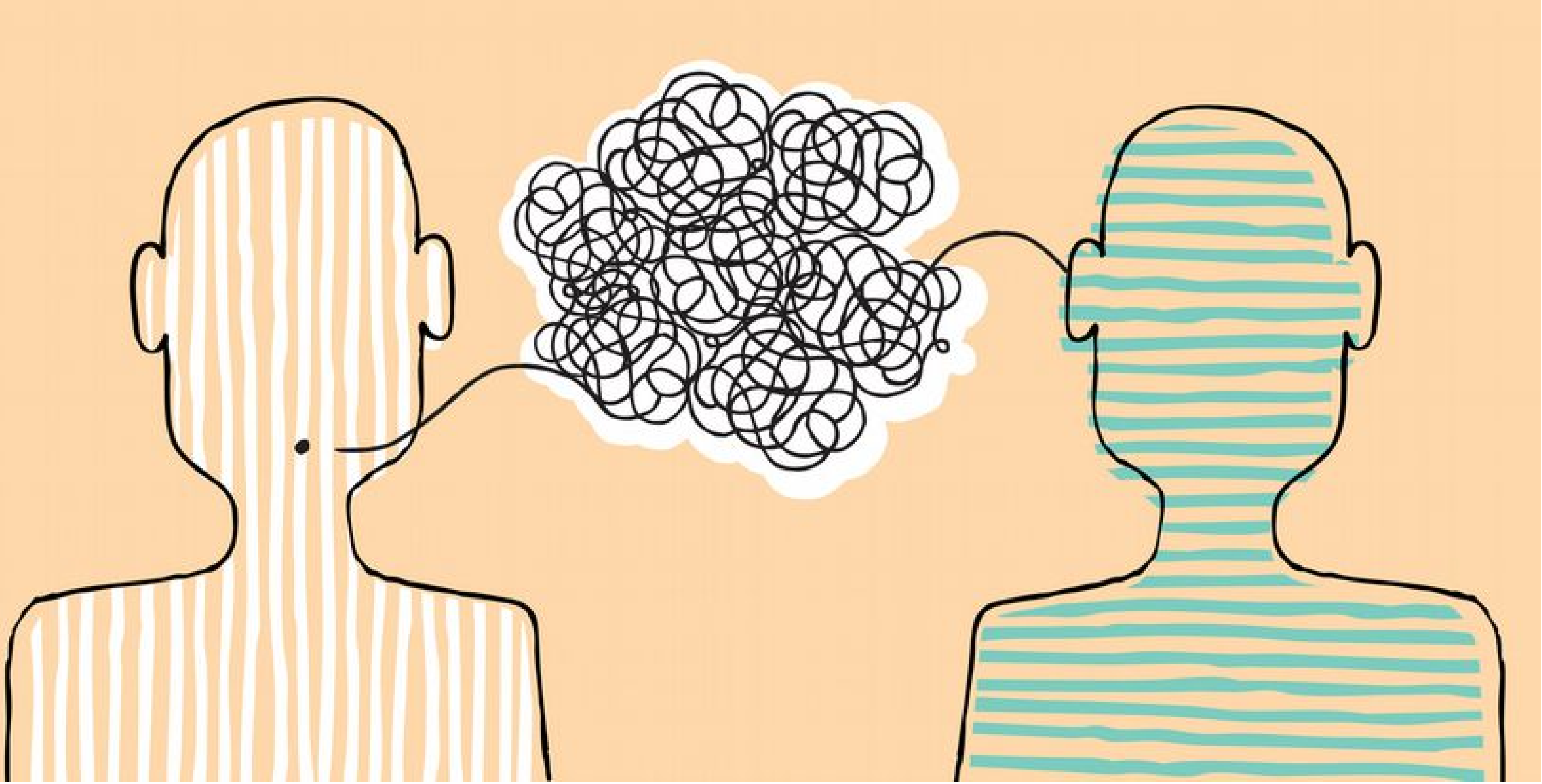
Shannon's Theory of Communication

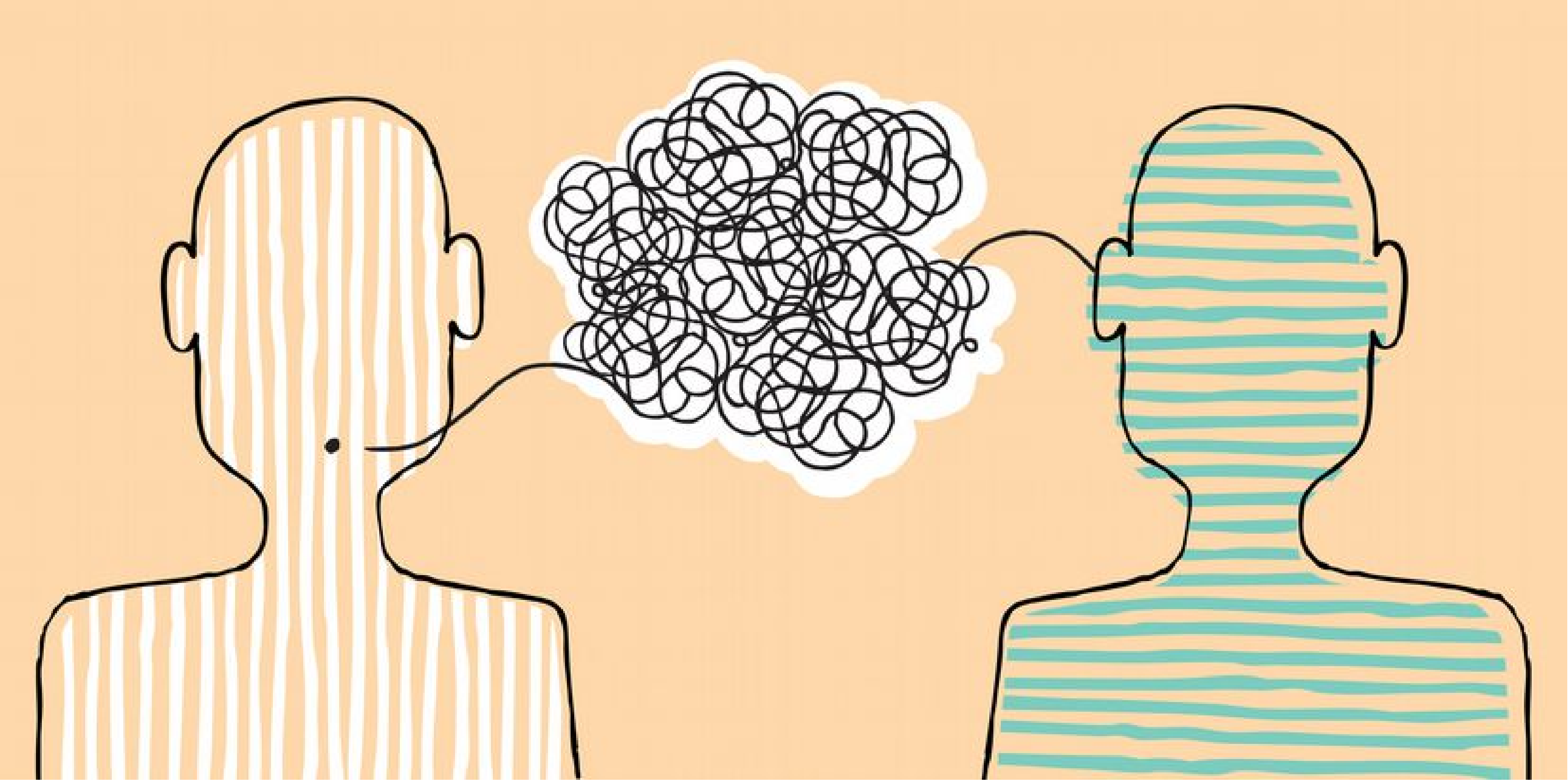
An operational introduction

5 September 2014,
Introduction to Information Systems

Giovanni Sileno g.sileno@uva.nl
Leibniz Center for Law
University of Amsterdam

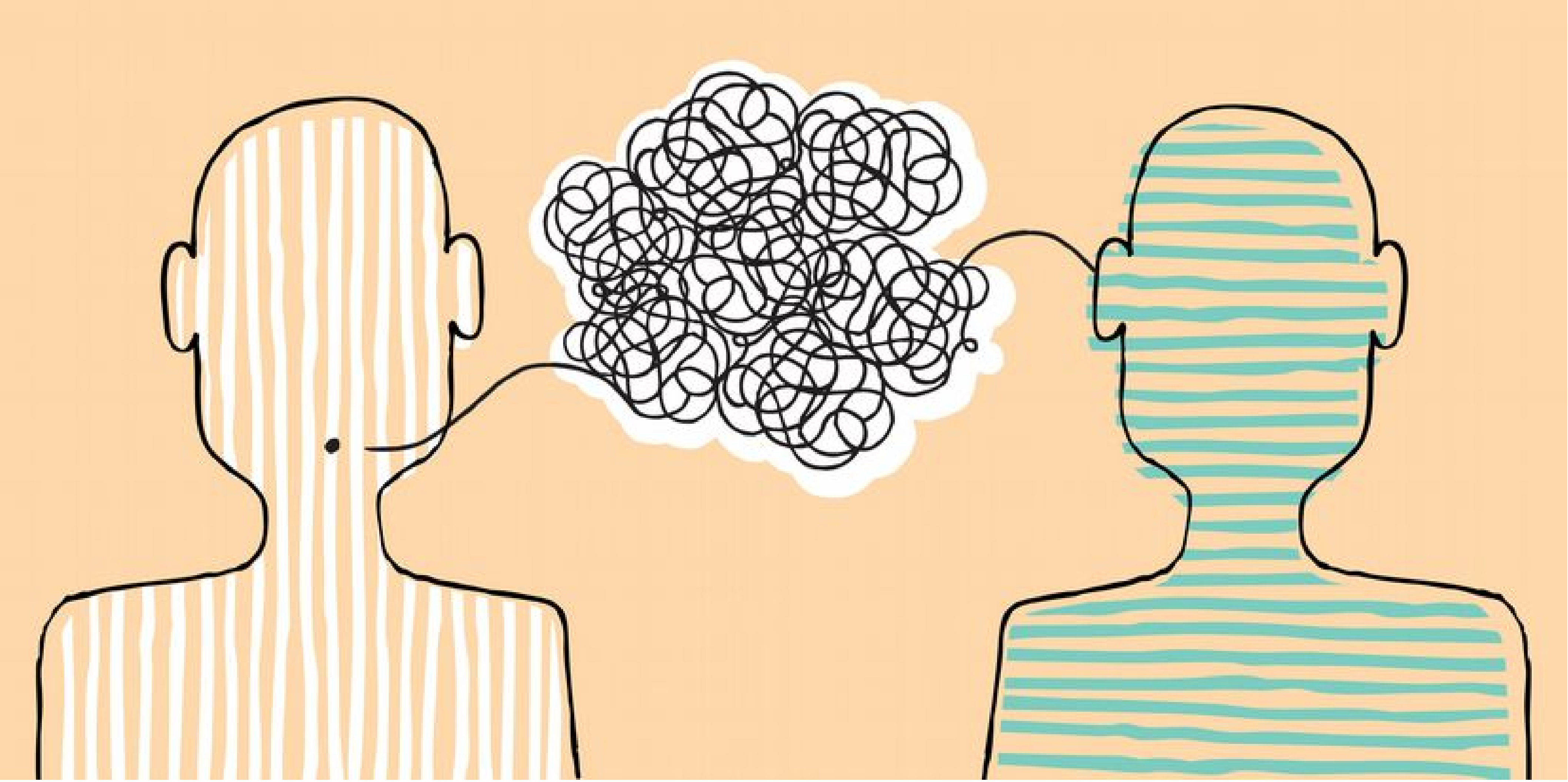


Leibniz Center for Law



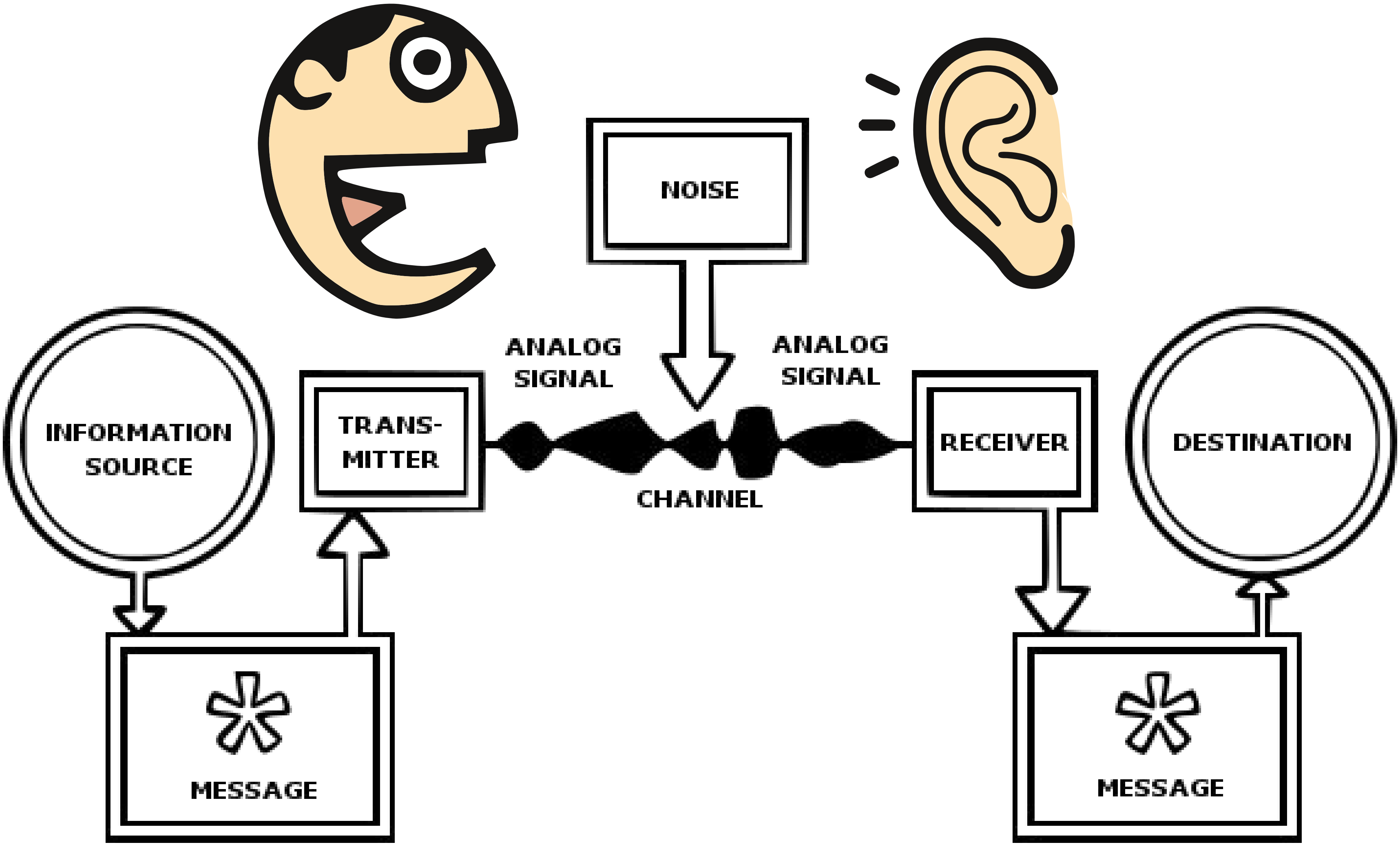
Fundamental basis of communication

- to reproduce at one point a **message** selected at another point.

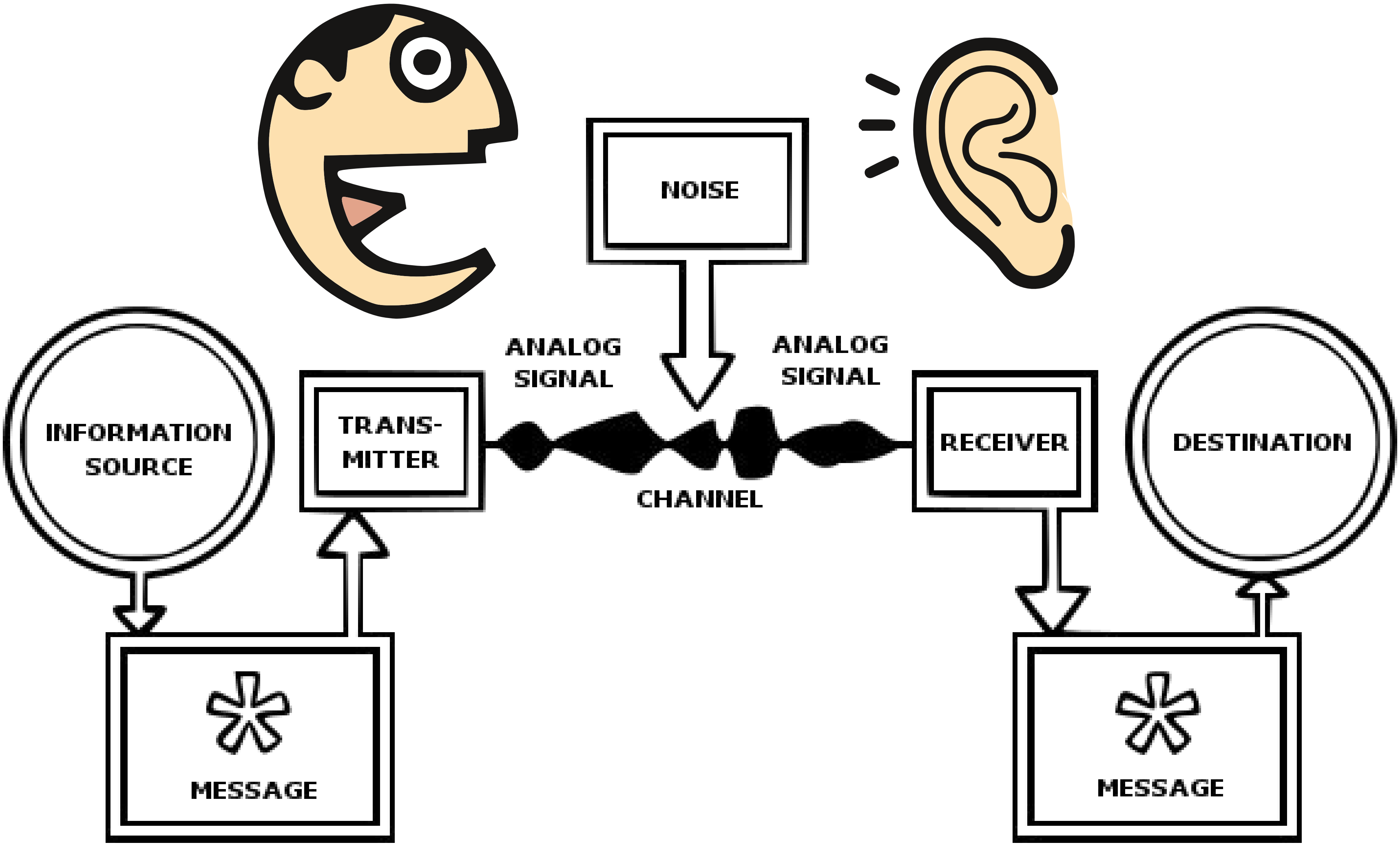


Fundamental basis of communication

- We focus on the transmission system: the **communication channel**.



Leibniz Center for Law



- Problem: **how much** "information" can we transmit reliably?

Quantification

Quantify "information"

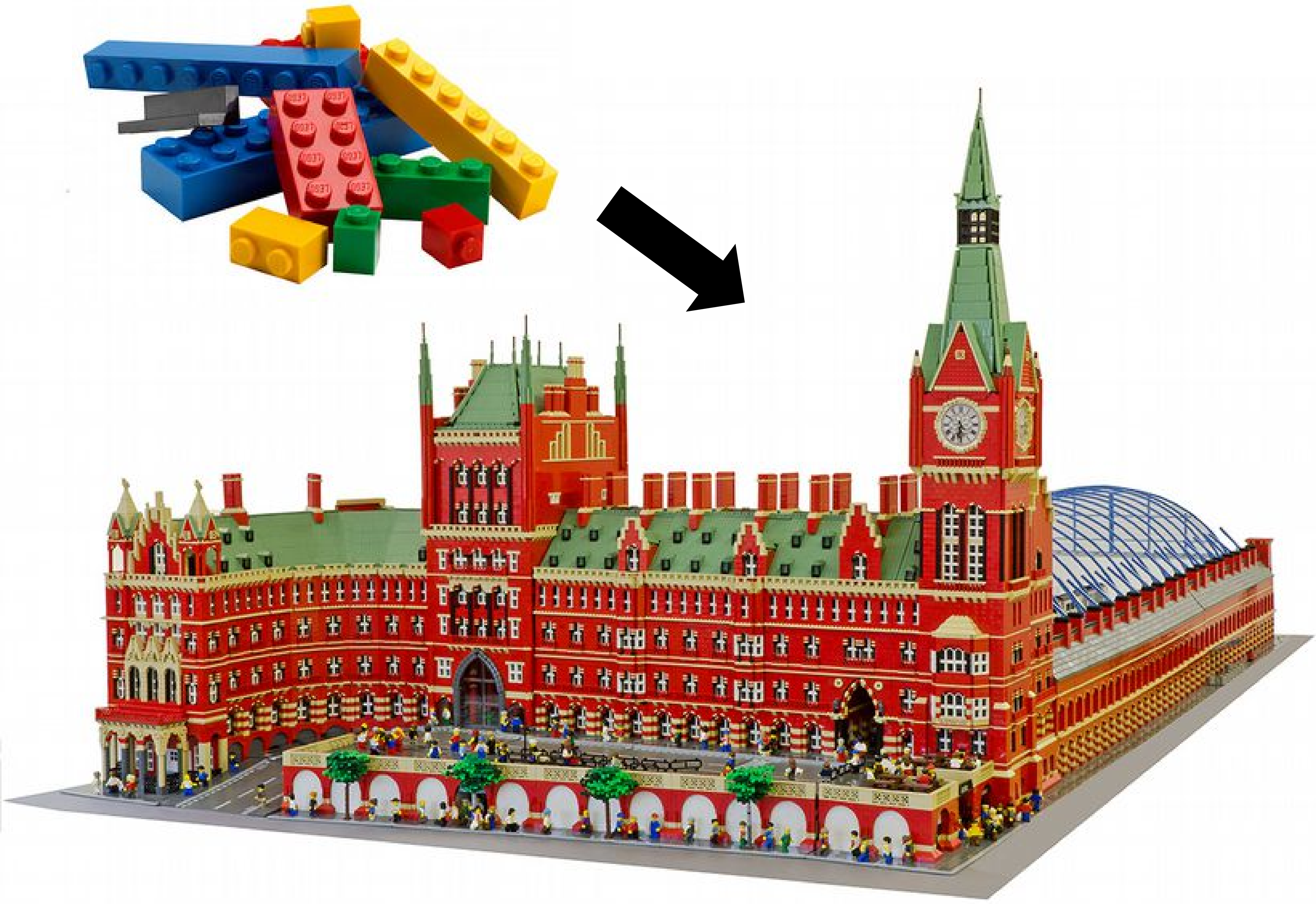
- What information are we talking about?
  - That concerning the **signal**, not the *interpreted* meaning of the signal.

Quantify "information"

- What information are we talking about?
  - That concerning the **signal**, not the *interpreted* meaning of the signal.
- Question:
  - how much information the next pages contain, taken as separate sources?

Quantify “information”

- A *datum* is reducible to a lack of uniformity.
  - **differentiation in the signal produces data**.
- How much can we differenciate the signal?
  - It depends on how the signal is constructed!



Quantification - 2

Exercise

- I choose randomly a card from a deck of 16 cards, ordered from 1 to 16. How many **yes/no** questions should you ask to discover which card I have in my hands?

Exercise

- I choose randomly a card from a deck of 16 cards, ordered from 1 to 16. How many **yes/no** questions should you ask to discover which card I have in my hands?

01. 02. 03. 04. 05. 06. 07. 08. 09. 10. 11. 12. 13. 14. 15. 16.

Exercise

- I choose randomly a card from a deck of 16 cards, ordered from 1 to 16. How many **yes/no** questions should you ask to discover which card I have in my hands?

01. 02. 03. 04. 05. 06. 07. 08. 09. 10. 11. 12. 13. 14. 15. 16.

- Dichotomy method:
  Is it greater than 8?

Exercise

- I choose randomly a card from a deck of 16 cards, ordered from 1 to 16. How many **yes/no** questions should you ask to discover which card I have in my hands?

01. 02. 03. 04. 05. 06. 07. 08. 09. 10. 11. 12. 13. 14. 15. 16.

- Dichotomy method:
  Is it greater than 8? No. → 8 cards remaining

Exercise

- I choose randomly a card from a deck of 16 cards, ordered from 1 to 16. How many **yes/no** questions should you ask to discover which card I have in my hands?

01. 02. 03. 04. 05. 06. 07. 08. 09. 10. 11. 12. 13. 14. 15. 16.

- Dichotomy method:
  Is it greater than 8? No. → 8 cards remaining
  Is it greater then 4? Yes. → 4 cards remaining

Exercise

- I choose randomly a card from a deck of 16 cards, ordered from 1 to 16. How many **yes/no** questions should you ask to discover which card I have in my hands?

01. 02. 03. 04. 05. 06. 07. 08. 09. 10. 11. 12. 13. 14. 15. 16.

- Dichotomy method:
  Is it greater than 8? No. → 8 cards remaining
  Is it greater then 4? Yes. → 4 cards remaining
  Is it greater then 6? No. → 2 cards remaining

Exercise

- I choose randomly a card from a deck of 16 cards, ordered from 1 to 16. How many **yes/no** questions should you ask to discover which card I have in my hands?

01. 02. 03. 04. 05. 06. 07. 08. 09. 10. 11. 12. 13. 14. 15. 16.

- Dichotomy method:

  Is it greater than 8? No. → 8 cards remaining
  Is it greater then 4? Yes. → 4 cards remaining
  Is it greater then 6? No. → 2 cards remaining
  Is it greater then 5? No. → 1 cards remaining

Unit of measuring "information"

- The cards correspond to the indexed **symbols** available at the source.

- **BIT**, for *binary digit*, corresponds to *one answer* to YES/NO questions

Unit of measuring “information”

- The cards correspond to the indexed **symbols** available at the source.

- **BIT**, for *binary digit*, corresponds to *one answer* to YES/NO questions

- Basic formulas:
  - # questions = Log_2 (# symbols)
  - # symbols = $2^{\#\text{questions}}$

"Trick" for Log_2 without calc

Write the series of multiples of 2

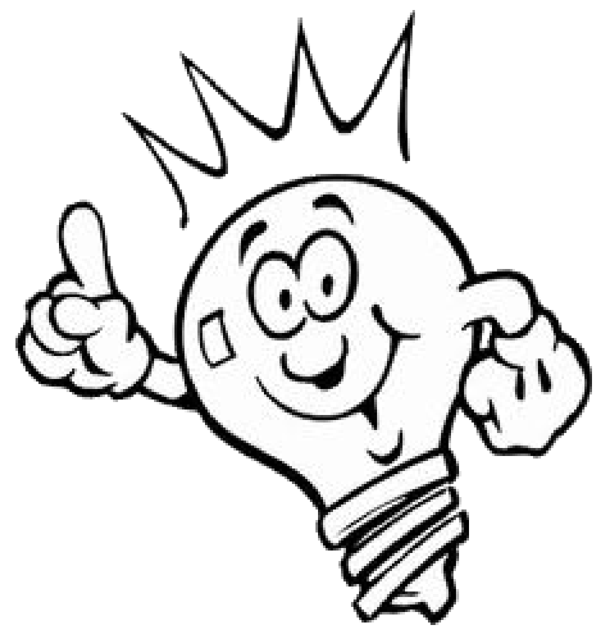
2 . 4 . 8 . 16 . 32 . 64 . 128 . 256 . 512 . 1024 ...

1 . 2 . 3 . 4 . 5 . 6 . 7 . 8 . 9 . 10 ...

From this sequence we read that:

$2^7 = 128$ $\qquad$ $\text{Log}_2 32 = 5$

$2^{-4} = 1/16$ $\qquad$ $\text{Log}_2 1/16 = -4$



Leibniz Center for Law

Exercise

- How many symbols do we have in this class?
  - as individuals
  - as genders

- How many symbols in this question?
  - as letters
  - as words

- Calculate how many bits we need to index them.

Exercise

- How many symbols do we have in this class?
  - as individuals
  - as genders
- How many symbols in this question?
  - as letters
  - as words
- NOTA BENE: the answers depend on the **referent** and on the “**filter**” chosen!

From binary to decimal code

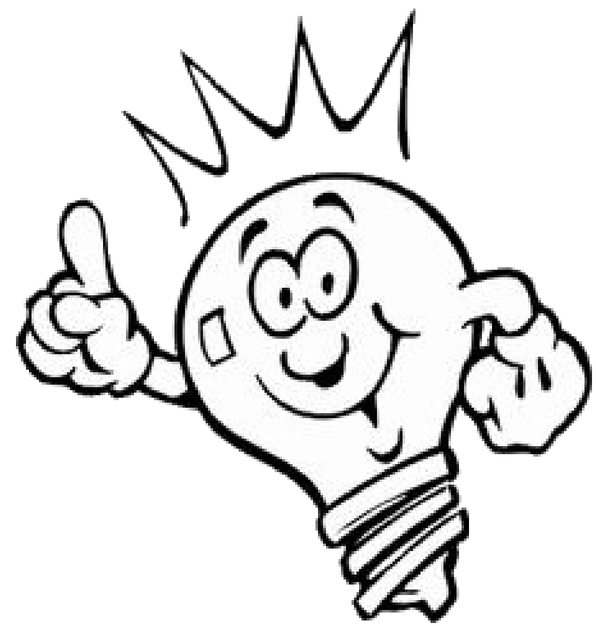
..	8	4	2	1
..	2^3	2^2	2^1	$\mathbf{2^0}$

0 0 0 0 = 0

0 1 0 0 = 4

0 1 0 1 = 4 + 1 = 5

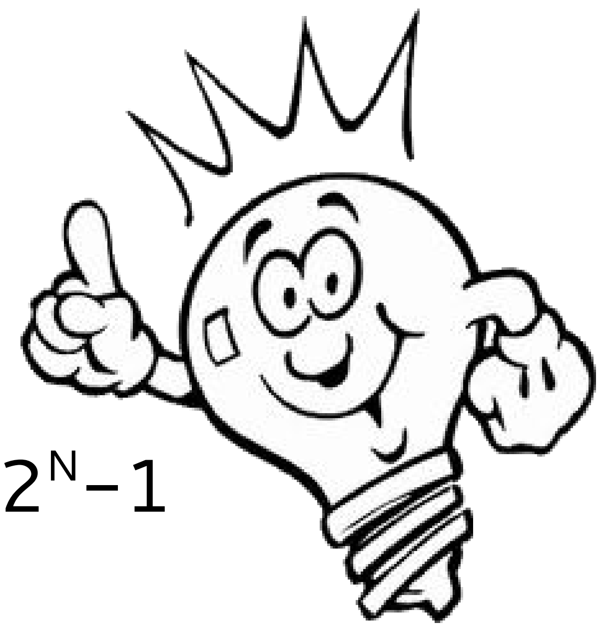
1 1 1 1 = 8 + 4 + 2 + 1 = 15



Leibniz Center for Law

From binary to decimal code

..	8	4	2	1	
..	2^3	2^2	2^1	$\mathbf{2^0}$	
	0	0	0	0	= 0
	0	1	0	0	= 4
	0	1	0	1	= 4 + 1 = 5
	1	1	1	1	= 8 + 4 + 2 + 1 = 15



- N bits index 2^N integers, from 0 to 2^N-1

From binary to decimal code

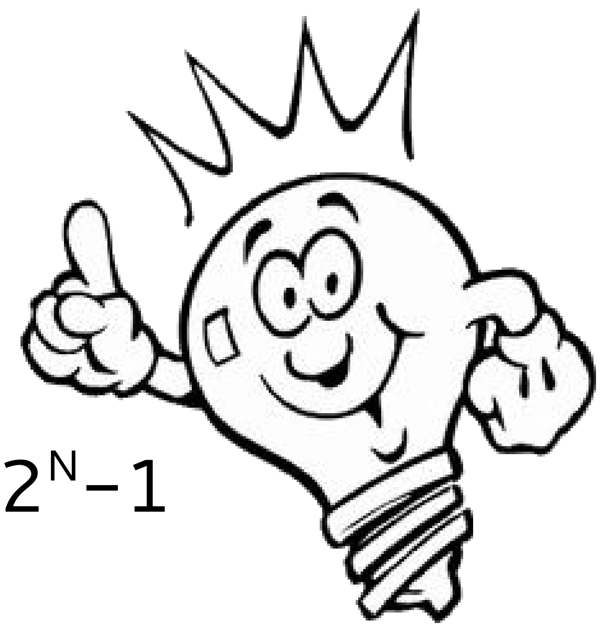
..	8	4	2	1
..	2^3	2^2	2^1	**2^0**

0 0 0 0 = 0

0 1 0 0 = 4

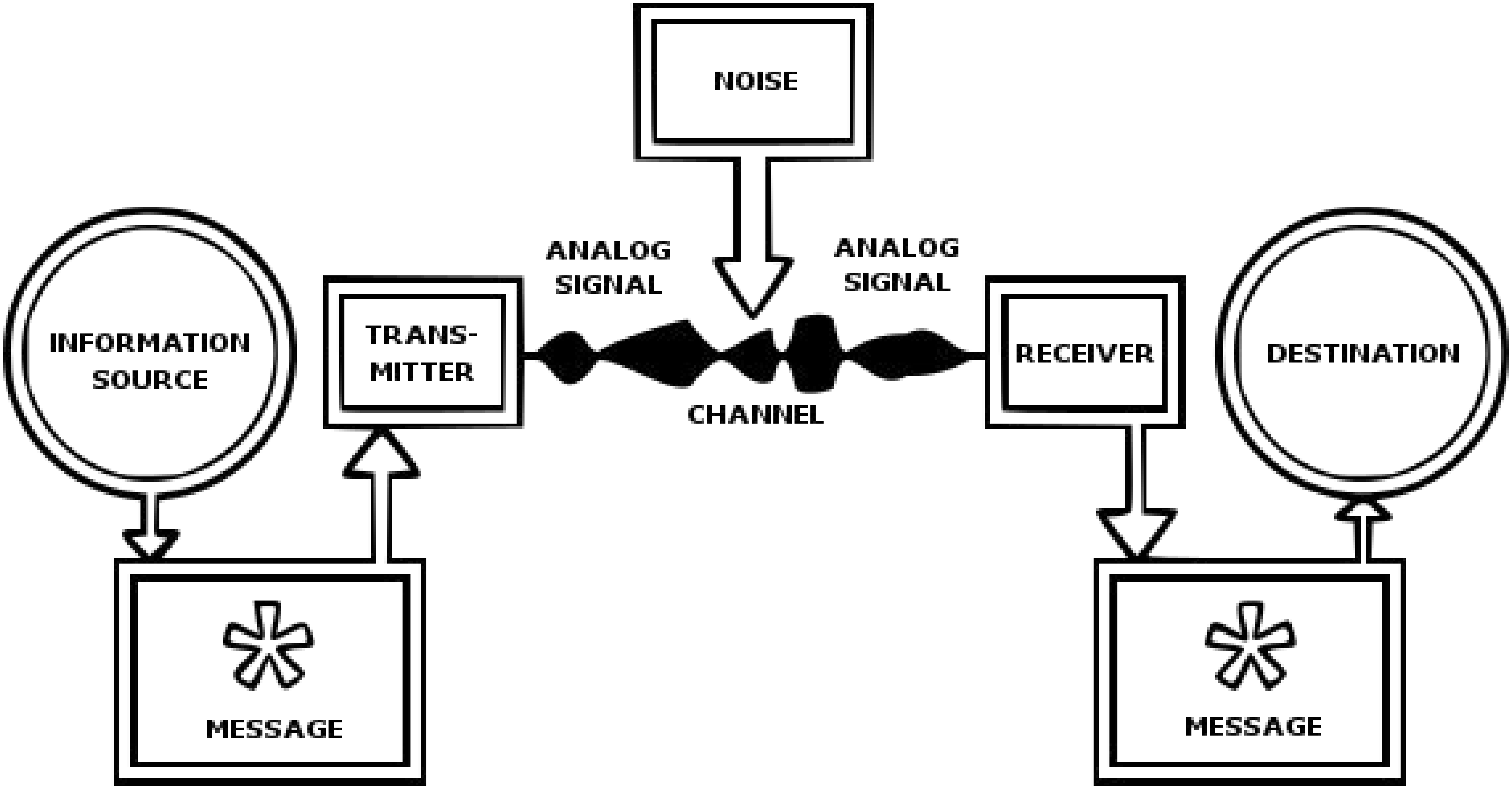
0 1 0 1 = 4 + 1 = 5

1 1 1 1 = 8 + 4 + 2 + 1 = 15

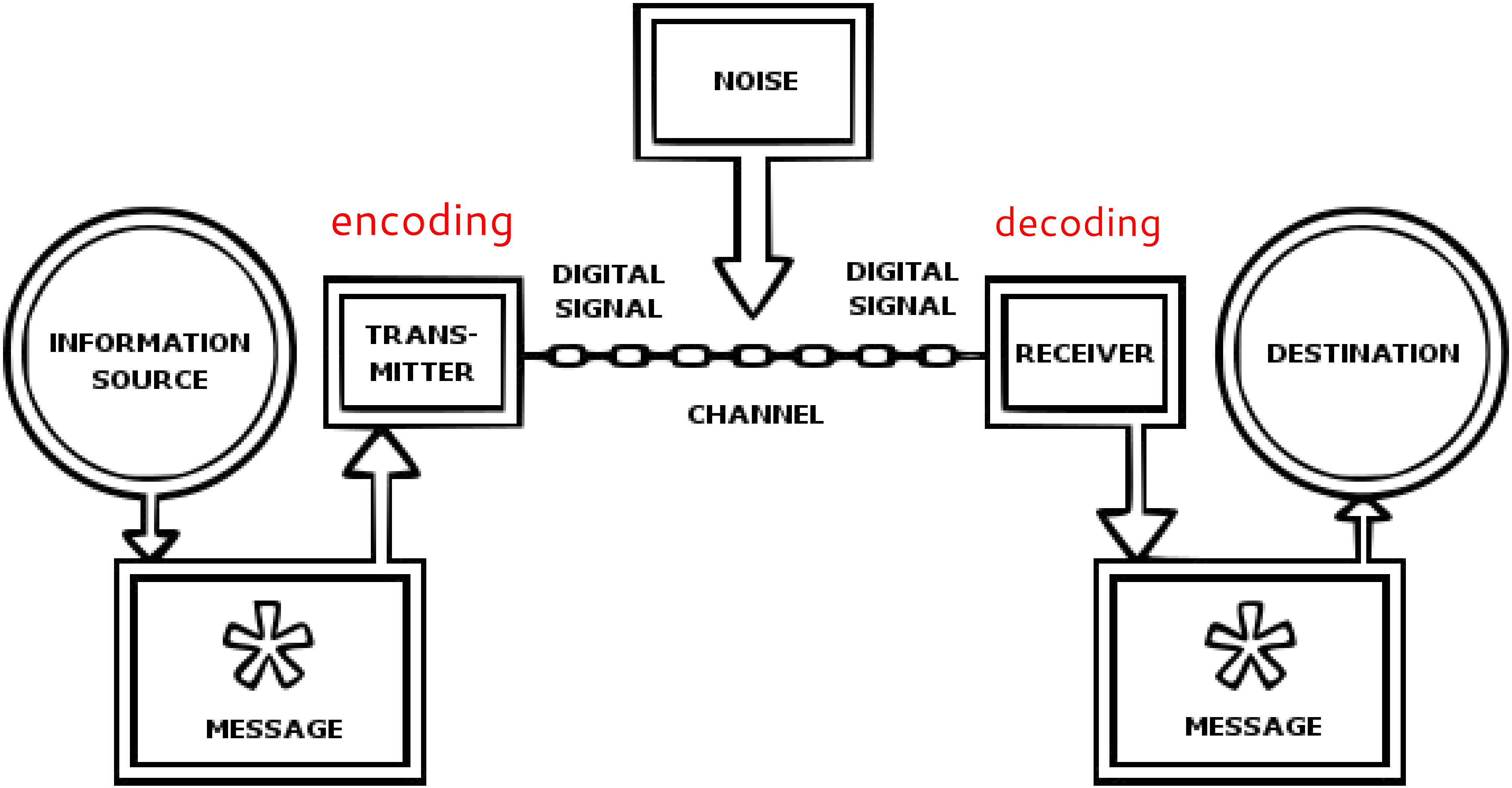


- N bits index 2^N integers, from 0 to 2^N-1
- Exercise:
  - write 9 with 4 bits.
  - write 01010 in decimal.
  - how many bits we need to write 5632?

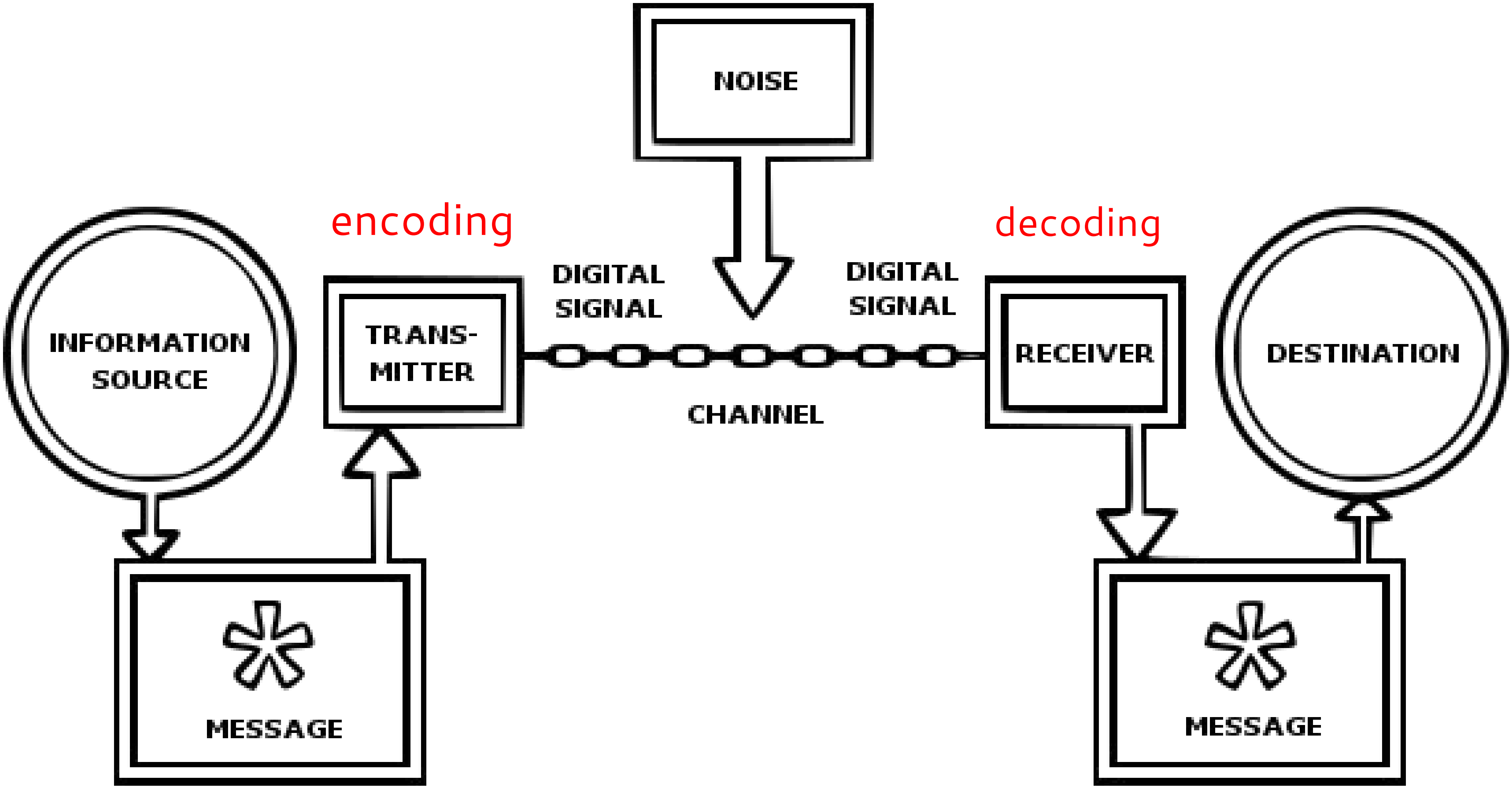
Encoding



Idea: instead of transmitting the *original* **analog signal**…



...what if we transmit the correspondent **digital signal**?



..as we are doing this trasformation,
we can choose if there is an encoding
more efficient than others!

Information

- For instance, we could use the *statistical information* we know about the source.

Information

- For instance, we could use the *statistical information* we know about the source.

- Intuitively
  - **common** symbols transport **less** information
  - **rare** symbols transport **more** information

- It can be interpreted as the *surprise*, the *unexpectedness* of x.

Information

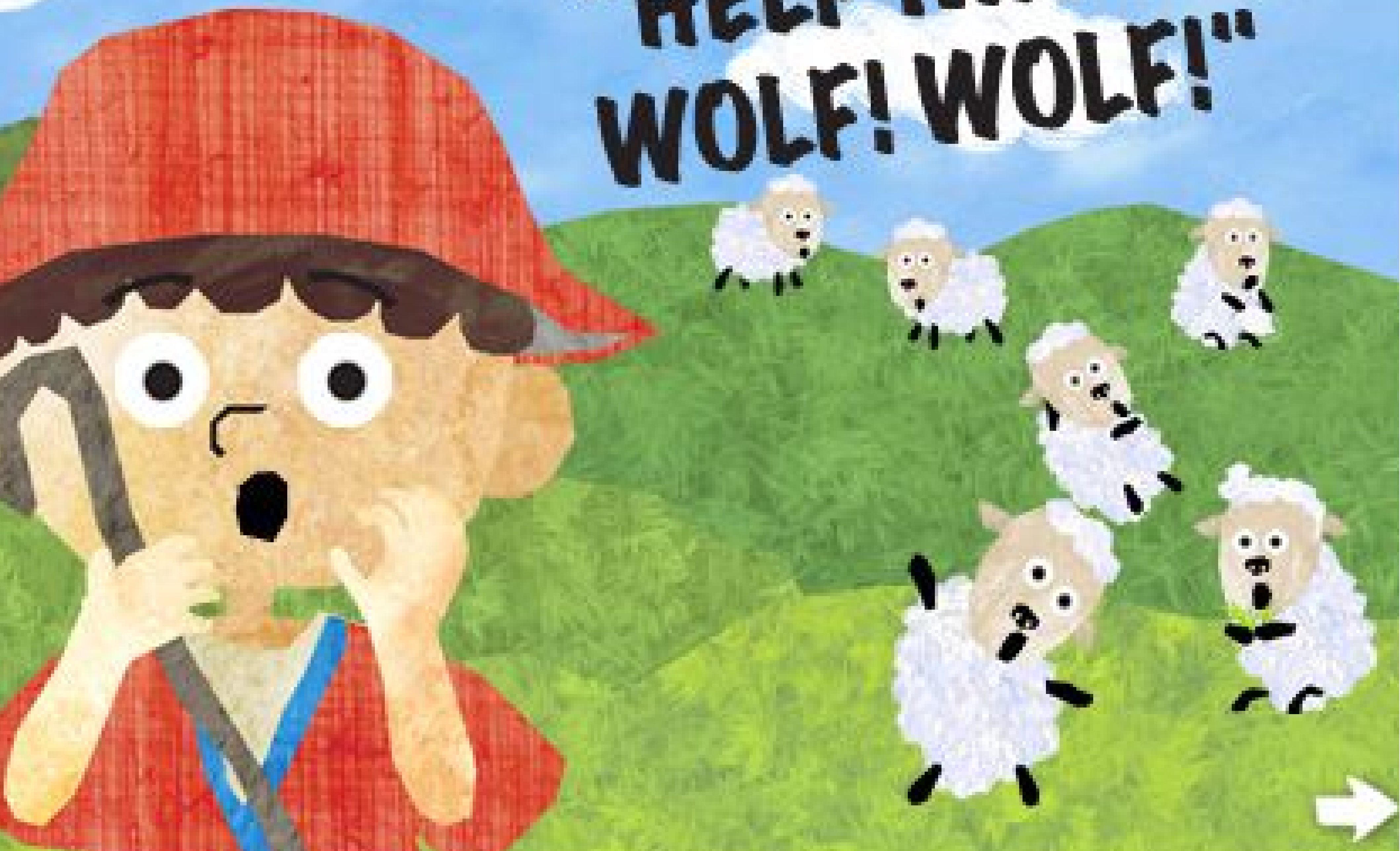
- For instance, we could use the *statistical information* we know about the source.

- Intuitively
  - **common** symbols transport **less** information
  - **rare** symbols transport **more** information

- common/rare to WHO?

He shouted,

"HELP ME! WOLF! WOLF! WOLF!"



Information

- Suppose each symbol **x** is extracted from the source with a certain probability **p**
- We define the **information** associated to the symbol **x**, in respect to the source:

  $$\mathbf{I(x) = -\,Log_2(p)}$$

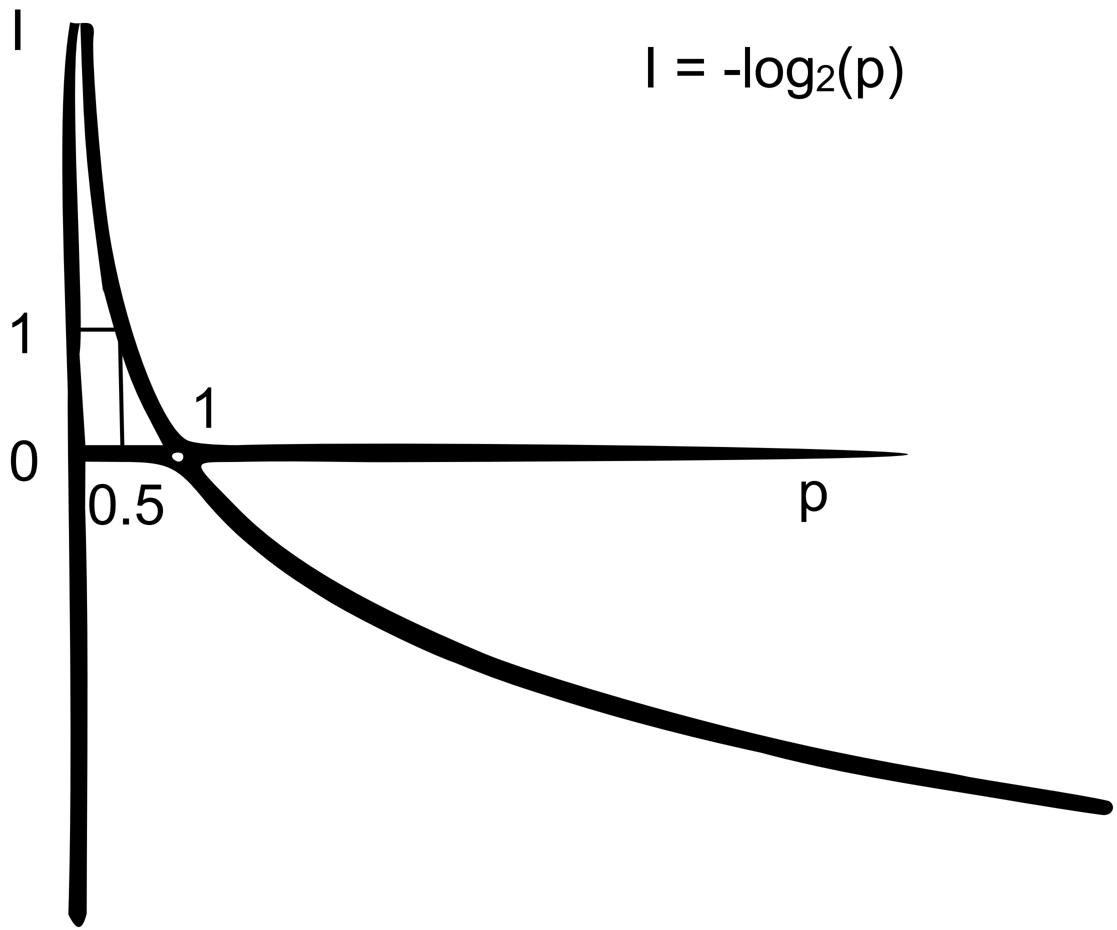
Information

- Suppose each symbol **x** is extracted from the source with a certain probability **p**
- We define the **information** associated to the symbol **x**, in respect to the source:

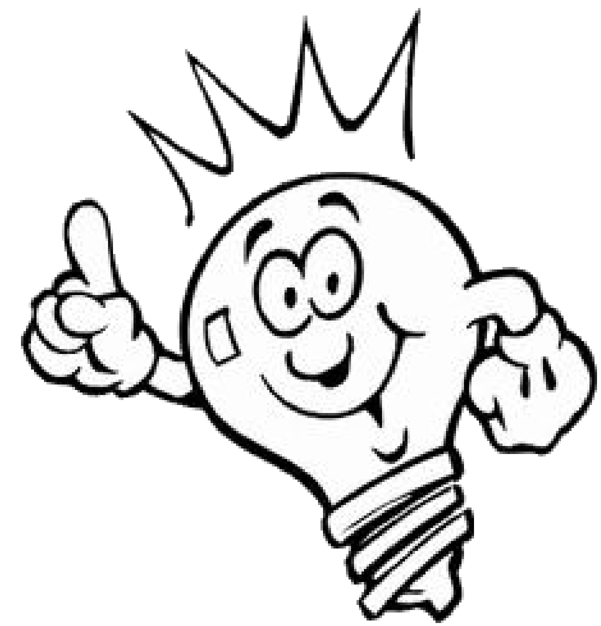
  $$\mathbf{I(x) = -\,Log_2(p)}$$

- Unit of measure: *bit*

How to draw the Log



- $0 \leqslant p \leqslant 1$



- Ex. fair coin: 2 symbols (head, tail)
  p(head) = 1/2, I(head) = 1 bit

Multiple extractions

- Assuming the symbols are **statistically independent**, we can calculate the probability of multiple extractions in this way:

  p(x AND y) = p(x) * p(y)

  I(x AND y) = I(x) + I(y)

Entropy

- Entropy is the **average quantity of information** of the source

$$
\begin{aligned}
H &= p(x) * I(x) + p(y) * I(y) \ldots \\
&= -\,p(x) * Log_2(p(x)) - p(y) * Log_2(p(y)) \ldots
\end{aligned}
$$

  It can be interpreted as the *average missing information* (required to specify an outcome *x* when we now the source probability distribution).

- Unit of measure: *bit/symbol*

Entropy

- Entropy is the **average quantity of information** of the source

$$\begin{aligned} \mathbf{H} &= \mathbf{p(x) * I(x) + p(y) * I(y) \ldots} \\ &= \mathbf{- p(x) * Log_2(p(x)) - p(y) * Log_2(p(y)) \ldots} \end{aligned}$$

- It depends on the *alphabet of symbols* of the source and the associated *probability distribution*.

- Ex. fair coin:
  $H = 0.5 * 1 + 0.5 * 1 = 1$

Maximum Entropy

- The maximum entropy of a source of N symbols is:

  $$\textbf{max H} = \textbf{Log}_2\textbf{(N)}$$

- It is obtained only when all symbols have the *same probability*.

Leibniz Center for Law

Maximum Entropy and Redundancy

- The maximum entropy of a source of N symbols is:

  max H = $\mathbf{Log_2(N)}$

- It is obtained only when all symbols have the *same probability*.

- A measure for *redundancy*:

  Redundancy = (max H – actual H)/max H

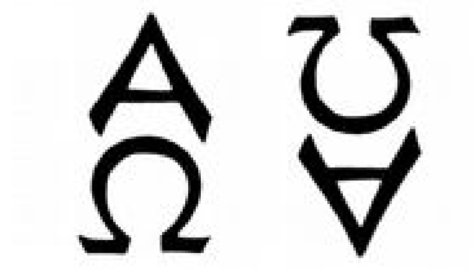
Exercise

- Calculate the information per symbol for a source with an alphabet of statistically independent symbols, with equal probability,

  - consisting of 1 symbol.
  - consisting of 2 symbols.
  - consisting of 16 symbols.

- How much the entropy?

Exercise



- Calculate the entropy of a source with an alphabet consisting of 4 symbols with these probability:

  $p_1 = 1/2$

  $p_2 = 1/4$

  $p_3 = p_4 = 1/8$

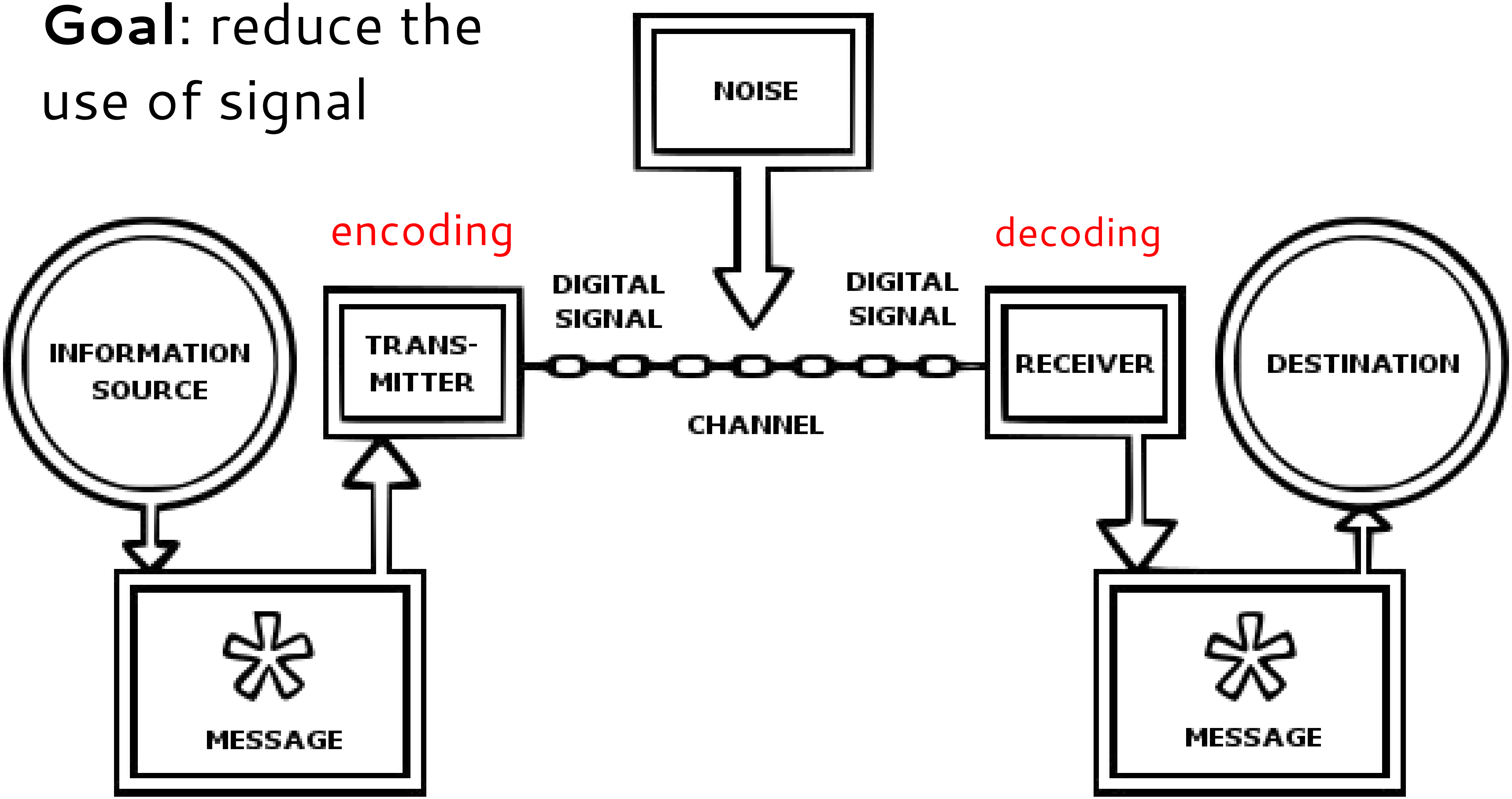
- What is the entropy if the symbols are equally probable?
- What is the redundancy?

Compression

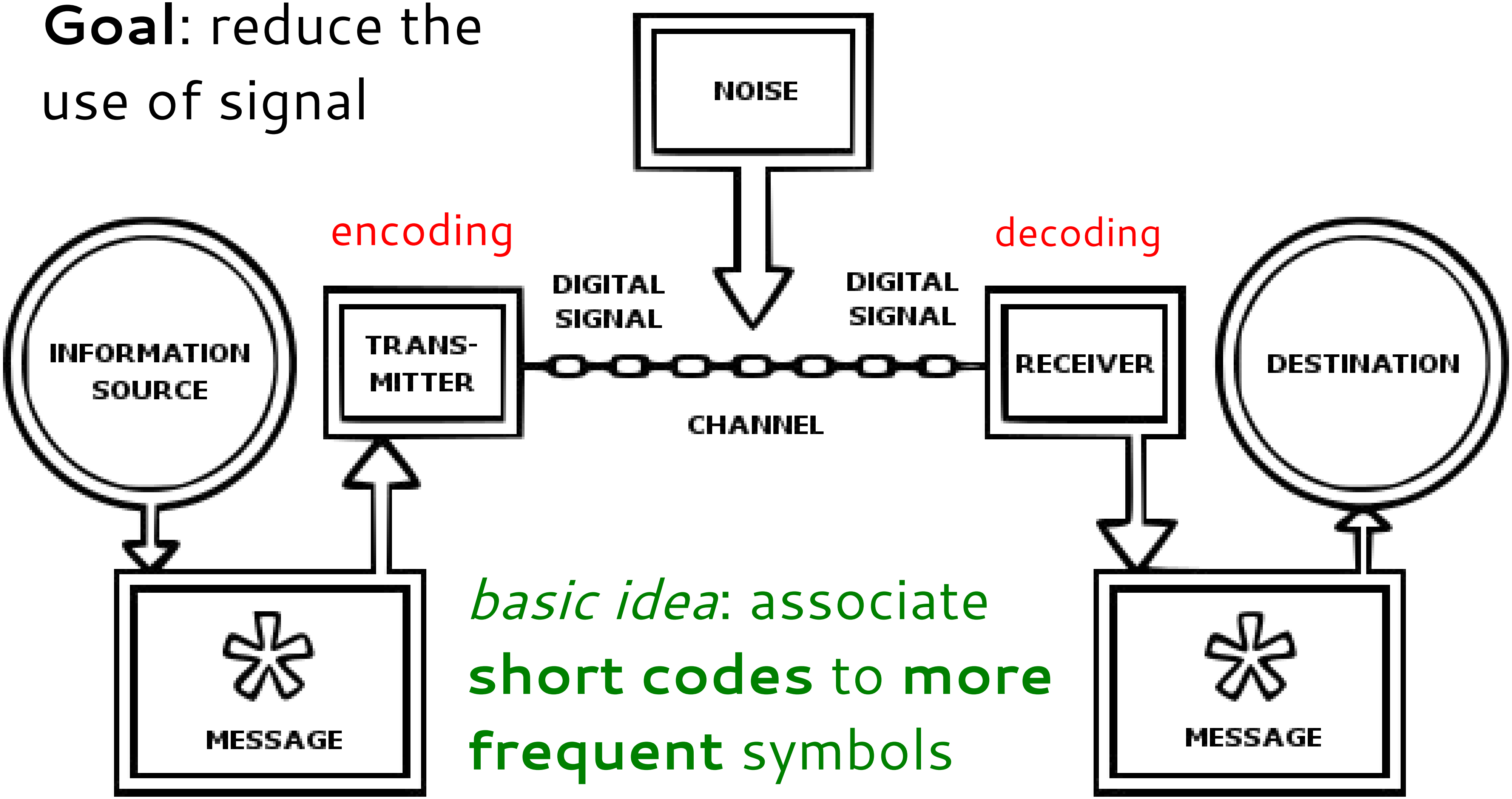
Source encoding

- In order to reduce the resource requirements (bandwith, power, etc.) we are interested to minimize the average length of the words of the code.

Goal: reduce the use of signal

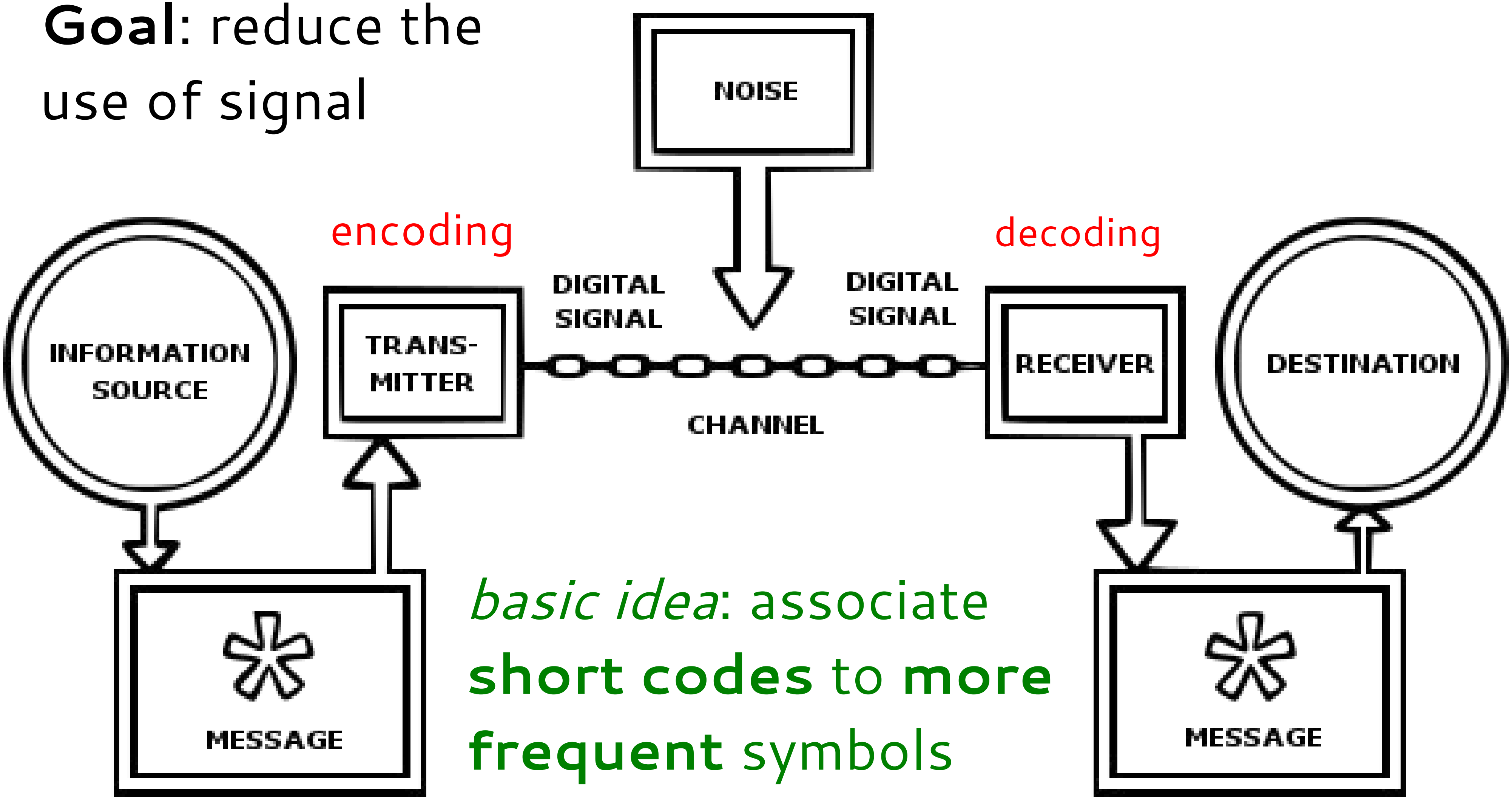


Goal: reduce the use of signal



basic idea: associate **short codes** to **more frequent** symbols

Goal: reduce the use of signal



encoding

decoding

basic idea: associate **short codes** to **more frequent** symbols

we are operating a **compression** on the messages!

Huffman algorithm

- produces an optimal *encoding* (as average length).
- 3-steps:
  - **order** symbols according to their probability
  - **group by two** the least probable symbols, and **sum** up their probabilities associated to a *new equivalent compound symbol*
  - **repeat** until you obtain *only one* equivalent compound symbol with probability 1

Example: Huffman algorithm

- Source counting 5 symbols with probability 1/3, 1/4, 1/6, 1/6, 1/12.

Example: Huffman algorithm

A] 1/3

B] 1/4

C] 1/6

D] 1/6

E] 1/12

Example: Huffman algorithm

A] 1/3

B] 1/4

C] 1/6

D] 1/6

E] 1/12

Example: Huffman algorithm

A] 1/3

B] 1/4

C] 1/6

D] 1/6

0 1/4

E] 1/12

1

Example: Huffman algorithm

A] 1/3

B] 1/4

C] 1/6

D] 1/6

E] 1/12

0 1/4

1

Example: Huffman algorithm

A] 1/3

B] 1/4

C] 1/6

0 5/12

1

D] 1/6

0 1/4

1

E] 1/12

Example: Huffman algorithm

A] 1/3 = 4/12

B] 1/4 = 3/12

C] 1/6

0 5/12

1

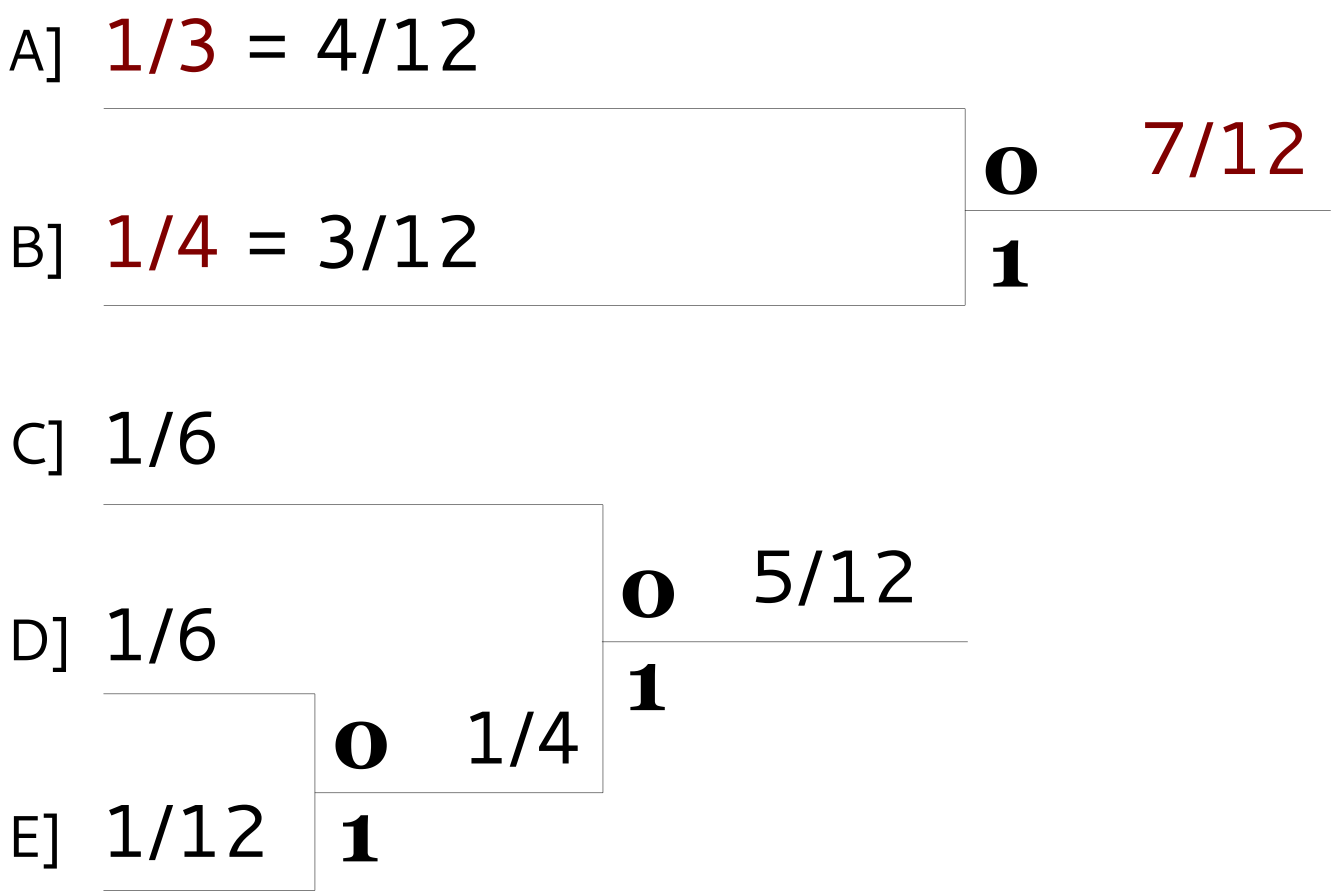
D] 1/6

0 1/4

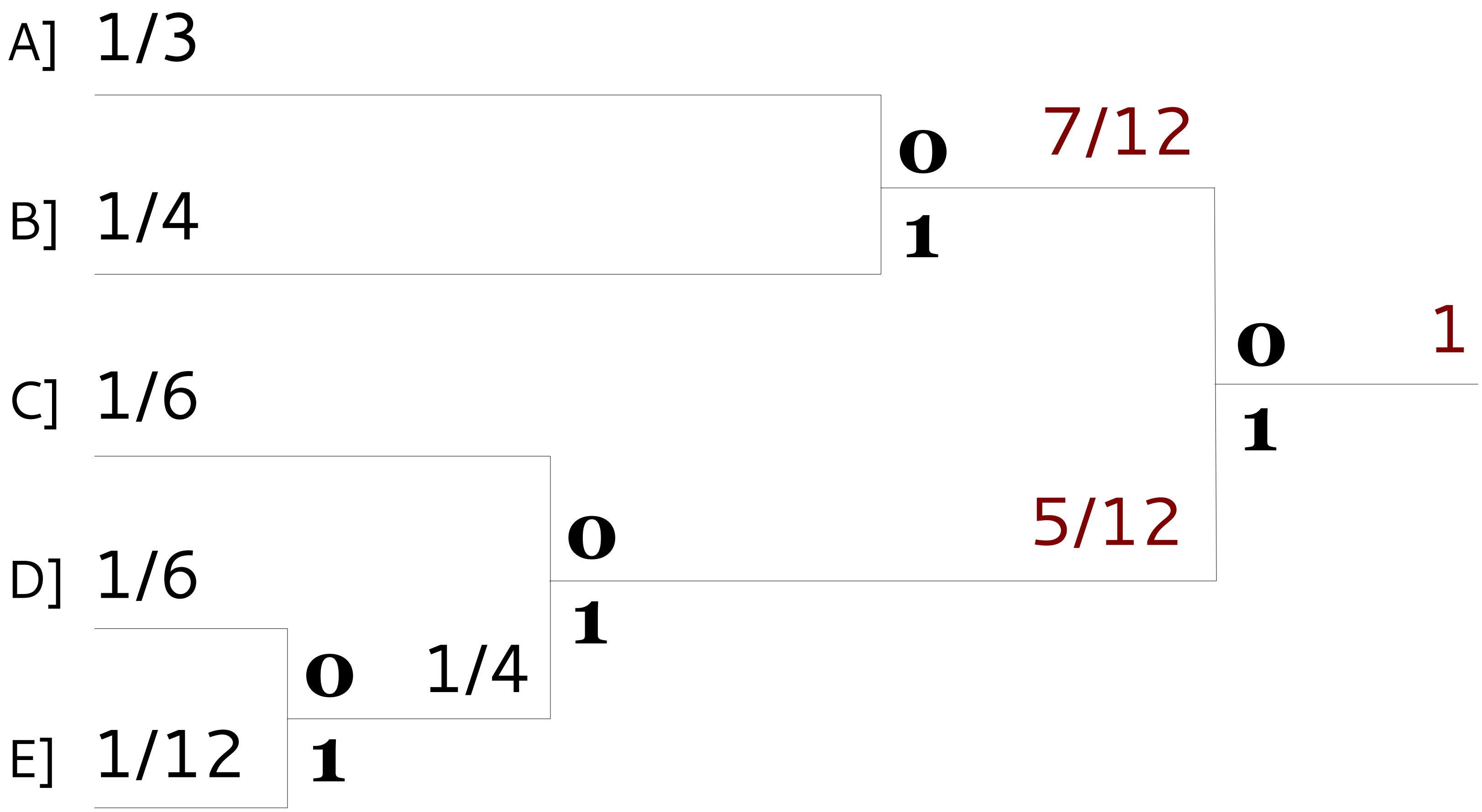
1

E] 1/12

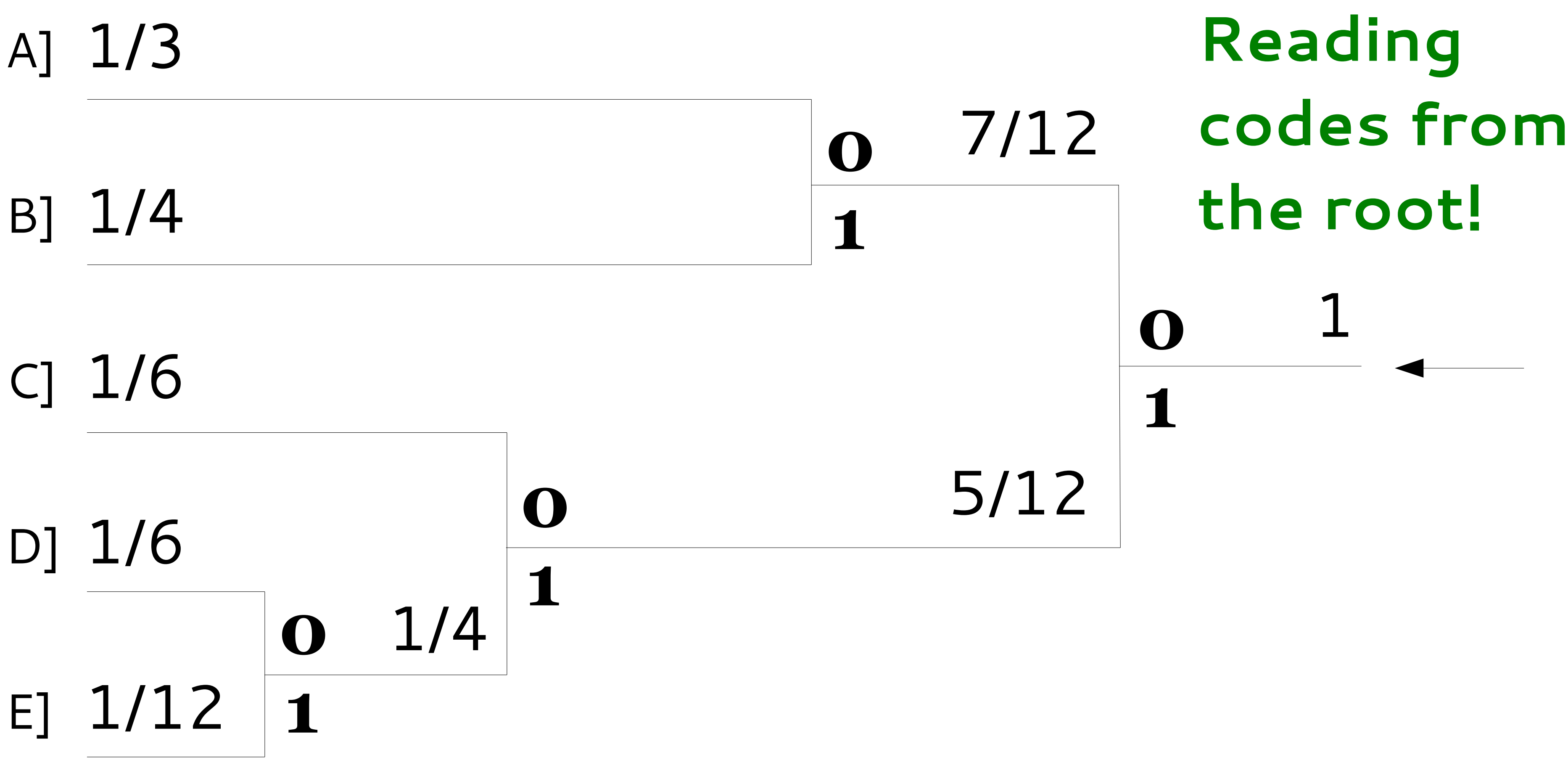
Example: Huffman algorithm



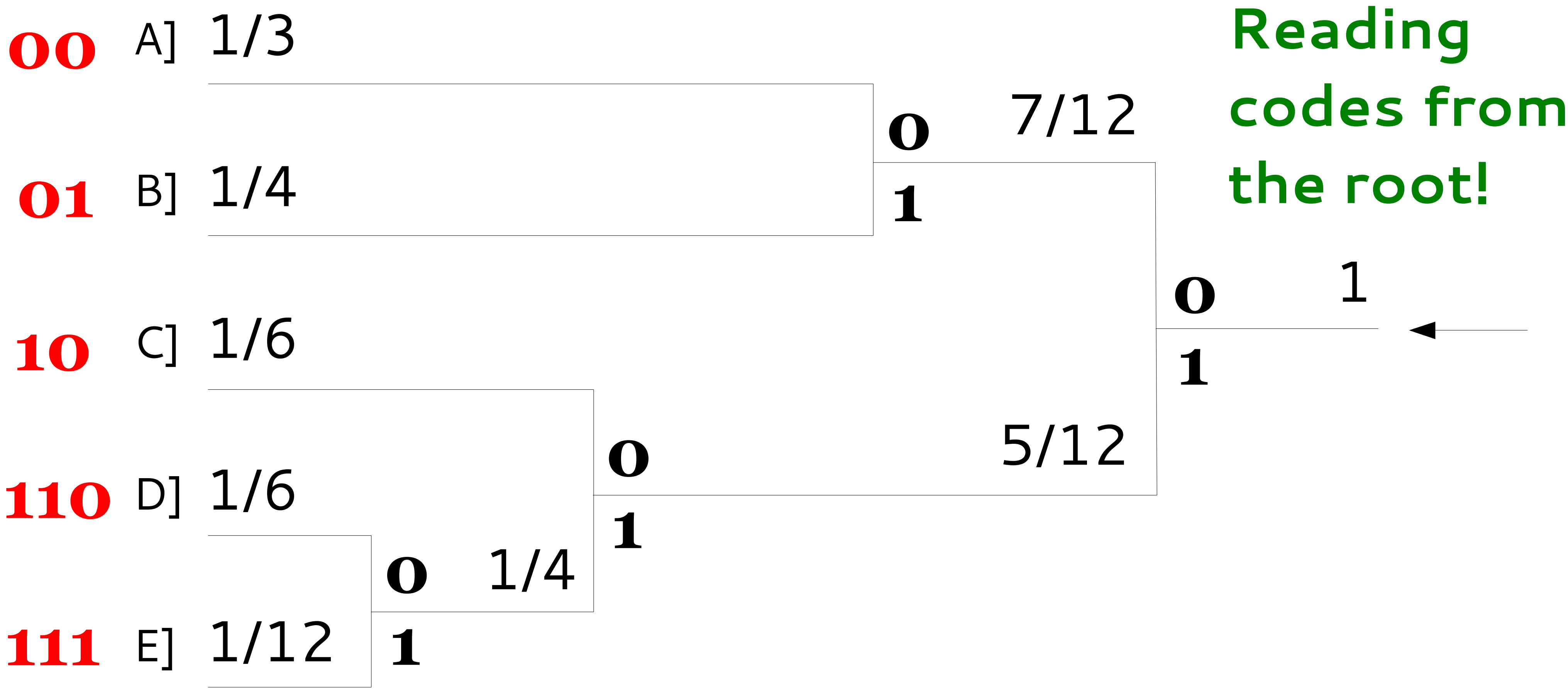
Example: Huffman algorithm



Example: Huffman algorithm



Example: Huffman algorithm



Source Coding Theorem

- Given a source with entropy H, it is always possible to find an encoding which satisfies:

  H ≤ average code length < H + 1

Source Coding Theorem

- Given a source with entropy H, it is always possible to find an encoding which satisfies:

  H ≤ average code length < H + 1

In the previous exercise:

$H = - 1/3 * \mathrm{Log}_2 (1/3) - 1/4 * \mathrm{Log}_2 (1/4) - \ldots$

$ACL = 2 * 1/3 + 2 * 1/4 + .. + 3 * 1/12$

H = 2.19, ACL = 2.25

Exercise

- Propose an encoding for a communication system associated to a sensor placed in a rainforest.
- The sensor recognizes the warbles/tweets of birds from several species..


toucan


parrot


hornbill
mongabay.com


eagle

Exercise

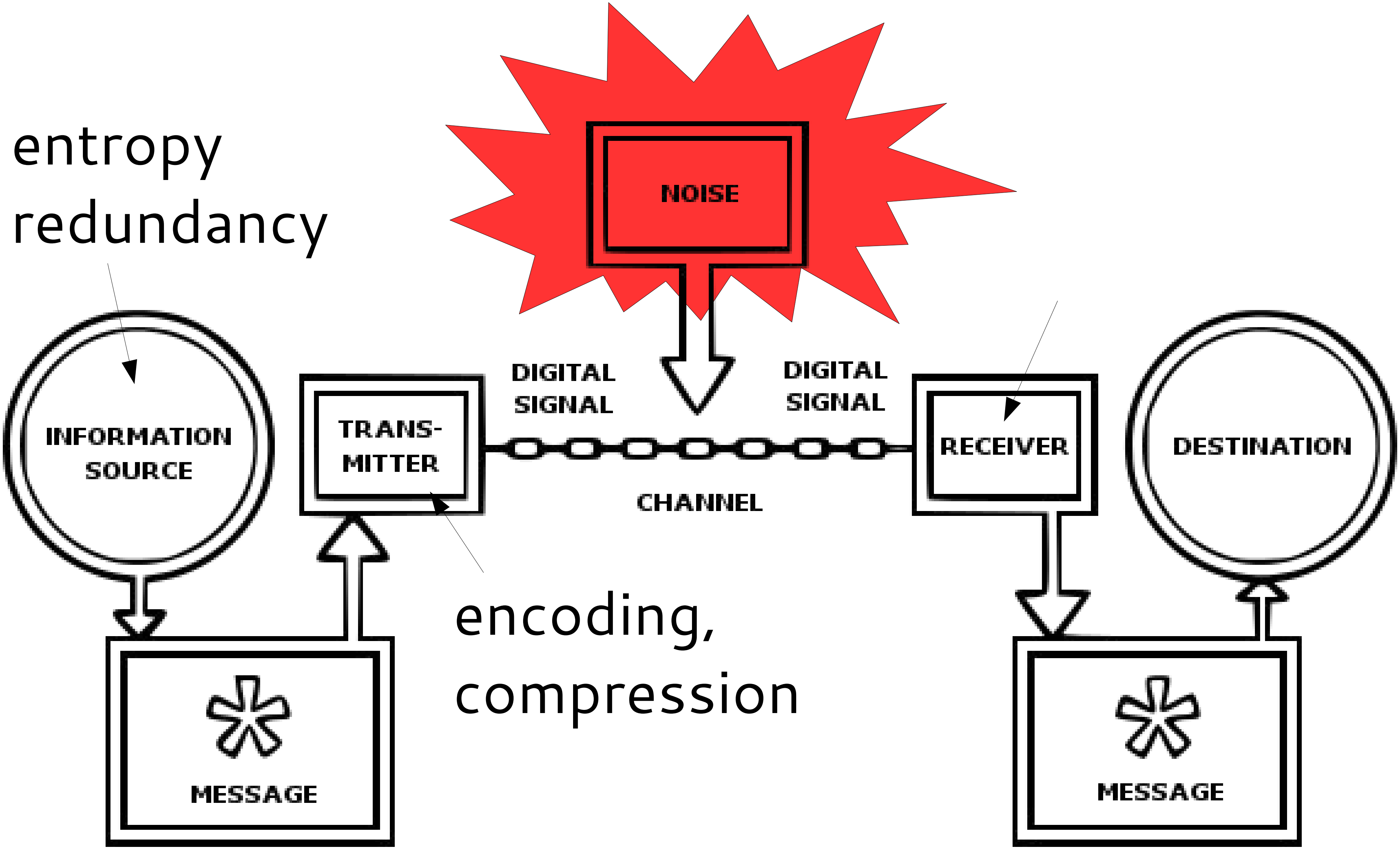
- Propose an encoding for a communication system associated to a sensor placed in a rainforest.
- The sensor recognizes the warbles/tweets of birds from several species, whose presence is described by these statistics:

p(toucan) = 1/3	p(parrot) = 1/2
p(eagle) = 1/24	p(hornbill) = 1/8

- Which of the assumptions you have used may be critical in this scenario?

Noise

entropy
redundancy



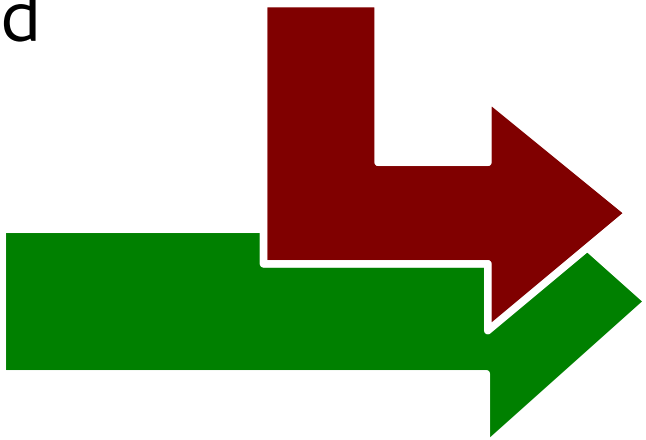
encoding,
compression

Type of noise

- Noise can be seen as an unintented source which interferes with the intented one.

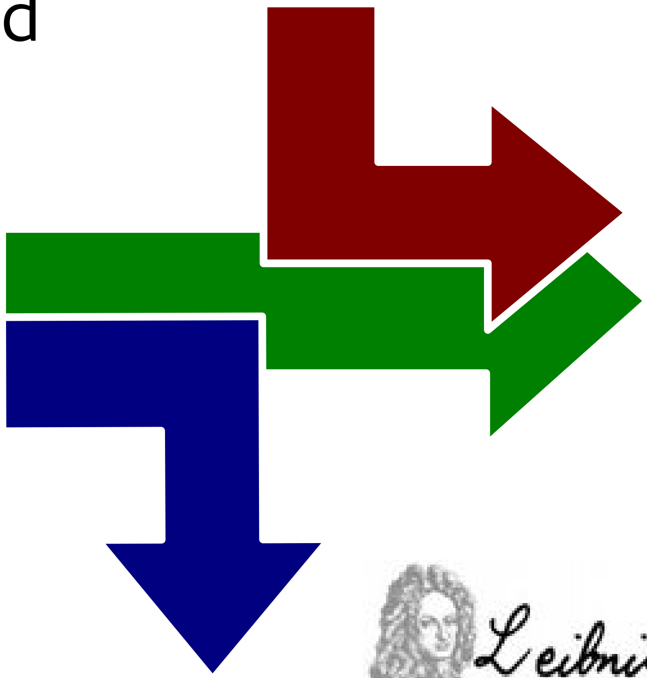
Type of noise

- Noise can be seen as an unintented source which interferes with the intented one.
- In terms of the outcomes, the communication channel may suffer of two types of interferences:
  - data received but unwanted



Type of noise

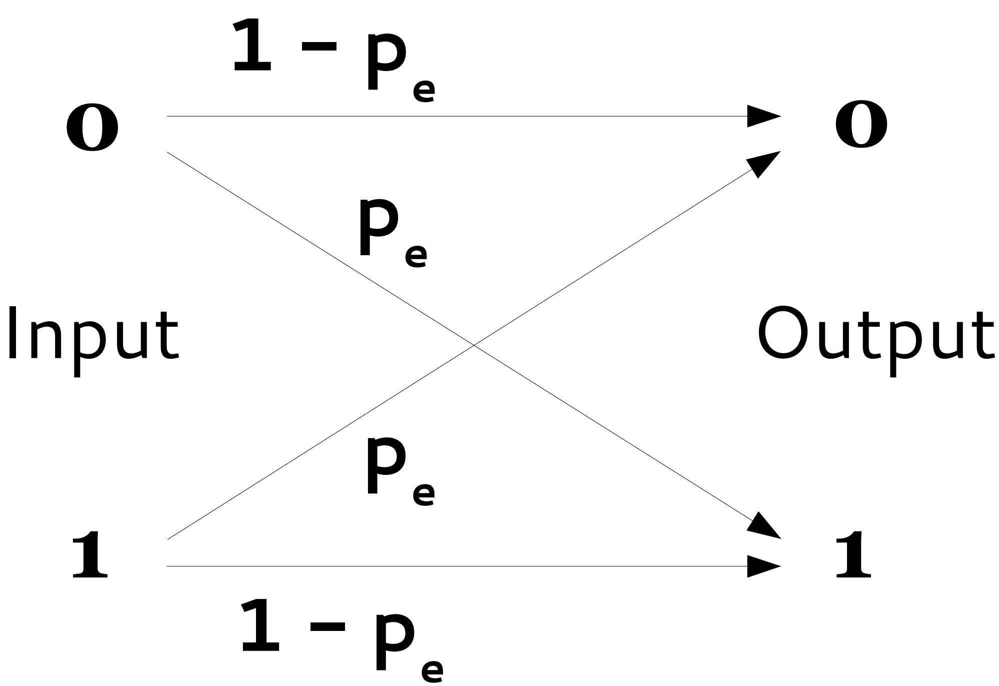
- Noise can be seen as an unintented source which interferes with the intented one.
- In terms of the outcomes, the communication channel may suffer of two types of interferences:
  - data received but unwanted
  - data sent never received



Leibniz Center for Law

Binary Simmetric Channel

- A *binary symmetric channel* (BSC) models the case that a binary input is flipped before the output.



p_e = error probability

$1 - p_e$ = probability of correct transmission

Binary Simmetric Channel

- Probability of transmissions on 1 bit

Input	Output	
0	0	$\rightarrow \mathbf{1 - p_e}$
0	1	$\rightarrow \mathbf{p_e}$

- Probability of transmissions on 2 bit

Input	Output	
0 0	0 0	$\rightarrow \mathbf{(1 - p_e) * (1 - p_e)}$
0 0	0 1	$\rightarrow \mathbf{(1 - p_e) * p_e}$
0 0	1 0	$\rightarrow \mathbf{p_e * (1 - p_e)}$
0 0	1 1	$\rightarrow \mathbf{p_e * p_e}$

Exercise

- Consider messages of 3 bits,
  - what is the probability of 2 bits inversion?
  - what is the probability of error?

Error detection

Simple error detection

- **Parity check**
  - A *parity* bit is added at the end of a of a string of bits (eg. 7): 0 if the number of 1 is even, 1 if odd

Coding :

0000000 → 00000000

1001001 → 10010011

0111111 → 01111110

Example of error detection

- **Parity check**
  - A *parity bit* is added at the end of a of a string of bits (eg. 7): 0 if the number of 1 is even, 1 if odd

Decoding while detecting errors

01111110 → ok

00100000 → error detected

10111011 → error not detected!

Exercise

- Add the parity bit

01100011010?

01011100?

0010010001?

1111011100100?

Perform the parity check

1110001010111

0001110011

1001110100

11011

Exercise

- Consider messages of 2 bits + 1 parity bit.
- What is the probability to detect the error?

Error correction

Simple error correction

- *Forward Error Correction* with (3, 1) repetition, each bit is repeated two times more.

Coding :

0	→	000
1	→	111
11	→	111111
010	→	000111000

Simple error correction

- *Forward Error Correction* with (3, 1) repetition, each bit is repeated two times more.

Decoding (while correcting errors)

010	→	0
011	→	1
111101	→	11
100011000	→	010

Exercise

- Decode and identify the errors on the following encoding:

  011000110101

  010111001000

  001001000011

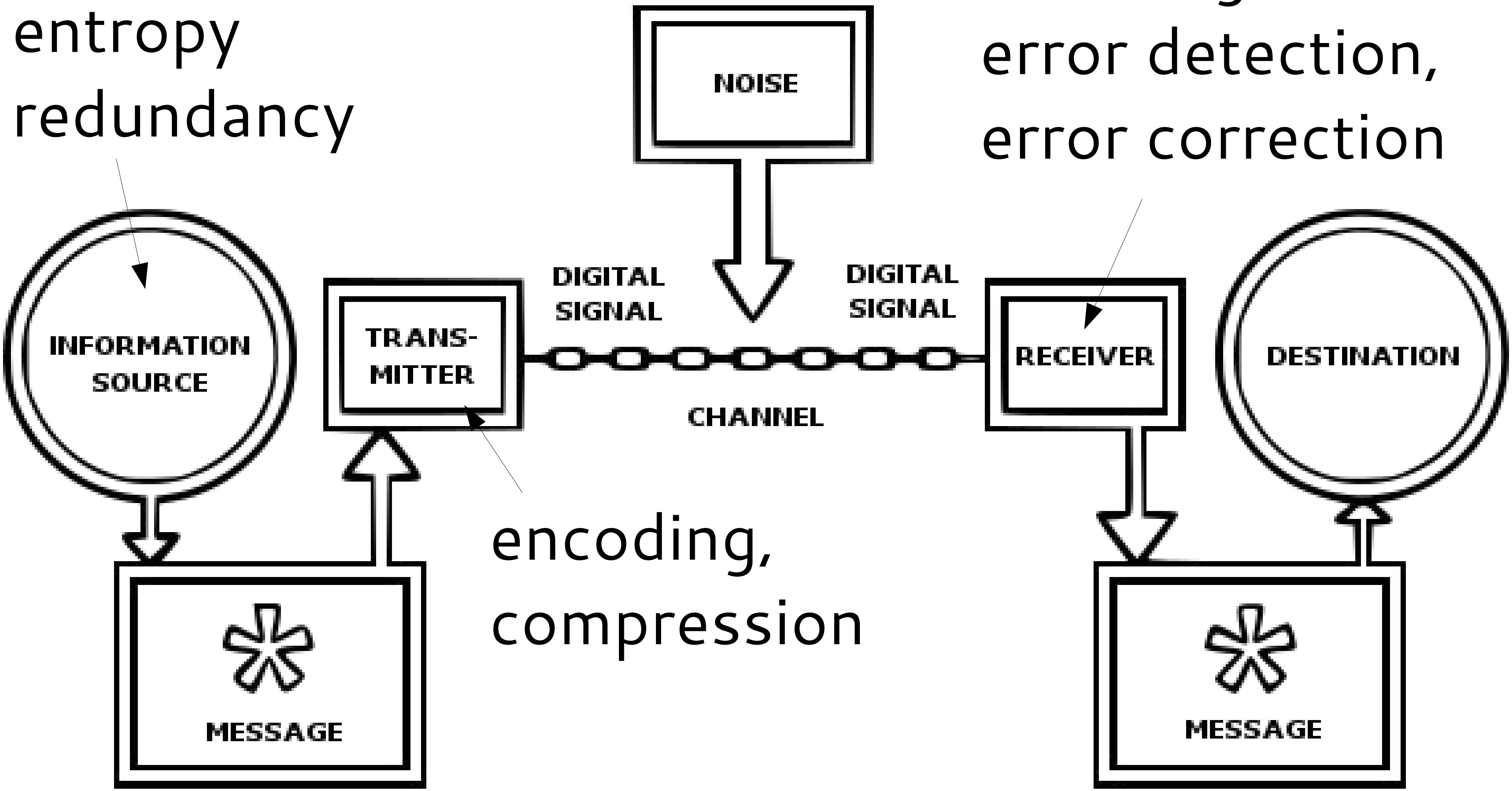
  111011001001

Summary

channel capacity

entropy
redundancy

decoding,
error detection,
error correction



encoding,
compression

Main points - Entropy

- In Information Science, Entropy is a measure of the **uncertainty** at the *reception point* of messages generated by a source.
- Greater entropy, greater signal randomness
- Less entropy, more redundancy.

Main points - Entropy

- In Information Science, Entropy is a measure of the **uncertainty** at the *reception point* of messages generated by a source.
- Greater entropy, greater signal randomness
- Less entropy, more redundancy.

- It depends on what counts as symbol and their probability distributions, which are always taken by an <u>observer</u>.

Side comment - Entropy

- In Physics, Entropy is a function related to the **amount of disorder**. It always increases (even if locally may decrease).

Side comment - Entropy

- In Physics, Entropy is a function related to the **amount of disorder**. It always increases (even if locally may decrease).

Main points – Redundancy & Noise

- As all communications suffer to a certain extent from noise, adding some **redundancy** is good for transmission, as it helps in detecting or even correcting certain errors.

Main points – Redundancy & Noise

- As all communications suffer to a certain extent from noise, adding some **redundancy** is good for transmission, as it helps in detecting or even correcting certain errors.
- Example:

(Some People Can Read This)

Somr Peoplt Cat Rea Tis

Main points – Redundancy & Noise

- As all communications suffer to a certain extent from noise, adding some **redundancy** is good for transmission, as it helps in detecting or even correcting certain errors.
- Example:

(SM PPL CN RD THS)

SMR PPLT CT R TS

Main points – Redundancy & Noise

- As all communications suffer to a certain extent from noise, adding some **redundancy** is good for transmission, as it helps in detecting or even correcting certain errors.
- Example:

  Some People Can Read This → Somr Peoplt Cat Rea Tis

  SM PPL CN RD THS → SMR PPLT CT R TS

- The redundancy is expressed by the correlation between letters composing words! (independent probability assumption not valid!)

Literature

Shannon, C. E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27, 379–423/623–656.

Weaver, W. (1949). Recent contributions to the mathematical theory of communication. The mathematical theory of communication.

Floridi, L. (2009). Philosophical conceptions of information. Formal Theories of Information, (2), 13–53.

Guizzo, E. M. (2003). The essential message: Claude Shannon and the making of information theory. Massachusetts Institute of Technology.

Lesne, A. (2014). Shannon entropy: a rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics. Mathematical Structures in Computer Science, 24(3).