no algorithm for deciding as a function of n whether or not $P = P'$ has a solution, and thus there cannot be any complete proof system for settling such questions either.

Up to now we have followed Turing's original approach, but now we will set off into new territory. Our point of departure is a remark of Courant and Robbins [6] that another way of obtaining a real number that is not on the list r_1, r_2, r_3, \dots is by tossing a coin. Here is their measure-theoretic argument that the real numbers are uncountable. Recall that r_1, r_2, r_3, \dots are the computable reals between zero and one. Cover r_1 with an interval of length $\epsilon/2$, cover r_2 with an interval of length $\epsilon/4$, cover r_3 with an interval of length $\epsilon/8$, and in general cover r_k with an interval of length $\epsilon/2^k$. Thus all computable reals in the unit interval are covered by this infinite set of intervals, and the total length of the covering intervals is

$$\sum_{k=1}^{\infty} \frac{\epsilon}{2^k} = \epsilon.$$

Hence if we take ϵ sufficiently small, the total length of the covering is arbitrarily small. In summary, the reals between zero and one constitute an interval of length one, and the subset that are computable can be covered by intervals whose total length is arbitrarily small. In other words, the computable reals are a set of measure zero, and if we choose a real in the unit interval at random, the probability that it is computable is zero. Thus one way to get an uncomputable real with probability one is to flip a fair coin, using independent tosses to obtain each bit of the binary expansion of its base-two representation.

If this train of thought is pursued, it leads one to the notion of a random real number, which can never be a computable real. Following Martin-Löf [12], we give a definition of a random real using constructive measure theory. We say that a set of real numbers X is a constructive measure zero set if there is an algorithm A which given n generates a (possibly infinite) set of intervals whose total length is less than or equal to 2^{-n} and which covers the set X . More precisely, the covering is in the form of a set C of finite binary strings s such that

$$\sum_{s \in C} 2^{-|s|} \leq 2^{-n}$$

(here $|s|$ denotes the length of the string s), and each real in the covered set X has a member of C as the initial part of its base-two expansion.

In other words, we consider sets of real numbers with the property that there is an algorithm A for producing arbitrarily small coverings of the set. Such sets of reals are constructively of measure zero. Since there are only countably many algorithms A for constructively covering measure zero sets, it follows that almost all real numbers are not contained in any set of constructive measure zero. Such reals are called (Martin-Löf) random reals. In fact, if the successive bits of a real number are chosen by coin flipping, with probability one it will not be contained in any set of constructive measure zero, and hence will be a random real number.

Note that no computable real number r is random. Here is how we get a constructive covering of arbitrarily small measure. The covering algorithm, given n , yields the n -bit initial sequence of the binary digits of r . This covers r and has total length or measure equal to 2^{-n} . Thus there is an algorithm for obtaining arbitrarily small coverings of the set consisting of the computable real r , and r is not a random real number. We leave to the reader the adaptation of the argument in Feller [9] proving the strong law of large numbers to show that reals in which all digits do not have equal limiting frequency have constructive measure zero. It follows that random reals are normal in Borel's sense, that is, in any base all digits have equal limiting frequency.

Let us consider the real number p whose n th bit in base-two notation is a zero or a one depending on whether or not the exponential diophantine equation

$$P(n, x_1, \dots, x_m) = P'(n, x_1, \dots, x_m)$$

has a solution in natural numbers x_1, \dots, x_m . We will show that p is not a random real. In fact, we will give an algorithm for producing coverings of measure $(n + 1)2^{-n}$, which can obviously be changed to one for producing coverings of measure not greater than 2^{-n} . Consider the first N values of the parameter n . If one knows for how many of these values of n , $P = P'$ has a solution, then one can find for which values of $n < N$ there are solutions. This is because the set of solutions of $P = P'$ is recursively enumerable, that is, one can try more and more solutions and eventually find each value of the parameter n for which there is a solution. The only problem is to decide when to give up further searches because all values of $n < N$ for which there are

solutions have been found. But if one is told how many such n there are, then one knows when to stop searching for solutions. So one can assume each of the $N + 1$ possibilities ranging from p has all of its initial N bits off to p has all of them on, and each one of these assumptions determines the actual values of the first N bits of p . Thus we have determined $N + 1$ different possibilities for the first N bits of p , that is, the real number p is covered by a set of intervals of total length $(N + 1)2^{-N}$, and hence is a set of constructive measure zero, and p cannot be a random real number.

Thus asking whether an exponential diophantine equation has a solution as a function of a parameter cannot give us a random real number. However asking whether or not the number of solutions is infinite can give us a random real. In particular, there is an exponential diophantine equation $Q = Q'$ such that the real number q is random whose n th bit is a zero or a one depending on whether or not there are infinitely many natural numbers x_1, \dots, x_m such that

$$Q(n, x_1, \dots, x_m) = Q'(n, x_1, \dots, x_m).$$

The equation $P = P'$ that we considered before encoded the halting problem, that is, the n th bit of the real number p was zero or one depending on whether the n th computer program ever outputs an n th digit. To construct an equation $Q = Q'$ such that q is random is somewhat more difficult; we shall limit ourselves to giving an outline of the proof:¹

1. First show that if one had an oracle for solving the halting problem, then one could compute the successive bits of the base-two representation of a particular random real number q .
2. Then show that if a real number q can be computed using an oracle for the halting problem, it can be obtained without using an oracle as the limit of a computable sequence of dyadic rational numbers (rationals of the form $K/2^L$).

¹The full proof is given later in this paper (Theorems R6 and R7), but is slightly different; it uses a particular random real number, Ω , that arises naturally in algorithmic information theory.

3. Finally show that any real number q that is the limit of a computable sequence of dyadic rational numbers can be encoded into an exponential diophantine equation $Q = Q'$ in such a manner that

$$Q(n, x_1, \dots, x_m) = Q'(n, x_1, \dots, x_m)$$

has infinitely many solutions x_1, \dots, x_m if and only if the n th bit of the real number q is a one. This is done using the fact “that every r.e. set has a *singlefold* exponential diophantine representation” (Jones and Matijasevič [11]).

$Q = Q'$ is quite a remarkable equation, as it shows that there is a kind of uncertainty principle even in pure mathematics, in fact, even in the theory of whole numbers. Whether or not $Q = Q'$ has infinitely many solutions jumps around in a completely unpredictable manner as the parameter n varies. It may be said that the truth or falsity of the assertion that there are infinitely many solutions is indistinguishable from the result of independent tosses of a fair coin. In other words, these are independent mathematical facts with probability one-half! This is where our search for a probabilistic proof of Turing’s theorem that there are uncomputable real numbers has led us, to a dramatic version of Gödel’s incompleteness theorem.

In Section 2 we define the real number Ω , and we develop as much of algorithmic information theory as we shall need in the rest of the paper. In Section 3 we compare a number of definitions of randomness, we show that Ω is random, and we show that Ω can be encoded into an exponential diophantine equation. In Section 4 we develop incompleteness theorems for Ω and for its exponential diophantine equation.

2. Algorithmic Information Theory [3]

First a piece of notation. By $\log x$ we mean the integer part of the base-two logarithm of x . That is, if $2^n \leq x < 2^{n+1}$, then $\log x = n$. Thus $2^{\log x} \leq x$, even if $x < 1$.

Our point of departure is the observation that the series

$$\sum \frac{1}{n}, \sum \frac{1}{n \log n}, \sum \frac{1}{n \log n \log \log n} \dots$$

all diverge. On the other hand,

$$\sum \frac{1}{n^2}, \sum \frac{1}{n(\log n)^2}, \sum \frac{1}{n \log n (\log \log n)^2} \cdots$$

all converge. To show this we use the Cauchy condensation test (Hardy [10]): if $\phi(n)$ is a nonincreasing function of n , then the series $\sum \phi(n)$ is convergent or divergent according as $\sum 2^n \phi(2^n)$ is convergent or divergent.

Here is a proof of the Cauchy condensation test

$$\begin{aligned} \sum \phi(k) &\geq \sum [\phi(2^n + 1) + \cdots + \phi(2^{n+1})] \\ &\geq \sum 2^n \phi(2^{n+1}) = \frac{1}{2} \sum 2^{n+1} \phi(2^{n+1}). \end{aligned}$$

$$\sum \phi(k) \leq \sum [\phi(2^n) + \cdots + \phi(2^{n+1} - 1)] \leq \sum 2^n \phi(2^n).$$

Thus $\sum \frac{1}{n}$ behaves the same as $\sum 2^n \frac{1}{2^n} = \sum 1$, which diverges. $\sum \frac{1}{n \log n}$ behaves the same as $\sum 2^n \frac{1}{2^n n} = \sum \frac{1}{n}$, which diverges.

$$\sum \frac{1}{n \log n \log \log n}$$

behaves the same as $\sum 2^n \frac{1}{2^n n \log n} = \sum \frac{1}{n \log n}$, which diverges, etc.

On the other hand, $\sum \frac{1}{n^2}$ behaves the same as $\sum 2^n \frac{1}{2^{2n}} = \sum \frac{1}{2^n}$, which converges. $\sum \frac{1}{n(\log n)^2}$ behaves the same as $\sum 2^n \frac{1}{2^n n^2} = \sum \frac{1}{n^2}$, which converges.

$$\sum \frac{1}{n \log n (\log \log n)^2}$$

behaves the same as $\sum 2^n \frac{1}{2^n n (\log n)^2} = \sum \frac{1}{n (\log n)^2}$, which converges, etc.

For the purposes of this paper, it is best to think of the algorithmic information content H , which we shall now define, as the borderline between $\sum 2^{-f(n)}$ converging and diverging!

Definition. Define an *information content measure* $H(n)$ to be a function of the natural number n having the property that

$$\Omega \equiv \sum 2^{-H(n)} \leq 1, \quad (1)$$

and that $H(n)$ is computable as a limit from above, so that the set

$$\{\text{"}H(n) \leq k\text{"}\} \tag{2}$$

of all upper bounds is r.e. We also allow $H(n) = +\infty$, which contributes zero to the sum (1) since $2^{-\infty} = 0$. It contributes no elements to the set of upper bounds (2).

Note. If H is an information content measure, then it follows immediately from $\sum 2^{-H(n)} = \Omega \leq 1$ that

$$\#\{k | H(k) \leq n\} \leq 2^n.$$

That is, there are at most 2^n natural numbers with information content less than or equal to n .

Theorem I. There is a minimal information content measure H , i.e., an information content measure with the property that for any other information content measure H' , there exists a constant c depending only on H and H' but not on n such that

$$H(n) \leq H'(n) + c.$$

That is, H is smaller, within $O(1)$, than any other information content measure.

Proof. Define H as

$$H(n) = \min_{k \geq 1} [H_k(n) + k], \tag{3}$$

where H_k denotes the information content measure resulting from taking the k th ($k \geq 1$) computer algorithm and patching it, if necessary, so that it gives limits from above and does not violate the $\Omega \leq 1$ condition (1). Then (3) gives H as a computable limit from above, and

$$\Omega = \sum_n 2^{-H(n)} \leq \sum_{k \geq 1} [2^{-k} \sum_n 2^{-H_k(n)}] \leq \sum_{k \geq 1} 2^{-k} = 1.$$

Q.E.D.

Definition. Henceforth we use this minimal information content measure H , and we refer to $H(n)$ as the *information content* of n . We also consider each natural number n to correspond to a bit string s and

vice versa, so that H is defined for strings as well as numbers.² In addition, let $\langle n, m \rangle$ denote a fixed computable one-to-one correspondence between natural numbers and ordered pairs of natural numbers. We define the *joint information content* of n and m to be $H(\langle n, m \rangle)$. Thus H is defined for ordered pairs of natural numbers as well as individual natural numbers. We define the *relative information content* $H(m|n)$ of m relative to n by the equation

$$H(\langle n, m \rangle) \equiv H(n) + H(m|n).$$

That is,

$$H(m|n) \equiv H(\langle n, m \rangle) - H(n).$$

And we define the *mutual information content* $I(n : m)$ of n and m by the equation

$$I(n : m) \equiv H(m) - H(m|n) \equiv H(n) + H(m) - H(\langle n, m \rangle).$$

Note. $\Omega = \sum 2^{-H(n)}$ is just on the borderline between convergence and divergence:

- $\sum 2^{-H(n)}$ converges.
- If $f(n)$ is computable and unbounded, then $\sum 2^{-H(n)+f(n)}$ diverges.
- If $f(n)$ is computable and $\sum 2^{-f(n)}$ converges, then $H(n) \leq f(n) + O(1)$.
- If $f(n)$ is computable and $\sum 2^{-f(n)}$ diverges, then $H(n) \geq f(n)$ infinitely often.

Let us look at a real-valued function $\rho(n)$ that is computable as a limit of rationals from below. And suppose that $\sum \rho(n) \leq 1$. Then $H(n) \leq -\log \rho(n) + O(1)$. So $2^{-H(n)}$ can be thought of as a *maximal* function $\rho(n)$ that is computable in the limit from *below* and has

²It is important to distinguish between the length of a string and its information content! However, a possible source of confusion is the fact that the “natural unit” for both length and information content is the “bit.” Thus one often speaks of an n -bit string, and also of a string whose information content is $\leq n$ bits.

$\sum \rho(n) \leq 1$, instead of thinking of $H(n)$ as a *minimal* function $f(n)$ that is computable in the limit from *above* and has $\sum 2^{-f(n)} \leq 1$.

Lemma I. For all n ,

$$\begin{aligned} H(n) &\leq 2 \log n + c, \\ &\leq \log n + 2 \log \log n + c', \\ &\leq \log n + \log \log n + 2 \log \log \log n + c'' \dots \end{aligned}$$

For infinitely many values of n ,

$$\begin{aligned} H(n) &\geq \log n, \\ &\geq \log n + \log \log n, \\ &\geq \log n + \log \log n + \log \log \log n \dots \end{aligned}$$

Lemma I2. $H(s) \leq |s| + H(|s|) + O(1)$. $|s|$ = the length in bits of the string s .

Proof.

$$\begin{aligned} 1 \geq \Omega &= \sum_n 2^{-H(n)} = \sum_n [2^{-H(n)} \sum_{|s|=n} 2^{-n}] \\ &= \sum_n \sum_{|s|=n} 2^{-[n+H(n)]} \\ &= \sum_s 2^{-[|s|+H(|s|)]}. \end{aligned}$$

The lemma follows by the minimality of H . Q.E.D.

Lemma I3. There are $< 2^{n-k+c}$ n -bit strings s such that $H(s) < n + H(n) - k$. Thus there are $< 2^{n-H(n)-k+c}$ n -bit strings s such that $H(s) < n - k$.

Proof.

$$\sum_n \sum_{|s|=n} 2^{-H(s)} = \sum_s 2^{-H(s)} = \Omega \leq 1.$$

Hence by the minimality of H

$$2^{-H(n)+c} \geq \sum_{|s|=n} 2^{-H(s)},$$

which yields the lemma. Q.E.D.

Lemma I4. If $\psi(n)$ is a computable partial function, then

$$H(\psi(n)) \leq H(n) + c_\psi.$$

Proof.

$$1 \geq \Omega = \sum_n 2^{-H(n)} \geq \sum_y \sum_{\psi(x)=y} 2^{-H(x)}.$$

Note that

$$2^{-a} \geq \sum_i 2^{-b_i} \Rightarrow a \leq \min b_i. \quad (4)$$

The lemma follows by the minimality of H . Q.E.D.

Lemma I5. $H(\langle n, m \rangle) = H(\langle m, n \rangle) + O(1)$.

Proof.

$$\sum_{\langle n, m \rangle} 2^{-H(\langle n, m \rangle)} = \sum_{\langle m, n \rangle} 2^{-H(\langle m, n \rangle)} = \Omega \leq 1.$$

The lemma follows by using the minimality of H in both directions. Q.E.D.

Lemma I6. $H(\langle n, m \rangle) \leq H(n) + H(m) + O(1)$.

Proof.

$$\sum_{\langle n, m \rangle} 2^{-[H(n)+H(m)]} = \Omega^2 \leq 1^2 \leq 1.$$

The lemma follows by the minimality of H . Q.E.D.

Lemma I7. $H(n) \leq H(\langle n, m \rangle) + O(1)$.

Proof.

$$\sum_n \sum_{\langle n, m \rangle} 2^{-H(\langle n, m \rangle)} = \sum_{\langle n, m \rangle} 2^{-H(\langle n, m \rangle)} = \Omega \leq 1.$$

The lemma follows from (4) and the minimality of H . Q.E.D.

Lemma I8. $H(\langle n, H(n) \rangle) = H(n) + O(1)$.

Proof. By Lemma I7,

$$H(n) \leq H(\langle n, H(n) \rangle) + O(1).$$

On the other hand, consider

$$\begin{aligned} \sum_{\substack{\langle n, i \rangle \\ H(n) \leq i}} 2^{-i-1} &= \sum_{\langle n, H(n)+j \rangle} 2^{-H(n)-j-1} \\ &= \sum_n \sum_{k \geq 1} 2^{-H(n)-k} = \sum_n 2^{-H(n)} = \Omega \leq 1. \end{aligned}$$

By the minimality of H ,

$$H(\langle n, H(n) + j \rangle) \leq H(n) + j + O(1).$$

Take $j = 0$. Q.E.D.

Lemma I9. $H(\langle n, n \rangle) = H(n) + O(1)$.

Proof. By Lemma I7,

$$H(n) \leq H(\langle n, n \rangle) + O(1).$$

On the other hand, consider $\psi(n) = \langle n, n \rangle$. By Lemma I4,

$$H(\psi(n)) \leq H(n) + c_\psi.$$

That is,

$$H(\langle n, n \rangle) \leq H(n) + O(1).$$

Q.E.D.

Lemma I10. $H(\langle n, 0 \rangle) = H(n) + O(1)$.

Proof. By Lemma I7,

$$H(n) \leq H(\langle n, 0 \rangle) + O(1).$$

On the other hand, consider $\psi(n) = \langle n, 0 \rangle$. By Lemma I4,

$$H(\psi(n)) \leq H(n) + c_\psi.$$

That is,

$$H(\langle n, 0 \rangle) \leq H(n) + O(1).$$

Q.E.D.

Lemma I11. $H(m|n) \equiv H(\langle n, m \rangle) - H(n) \geq -c$.

(Proof: use Lemma I7.)

Lemma I12. $I(n : m) \equiv H(n) + H(m) - H(\langle n, m \rangle) \geq -c$.

(Proof: use Lemma I6.)

Lemma I13. $I(n : m) = I(m : n) + O(1)$.

(Proof: use Lemma I5.)

Lemma I14. $I(n : n) = H(n) + O(1)$.

(Proof: use Lemma I9.)

Lemma I15. $I(n : 0) = O(1)$.

(Proof: use Lemma I10.)

Note. The further development of this algorithmic version of information theory³ requires the notion of the size in bits of a self-delimiting computer program (Chaitin [3]), which, however, we can do without in this paper.

3. Random Reals

Definition (Martin-Löf [12]). Speaking geometrically, a real r is Martin-Löf random if it is never the case that it is contained in each set of an r.e. infinite sequence A_i of sets of intervals with the property that the measure⁴ of the i th set is always less than or equal to 2^{-i} ,

$$\mu(A_i) \leq 2^{-i}. \quad (5)$$

Here is the definition of a Martin-Löf random real r in a more compact notation:

$$\forall i [\mu(A_i) \leq 2^{-i}] \Rightarrow \neg \forall i [r \in A_i].$$

An equivalent definition, if we restrict ourselves to reals in the unit interval $0 \leq r \leq 1$, may be formulated in terms of bit strings rather than geometrical notions, as follows. Define a *covering* to be an r.e. set of ordered pairs consisting of a natural number i and a bit string s ,

$$\text{Covering} = \{\langle i, s \rangle\},$$

with the property that if $\langle i, s \rangle \in \text{Covering}$ and $\langle i, s' \rangle \in \text{Covering}$, then it is not the case that s is an extension of s' or that s' is an extension

³Compare the original ensemble version of information theory given in Shannon and Weaver [13].

⁴I.e., the sum of the lengths of the intervals, being careful to avoid counting overlapping intervals twice.

of s .⁵ We simultaneously consider A_i to be a set of (finite) bit strings

$$\{s \mid \langle i, s \rangle \in \text{Covering}\}$$

and to be a set of real numbers, namely those which in base-two notation have a bit string in A_i as an initial segment.⁶ Then condition (5) becomes

$$\mu(A_i) = \sum_{\langle i, s \rangle \in \text{Covering}} 2^{-|s|} \leq 2^{-i}, \quad (6)$$

where $|s|$ = the length in bits of the string s .

Note. This is equivalent to stipulating the existence of an arbitrary “regulator of convergence” $f \rightarrow \infty$ that is computable and nondecreasing such that $\mu(A_i) \leq 2^{-f(i)}$. A_0 is only required to have measure ≤ 1 and is sort of useless, since we are working within the unit interval $0 \leq r \leq 1$.⁷

Any real number, considered as a singleton set, is a set of measure zero, but not constructively so! Similarly, the notion of a von Mises’ collective,⁸ which is an infinite bit string such that any place selection rule based on the preceding bits picks out a substring with the same limiting frequency of 0’s and 1’s as the whole string has, is contradictory. But Alonzo Church’s idea, to allow only computable place selection rules, saves the concept.

⁵This is to avoid overlapping intervals and enable us to use the formula (6). It is easy to convert a covering which does not have this property into one that covers exactly the same set and does have this property. How this is done depends on the order in which overlaps are discovered: intervals which are subsets of ones which have already been included in the enumeration of A_i are eliminated, and intervals which are supersets of ones which have already been included in the enumeration must be split into disjoint subintervals, and the common portion must be thrown away.

⁶I.e., the geometrical statement that a point is covered by (the union of) a set of intervals, corresponds in bit string language to the statement that an initial segment of an infinite bit string is contained in a set of finite bit strings. The fact that some reals correspond to two infinite bit strings, e.g., $.100000\dots = .011111\dots$, causes no problems. We are working with closed intervals, which include their endpoints.

⁷It makes $\sum \mu(A_i) \leq 2$ instead of what it should be, namely, ≤ 1 . So A_0 really ought to be abolished!

⁸See Feller [9].

Definition (Solovay [14]). A real r is Solovay random if for any r.e. infinite sequence A_i of sets of intervals with the property that the sum of the measures of the A_i converges

$$\sum \mu(A_i) < \infty,$$

r is contained in at most finitely many of the A_i . In other words,

$$\sum \mu(A_i) < \infty \Rightarrow \exists N \forall (i > N) [r \notin A_i].$$

A real r is weakly Solovay random (“Solovay random with a regulator of convergence”) if for any r.e. infinite sequence A_i of sets of intervals with the property that the sum of the measures of the A_i converges constructively, then r is contained in at most finitely many of the A_i . In other words, a real r is weakly Solovay random if the existence of a computable function $f(n)$ such that for each n ,

$$\sum_{i \geq f(n)} \mu(A_i) \leq 2^{-n},$$

implies that r is contained in at most finitely many of the A_i . That is to say,

$$\forall n \left[\sum_{i \geq f(n)} \mu(A_i) \leq 2^{-n} \right] \Rightarrow \exists N \forall (i > N) [r \notin A_i].$$

Definition (Chaitin [3]). A real r is Chaitin random if (the information content of the initial segment r_n of length n of the base-two expansion of r) does not drop arbitrarily far below n : $\liminf H(r_n) - n > -\infty$.⁹ In other words,

$$\exists c \forall n [H(r_n) \geq n - c].$$

A real r is strongly Chaitin random if (the information content of the initial segment r_n of length n of the base-two expansion of r) eventually

⁹Thus

$$\begin{aligned} n - c \leq H(r_n) &\leq n + H(n) + c' \\ &\leq n + \log n + 2 \log \log n + c'' \end{aligned}$$

by Lemmas I2 and I.

becomes and remains arbitrarily greater than n : $\liminf H(r_n) - n = \infty$. In other words,

$$\forall k \exists N_k \forall (n \geq N_k) [H(r_n) \geq n + k].$$

Note. All these definitions hold with probability one (see Theorem R4).

Theorem R1. Martin-Löf random \Leftrightarrow Chaitin random.

Proof. \neg Martin-Löf \Rightarrow \neg Chaitin. Suppose that a real number r has the property that

$$\forall i [\mu(A_i) \leq 2^{-i} \ \& \ r \in A_i].$$

The series

$$\sum 2^n / 2^{n^2} = \sum 2^{-n^2+n} = 2^{-0} + 2^{-1} + 2^{-2} + 2^{-6} + 2^{-12} + 2^{-20} + \dots$$

obviously converges, and define N so that

$$\sum_{n \geq N} 2^{-n^2+n} \leq 1.$$

(In fact, we can take $N = 2$.) Let the variable s range over bit strings, and consider

$$\sum_{n \geq N} \sum_{s \in A_{n^2}} 2^{-[|s|-n]} = \sum_{n \geq N} 2^n \mu(A_{n^2}) \leq \sum_{n \geq N} 2^{-n^2+n} \leq 1.$$

It follows from the minimality of H that

$$s \in A_{n^2} \text{ and } n \geq N \Rightarrow H(s) \leq |s| - n + c.$$

Thus, since $r \in A_{n^2}$ for all $n \geq N$, there will be infinitely many initial segments r_k of length k of the base-two expansion of r with the property that $r_k \in A_{n^2}$ and $n \geq N$, and for each of these r_k we have

$$H(r_k) \leq |r_k| - n + c.$$

Thus the information content of an initial segment of the base-two expansion of r can drop arbitrarily far below its length.

Proof. \neg Chaitin \Rightarrow \neg Martin-Löf. Suppose that $H(r_n) - n$ can go arbitrarily negative. There are $< 2^{n-k+c}$ n -bit strings s such that $H(s) < n + H(n) - k$ (Lemma I3). Thus there are $< 2^{n-H(n)-k}$ n -bit strings s such that $H(s) < n - k - c$. That is, the probability that an n -bit string s has $H(s) < n - k - c$ is $< 2^{-H(n)-k}$. Summing this over all n , we get

$$\sum_n 2^{-H(n)-k} = 2^{-k} \sum_n 2^{-H(n)} = 2^{-k} \Omega \leq 2^{-k},$$

since $\Omega \leq 1$. Thus if a real r has the property that $H(r_n)$ dips below $n - k - c$ for even one value of n , then r is covered by an r.e. set A_k of intervals & $\mu(A_k) \leq 2^{-k}$. Thus if $H(r_n) - n$ goes arbitrarily negative, for each k we can compute an A_k with $\mu(A_k) \leq 2^{-k}$ & $r \in A_k$, and r is not Martin-Löf random. Q.E.D.

Theorem R2. Solovay random \Leftrightarrow strong Chaitin random.

Proof. \neg Solovay \Rightarrow \neg (strong Chaitin). Suppose that a real number r has the property that it is in infinitely many A_i and

$$\sum \mu(A_i) < \infty.$$

Then there must be an N such that

$$\sum_{i \geq N} \mu(A_i) \leq 1.$$

Hence

$$\sum_{i \geq N} \sum_{s \in A_i} 2^{-|s|} = \sum_{i \geq N} \mu(A_i) \leq 1.$$

It follows from the minimality of H that

$$s \in A_i \text{ and } i \geq N \Rightarrow H(s) \leq |s| + c,$$

i.e., if a bit string s is in A_i and $i \geq N$, then its information content is less than or equal to its size in bits $+c$. Thus $H(r_n) \leq |r_n| + c = n + c$ for infinitely many initial segments r_n of length n of the base-two expansion of r , and it is not the case that $H(r_n) - n \rightarrow \infty$.

Proof. \neg (strong Chaitin) \Rightarrow \neg Solovay. \neg (strong Chaitin) says that there is a k such that for infinitely many values of n we have $H(r_n) -$

$n < k$. The probability that an n -bit string s has $H(s) < n + k$ is $< 2^{-H(n)+k+c}$ (Lemma I3). Let A_n be the r.e. set of all n -bit strings s such that $H(s) < n + k$.

$$\sum \mu(A_n) \leq \sum_n 2^{-H(n)+k+c} = 2^{k+c} \sum_n 2^{-H(n)} = 2^{k+c} \Omega \leq 2^{k+c},$$

since $\Omega \leq 1$. Hence $\sum \mu(A_n) < \infty$ and r is in infinitely many of the A_n , and thus r is not Solovay random. Q.E.D.

Theorem R3. Martin-Löf random \Leftrightarrow weak Solovay random.

Proof. \neg Martin-Löf $\Rightarrow \neg$ (weak Solovay). We are given that $\forall i [r \in A_i]$ and $\forall i [\mu(A_i) \leq 2^{-i}]$. Hence $\sum \mu(A_i)$ converges and the inequality

$$\sum_{i>N} \mu(A_i) \leq 2^{-N}$$

gives us a regulator of convergence.

Proof. \neg (weak Solovay) $\Rightarrow \neg$ Martin-Löf. Suppose

$$\sum_{i \geq f(n)} \mu(A_i) \leq 2^{-n}$$

and the real number r is in infinitely many of the A_i . Let

$$B_n = \bigcup_{i \geq f(n)} A_i.$$

Then $\mu(B_n) \leq 2^{-n}$ and $r \in B_n$, so r is not Martin-Löf random. Q.E.D.

Note. In summary, the five definitions of randomness reduce to at most two:

- Martin-Löf random \Leftrightarrow Chaitin random \Leftrightarrow weak Solovay random.¹⁰
- Solovay random \Leftrightarrow strong Chaitin random.¹¹
- Solovay random \Rightarrow Martin-Löf random.¹²
- Martin-Löf random \Rightarrow Solovay random???

¹⁰Theorems R1 and R3.

¹¹Theorem R2.

¹²Because strong Chaitin \Rightarrow Chaitin.

Theorem R4. With probability one, a real number r is Martin-Löf random and Solovay random.

Proof 1. Since Solovay random \Rightarrow Martin-Löf random (is the converse true?), it is sufficient to show that r is Solovay random with probability one. Suppose

$$\sum \mu(A_i) < \infty,$$

where the A_i are an r.e. infinite sequence of sets of intervals. Then (this is the Borel–Cantelli lemma (Feller [9])),

$$\lim_{N \rightarrow \infty} \Pr\left\{\bigcup_{i \geq N} A_i\right\} \leq \lim_{N \rightarrow \infty} \sum_{i \geq N} \mu(A_i) = 0,$$

and the probability is zero that a real r is in infinitely many of the A_i . But there are only countably many choices for the r.e. sequence of A_i , since there are only countably many algorithms. Since the union of a countable number of sets of measure zero is also of measure zero, it follows that with probability one r is Solovay random.

Proof 2. We use the Borel–Cantelli lemma again. This time we show that the strong Chaitin criterion for randomness, which is equivalent to the Solovay criterion, is true with probability one. Since for each k ,

$$\sum_n \Pr\{H(r_n) < n + k\} \leq 2^{k+c}$$

and thus converges,¹³ it follows that for each k with probability one $H(r_n) < n + k$ only finitely often. Thus, with probability one,

$$\lim_{n \rightarrow \infty} H(r_n) - n = \infty.$$

Q.E.D.

Theorem R5. r Martin-Löf random $\Rightarrow H(r_n) - n$ is unbounded. (Does r Martin-Löf random $\Rightarrow \lim H(r_n) - n = \infty$?)

Proof. We shall prove the theorem by assuming that $H(r_n) - n < c$ for all n and deducing that r cannot be Martin-Löf random. Let c' be the constant of Lemma I3, so that the number of k -bit strings s with $H(s) < k + H(k) - i$ is $< 2^{k-i+c'}$

¹³See the second half of the proof of Theorem R2.

Consider r_k for $k = 1$ to $2^{n+c+c'}$. We claim that the probability of the event A_n that r simultaneously satisfies the $2^{n+c+c'}$ inequalities

$$H(r_k) < k + c \quad (k = 1, \dots, 2^{n+c+c'})$$

is $< 2^{-n}$. (See the next paragraph for the proof of this claim.) Thus we have an r.e. infinite sequence A_n of sets of intervals with measure $\mu(A_n) \leq 2^{-n}$ which all contain r . Hence r is not Martin-Löf random.

Proof of Claim. Since $\sum 2^{-H(k)} = \Omega \leq 1$, there is a k between 1 and $2^{n+c+c'}$ such that $H(k) \geq n + c + c'$. For this value of k ,

$$\Pr\{H(r_k) < k + c\} \leq 2^{-H(k)+c+c'} \leq 2^{-n},$$

since the number of k -bit strings s with $H(s) < k + H(k) - i$ is $< 2^{k-i+c'}$ (Lemma I3). Q.E.D.

Theorem R6. Ω is a Martin-Löf–Chaitin–weak Solovay random real number. More generally, if N is an infinite r.e. set of natural numbers, then

$$\theta = \sum_{n \in N} 2^{-H(n)}$$

is a Martin-Löf–Chaitin–weak Solovay random real.¹⁴

Proof. Since $H(n)$ can be computed as a limit from above, $2^{-H(n)}$ can be computed as a limit from below. It follows that given θ_k , the first k bits of the base-two expansion *without infinitely many consecutive trailing zeros*¹⁵ of the real number θ , one can calculate the finite set of all $n \in N$ such that $H(n) \leq k$, and then, since N is infinite, one can calculate an $n \in N$ with $H(n) > k$. That is, there is a computable partial function ψ such that

$$\psi(\theta_k) = \text{a natural number } n \text{ with } H(n) > k.$$

¹⁴Incidentally, this implies that θ is not a computable real number, from which it follows that $0 < \theta < 1$, that θ is irrational, and even that θ is transcendental.

¹⁵If there is a choice between ending the base-two expansion of θ with infinitely many consecutive zeros or with infinitely many consecutive ones (i.e., if θ is a dyadic rational), then we must choose the infinity of consecutive ones. This is to ensure that considered as real numbers

$$\theta_k < \theta < \theta_k + 2^{-k}.$$

Of course, it will follow from this theorem that θ must be an irrational number, so this situation cannot actually occur, but we don't know that yet!

But by Lemma I4,

$$H(\psi(\theta_k)) \leq H(\theta_k) + c_\psi.$$

Hence

$$k < H(\psi(\theta_k)) \leq H(\theta_k) + c_\psi$$

and

$$H(\theta_k) > k - c_\psi.$$

Thus θ is Chaitin random, and by Theorems R1 and R3 it is also Martin-Löf random and weakly Solovay random. Q.E.D.

Theorem R7. There is an exponential diophantine equation

$$L(n, x_1, \dots, x_m) = R(n, x_1, \dots, x_m)$$

which has only finitely many solutions x_1, \dots, x_m if the n th bit of Ω is a 0, and which has infinitely many solutions x_1, \dots, x_m if the n th bit of Ω is a 1.

Proof. Since $H(n)$ can be computed as a limit from above, $2^{-H(n)}$ can be computed as a limit from below. It follows that

$$\Omega = \sum 2^{-H(n)}$$

is the limit from below of a computable sequence $\omega_1 \leq \omega_2 \leq \omega_3 \leq \dots$ of rational numbers

$$\Omega = \lim_{k \rightarrow \infty} \omega_k.$$

This sequence converges extremely slowly! The exponential diophantine equation $L = R$ is constructed from the sequence ω_k by using the theorem that “every r.e. relation has a *singlefold* exponential diophantine representation” (Jones and Matijasevič [11]). Since the assertion that

“the n th bit of ω_k is a 1”

is an r.e. relation between n and k (in fact, it is a recursive relation), the theorem of Jones and Matijasevič yields an equation

$$L(n, k, x_2, \dots, x_m) = R(n, k, x_2, \dots, x_m)$$

involving only additions, multiplications, and exponentiations of natural number constants and variables, and this equation has exactly one

solution x_2, \dots, x_m in natural numbers if the n th bit of the base-two expansion of ω_k is a 1, and it has no solution x_2, \dots, x_m in natural numbers if the n th bit of the base-two expansion of ω_k is a 0. The number of different m -tuples x_1, \dots, x_m of natural numbers which are solutions of the equation

$$L(n, x_1, \dots, x_m) = R(n, x_1, \dots, x_m)$$

is therefore infinite if the n th bit of the base-two expansion of Ω is a 1, and it is finite if the n th bit of the base-two expansion of Ω is a 0. Q.E.D.

4. Incompleteness Theorems

Having developed the necessary information-theoretic formalism in Section 2, and having studied the notion of a random real in Section 3, we can now begin to derive incompleteness theorems.

The setup is as follows. The axioms of a formal theory are considered to be encoded as a single finite bit string, the rules of inference are considered to be an algorithm for enumerating the theorems given the axioms, and in general we shall fix the rules of inference and vary the axioms. More formally, the rules of inference F may be considered to be an r.e. set of propositions of the form

$$\text{“Axioms } \vdash_F \text{ Theorem.”}$$

The r.e. set of theorems deduced from the axiom A is determined by selecting from the set F the theorems in those propositions which have the axiom A as an antecedent. In general we will consider the rules of inference F to be fixed and study what happens as we vary the axioms A . By an n -bit theory we shall mean the set of theorems deduced from an n -bit axiom.

4.1. Incompleteness Theorems for Lower Bounds on Information Content

Let us start by rederiving within our current formalism an old and very basic result, which states that even though most strings are random, one can never prove that a specific string has this property.

If one produces a bit string s by tossing a coin n times, 99.9% of the time it will be the case that $H(s) \approx n + H(n)$ (Lemmas I2 and I3). In fact, if one lets n go to infinity, with probability one $H(s) > n$ for all but finitely many n (Theorem R4). However,

Theorem LB (Chaitin [1,2,4]). Consider a formal theory all of whose theorems are assumed to be true. Within such a formal theory a specific string cannot be proven to have information content more than $O(1)$ greater than the information content of the axioms of the theory. That is, if “ $H(s) \geq n$ ” is a theorem only if it is true, then it is a theorem only if $n \leq H(\text{axioms}) + O(1)$. Conversely, there are formal theories whose axioms have information content $n + O(1)$ in which it is possible to establish all true propositions of the form “ $H(s) \geq n$ ” and of the form “ $H(s) = k$ ” with $k < n$.

Proof. Consider the enumeration of the theorems of the formal axiomatic theory in order of the size of their proofs. For each natural number k , let s^* be the string in the theorem of the form “ $H(s) \geq n$ ” with $n > H(\text{axioms}) + k$ which appears first in the enumeration. On the one hand, if all theorems are true, then

$$H(\text{axioms}) + k < H(s^*).$$

On the other hand, the above prescription for calculating s^* shows that

$$s^* = \psi(\langle \langle \text{axioms}, H(\text{axioms}) \rangle, k \rangle) \quad (\psi \text{ partial recursive}),$$

and thus

$$H(s^*) \leq H(\langle \langle \text{axioms}, H(\text{axioms}) \rangle, k \rangle) + c_\psi \leq H(\text{axioms}) + H(k) + O(1).$$

Here we have used the subadditivity of information $H(\langle s, t \rangle) \leq H(s) + H(t) + O(1)$ (Lemma I6) and the fact that $H(\langle s, H(s) \rangle) \leq H(s) + O(1)$ (Lemma I8). It follows that

$$H(\text{axioms}) + k < H(s^*) \leq H(\text{axioms}) + H(k) + O(1),$$

and thus

$$k < H(k) + O(1).$$

However, this inequality is false for all $k \geq k_0$, where k_0 depends only on the rules of inference. A contradiction is avoided only if s^* does not

exist for $k = k_0$, i.e., it is impossible to prove in the formal theory that a specific string has H greater than $H(\text{axioms}) + k_0$.

Proof of Converse. The set T of all true propositions of the form “ $H(s) \leq k$ ” is r.e. Choose a fixed enumeration of T without repetitions, and for each natural number n , let s^* be the string in the last proposition of the form “ $H(s) \leq k$ ” with $k < n$ in the enumeration. Let

$$\Delta = n - H(s^*) > 0.$$

Then from $s^*, H(s^*), \&\Delta$ we can calculate $n = H(s^*) + \Delta$, then all strings s with $H(s) < n$, and then a string s_n with $H(s_n) \geq n$. Thus

$$n \leq H(s_n) = H(\psi(\langle\langle s^*, H(s^*) \rangle\rangle, \Delta)) \quad (\psi \text{ partial recursive}),$$

and so

$$\begin{aligned} n \leq H(\langle\langle s^*, H(s^*) \rangle\rangle, \Delta) + c_\psi &\leq H(s^*) + H(\Delta) + O(1) \\ &\leq n + H(\Delta) + O(1) \end{aligned} \quad (7)$$

by Lemmas I6 and I8. The first line of (7) implies that

$$\Delta \equiv n - H(s^*) \leq H(\Delta) + O(1),$$

which implies that Δ and $H(\Delta)$ are both bounded. Then the second line of (7) implies that

$$H(\langle\langle s^*, H(s^*) \rangle\rangle, \Delta) = n + O(1).$$

The triple $\langle\langle s^*, H(s^*) \rangle\rangle, \Delta$ is the desired axiom: it has information content $n + O(1)$, and by enumerating T until all true propositions of the form “ $H(s) \leq k$ ” with $k < n$ have been discovered, one can immediately deduce all true propositions of the form “ $H(s) \geq n$ ” and of the form “ $H(s) = k$ ” with $k < n$. Q.E.D.

4.2. Incompleteness Theorems for Random Reals: First Approach

In this section we begin our study of incompleteness theorems for random reals. We show that any particular formal theory can enable one

to determine at most a finite number of bits of Ω . In the following sections (4.3 and 4.4) we express the upper bound on the number of bits of Ω which can be determined, in terms of the axioms of the theory; for now, we just show that an upper bound exists. We shall not use any ideas from algorithmic information theory until Section 4.4; for now (Sections 4.2 and 4.3) we only make use of the fact that Ω is Martin-Löf random.

If one tries to guess the bits of a random sequence, the average number of correct guesses before failing is exactly 1 guess! Reason: if we use the fact that the expected value of a sum is equal to the sum of the expected values, the answer is the sum of the chance of getting the first guess right, plus the chance of getting the first and the second guesses right, plus the chance of getting the first, second and third guesses right, etc.,

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots = 1.$$

Or if we directly calculate the expected value as the sum of (the # right till first failure) \times (the probability),

$$\begin{aligned} & 0 \times \frac{1}{2} + 1 \times \frac{1}{4} + 2 \times \frac{1}{8} + 3 \times \frac{1}{16} + 4 \times \frac{1}{32} + \dots \\ &= 1 \times \sum_{k>1} 2^{-k} + 1 \times \sum_{k>2} 2^{-k} + 1 \times \sum_{k>3} 2^{-k} + \dots \\ &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = 1. \end{aligned}$$

On the other hand (see the next section), if we are allowed to try 2^n times a series of n guesses, one of them will always get it right, if we try all 2^n different possible series of n guesses.

Theorem X. Any given formal theory T can yield only finitely many (scattered) bits of (the base-two expansion of) Ω .

When we say that a theory yields a bit of Ω , we mean that it enables us to determine its position and its 0/1 value.

Proof. Consider a theory T , an r.e. set of true assertions of the form

“The n th bit of Ω is 0.”

“The n th bit of Ω is 1.”

Here n denotes specific natural numbers.

If T provides k different (scattered) bits of Ω , then that gives us a covering A_k of measure 2^{-k} which includes Ω : Enumerate T until k bits of Ω are determined, then the covering is all bit strings up to the last determined bit with all determined bits okay. If n is the last determined bit, this covering will consist of 2^{n-k} n -bit strings, and will have measure $2^{n-k}/2^n = 2^{-k}$.

It follows that if T yields infinitely many different bits of Ω , then for any k we can produce by running through all possible proofs in T a covering A_k of measure 2^{-k} which includes Ω . But this contradicts the fact that Ω is Martin-Löf random. Hence T yields only finitely many bits of Ω . Q.E.D.

Corollary X. Since by Theorem R7 Ω can be encoded into an exponential diophantine equation

$$L(n, x_1, \dots, x_m) = R(n, x_1, \dots, x_m), \quad (8)$$

it follows that any given formal theory can permit one to determine whether (8) has finitely or infinitely many solutions x_1, \dots, x_m , for only finitely many specific values of the parameter n .

4.3. Incompleteness Theorems for Random Reals: |Axioms|

Theorem A. If $\sum 2^{-f(n)} \leq 1$ and f is computable, then there is a constant c_f with the property that no n -bit theory ever yields more than $n + f(n) + c_f$ bits of Ω .

Proof. Let A_k be the event that there is at least one n such that there is an n -bit theory that yields $n + f(n) + k$ or more bits of Ω .

$$\begin{aligned} \Pr\{A_k\} &\leq \sum_n \left[\left(\begin{array}{c} 2^n \\ n\text{-bit} \\ \text{theories} \end{array} \right) \left(\begin{array}{c} 2^{-[n+f(n)+k]} \\ \text{probability that yields} \\ n + f(n) + k \text{ bits of } \Omega \end{array} \right) \right] \\ &= 2^{-k} \sum_n 2^{-f(n)} \leq 2^{-k} \end{aligned}$$

since $\sum 2^{-f(n)} \leq 1$. Hence $\Pr\{A_k\} \leq 2^{-k}$, and $\sum \Pr\{A_k\}$ also converges. Thus only finitely many of the A_k occur (Borel–Cantelli lemma (Feller

[9])). That is,

$$\lim_{N \rightarrow \infty} \Pr\left\{ \bigcup_{k > N} A_k \right\} \leq \sum_{k > N} \Pr\{A_k\} \leq 2^{-N} \rightarrow 0.$$

More Detailed Proof. Assume the opposite of what we want to prove, namely that for every k there is at least one n -bit theory that yields $n + f(n) + k$ bits of Ω . From this we shall deduce that Ω cannot be Martin-Löf random, which is impossible.

To get a covering A_k of Ω with measure $\leq 2^{-k}$, consider a specific n and all n -bit theories. Start generating theorems in each n -bit theory until it yields $n + f(n) + k$ bits of Ω (it does not matter if some of these bits are wrong). The measure of the set of possibilities for Ω covered by the n -bit theories is thus $\leq 2^n 2^{-n-f(n)-k} = 2^{-f(n)-k}$. The measure $\mu(A_k)$ of the union of the set of possibilities for Ω covered by n -bit theories with any n is thus

$$\leq \sum_n 2^{-f(n)-k} = 2^{-k} \sum_n 2^{-f(n)} \leq 2^{-k} \quad (\text{since } \sum_n 2^{-f(n)} \leq 1).$$

Thus Ω is covered by A_k and $\mu(A_k) \leq 2^{-k}$ for every k if there is always an n -bit theory that yields $n + f(n) + k$ bits of Ω , which is impossible. Q.E.D.

Corollary A. If $\sum 2^{-f(n)}$ converges and f is computable, then there is a constant c_f with the property that no n -bit theory ever yields more than $n + f(n) + c_f$ bits of Ω .

Proof. Choose c so that $\sum 2^{-f(n)} \leq 2^c$. Then $\sum 2^{-[f(n)+c]} \leq 1$, and we can apply Theorem A to $f'(n) = f(n) + c$. Q.E.D.

Corollary A2. Let $\sum 2^{-f(n)}$ converge and f be computable as before. If $g(n)$ is computable, then there is a constant $c_{f,g}$ with the property that no $g(n)$ -bit theory ever yields more than $g(n) + f(n) + c_{f,g}$ bits of Ω . For example, consider N of the form 2^{2^n} . For such N , no N -bit theory ever yields more than $N + f(\log \log N) + c_{f,g}$ bits of Ω .

Note. Thus for n of special form, i.e., which have concise descriptions, we get better upper bounds on the number of bits of Ω which are yielded by n -bit theories. This is a foretaste of the way algorithmic information theory will be used in Theorem C and Corollary C2 (Sect. 4.4).

Lemma for Second Borel–Cantelli Lemma! For any finite set $\{x_k\}$ of non-negative real numbers,

$$\prod(1 - x_k) \leq \frac{1}{\sum x_k}.$$

Proof. If x is a real number, then

$$1 - x \leq \frac{1}{1 + x}.$$

Thus

$$\prod(1 - x_k) \leq \frac{1}{\prod(1 + x_k)} \leq \frac{1}{\sum x_k},$$

since if all the x_k are non-negative

$$\prod(1 + x_k) \geq \sum x_k.$$

Q.E.D.

Second Borel–Cantelli Lemma (Feller [9]). Suppose that the events A_n have the property that it is possible to determine whether or not the event A_n occurs by examining the first $f(n)$ bits of Ω , where f is a computable function. If the events A_n are mutually independent and $\sum \Pr\{A_n\}$ diverges, then Ω has the property that infinitely many of the A_n must occur.

Proof. Suppose on the contrary that Ω has the property that only finitely many of the events A_n occur. Then there is an N such that the event A_n does not occur if $n \geq N$. The probability that none of the events $A_N, A_{N+1}, \dots, A_{N+k}$ occur is, since the A_n are mutually independent, precisely

$$\prod_{i=0}^k (1 - \Pr\{A_{N+i}\}) \leq \frac{1}{\left[\sum_{i=0}^k \Pr\{A_{N+i}\}\right]},$$

which goes to zero as k goes to infinity. This would give us arbitrarily small covers for Ω , which contradicts the fact that Ω is Martin-Löf random. Q.E.D.

Theorem B. If $\sum 2^{n-f(n)}$ diverges and f is computable, then infinitely often there is a run of $f(n)$ zeros between bits 2^n & 2^{n+1} of Ω

($2^n \leq \text{bit} < 2^{n+1}$). Hence there are rules of inference which have the property that there are infinitely many N -bit theories that yield (the first) $N + f(\log N)$ bits of Ω .

Proof. We wish to prove that infinitely often Ω must have a run of $k = f(n)$ consecutive zeros between its 2^n th & its 2^{n+1} th bit position. There are 2^n bits in the range in question. Divide this into nonoverlapping blocks of $2k$ bits each, giving a total of $2^n/2k$ blocks. The chance of having a run of k consecutive zeros in each block of $2k$ bits is

$$\geq \frac{k2^{k-2}}{2^{2k}}. \quad (9)$$

Reason:

- There are $2k - k + 1 \geq k$ different possible choices for where to put the run of k zeros in the block of $2k$ bits.
- Then there must be a 1 at each end of the run of 0's, but the remaining $2k - k - 2 = k - 2$ bits can be anything.
- This may be an underestimate if the run of 0's is at the beginning or end of the $2k$ bits, and there is no room for endmarker 1's.
- There is no room for another 10^k1 to fit in the block of $2k$ bits, so we are not overestimating the probability by counting anything twice.

Summing (9) over all $2^n/2k$ blocks and over all n , we get

$$\geq \sum_n \left[\frac{k2^{k-2}}{2^{2k}} \frac{2^n}{2k} \right] = \frac{1}{8} \sum_n 2^{n-k} = \frac{1}{8} \sum 2^{n-f(n)} = \infty.$$

Invoking the second Borel–Cantelli lemma (if the events A_i are independent and $\sum \Pr\{A_i\}$ diverges, then infinitely many of the A_i must occur), we are finished. Q.E.D.

Corollary B. If $\sum 2^{-f(n)}$ diverges and f is computable and nondecreasing, then infinitely often there is a run of $f(2^{n+1})$ zeros between bits 2^n & 2^{n+1} of Ω ($2^n \leq \text{bit} < 2^{n+1}$). Hence there are infinitely many N -bit theories that yield (the first) $N + f(N)$ bits of Ω .

Proof. If $\sum 2^{-f(n)}$ diverges and f is computable and nondecreasing, then by the Cauchy condensation test

$$\sum 2^n 2^{-f(2^n)}$$

also diverges, and therefore so does

$$\sum 2^n 2^{-f(2^{n+1})}.$$

Hence, by Theorem B, infinitely often there is a run of $f(2^{n+1})$ zeros between bits 2^n and 2^{n+1} . Q.E.D.

Corollary B2. If $\sum 2^{-f(n)}$ diverges and f is computable, then infinitely often there is a run of $n + f(n)$ zeros between bits 2^n & 2^{n+1} of Ω ($2^n \leq \text{bit} < 2^{n+1}$). Hence there are infinitely many N -bit theories that yield (the first) $N + \log N + f(\log N)$ bits of Ω .

Proof. Take $f(n) = n + f'(n)$ in Theorem B. Q.E.D.

Theorem AB. (a) There is a c with the property that no n -bit theory ever yields more than $n + \log n + 2 \log \log n + c$ (scattered) bits of Ω .

(b) There are infinitely many n -bit theories that yield (the first) $n + \log n + \log \log n$ bits of Ω .

Proof. Using the Cauchy condensation test, we have seen (beginning of Sect. 2) that

$$\begin{aligned} \text{(a)} \quad & \sum \frac{1}{n(\log n)^2} \text{ converges and} \\ \text{(b)} \quad & \sum \frac{1}{n \log n} \text{ diverges.} \end{aligned}$$

The theorem follows immediately from Corollaries A and B. Q.E.D.

4.4. Incompleteness Theorems for Random Reals: H(Axioms)

Theorem C is a remarkable extension of Theorem R6:

- We have seen that the information content of [knowing the first n bits of Ω] is $\geq n - c$.

- Now we show that the information content of [knowing any n bits of Ω (their positions and 0/1 values)] is $\geq n - c$.

Lemma C. $\sum_n \#\{s|H(s) < n\}2^{-n} = \Omega \leq 1$.

Proof.

$$\begin{aligned}
1 \geq \Omega &= \sum_s 2^{-H(s)} = \sum_n \#\{s|H(s) = n\}2^{-n} \\
&= \sum_n \#\{s|H(s) = n\}2^{-n} \sum_{k \geq 1} 2^{-k} \\
&= \sum_n \sum_{k \geq 1} \#\{s|H(s) = n\}2^{-n-k} \\
&= \sum_n \#\{s|H(s) < n\}2^{-n}.
\end{aligned}$$

Q.E.D.

Theorem C. If a theory has $H(\text{axiom}) < n$, then it can yield at most $n + c$ (scattered) bits of Ω .

Proof. Consider a particular k and n . If there is an axiom with $H(\text{axiom}) < n$ which yields $n + k$ scattered bits of Ω , then even without knowing which axiom it is, we can cover Ω with an r.e. set of intervals of measure

$$\leq \begin{pmatrix} \#\{s|H(s) < n\} \\ \# \text{ of axioms} \\ \text{with } H < n \end{pmatrix} \begin{pmatrix} 2^{-n-k} \\ \text{measure of set of} \\ \text{possibilities for } \Omega \end{pmatrix} = \#\{s|H(s) < n\}2^{-n-k}.$$

But by the preceding lemma, we see that

$$\sum_n \#\{s|H(s) < n\}2^{-n-k} = 2^{-k} \sum_n \#\{s|H(s) < n\}2^{-n} \leq 2^{-k}.$$

Thus if even one theory with $H < n$ yields $n + k$ bits of Ω , for any n , we get a cover for Ω of measure $\leq 2^{-k}$. This can only be true for finitely many values of k , or Ω would not be Martin-Löf random. Q.E.D.

Corollary C. No n -bit theory ever yields more than $n + H(n) + c$ bits of Ω .

(Proof: Theorem C and by Lemma I2, $H(\text{axiom}) \leq |\text{axiom}| + H(|\text{axiom}|) + c$.)

Lemma C2. If $g(n)$ is computable and unbounded, then $H(n) < g(n)$ for infinitely many values of n .

Proof. Define the inverse of g as

$$g^{-1}(n) = \min_{g(k) \geq n} k.$$

Then using Lemmas I and I4 we see that for all sufficiently large values of n ,

$$H(g^{-1}(n)) \leq H(n) + O(1) \leq O(\log n) < n \leq g(g^{-1}(n)).$$

That is, $H(k) < g(k)$ for all $k = g^{-1}(n)$ and n sufficiently large. Q.E.D.

Corollary C2. Let $g(n)$ be computable and unbounded. For infinitely many n , no n -bit theory yields more than $n + g(n) + c$ bits of Ω .

(Proof: Corollary C and Lemma C2.)

Note. In appraising Corollaries C and C2, the trivial formal systems in which there is always an n -bit axiom that yields the first n bits of Ω should be kept in mind. Also, compare Corollaries C and A, and Corollaries C2 and A2.

In summary,

Theorem D. There is an exponential diophantine equation

$$L(n, x_1, \dots, x_m) = R(n, x_1, \dots, x_m) \tag{10}$$

which has only finitely many solutions x_1, \dots, x_m if the n th bit of Ω is a 0, and which has infinitely many solutions x_1, \dots, x_m if the n th bit of Ω is a 1. Let us say that a formal theory “settles k cases” if it enables one to prove that the number of solutions of (10) is finite or that it is infinite for k specific values (possibly scattered) of the parameter n . Let $f(n)$ and $g(n)$ be computable functions.

- $\sum 2^{-f(n)} < \infty \Rightarrow$ all n -bit theories settle $\leq n + f(n) + O(1)$ cases.
- $\sum 2^{-f(n)} = \infty$ and $f(n) \leq f(n+1) \Rightarrow$ for infinitely many n , there is an n -bit theory that settles $\geq n + f(n)$ cases.
- $H(\text{theory}) < n \Rightarrow$ it settles $\leq n + O(1)$ cases.

- n -bit theory \Rightarrow it settles $\leq n + H(n) + O(1)$ cases.
- g unbounded \Rightarrow for infinitely many n , all n -bit theories settle $\leq n + g(n) + O(1)$ cases.

Proof. The theorem combines Theorem R7, Corollaries A and B, Theorem C, and Corollaries C and C2. Q.E.D.

5. Conclusion

In conclusion, we have seen that proving whether particular exponential diophantine equations have finitely or infinitely many solutions, is absolutely intractable. Such questions escape the power of mathematical reasoning. This is a region in which mathematical truth has no discernible structure or pattern and appears to be completely random. These questions are completely beyond the power of human reasoning. Mathematics cannot deal with them.

Quantum physics has shown that there is randomness in nature. I believe that we have demonstrated in this paper that randomness is already present in pure mathematics. This does not mean that the universe and mathematics are lawless, it means that laws of a different kind apply: statistical laws.

References

- [1] G. J. CHAITIN, Information-theoretic computational complexity, *IEEE Trans. Inform. Theory* **20** (1974), 10–15.
- [2] G. J. CHAITIN, Randomness and mathematical proof, *Sci. Amer.* **232**, No. 5 (1975), 47–52.
- [3] G. J. CHAITIN, A theory of program size formally identical to information theory, *J. Assoc. Comput. Mach.* **22** (1975), 329–340.
- [4] G. J. CHAITIN, Gödel's theorem and information, *Internat. J. Theoret. Phys.* **22** (1982), 941–954.

- [5] G. J. CHAITIN, Randomness and Gödel's theorem, "Mondes en Développement," Vol. 14, No. 53, in press.
- [6] R. COURANT and H. ROBBINS, "What is Mathematics?," Oxford Univ. Press, London, 1941.
- [7] M. DAVIS, H. PUTNAM, and J. ROBINSON, The decision problem for exponential diophantine equations, *Ann. Math.* **74** (1961), 425–436.
- [8] M. DAVIS, "The Undecidable—Basic Papers on Undecidable Propositions, Unsolvability Problems and Computable Functions," Raven, New York, 1965.
- [9] W. FELLER, "An Introduction to Probability Theory and Its Applications, I," Wiley, New York, 1970.
- [10] G. H. HARDY, "A Course of Pure Mathematics," 10th ed., Cambridge Univ. Press, London, 1952.
- [11] J. P. JONES and Y. V. MATIJASEVIČ, Register machine proof of the theorem on exponential diophantine representation of enumerable sets, *J. Symbolic Logic* **49** (1984), 818–829.
- [12] P. MARTIN-LÖF, The definition of random sequences, *Inform. Control* **9** (1966), 602–619.
- [13] C. E. SHANNON and W. WEAVER, "The Mathematical Theory of Communication," Univ. of Illinois Press, Urbana, 1949.
- [14] R. M. SOLOVAY, Private communication, 1975.
- [15] A. M. TURING, On computable numbers, with an application to the Entscheidungsproblem, *Proc. London Math. Soc.* **42** (1937), 230–265; also in [8].