Randomness
and
Mathematical Proof*

Behdad Esfahbod
behdad@behdad.org

Omid Etesami
etesami@ce.sharif.edu

Computer Engineering Department
Sharif University of Technology
Tehran, Iran

November 30, 2002

*by Gregory J. Chaitin, Scientific American **232**-5 (May 1975), pp. 47–52, available from http://www.cs.auckland.ac.nz/CDMTCS/chaitin/
slides from http://behdad.org/download/presentation/chaitin75

Although randomness can be precisely defined and can even be measured, a given number cannot be proved to be random. This enigma establishes a limit to what is possible in mathematics.

What is a Random Sequence?

01010101010101010101
01101100110111100010

- Our intuitive notion
- How we can predict the following numbers?
- The origin has been random?
- More sensible definition required, one that does not contradict the intuitive concept of a "patternless" number.

An Algorithmic Definition

- Roots in *information theory*

- A simple example:

  – You want to send a friend the trigonometric functions.
  You can just send him $e^{ix} = \cos x + i \sin x$, and how to use it.

  – You want to send her the scores of all the major-league games played in the last century.
  You have no alternative to sending the entire list of scores.

- Here we want to communicate with a digital computer.

An Algorithmic Definition (continued)

- Now there's a difference:
  - `print 01 ten times.`
    $\rightarrow$ `print 01 a million times.`
  - `print 011011001101111100010.`
    $\rightarrow$ `print 011011001101111100010....`
- All random numbers are *incompressible*.
- Definition proposed independently by A. N. Kolmogorov, and G. J. Chaitin in 1960s.

Solomonoff's Model of Inductive Method

- A scientist observes a series of binary digits.
- Then seeks to explain these observations through a theory.
- This theory, can be regarded as an algorithm.
- Which can generate the series, and extend them, that is predicting.
- There are always several competing theories.
- The scientist is to choose among them.
- The model demands that the smallest algorithm be selected.

Example of Inductive Reasoning

- **Observations:** 0101010101

- **Predictions:** 01010101010101010101
  Theory: Ten repetitions of 01
  Size of Theory: 21 characters

- **Predictions:** 01010101010000000000
  Theory: Five repetitions of 01 followed by ten 0's
  Size of Theory: 42 characters

Given differing theories of apparently equal merit, the simplest is to be preferred.

The Machine

- Different machines communicate through different computer languages.
- Sets of instructions expressed in a specific language might require more or fewer bits than in another language.
- Any machine can simulate another machine with just a fixed number of bits.
- So, the choice of computer matters very little.
- We choose for our calculations an ideal computer. Input and output are in binary.

Minimal Programs

- Infinite number of algorithms for generating any specified series of numbers.
- The programs of greatest interest, are the smallest ones.
- The smallest programs are called minimal programs.
- For a given series, there may be only one minimal program, or there may be many.
- What is important is that there always *exists* a minimal program.

Minimal Programs (continued)

- Any minimal program is necessarily random, independent of the series it generates (as we have defined randomness this way).

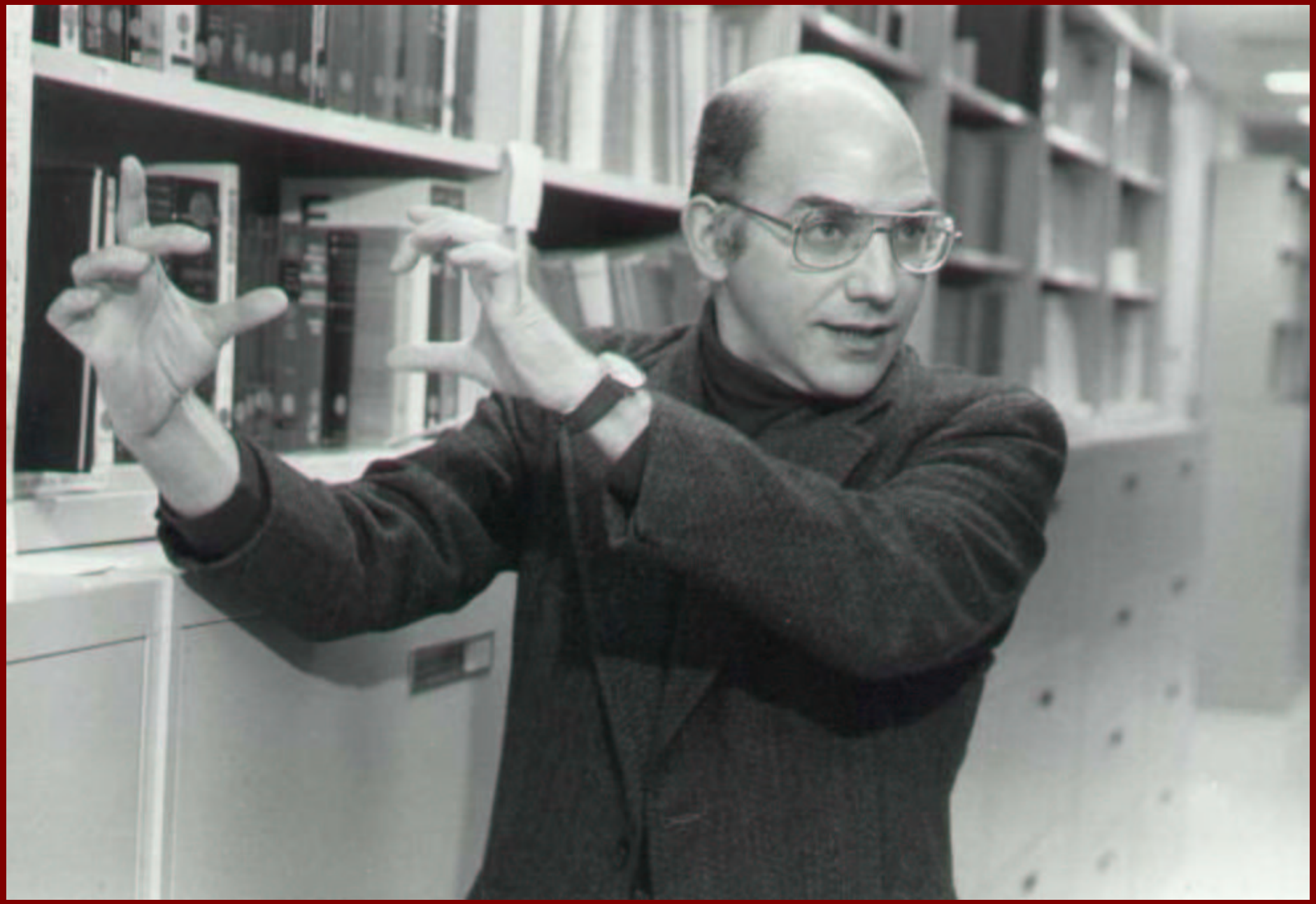
- Scenario:
  - Program P is a minimal program for series S.
  - If P is not random, by definition there's another program $\mathcal{P}$, substantially smaller than P, that will generate it.
  - Now we can produce S this way: *From* $\mathcal{P}$ *calculate* P, *then from* P *calculate* S.
  - This program is only a few bits longer than $\mathcal{P}$.
  - P is not a minimal program.

Complexity

- The complexity of a series is the number of input bits to a machine.
- It's therefore equal to the size of the minimal programs of the series.
- In our new terms: **A random series of digits is one whose complexity is approximately equal to its size in bits.**
- A number with n digits, may be of complexity $n-1$, $n-10$, $n-100$, ... The exact border between random and non-random numbers remains somewhat arbitrary.

Properties of Random Numbers

- A sequence of length n of all 0's, has complexity of about $\log_2 n$.
- A sequence of length n, which the relative frequencies of 1's to 0's is three to one, has complexity of about $4n/5$.
- In any random binary series, the frequencies of 0's and 1's must be very close to one-half.
- With a simple calculation, more that 99.9% of all n-digit numbers are random, for large n's



Formal Systems

- It's fairly easy to prove that a specific series is not random; just to find a small program that generates it.

- To show that a particular series is random, must prove that no small program exists for calculating it.

- Chaitin proved that no such proof exists.

- But, **what constitutes a valid proof in mathematics, how is such a proof to be recognized?**

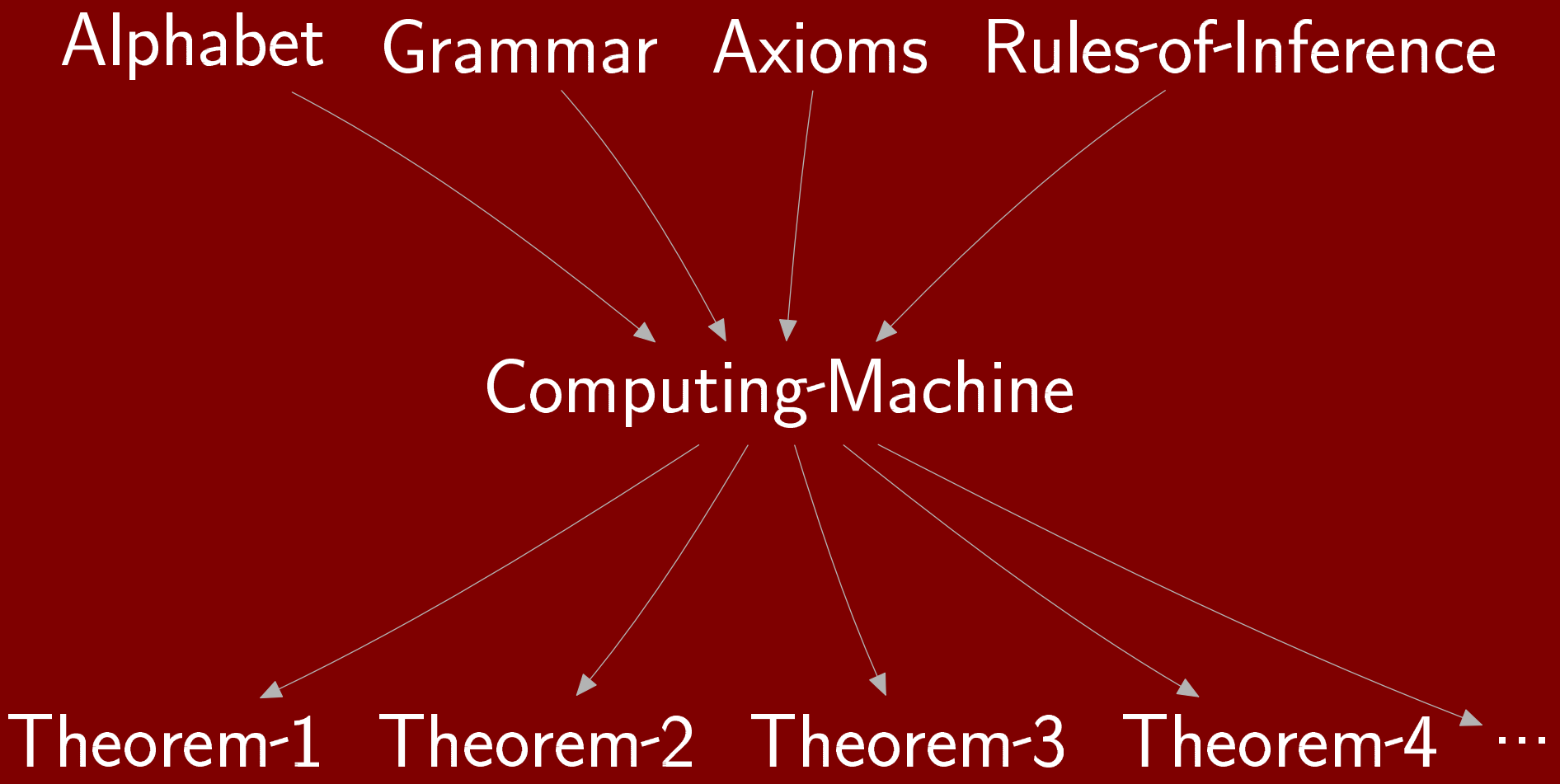
Formal Systems (continued)

- David Hilbert defined an artificial language, in which valid proofs could be found mechanically.

- Gödel proved that there is no such perfect language.

- Hilbert's model of a Formal System:
  - A finite alphabet of symbols
  - An unambiguous grammer defining meaningful statements
  - A finite list of *axioms*, which are initial assumptions
  - A finite list of *rules of inference*, which theorems can be deduced from the axioms or other theorems with

A Formal System



Unprovable Statements

- Gödel showed in 1931 that Hilbert's plan for a completely systematic mathematics cannot be fulfilled.

- Did this by constructing an assertion in Hilbert's formal system, that is true, but cannot be proved in the system.

- No matter how large and complicated the system is.

- Any such system is **incomplete**.

- There can be no definitive answer to the question "What is a valid proof?"

Gödel's Incompleteness Theorem

- Based on Epimenides Paradox: *"This statement is false."* This assertion is neither true nor false.
- Gödel replaced truth with provability: *"This statement is unprovable."* This assertion is provable if and only if it is false.
- Either a falsehood is provable, which is forbidden by the system, or a true statement is unprovable.
- So the formal system is incomplete!
- Gödel also converted this sentence into an assertion in Hilbert's formal system.

Chaitin's Proof to Gödel's Theorem

- Based on Berry Paradox: *"Find the smallest positive integer which to be specified requires more characters than there are in this sentence."*

- Chaitin replaced truth with provability again: *"Find the smallest positive integer which can be proved to require more characters than there are in this sentence."*

- In computer terms, it means: *"Find a series of binary digits that can be proved to be of a complexity greater than the number of bits in this program."*. And what this program actually does?

- Testing all possible proofs in the formal system in order of their size, until it finds an answer. So it finds the first number that it can be proved incapable of finding!

What Does it Really Mean?

- The limiting factor is the number of bits in the formal system as a whole. The information content of axioms and rules is the complexity of the formal system.

- **In a formal system of complexity n, it is impossible to prove that a particular series of binary digits is of complexity greater than $n + c$, where c is a constant that is independent of the particular system employed.**

- Also, **no number(program) can be proved to be random(minimal), unless the complexity of the number(program) is less than that of the system itself.**

- Suggests that in order to progress, mathematicians, like investigators in other sciences, must search for new axioms.

Three Paradoxes

- **Russell Paradox:** *Consider the set of all sets that are not members of themselves.* Is this set a member of itself?

- **Epimenides Paradox:** *Consider this statement: "This statement is false."* Is this statement true?

- **Berry Paradox:** *Consider this sentence: "Find the smallest positive integer which to be specified requires more characters than there are in this sentence."* Does this sentence specify a positive integer?

Unprovable Statements

- This statement is unprovable.
- The complexity of 011011001101111100010 is greater than 15 bits.
- The series of digits 011011001101111100010 is random.
- 10100 is a minimal program for the series 11111111111111111111.

THE END