In search of the shortest description

Damian Niwiński

Faculty of Mathematics, Informatics, and Mechanics

University of Warsaw

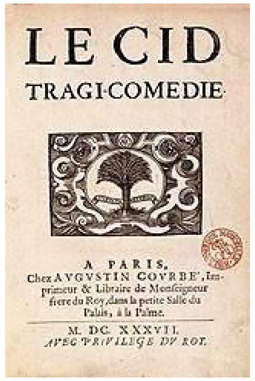
Philosophers' Rally, Gdańsk, June 2012

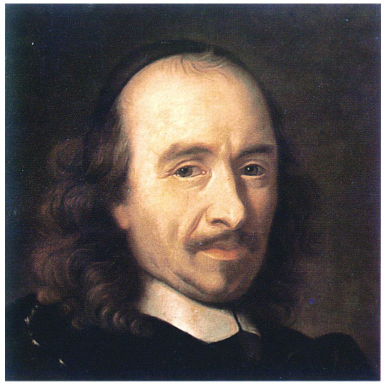
Je n'ai fait celle-ci plus longue que parce que je n'ai pas eu le loisir de la faire plus courte.

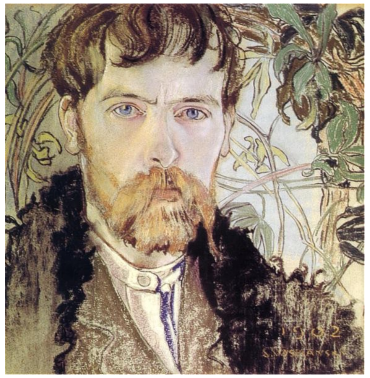
I have made this [letter] longer, because I have not had the time to make it shorter.

Blaise Pascal, *Lettres provinciales*, 1657

Shortening is art


LE CID
TRAGI-COMEDIE.
A PARIS,
Chez AVGVSTIN COVRBE', Imprimeur & Libraire de Monſeigneur frere du Roy, dans la petite Salle du Palais, à la Palme.
M. DC. XXXVII.
AVEC PRIVILEGE DV ROY.





SCÈNE PREMIÈRE – CHIMÈNE, ELVIRE

CHIMÈNE

Elvire, m’as tu fait un rapport bien sincère ?

Ne déguises-tu rien de ce qu’a dit mon père ?

In translation by Stanisław Wyspiański:

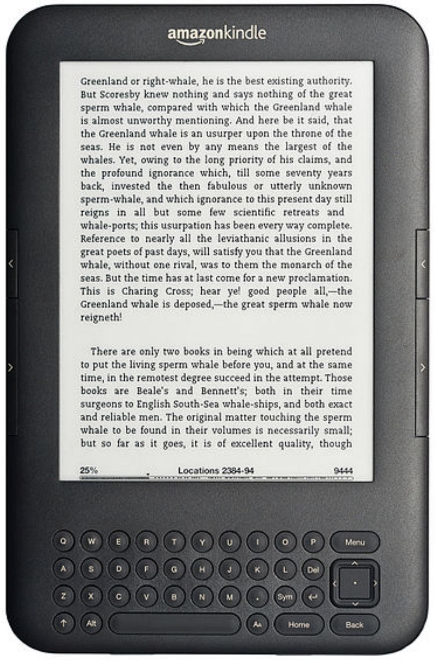
SZIMENA

Więc mówił. . . ?

After Tadeusz Boy Żeleński

Shortening is technology

Biblioteca Joanina, Coimbra, Portugal


Greenland or right-whale, he is the best existing authority. But Scoresby knew nothing and says nothing of the great sperm whale, compared with which the Greenland whale is almost unworthy mentioning. And here be it said, that the Greenland whale is an usurper upon the throne of the seas. He is not even by any means the largest of the whales. Yet, owing to the long priority of his claims, and the profound ignorance which, till some seventy years back, invested the then fabulous or utterly unknown sperm-whale, and which ignorance to this present day still reigns in all but some few scientific retreats and whale-ports; this usurpation has been every way complete. Reference to nearly all the leviathanic allusions in the great poets of past days, will satisfy you that the Greenland whale, without one rival, was to them the monarch of the seas. But the time has at last come for a new proclamation. This is Charing Cross; hear ye! good people all,—the Greenland whale is deposed,—the great sperm whale now reigneth!

There are only two books in being which at all pretend to put the living sperm whale before you, and at the same time, in the remotest degree succeed in the attempt. Those books are Beale's and Bennett's; both in their time surgeons to English South-Sea whale-ships, and both exact and reliable men. The original matter touching the sperm whale to be found in their volumes is necessarily small; but so far as it goes, it is of excellent quality, though

25% Locations 2384-94 9444

Photo NotFromUtrecht

Shortening is mathematics

7, 625, 597, 484, 987

3^{3^3}

0110100110010110100101100110100110010110011010010110100110010110. . .

$0 \to 01$

$1 \to 10$

$$\begin{array}{cccccccccccccccc}
0 & & & & & & & & 1 & & & & & & & \\
0 & & & & 1 & & & & 1 & & & & 0 & & & \\
0 & & 1 & & 1 & & 0 & & 1 & & 0 & & 0 & & 1 & \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
. & . & . & . & . & . & . & . & . & . & . & . & . & . & . & .
\end{array}$$

3.14159 26535 89793 23846 26433 83279 50288 41971 69399 37510 58209 . . .

program generating π

To find a short description, we need to understand an *idea* behind an object.

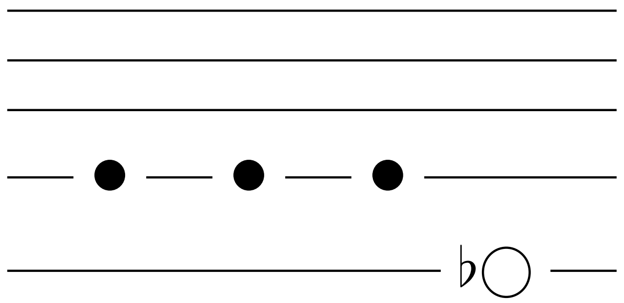


Photo MITO SettembreMusica

But is it always possible to shorten a number ?

Not always **!**

Lemma. If, for each $n \in \mathbb{N}$, $\text{short}(n)$ is a word over r symbols and $\text{short}(n) \neq \text{short}(n')$ whenever $n \neq n'$, then for **infinitely many** n's, we have

$$\text{short}(n) \quad \geq \quad \lfloor \log_r n \rfloor.$$

In this case short is not better than the standard positional notation.

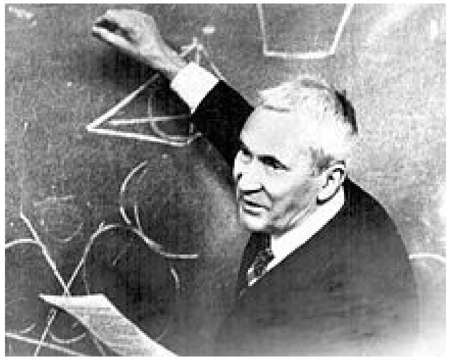
This comes from simple calculation.

If r^k objects are described by r symbols then at least one of them has a description of length at least k.

This is because the number of words shorter than k is

$$1 + r + r^2 + \ldots + r^{k-1} = \frac{r^k - 1}{r - 1} < r^k.$$

$$\begin{array}{cccccccc}
\bigcirc & \square & \diamond & \triangle & \nabla & \clubsuit & \heartsuit & \spadesuit \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\varepsilon & 0 & 1 & 00 & 01 & 10 & 11 & \underline{000} \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
10 & 0 & \varepsilon & 11 & \underline{101} & 00 & \underline{011} & 01
\end{array}$$



Andrey N. Kolmogorov introduced (in 1960s) the following concept:

number n $\longmapsto$

```
$x := 10$
For $i = 1..10$ do
$x := x \star x$
Return $x$
```

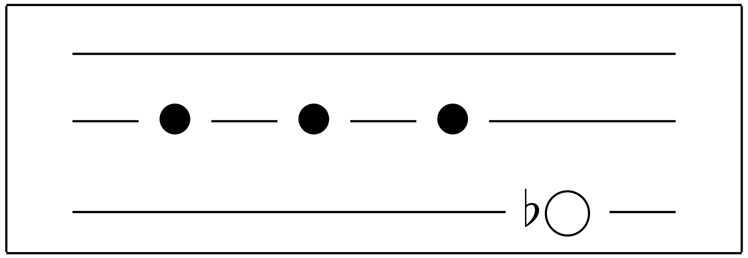
the shortest program generating n

(in a fixed programming language, e.g., **Pascal**)

This program can be viewed as an *ideal* shortening of n .

Remark. The length of the program depends on the choice of a programming language, but only up to additive constant.

$$\begin{array}{l}31415926535897932\\38462643383279502\\88419716939937510\end{array} \longmapsto$$



By the Lemma,

$n \quad \longmapsto$

```
While $\min(a, b) > 0$ do
if $b > a$ then $b := b \bmod a$
else $a := a \bmod b$
Return $\max(a, b)$
```

is not always a shortening.

For some n's, the best what we can do is

$9346296618342597 \quad \longmapsto$ **Write** (9346296618342597)

i.e., the program is as long as the number itself.

Kolmogorov called such numbers **random**.

More precisely, the celebrated **Kolmogorov complexity** of a number n a is:

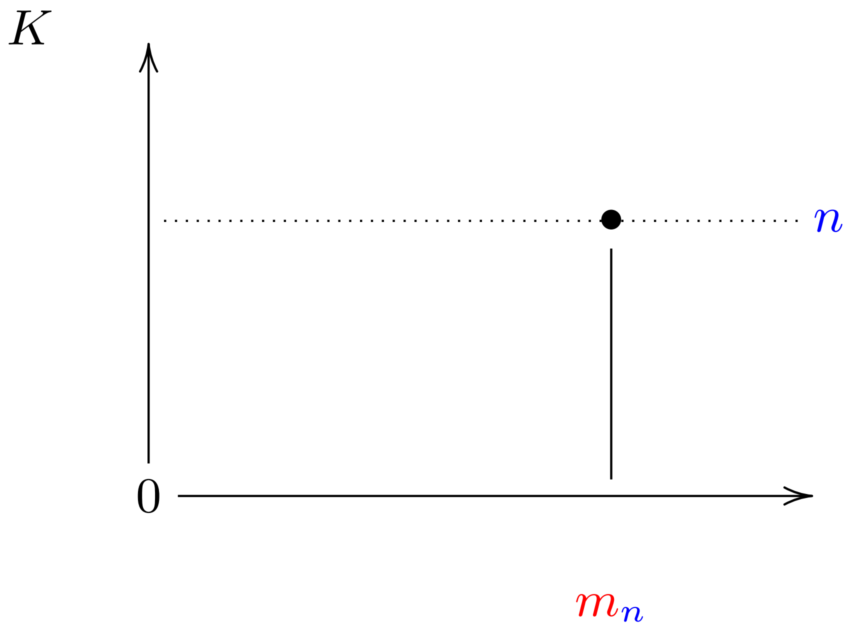
$$K(n) \quad = \quad \text{the length of the shortest binary program generating } n.$$

A number n is **random** whenever

$$K(n) \geq \lfloor \log_2 n \rfloor .$$

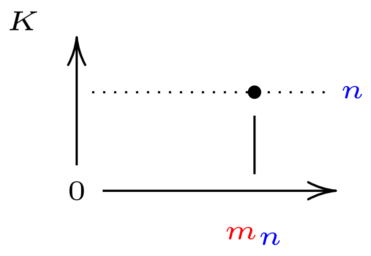
There are infinitely many random numbers.

Random numbers exist in Nature, but can neither be computed nor recognized.



Let m_n be the least such that $n \leq K\,(m_n)$. Then m_n is random !

(Indeed, some $i \in \{0, 1, \ldots, 2^n - 1\}$ must satisfy $\lfloor \log_2 i \rfloor \leq n \leq K(i)$, whereas $m_n \leq i$ implies $\lfloor \log_2 m_n \rfloor \leq \lfloor \log_2 i \rfloor \leq n \leq K\,(m_n)$.)



But no program $P(x)$ can compute the function $n \mapsto m_n$, for all n.

Otherwise, for given n, the program $P(\underline{n})$ computes m_n and has the length

$$\log_2 n + const \quad < \quad n$$

A **contradiction !**

Note a reminiscence of **Berry's paradox**:

Let m_{1000} be $\boxed{\text{the least natural number, which cannot be described in English with less than } 1000 \text{ symbols.}}$.

Consequently, the function $n \mapsto K(n)$ cannot be computed, as otherwise we could use it

```
$x := 0$
While $K(x) < n$ do $x := x + 1$
Return $m_n = x$
```

Neither there is a computable way to *verify* if a given number m is random:

The set of random numbers is not computable.

Otherwise

```
Input $n$
$count := i := 0$
While $count < n$ do
    if Random$(i)$ then $count := count + 1$ fi
    $i := i + 1$
Return $i$
```

computes the n th random number. Take, e.g., 2^{2^n} to obtain a contradiction.

In general, we can only prove **non**-randomness, e.g.,

3 14159265358979323846264338327950288419716939937510

58209749445923078164062862089986280348253421170679

82148086513282306647093844609550582231725359408128

48111745028410270193852110555964462294895493038196

98336733624406566430860213949463952247371907021798

60943702770539217176293176752384674818467669405132

00056812714526356082778577134275778960917363717872

is not random.

In about the same time (1960s), **Gregory Chaitin** (born 1947), then a high school student in New York, discovered independently the main concepts of the Kolmogorov complexity.

He found an alternative proof of the

Gödel Incompleteness Theorem.

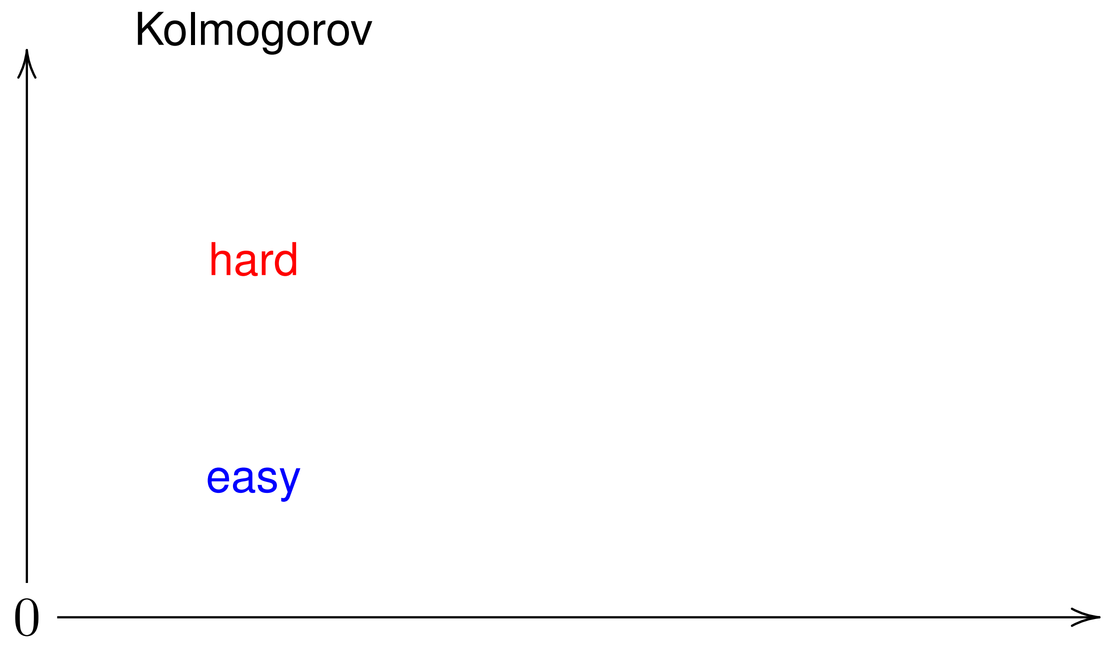
If a theory T is a consistent extension of the Peano arithmetics PA then T is incomplete.

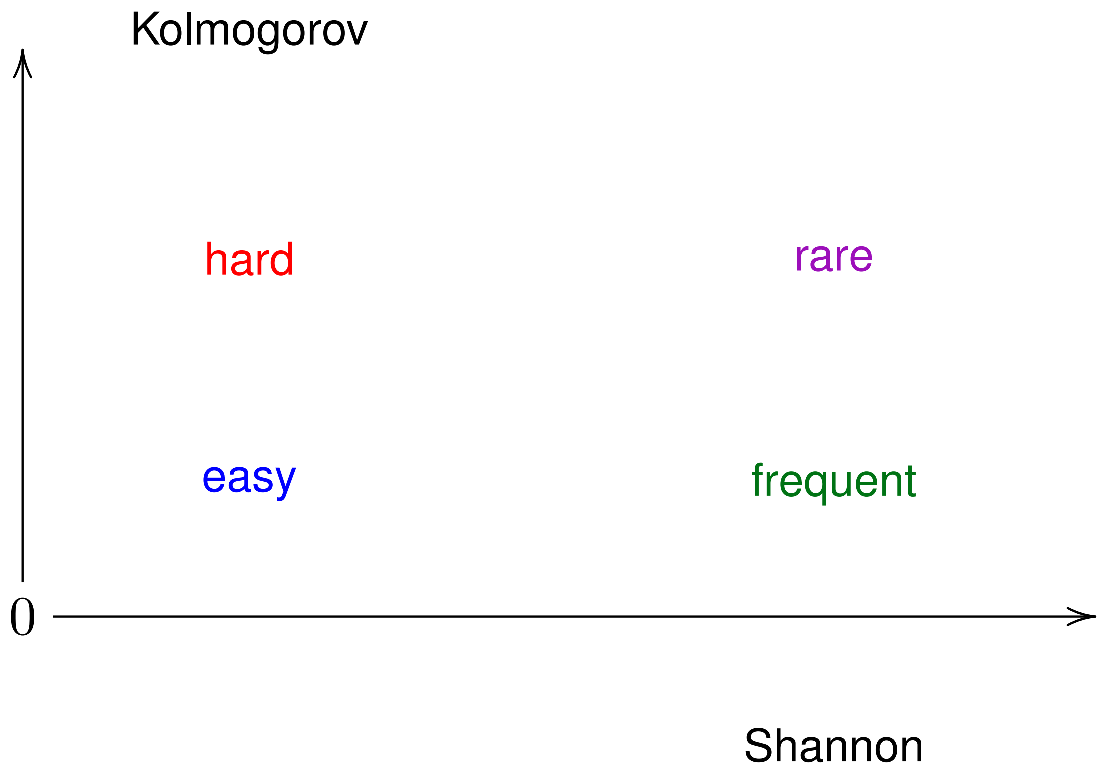
Proof. The property n is not random is semi-representable in PA.

But it is not possible that, for each $n \in \mathbb{N}$,

$T \vdash n$ is random or $T \vdash n$ is not random,

as otherwise the set of random numbers would computable.


Kolmogorov
hard
easy
0


Kolmogorov
hard
rare
easy
frequent
0
Shannon

Claude Shannon led the foundations of information theory in his landmark paper of **1948**

A Mathematical Theory of Communication.

In the subsequent paper of **1951**

Prediction and Entropy of Printed English,

he related his theory to computational linguistics.

Examples from Shannon’s original paper and Lucky’s book.

Quoted from T.M.Cover, J.A.Thomas, *Elements of Information Theory*.

The symbols are independent and equiprobable.

XFOML RXKHRJFFJUJ ZLPWCFWKCYJ

FFJEYVKCQSGYD QPAAMKBZAACIBZLHJQD

The symbols are independent. Frequency of letters matches English text.

OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI

ALHENHTTPA OOBTTVA NAH BRL

The frequency of pairs of letters matches English text.

ON IE ANTISOUTINYS ARE T INCTORE ST B S DEAMY

ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO

TIZIN ANDY TOBE SEACE CTISBE

The frequency of triplets of letters matches English text.

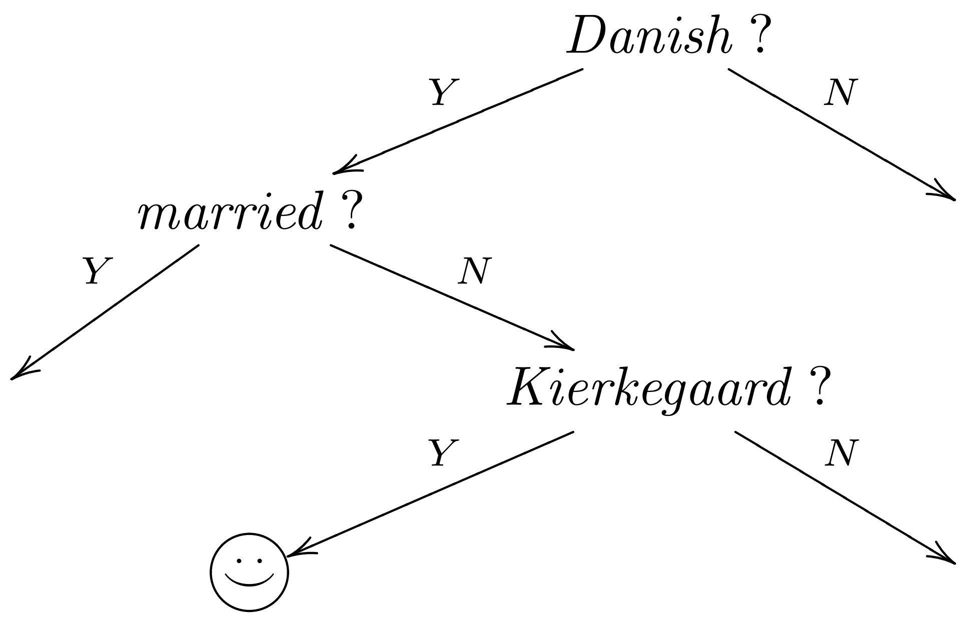
IN NO IST LAT WHEY CRATICT FROURE BERS GROCID PONDENOME OF DEMONSTURES OH THE REPTAGIN IS REOGACTIONA OF CRE

The frequency of quadruples of letters matches English text.

Each letter depends previous three letters.

THE GENERATED JOB PROVIDUAL BETTER TRAND THE DISPLAYED CODE, ABOVERY UPONDULTS WELL THE CODERST IN THESTICAL IT DO HOCK BOTHE MERG. (INSTATES CONS ERATION. NEVER ANY OF PUBLE AND TO THEORY. EVENTIAL CALLEGAND TO ELAST BENERATED IN WITH PIES AS IS WITH THE)

Example of 20 question game



Question game

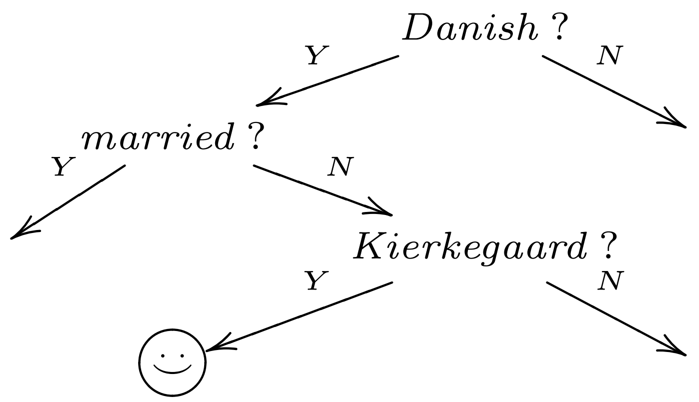
Answerer chooses a challenge from some known set of objects.

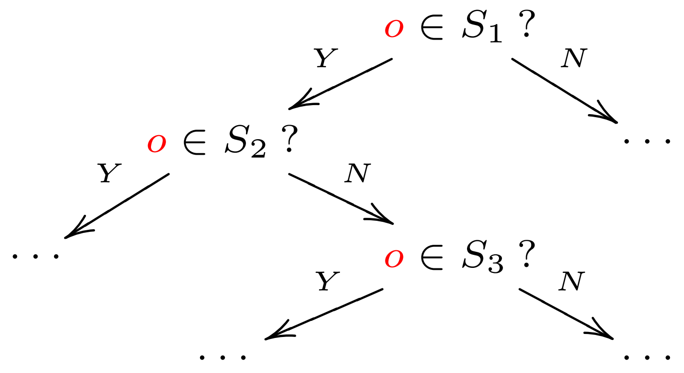
The challenge is kept secrete.

Questioner asks questions: *does the challenge belong to the set of objects* S ?

Answerer answers: *yes* or *no*.

Questioner wishes to guess the challenge ASP.





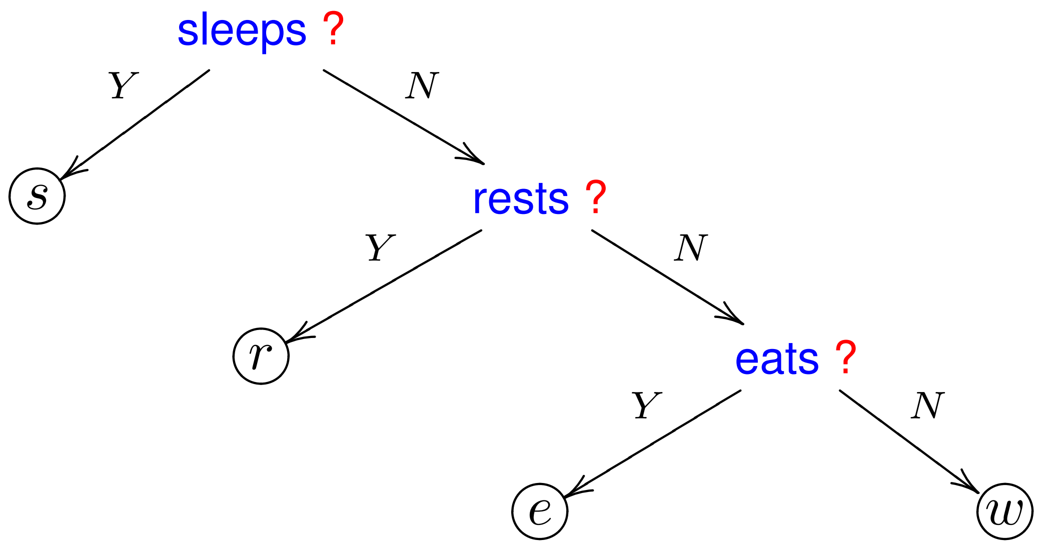
Default strategy

$$\begin{tikzcd}
&&&&&&& \{1,2,3,4\}? \arrow[dllll] \arrow[drrrr] &&&&&&& \\
&&& \{1,2\}? \arrow[dll] \arrow[drr] &&&&&&&& \{6,7\}? \arrow[dll] \arrow[drr] &&& \\
& 1? \arrow[dl] \arrow[dr] &&&& 3? \arrow[dl] \arrow[dr] &&&& 5? \arrow[dl] \arrow[dr] &&&& 7? \arrow[dl] \arrow[dr] & \\
1 && 2 && 3 && 4 && 5 && 6 && 7 && 8
\end{tikzcd}$$

Number of questions is $\log_2 |\mathit{Objects}|$.

Knowing the probability distribution, we can do better.

Pr (sleeps) = $\frac{1}{2}$, Pr (rests) = $\frac{1}{4}$, Pr (eats) = Pr (works) = $\frac{1}{8}$.

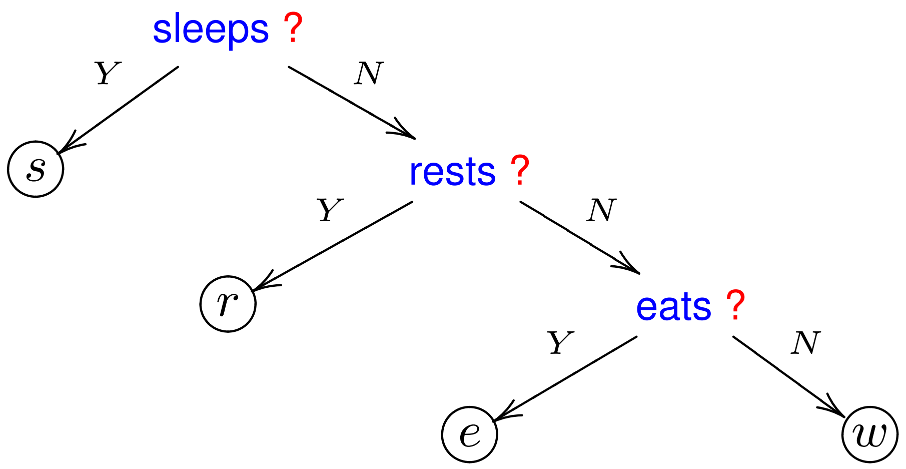


The expected number of questions:

$$1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \left(\frac{1}{8} + \frac{1}{8}\right) = \frac{7}{4} < 2 = \log_2 4.$$

Knowing the probability distribution, we can do better.

p (sleeps) = $\frac{1}{2}$, p (rests) = $\frac{1}{4}$, p (eats) = p (works) = $\frac{1}{8}$.



The expected number of questions:

$$1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \left(\frac{1}{8} + \frac{1}{8}\right) = \frac{7}{4} < 2 = \log_2 4.$$

The number of questions to guess an object with probability q is $\log_2 \frac{1}{q}$.

Shannon's entropy

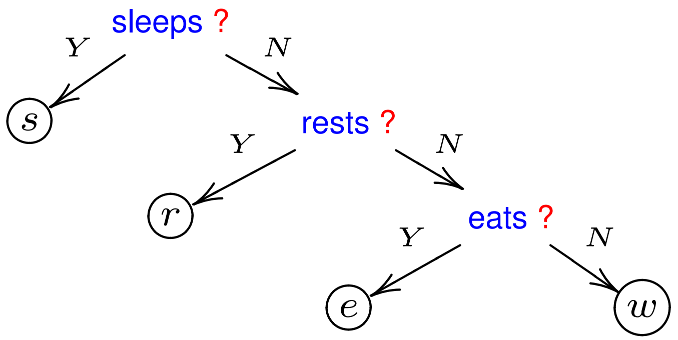
In general, if p is a probability distribution on the set of objects S, the *Shannon entropy* is

$$H(S) \quad = \quad \sum_{s \in S} p(s) \cdot \log_2 \frac{1}{p(s)}.$$

It can be viewed as the average number of questions to be asked in an optimal strategy in the question game.

The number of questions to identify an object s is $\log_2 \frac{1}{p(s)}$.

p (sleeps) = $\frac{1}{2}$, p (rests) = $\frac{1}{4}$, p (eats) = p (works) = $\frac{1}{8}$.



In the example above $H(S) =$

$$\begin{aligned}
&= p(s) \log \frac{1}{p(s)} + p(r) \log \frac{1}{p(r)} + p(e) \log \frac{1}{p(e)} + p(w) \log \frac{1}{p(w)} \\
&= \frac{1}{2}\ 1 + \frac{1}{4}\ 2 + \frac{1}{8}\ 3 + \frac{1}{8}\ 3 \\
&= \frac{7}{4}
\end{aligned}$$

$H(S) = \sum_{s \in S} p(s) \cdot \log_2 \frac{1}{p(s)}$.

The number of questions to identify an object s is $\log_2 \frac{1}{p(s)}$.

The definition of entropy conforms to the *Weber-Fechner law* of cognitive science:

the human *perception* (P) of the growth of a physical *stimuli* (S), is proportional to the **relative** growth of the stimuli rather than to its absolute growth,

$$\partial P \quad \approx \quad \frac{\partial S}{S}$$

consequently, after integration,

$$P \quad \approx \quad \log S.$$

This has been observed in perception of weight, brightness, sound (both intensity and height), and even one's economic status.

Good !

Good !

Good !

Good !

Good !

Intuitively, the Shannon entropy tells us how difficult is it to find an object.

The Kolmogorov complexity tells us, how difficult is it to create it.

Is there a link between them ?

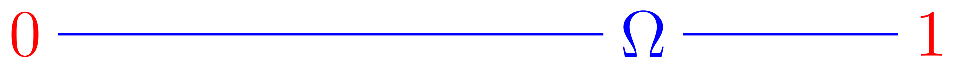
Chaitin's constant

$$\Omega \quad = \quad \sum_{P\downarrow} 2^{-|P|}$$

where P ranges over binary programs, and $P\downarrow$ means that P halts.

Ω can be interpreted as probability that by tossing a coin, we find a binary program, which moreover halts.

$$\underbrace{01101011011001101010101010101101001101010}_{\text{this is a program which halts ☺}}\ldots$$



Every prefix of Ω is Kolmogorov random.

Knowledge of Ω would suffice to reconstruct any computation.

For $n \in \mathbb{N}$, let $P \downarrow n$ means that P halts and produces n.

Let

$$p(n) \quad = \quad \sum_{P \downarrow n} 2^{-|P|}$$

This can be interpreted as probability that by tossing a coin, we find a binary program, which halts and produces n.

Theorem.

$$K(n) \approx \log_2 \frac{1}{p(n)}$$

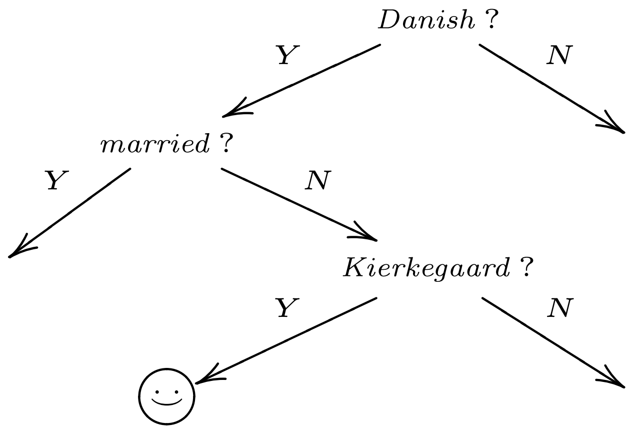
More precisely, there exists a constant c, such that, for all n,

$$K(n) - c \leq \log_2 \frac{1}{p(n)} \leq K(n) + c$$

$K(n) \approx \log_2 \frac{1}{p(n)}$

Kolmogorov Shannon

```
$x := 10$
For $i = 1..10$ do
$x := x \star x$
Return $x$
```



Researchers who have contributed to the theory: C.E.Shannon, A.N.Kolmogorov, R.Solomonoff, G.Chaitin, P.Martin–Löf, L.A.Levin, P.Gacs, and many others.

Literature.

Thomas M. Cover, and Joy A. Thomas,
Elements of Information Theory, John Wiley & Sons, 1991.

Ming Li and Paul Vitanyi,
An Introduction to Kolmogorov Complexity and Its Applications, Springer, 2008.

D.Niwiński, M.Strojnowski, M.Wojnarski, *Teoria informacji*,
`http://wazniak.mimuw.edu.pl/`