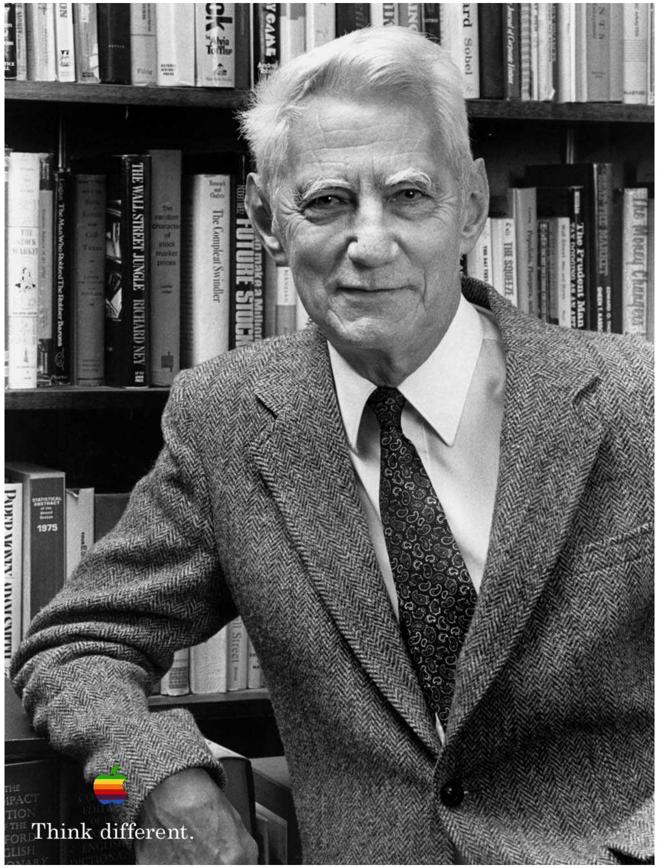
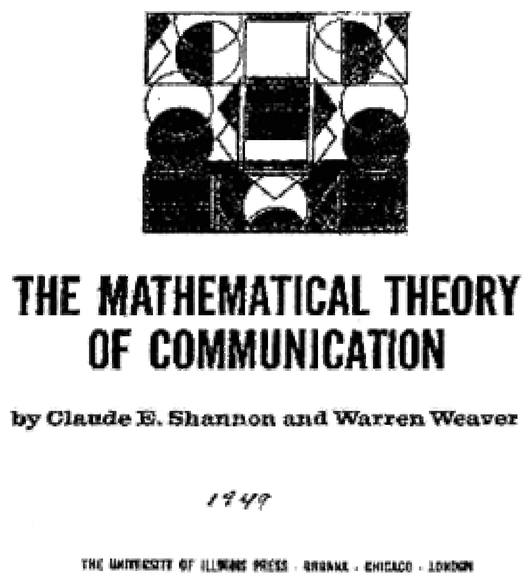
# Information Theory

Laura Kovács





THE MATHEMATICAL THEORY OF COMMUNICATION

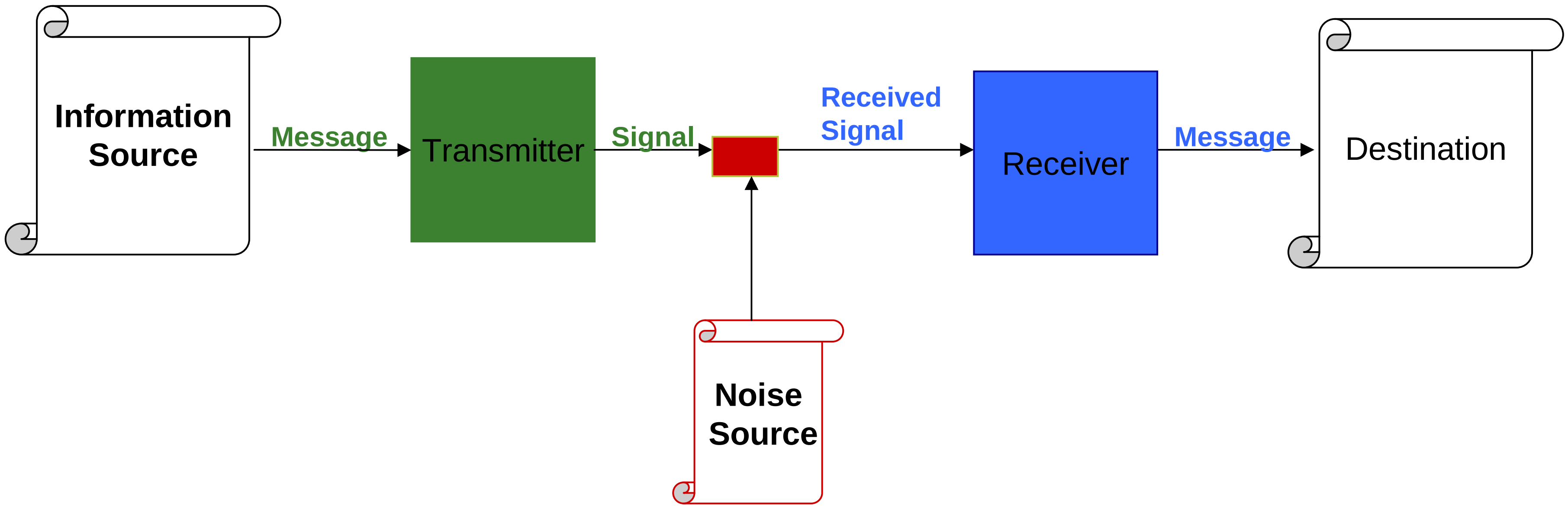
by Claude E. Shannon and Warren Weaver

1949

THE UNIVERSITY OF ILLINOIS PRESS · URBANA · CHICAGO · LONDON

**Claude Shannon (1916-2001)**

# A Communication System Model



# Information Theory by Claude Shannon & Warren Weaver

- Information (measured in *bits*):
  - the amount of uncertainty a message eliminates;
- Based on a linear model of communication;
- Noise :
  - reduces information by increasing uncertainty;
- Redundancy:
  - reduces loss of information due to noise.

# Information Theory by Claude Shannon & Warren Weaver

- Information (measured in *bits*):
  - the amount of uncertainty a message eliminates;
- Based on a linear model of communication;
- Noise :
  - reduces information by increasing uncertainty;
- Redundancy:
  - reduces loss of information due to noise.

e an i 0 du an

Ge an is 0% red nt

German is 50% redunda

# Information Content h(p)

(Informationgehalt)

= MEASURE OF UNCERTAINTY associated with probability p

- Independent of coding;
- Increases when probability p decreases (uncertainty increases);

# Information Content h(p)

(Informationgehalt)

= MEASURE OF UNCERTAINTY associated with probability p

- Independent of coding;
- Increases when probability p decreases (uncertainty increases);
- $h(p_1*p_2) = h(p_1)+h(p_2)$;
- $h(p)=ld(1/p)$ [bit].

# Examples of Information Content

- $h(1) = 0$ Message will be always expected
- $h(0) = \infty$ Message will be never expected
- $h(0.5) = 1$ Message will be by 50% expected
- $h(0.1) = 3.32$ Information content of a decimal digit

# Binary Logarithm ld x

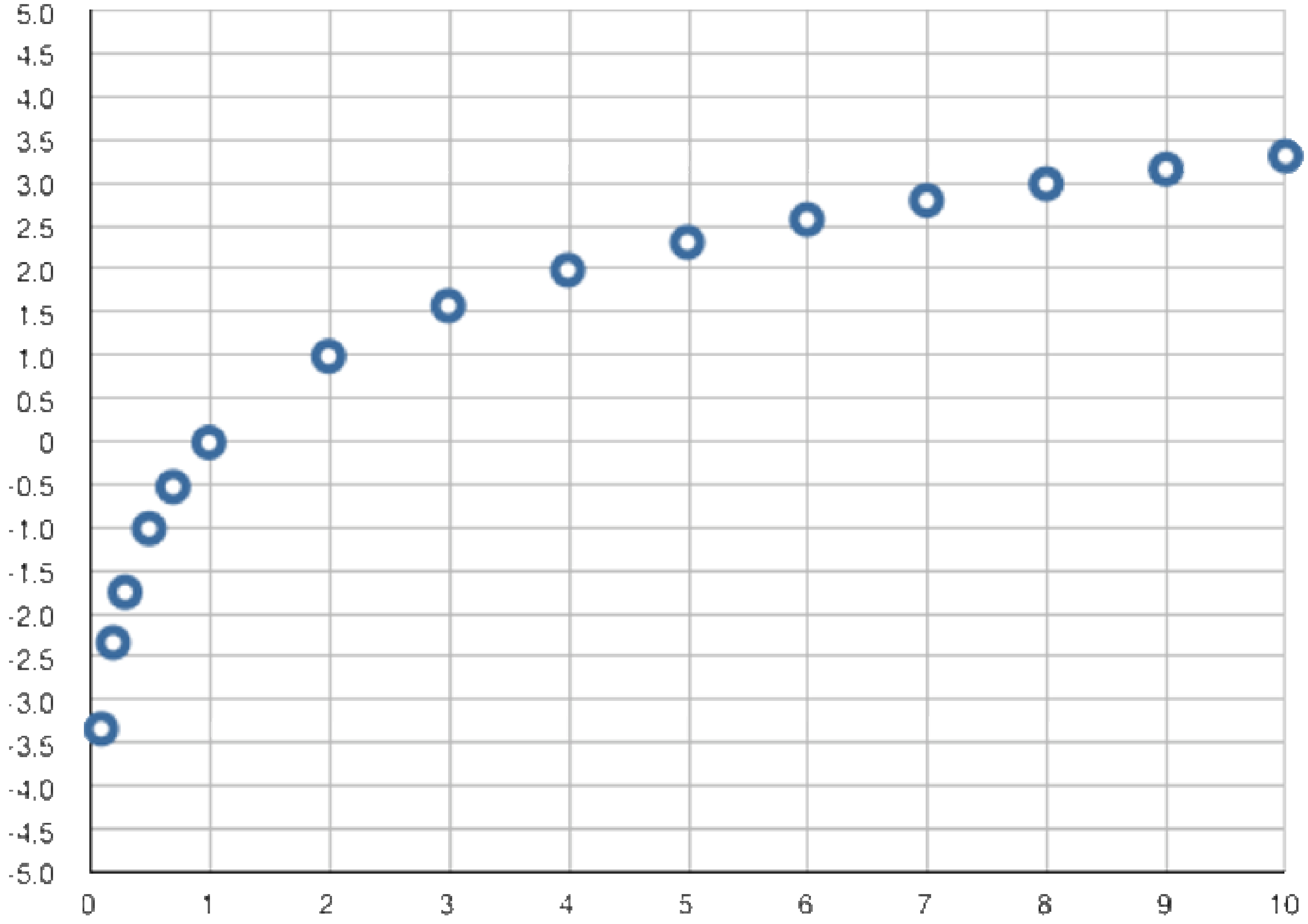
## Properties, identities

- $2^{\text{ld } x} = x$
- ld 1/x = - ld x
- ld (x*y) = ld x + ld y
- ld 2*x = 1 + ld x
- ld $x^n$ = n*ld x
- ld $x^2$ = 2*ld x
- ld x = ln x / ln 2 ~ 1.44*ln x
- ld x = log x / log 2 ~ 3.32*log x

# Examples

- ld 1 = 0
- ld 2 = 1
- ld 4 = ld ($2^2$) = 2*ld 2 = 2
- ld 8 = ld ($2^3$) = 3*ld 2 = 3
- ld 10 ~ 3.32*log 10 = 3.32*1 = 3.32
- ld 5 = ld 10 - 1 ~ 2.32
- ld 9 ~ (ld 8 + ld 10)/2 = (3 + 3.32)/2 = 3.16
- ld 3 = (ld 9)/2 ~ 1.58
- ld 7 = (ld 49)/2 ~ (ld 50)/2 = (ld 100 -1)/2 = (2*ld 10 - 1)/2 = 2.82

| x | ld x |
|---|---|
| 1 | 0.00 |
| 2 | 1.00 |
| 3 | 1.58 |
| 4 | 2.00 |
| 5 | 2.32 |
| 6 | 2.58 |
| 7 | 2.81 |
| 8 | 3.00 |
| 9 | 3.17 |
| 10 | 3.32 |


5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
-3.5
-4.0
-4.5
-5.0
0
1
2
3
4
5
6
7
8
9
10

- Average Information Content **(Entropy)**:
  [Mittlerer Informationsgehalt]

$$H = \sum_i \; p_i \, ld \, 1/p_i \text{ [bit]}$$

- Average Word Length:
  [Mittlere Wortlänge]

$$L = \sum_i \; p_i \, l_i \text{ [bit]}$$

- Redundancy:
  [Redundanz]

$$R = L - H \text{ [bit]}$$

**Shannon's Coding Theorem: L ≥ H**

# Shannon's Binary Entropy Function

[Shannon'sche Binäre Funktion]

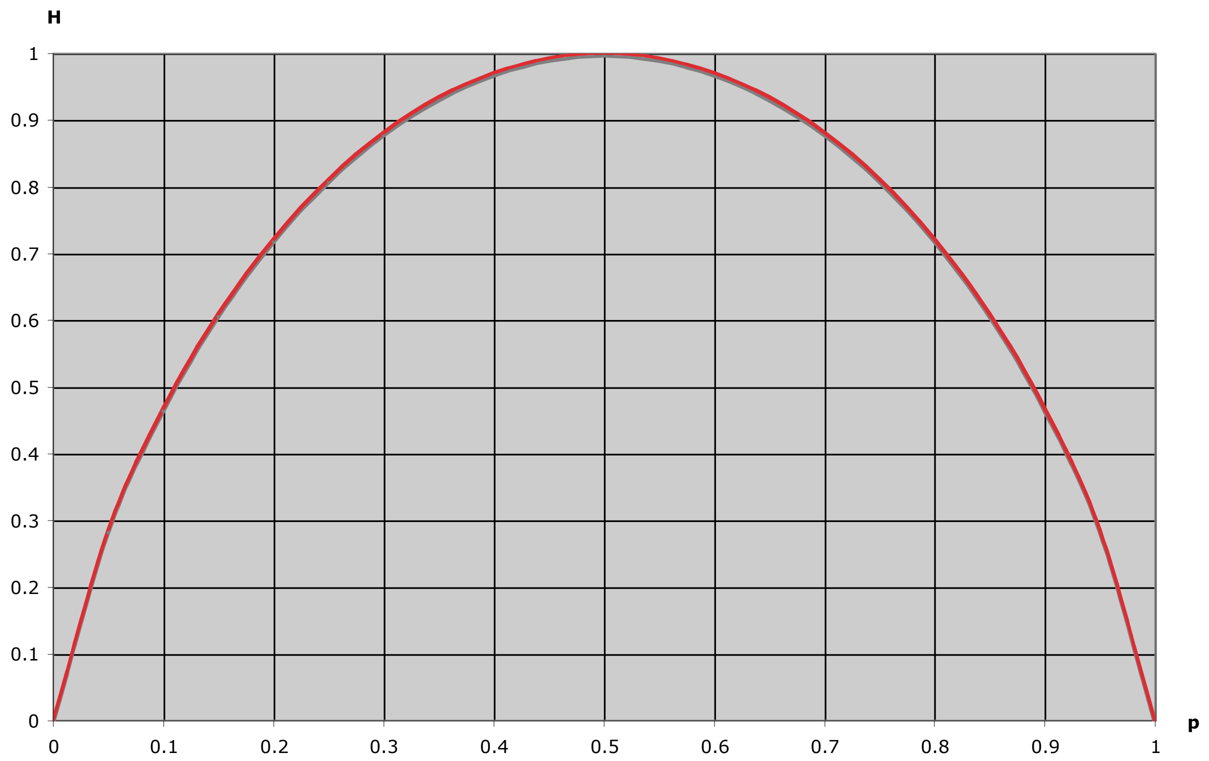
***Example 1:***

***Consider tossing a coin with known, not necessarily fair, probabilities of coming up heads or tails.***

# Shannon's Binary Entropy Function

[Shannon'sche Binäre Funktion]

**H= p ld 1/p + (1-p) ld 1/(1-p)**




*Entropy of a coin toss as a function of the probability p of it coming up heads.*

## *Example 2: Redundancy of a decimal digit.*

(All 10 digits have same probabilities.)

- Average Information Content (Entropy):

$$H = \text{ld } 10 = 3.32 \text{ bit}$$

- Word Length (each binary digit has the *same prefixed length* of 4 bits):

$$L = 4 \text{ bit}$$

- Redundancy:

$$R = L - H = 0.68 \text{ bit}$$

***Example 3: Average Information Content of a (German) Letter.***

(All 26 letters have same probabilities.)

- Average Information Content (Entropy):

$$H = \text{ld}\, 26 = 4.7 \text{ bit}$$

## *Example 4: Average Information Content of a (German) Letter.*

(Letters have different probabilities.)

- Average Information Content (Entropy):

$$H = 4.1 \text{ bit}$$

Note: More than a half of all German texts are formed by letters e, n, r, i and the space symbol.

## *Example 4: Average Information Content of a (German) Letter.*

(Sequence of [2] consecutive letters have different probabilities.)

- Average Information Content (Entropy):

$$H \sim 2 \text{ bit}$$

## *Example 5: Average Information Content of a (German) Word.*

(10 Million Words have different probabilities.)

- Average Information Content (Entropy):

$$H \sim 15.5 \text{ bit}$$

- Word Length:

$$L = 7.7 \text{ letters}$$

***Example 6: Information flow during reading.***

Reading: 25 letters per second

↓

Information flow: 50 bits/sec

↓

Storing: 25 bits/sec

↓

In 60 years a person could store approx. $3*10^{10}$ bits.

***Theoretically:*** **the brain's storage capacity ~ $10^{12}$ bits.**

**→ one could store read information for more than ~1900 years ☺**

# Reminder for Next Lecture

- Carry out the solution of Example 4 of this lecture!

- Next lecture on October 7th !
  - No lecture on September 30th .

- Homework sessions start today (September 23rd), at 12pm!

- This week's homework due by September 30th, 10am.

- Contact teaching assistants for help, if needed!