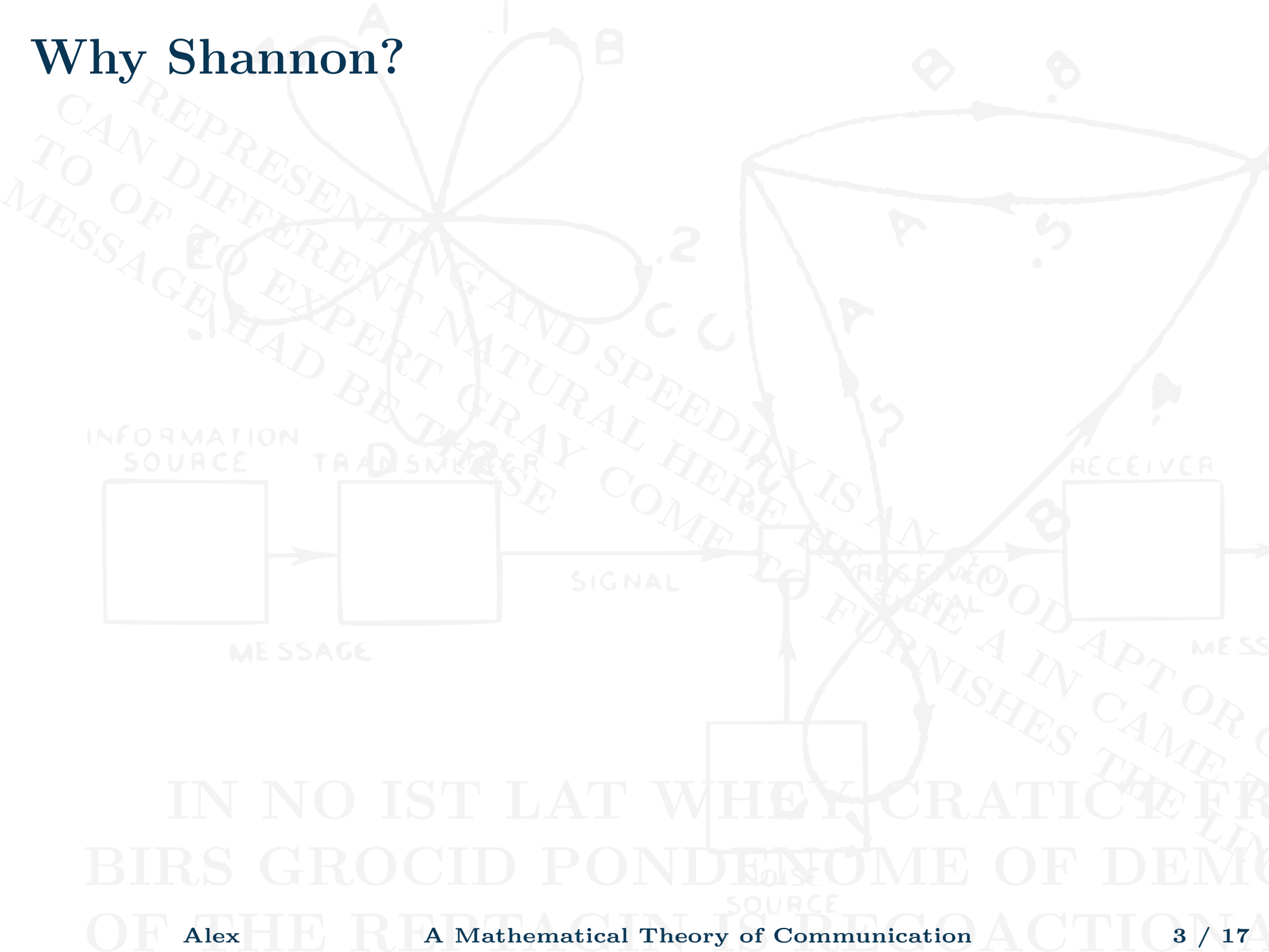
A Mathematical Theory of Communication

(after C. E. Shannon)

Alex Vlasiuk

Image: IEEE Information Theory Society

Why Shannon?


INFORMATION SOURCE
TRANSMITTER
SIGNAL
RECEIVER
MESSAGE
NOISE SOURCE

Why Shannon?

- "the father of information theory"

Why Shannon?

- "the father of information theory"
- ideas from the 1948 paper are ubiquitous

Why Shannon?

- "the father of information theory"
- ideas from the 1948 paper are ubiquitous
- (hopefully) some can be explained through handwaving

Why Shannon?

- "the father of information theory"
- ideas from the 1948 paper are ubiquitous
- (hopefully) some can be explained through handwaving:

©Jeff Portaro, Noun Project

Why Shannon?

- "the father of information theory"
- ideas from the 1948 paper are ubiquitous
- (hopefully) some can be explained through handwaving:

©Jeff Portaro, Noun Project

- was on my desktop

Why Shannon?

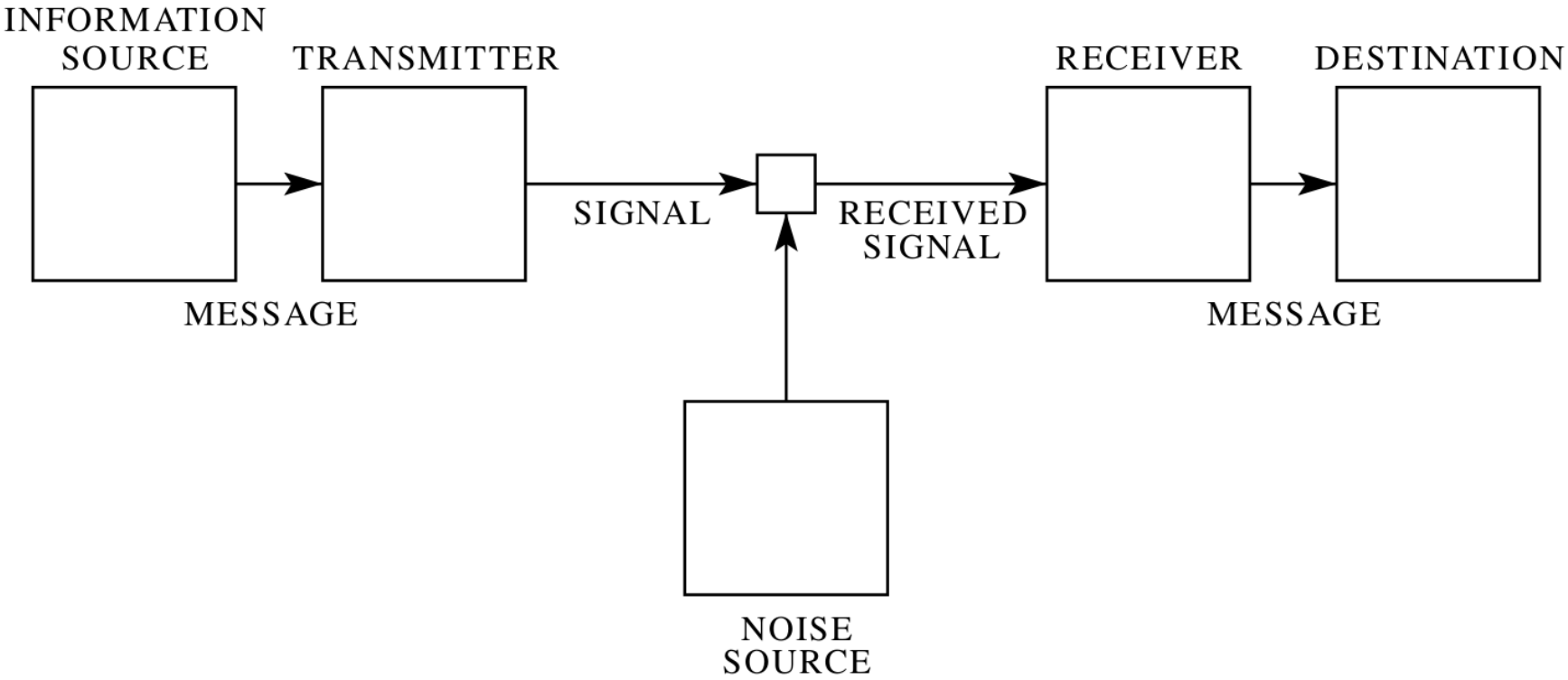
- "the father of information theory"
- ideas from the 1948 paper are ubiquitous
- (hopefully) some can be explained through handwaving:

©Jeff Portaro, Noun Project

- was on my desktop

Shannon, Claude Elwood. "A mathematical theory of communication." ACM SIGMOBILE Mobile Computing and Communications Review 5.1 (2001): 3-55.

Setting


INFORMATION SOURCE
TRANSMITTER
RECEIVER
DESTINATION
SIGNAL
RECEIVED SIGNAL
MESSAGE
MESSAGE
NOISE SOURCE

Capacity and states of a channel

Symbols: $S_1, \ldots, S_n$ with certain durations $t_1, \ldots, t_n$.

Capacity and states of a channel

Symbols: $S_1, \ldots, S_n$ with certain durations $t_1, \ldots, t_n$. Allowed combinations of symbols are signals.

Capacity and states of a channel

Symbols: $S_1, \ldots, S_n$ with certain durations $t_1, \ldots, t_n$. Allowed combinations of symbols are signals.

Capacity of a channel:

$$C = \lim_{T \to \infty} \frac{\log N(T)}{T},$$

$N(T)$ is the number of allowed signals of duration T.

Capacity and states of a channel

Symbols: $S_1, \ldots, S_n$ with certain durations $t_1, \ldots, t_n$. Allowed combinations of symbols are signals.

Capacity of a channel:

$$C = \lim_{T \to \infty} \frac{\log N(T)}{T},$$

$N(T)$ is the number of allowed signals of duration T. Units: bits per second.

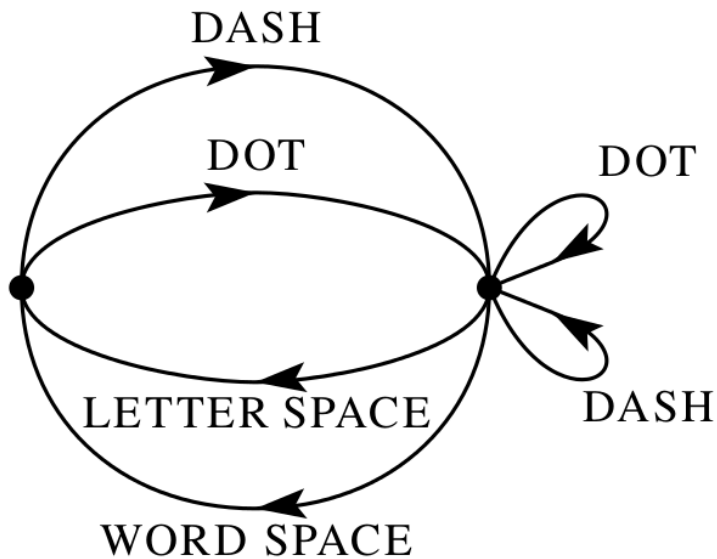
Capacity and states of a channel

Symbols: $S_1, \ldots, S_n$ with certain durations $t_1, \ldots, t_n$. Allowed combinations of symbols are signals.

Capacity of a channel:

$$C = \lim_{T \to \infty} \frac{\log N(T)}{T},$$

$N(T)$ is the number of allowed signals of duration T. Units: bits per second.



Graphical representation of a Markov process

Source is a stochastic (random) process.

Graphical representation of a Markov process

Source is a stochastic (random) process.
Example. Alphabet: A, B, C.

Graphical representation of a Markov process

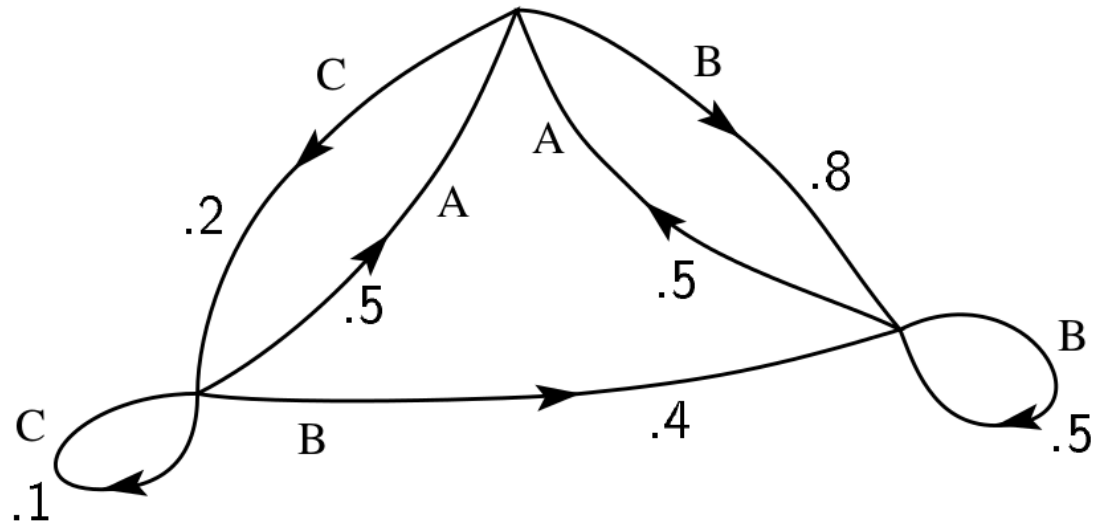
Source is a stochastic (random) process.
Example. Alphabet: A, B, C. Transition probabilities:

$p_i(j)$		j		
		A	B	C
i	A	0	$\frac{4}{5}$	$\frac{1}{5}$
	B	$\frac{1}{2}$	$\frac{1}{2}$	0
	C	$\frac{1}{2}$	$\frac{2}{5}$	$\frac{1}{10}$

Graphical representation of a Markov process

Source is a stochastic (random) process.
Example. Alphabet: A, B, C. Transition probabilities:

$p_i(j)$		j		
		A	B	C
i	A	0	$\frac{4}{5}$	$\frac{1}{5}$
	B	$\frac{1}{2}$	$\frac{1}{2}$	0
	C	$\frac{1}{2}$	$\frac{2}{5}$	$\frac{1}{10}$



Example: approximations to English

Using 27 (26+space) alphabet.

Example: approximations to English

Using 27 (26+space) alphabet.

- symbols independent and equiprobable:
  XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD
  QPAAMKBZAACIBZLHJQD

Example: approximations to English

Using 27 (26+space) alphabet.

- symbols independent and equiprobable:
  XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD
- symbols independent but with frequencies of English text:
  OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL

Example: approximations to English

Using 27 (26+space) alphabet.

- symbols independent and equiprobable:
  XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD
- symbols independent but with frequencies of English text:
  OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL
- digram structure as in English:
  ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE

Example: approximations to English

Using 27 (26+space) alphabet.

- symbols independent and equiprobable:
  XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD
- symbols independent but with frequencies of English text:
  OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL
- digram structure as in English:
  ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE

  "One opens a book at random and selects a letter at random on the page. This letter is recorded. The book is then opened to another page and one reads until this letter is encountered. The succeeding letter is then recorded."

- trigram structure as in English:
  IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE

- trigram structure as in English:
  IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE
- first-order word approximation:
  REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE

- trigram structure as in English:
  IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE
- first-order word approximation:
  REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE
- Second-order word approximation:
  THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED

Entropy

A set of possible events with probabilities

$$p_1, p_2, \ldots, p_n$$

Entropy

A set of possible events with probabilities

$$p_1, p_2, \ldots, p_n$$

Need: a measure of uncertainty in the outcome

Entropy

A set of possible events with probabilities

$$p_1, p_2, \ldots, p_n$$

Need: a measure of uncertainty in the outcome

$$\boldsymbol{H = -\sum_{i=1}^{n} p_i \log p_i}$$

Entropy

A set of possible events with probabilities

$$p_1, p_2, \ldots, p_n$$

Need: a measure of uncertainty in the outcome

$$\boldsymbol{H = -\sum_{i=1}^{n} p_i \log p_i}$$

Example: two possibilities with probabilities p and $q = 1 - p$.

Entropy

A set of possible events with probabilities

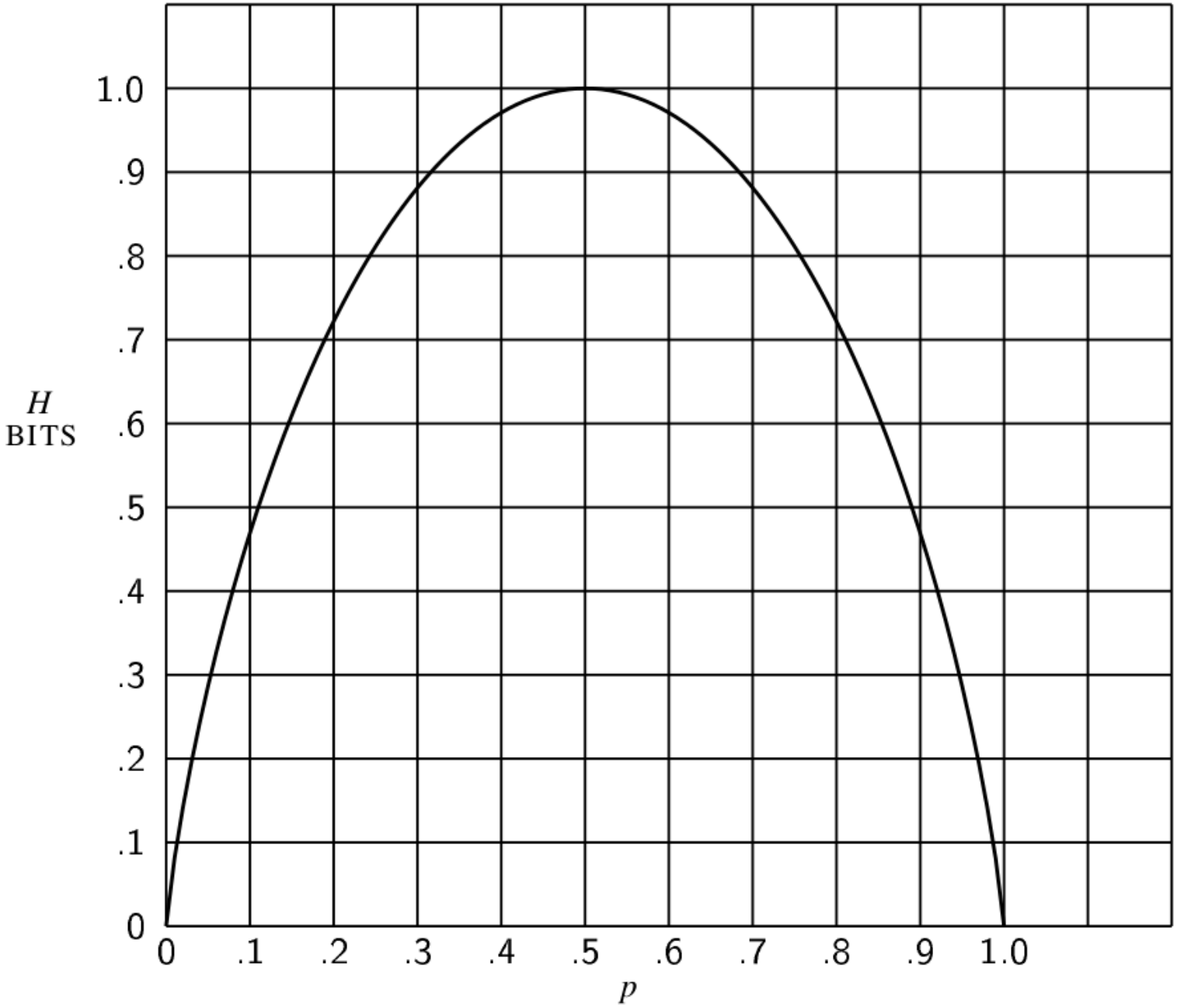
$$p_1, p_2, \ldots, p_n$$

Need: a measure of uncertainty in the outcome

$$\boldsymbol{H = -\sum_{i=1}^{n} p_i \log p_i}$$

Example: two possibilities with probabilities p and $q = 1 - p$.

$$H = -(p \log p + q \log q)$$


H
BITS
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
p

Conditional entropy and entropy of a source

$x,\ y$ - events

$$H(x) + H(y) \geq H(x, y)$$

Conditional entropy and entropy of a source

$x,\ y$ - events

$$H(x) + H(y) \geq H(x, y) = H(x) + H_x(y)$$

Conditional entropy and entropy of a source

$x,\ y$ - events

$$H(x) + H(y) \geq H(x, y) = H(x) + H_x(y)$$

A source has states with entropies H_i, transition probabilities are $p_i(j)$

Conditional entropy and entropy of a source

$x,\ y$ - events

$$H(x) + H(y) \geq H(x, y) = H(x) + H_x(y)$$

A source has states with entropies H_i, transition probabilities are $p_i(j)$, then

$$H = \sum_i P_i H_i = -\sum_{i,j} P_i\, p_i(j) \log p_i(j)$$

Conditional entropy and entropy of a source

$x,\ y$ - events

$$H(x) + H(y) \geq H(x, y) = H(x) + H_x(y)$$

A source has states with entropies H_i, transition probabilities are $p_i(j)$, then

$$H = \sum_i P_i H_i = -\sum_{i,j} P_i\, p_i(j) \log p_i(j)$$

Different units!

Conditional entropy and entropy of a source

$x,\ y$ - events

$$H(x) + H(y) \geq H(x, y) = H(x) + H_x(y)$$

A source has states with entropies H_i, transition probabilities are $p_i(j)$, then

$$H = \sum_i P_i H_i = -\sum_{i,j} P_i\, p_i(j) \log p_i(j)$$

Different units! Turns out,

$$\lim_{N \to \infty} \frac{\log n(q)}{N} = H$$

$n(q)$ – number of the most probable sequences of length N, total probability q with any $q \neq 0,\ 1$.

Conditional entropy and entropy of a source

$x,\ y$ - events

$$H(x) + H(y) \geq H(x, y) = H(x) + H_x(y)$$

A source has states with entropies H_i, transition probabilities are $p_i(j)$, then

$$H = \sum_i P_i H_i = -\sum_{i,j} P_i\, p_i(j) \log p_i(j)$$

Different units! Turns out,

$$\lim_{N \to \infty} \frac{\log n(q)}{N} = H$$

$n(q)$ – number of the most probable sequences of length N, total probability q with any $q \neq 0,\ 1$.

Entropy of source: bits per symbol

Noiseless case

Noiseless case

Theorem (the fundamental theorem for a noiseless channel)

Let a source have entropy $\boldsymbol{H}$ *bits/symbol and a channel have a capacity* $\boldsymbol{C}$ *bits/second. Then it is possible to encode the output of the source to transmit at the average rate* $\frac{\boldsymbol{C}}{\boldsymbol{H}} - \boldsymbol{\epsilon}$ *symbols/second over the channel where* ϵ *is arbitrarily small.*
It is not possible to transmit at an average rate greater than $\frac{C}{H}$.

Noiseless case

Theorem (the fundamental theorem for a noiseless channel)

Let a source have entropy $\boldsymbol{H}$ *bits/symbol and a channel have a capacity* $\boldsymbol{C}$ *bits/second. Then it is possible to encode the output of the source to transmit at the average rate* $\frac{\boldsymbol{C}}{\boldsymbol{H}} - \boldsymbol{\epsilon}$ *symbols/second over the channel where* ϵ *is arbitrarily small.*
It is not possible to transmit at an average rate greater than $\frac{C}{H}$.

$$\frac{[C]}{[H]} = \frac{\text{bits/s}}{\text{bits/sym}} = \text{sym/s}$$

Noiseless case

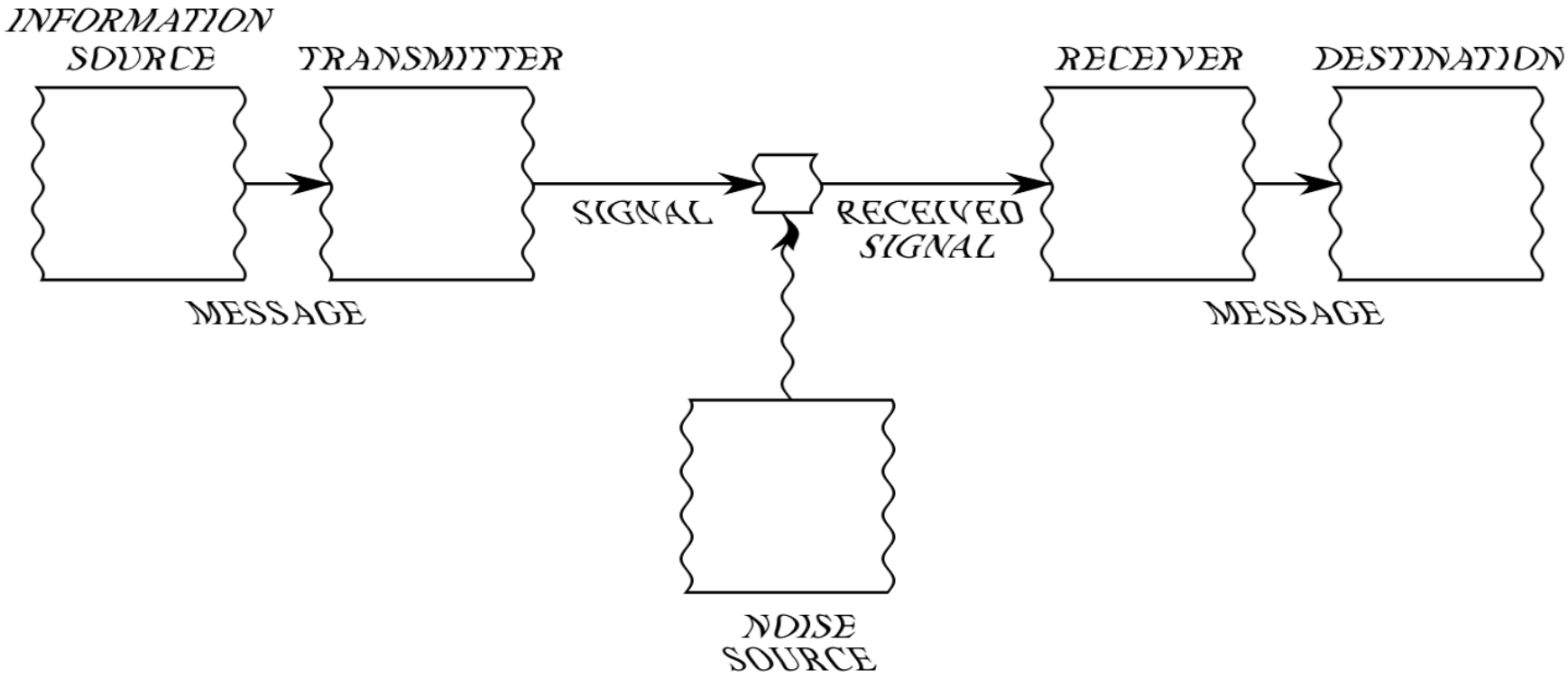
Theorem (the fundamental theorem for a noiseless channel)

Let a source have entropy $\boldsymbol{H}$ *bits/symbol and a channel have a capacity* $\boldsymbol{C}$ *bits/second. Then it is possible to encode the output of the source to transmit at the average rate* $\frac{\boldsymbol{C}}{\boldsymbol{H}} - \boldsymbol{\epsilon}$ *symbols/second over the channel where* ϵ *is arbitrarily small.*
It is not possible to transmit at an average rate greater than $\frac{C}{H}$.

$$\frac{[C]}{[H]} = \frac{\text{bits/s}}{\text{bits/sym}} = \text{sym/s}$$

The proof involves constructing an explicit code that achieves the required rate: Shannon-Fano coding.

Noisy case



Source output: x, decoded output: y. Noise: stochastic process as well.

Equivocation

$H_y(x)$ – equivocation.

Equivocation

$H_y(x)$ – equivocation. Actual transmission rate:

$$R = H(x) - H_y(x).$$

Equivocation

$H_y(x)$ – equivocation. Actual transmission rate:

$$R = H(x) - H_y(x).$$

Capacity of a noisy channel (maximum over all sources):

$$C = \max(H(x) - H_y(x)).$$

Transmitting: 1000 bits/second with probabilities $p_0 = p_1 = \frac{1}{2}$. On average, 1 in 100 is received incorrectly.

Equivocation

$H_y(x)$ – equivocation. Actual transmission rate:

$$R = H(x) - H_y(x).$$

Capacity of a noisy channel (maximum over all sources):

$$C = \max(H(x) - H_y(x)).$$

Transmitting: 1000 bits/second with probabilities $p_0 = p_1 = \frac{1}{2}$. On average, 1 in 100 is received incorrectly.
Saying that rate is 990 $(= 1000 \times 0.99)$ is not reasonable: don't know where the errors occur.

Equivocation

$H_y(x)$ – equivocation. Actual transmission rate:

$$R = H(x) - H_y(x).$$

Capacity of a noisy channel (maximum over all sources):

$$C = \max(H(x) - H_y(x)).$$

Transmitting: 1000 bits/second with probabilities $p_0 = p_1 = \frac{1}{2}$. On average, 1 in 100 is received incorrectly.
Saying that rate is 990 $(= 1000 \times 0.99)$ is not reasonable: don't know where the errors occur.
With the above definition of $H_y(x)$

Equivocation

$H_y(x)$ – equivocation. Actual transmission rate:

$$R = H(x) - H_y(x).$$

Capacity of a noisy channel (maximum over all sources):

$$C = \max(H(x) - H_y(x)).$$

Transmitting: 1000 bits/second with probabilities $p_0 = p_1 = \frac{1}{2}$. On average, 1 in 100 is received incorrectly.
Saying that rate is 990 $(= 1000 \times 0.99)$ is not reasonable: don't know where the errors occur.
With the above definition of $H_y(x)$ (if $y = 1$ is received, probability that $x = 1$ was sent is 0.99, etc):

$$H_y(x) = -(0.99 \log 0.99 + 0.01 \log 0.01) = 0.081 \text{ bits/symbol.}$$

Equivocation

$H_y(x)$ – equivocation. Actual transmission rate:

$$R = H(x) - H_y(x).$$

Capacity of a noisy channel (maximum over all sources):

$$C = \max(H(x) - H_y(x)).$$

Transmitting: 1000 bits/second with probabilities $p_0 = p_1 = \frac{1}{2}$. On average, 1 in 100 is received incorrectly.
Saying that rate is 990 $(= 1000 \times 0.99)$ is not reasonable: don't know where the errors occur.
With the above definition of $H_y(x)$ (if $y = 1$ is received, probability that $x = 1$ was sent is 0.99, etc):

$$H_y(x) = -(0.99 \log 0.99 + 0.01 \log 0.01) = 0.081 \text{ bits/symbol.}$$

Thus the actual transmission rate is

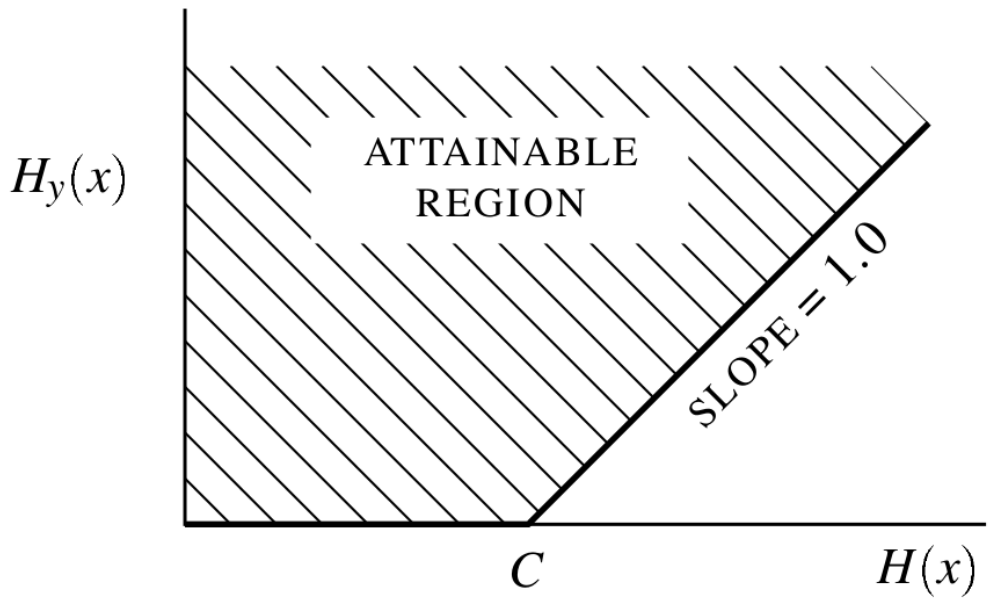
$$R = 1000 - 81 = 919 \text{ bits/second}$$

Theorem (the fundamental theorem for a discrete channel with noise)

Let a discrete channel have the capacity $\boldsymbol{C}$ *and a discrete source the entropy per second* $\boldsymbol{H}$*. If* $\boldsymbol{H} \leq \boldsymbol{C}$*, there* **exists** *a coding system with an arbitrarily small frequency of errors (or an arbitrarily small equivocation* $H_y(x)$*) during transmission. If* $\boldsymbol{H} > \boldsymbol{C}$ *it is possible to encode the source so that the equivocation is less than* $\boldsymbol{H} - \boldsymbol{C} + \boldsymbol{\epsilon}$ *where* ϵ *is arbitrarily small. There is no method of encoding which gives an equivocation less than* $H - C$*.*

Theorem (the fundamental theorem for a discrete channel with noise)

Let a discrete channel have the capacity $\boldsymbol{C}$ *and a discrete source the entropy per second* $\boldsymbol{H}$*. If* $\boldsymbol{H} \leq \boldsymbol{C}$*, there* **exists** *a coding system with an arbitrarily small frequency of errors (or an arbitrarily small equivocation* $H_y(x)$*) during transmission. If* $\boldsymbol{H} > \boldsymbol{C}$ *it is possible to encode the source so that the equivocation is less than* $\boldsymbol{H} - \boldsymbol{C} + \boldsymbol{\epsilon}$ *where* ϵ *is arbitrarily small. There is no method of encoding which gives an equivocation less than* $H - C$*.*



Shannon-Fano coding

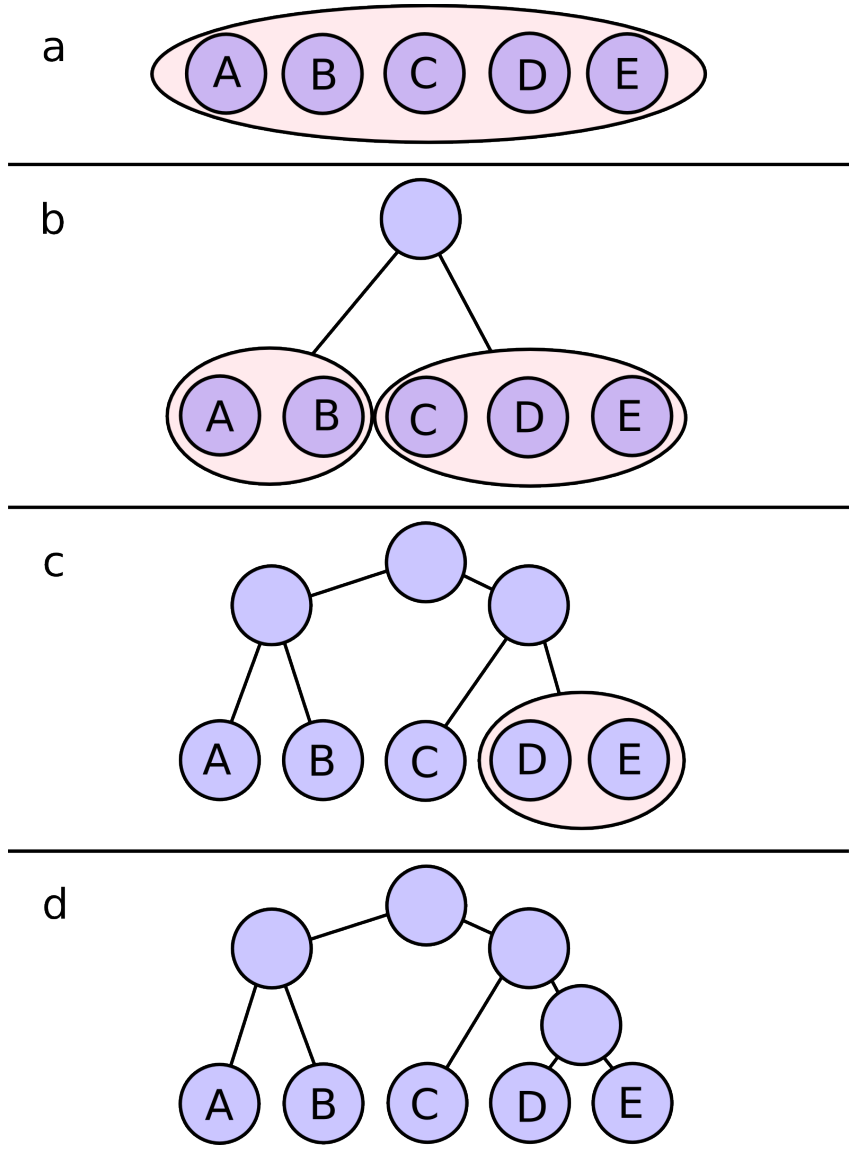
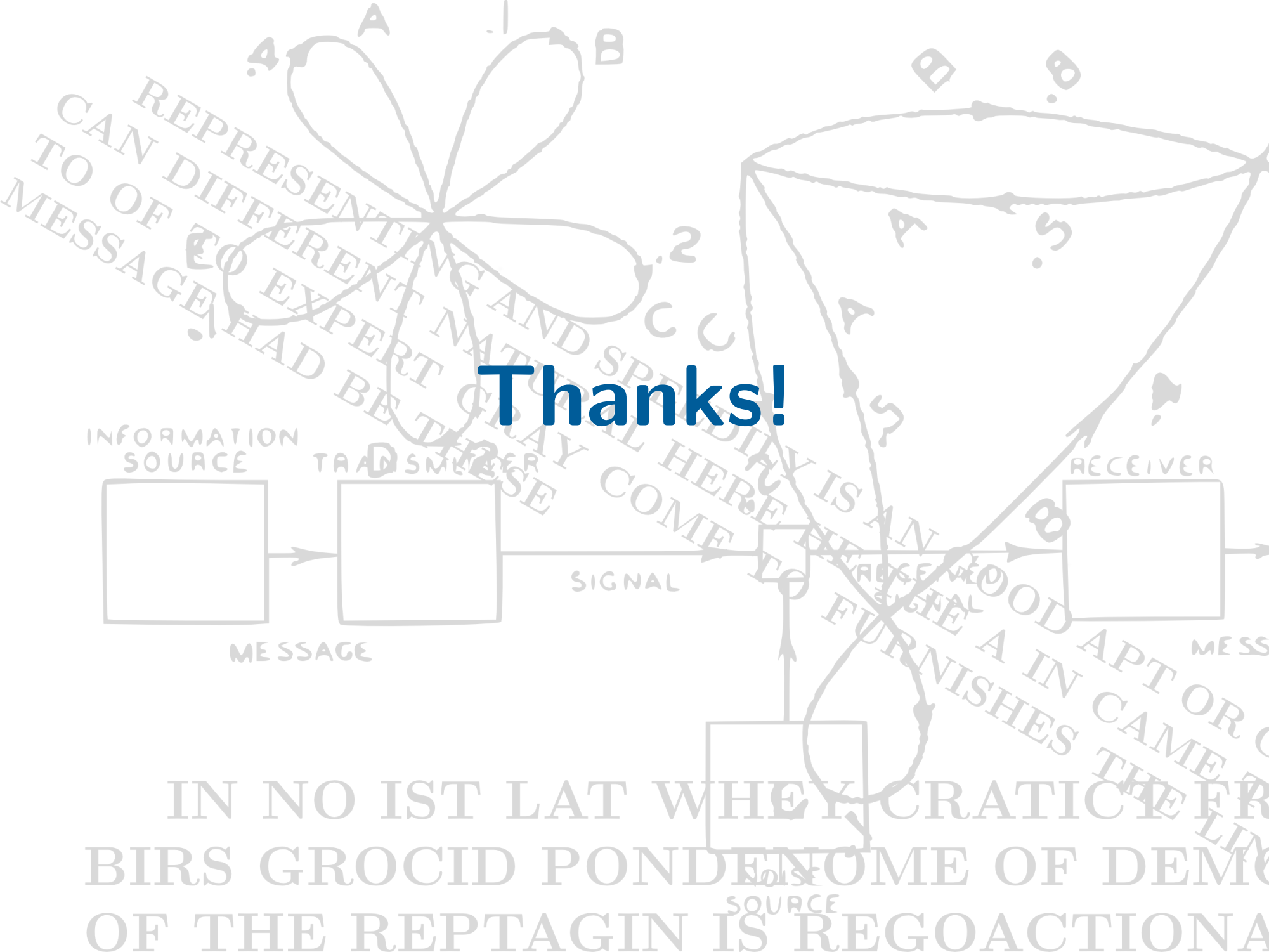


Image: Wikimedia

Thanks!


INFORMATION SOURCE
TRANSMITTER
SIGNAL
RECEIVER
MESSAGE
NOISE SOURCE