Information Theory

EE 276

INSTRUCTOR

Tsachy Weissman

TA

Basil N. Saeed

TA

Noah A. Huffman

goal

- expose the beauty and utility of the science of information (and, specifically, information theory)
- whet your appetite for subsequent learning
- information scientific thinking (seeing the world through the lens of information)

what is information?

information

noun | in·for·ma·tion | \ˌin-fər-ˈmā-shən\

Popularity: Top 1% of lookups | Updated on: 3 Sep 2018

TRENDING NOW: hirsute op-ed collegiality mistrial hogwash SEE ALL ›

Tip: Synonym Guide | Examples: INFORMATION in a Sentence

Definition of INFORMATION

1 : the communication or reception of knowledge or intelligence

2 a (1) : knowledge obtained from investigation, study, or instruction (2) : INTELLIGENCE, NEWS (3) : FACTS, DATA
b : the attribute inherent in and communicated by one of two or more alternative sequences or arrangements of something (such as nucleotides in DNA or binary digits in a computer program) that produce specific effects
c (1) : a signal or character (as in a communication system or computer) representing data (2) : something (such as a message, experimental data, or a picture) which justifies change in a construct (such as a plan or theory) that represents physical or mental experience or another construct
d : a quantitative measure of the content of information; *specifically* : a numerical quantity that measures the uncertainty in the outcome of an experiment to be performed

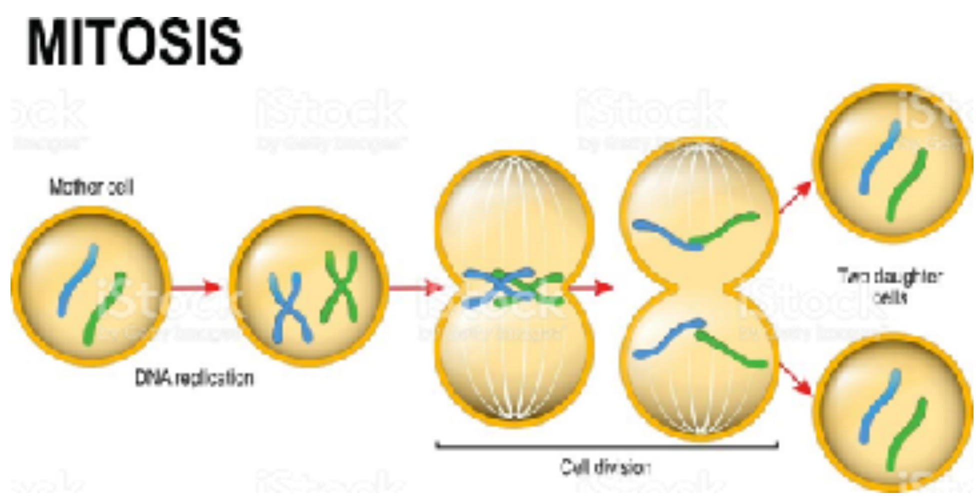
3 : the act of informing against a person

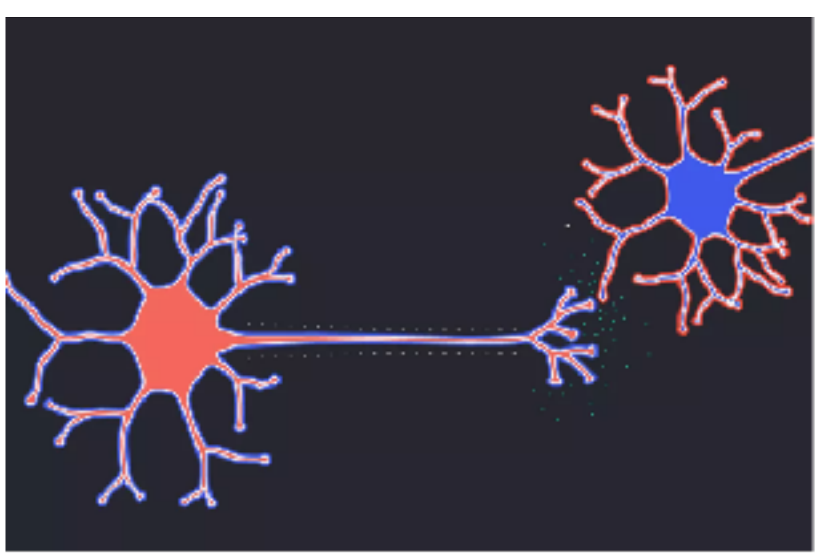
4 : a formal accusation of a crime made by a prosecuting officer as distinguished from an indictment presented by a grand jury

—**informational** \ˌin-fər-ˈmā-shnəl, -shə-nᵊl\ *adjective*

—**informationally** *adverb*

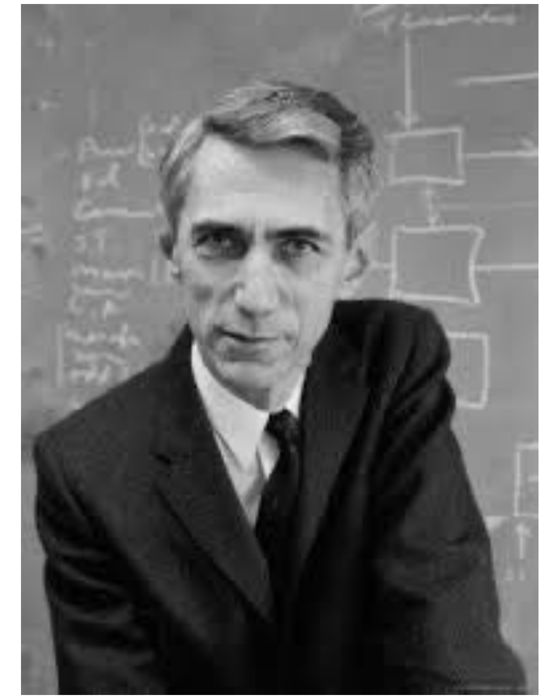
what is communication?

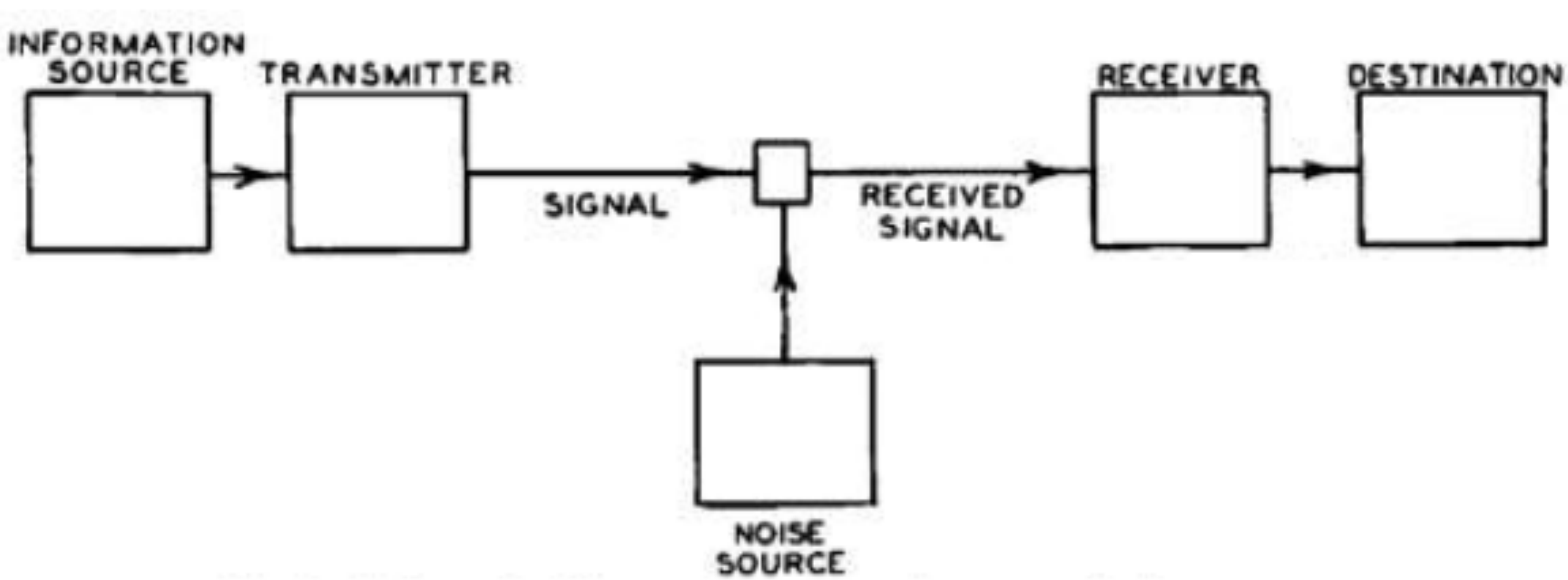


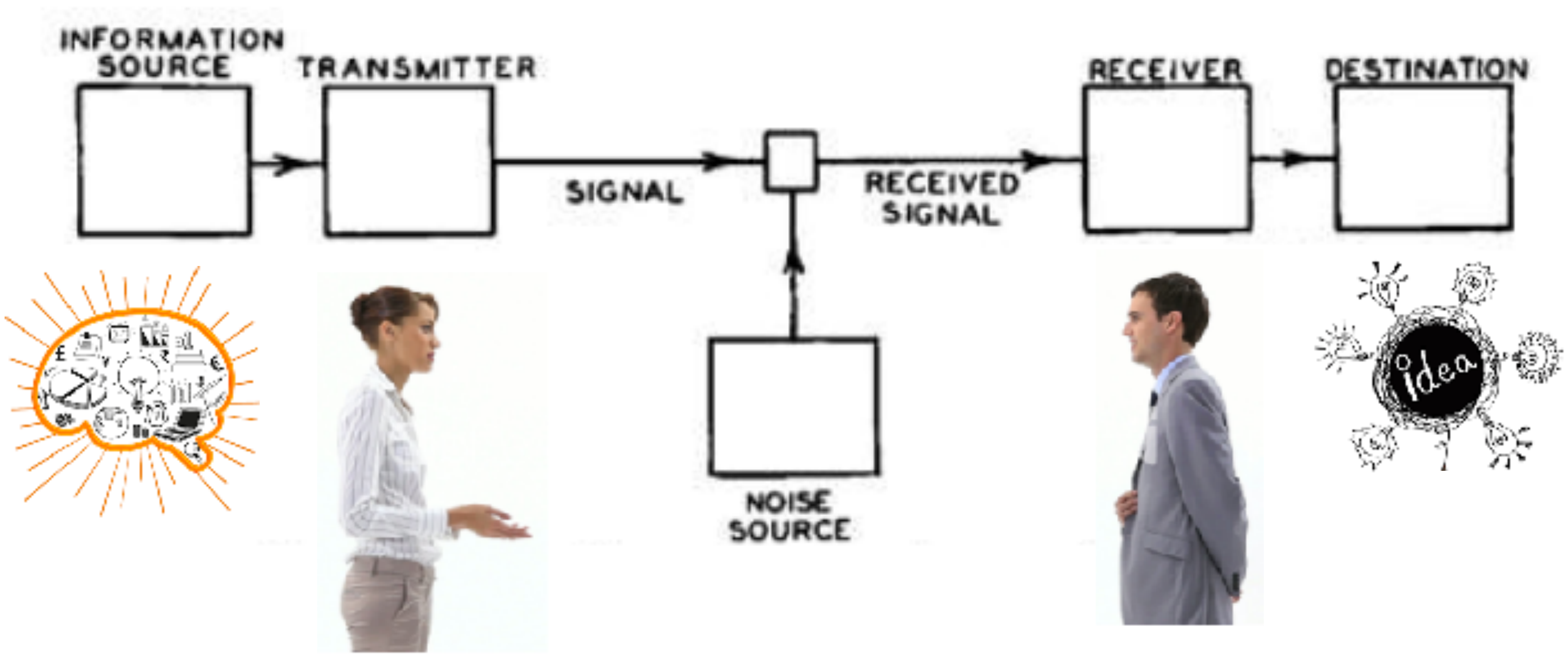


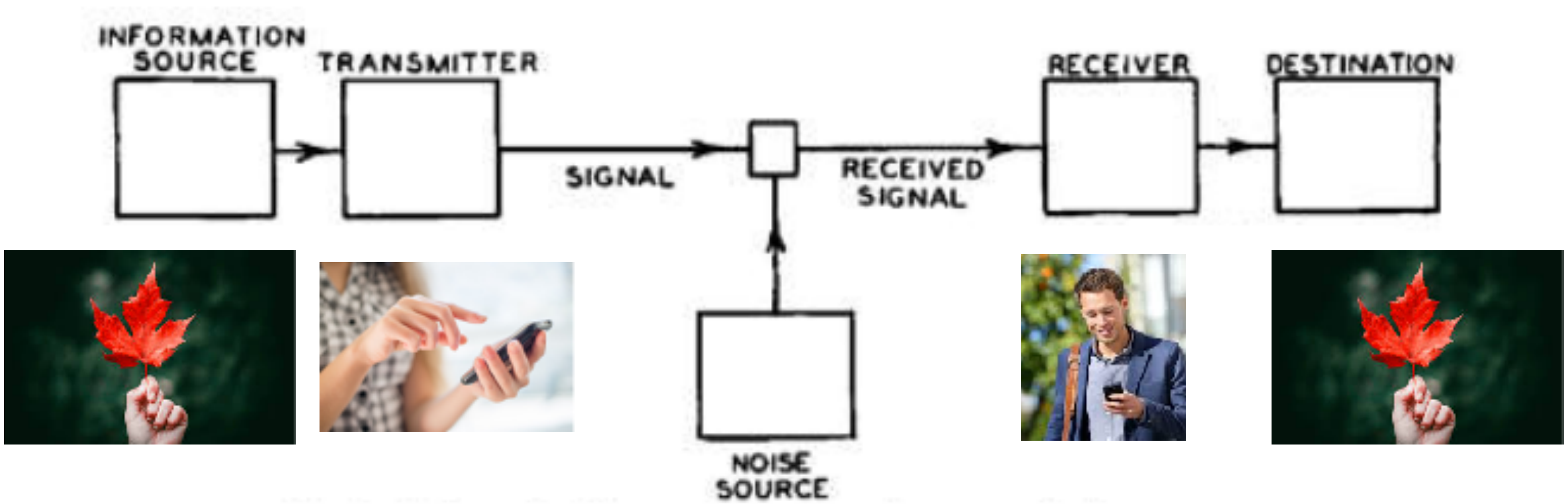
Claude Elwood Shannon
1948

"A Mathematical Theory of Communication"






INFORMATION SOURCE
TRANSMITTER
SIGNAL
RECEIVED SIGNAL
RECEIVER
DESTINATION
NOISE SOURCE
idea

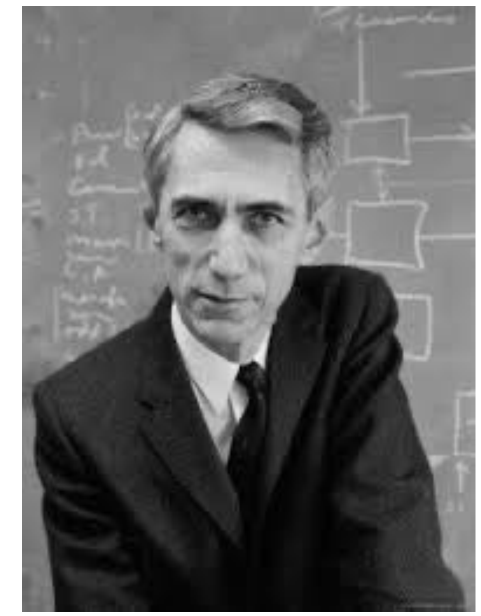

INFORMATION SOURCE
TRANSMITTER
SIGNAL
RECEIVED SIGNAL
RECEIVER
DESTINATION
NOISE SOURCE

Shannon’s genius

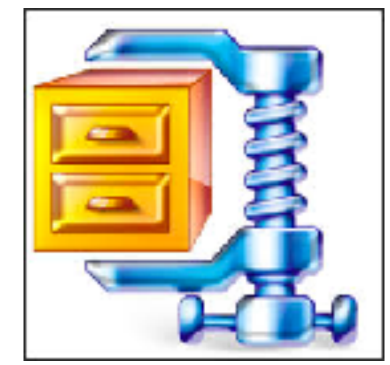
- the question
- the answer

a bit about the bit

0 **or** **1**

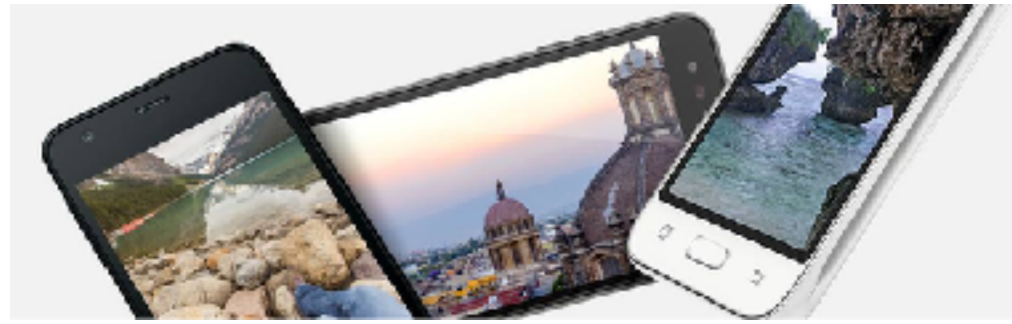


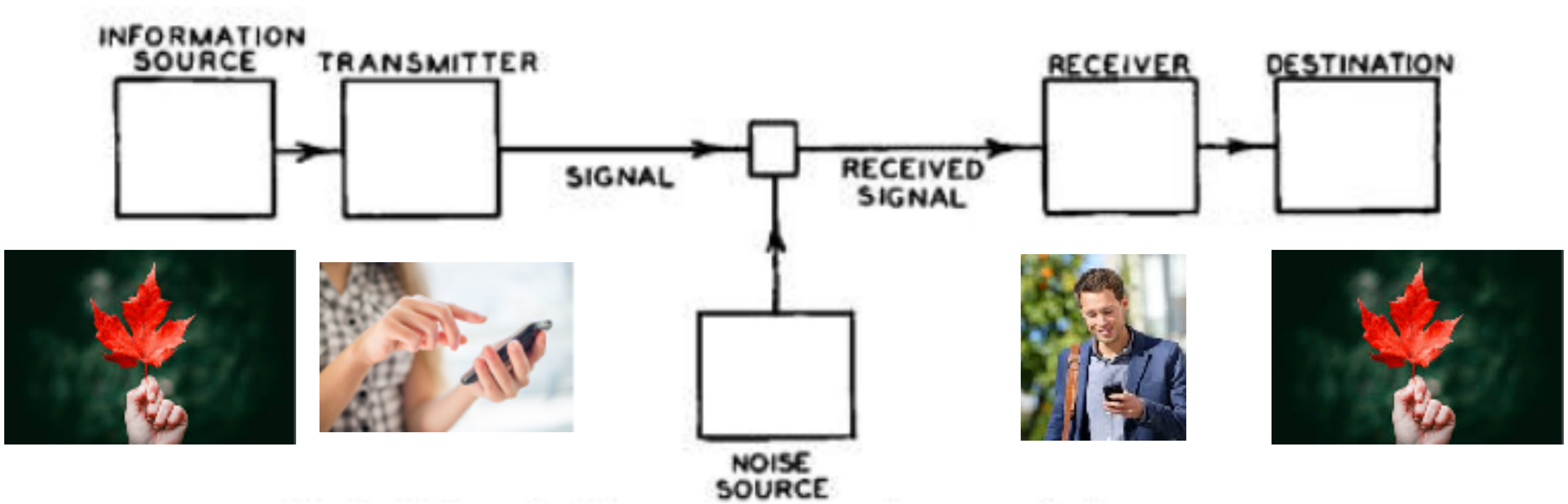
2 pillars of the science of information



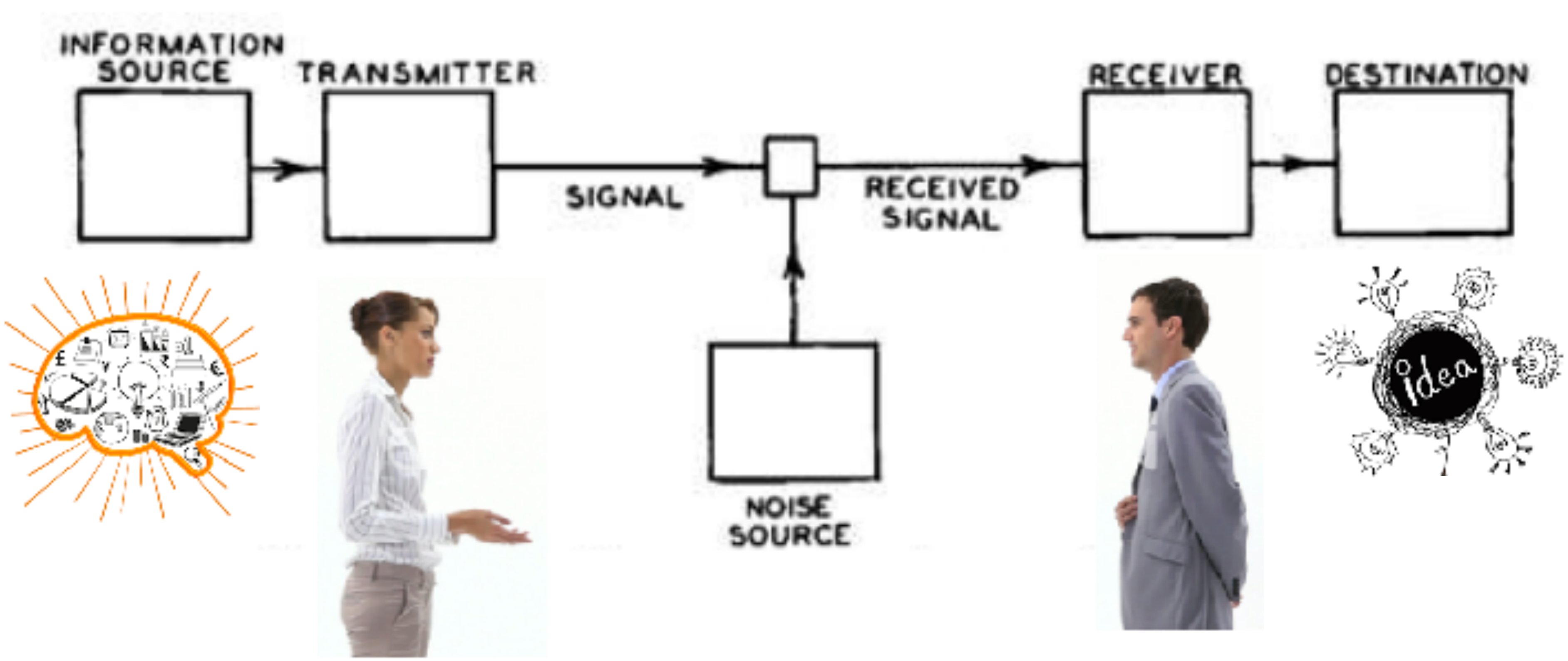
- succinct representation of the information source in bits (compression)
- effective and reliable communication of bits (across unreliable media)

Shannon discovered the two,
showed reliable communication of bits is generally possible,
and that combining the two is optimal





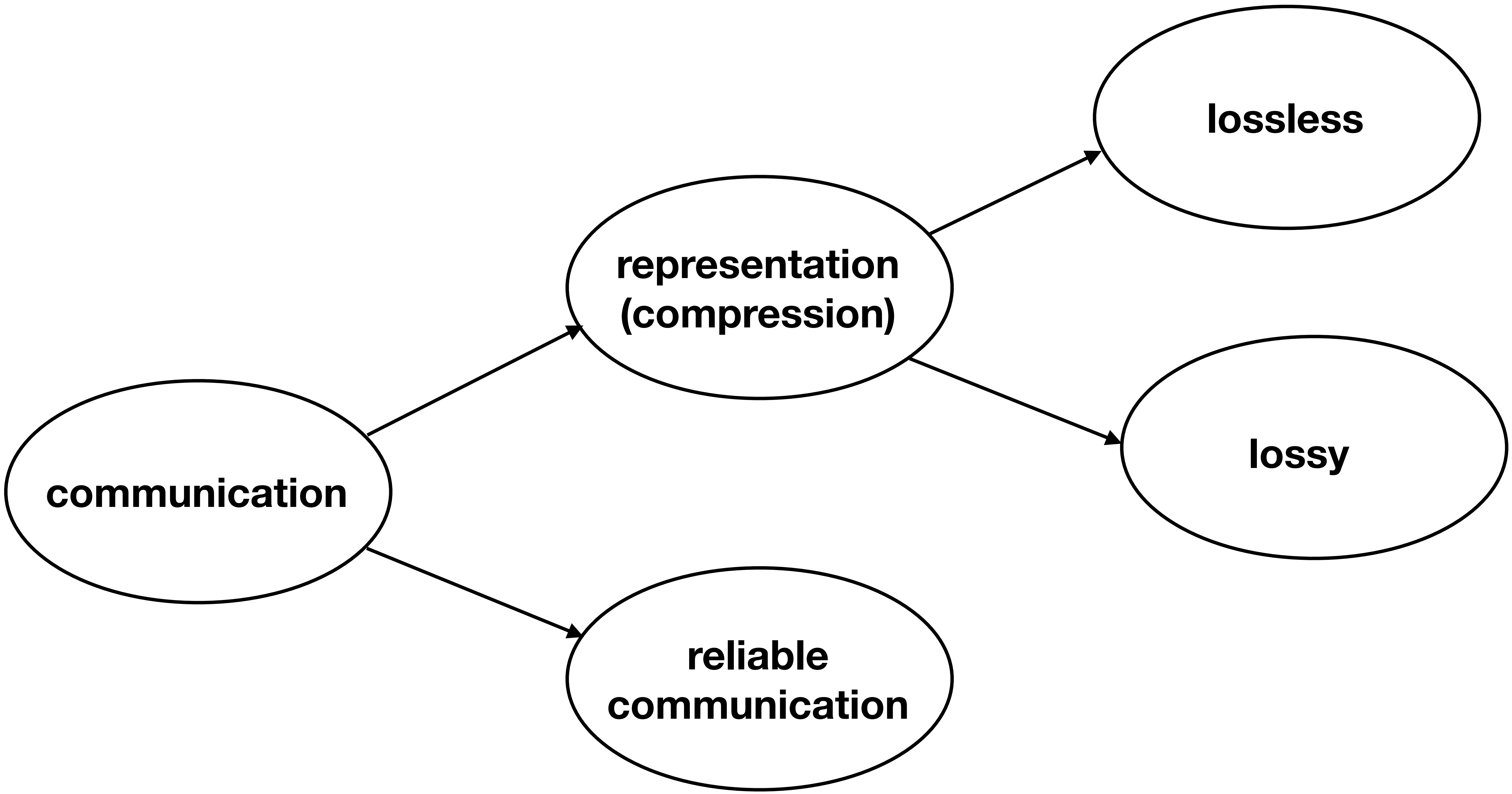
INFORMATION SOURCE
TRANSMITTER
SIGNAL
RECEIVED SIGNAL
RECEIVER
DESTINATION
NOISE SOURCE


INFORMATION SOURCE
TRANSMITTER
SIGNAL
RECEIVED SIGNAL
RECEIVER
DESTINATION
NOISE SOURCE
idea

and everything else

- neurons
- genetics/genomics
- language
- matter
- etc.

course theme I: communication



course theme II: concrete schemes

- Shannon
- Huffman
- Arithmetic
- Lempel-Ziv (GZIP)
- JPEG
- Polar codes for reliable communication (5G)

course theme III: measures of information

- entropy
- relative entropy
- mutual information
- Shannon capacity
- rate-distortion function

course theme IV: a bit on relations to and manifestations in other areas

- genomics
- machine learning, data science, statistics
- neuroscience, human-inspired compression

approximate lecture schedule

- Introduction and motivating examples
- Information measures: entropy, relative entropy and MI
- AEP and typicality
- Variable length lossless compression: prefix and Shannon codes
- Kraft inequality and Huffman coding
- Lempel Ziv compression
- Reliable communication and channel capacity
- Information measures for continuous random variables
- AWGN channel
- Joint AEP and Channel coding theorem
- Channel coding theorem converse
- Polar codes
- Lossy compression and rate distortion
- Method of types and Sanov's theorem
- Strong, conditional and joint typicality
- Direct and converse in rate distortion theorem
- Joint source-channel coding and the separation theorem
- Distributed compression and Slepian-Wolf coding
- Compression and learning, directed information and its estimation

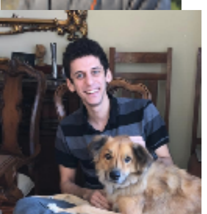
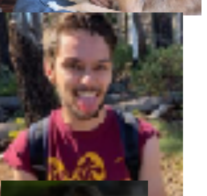
course elements

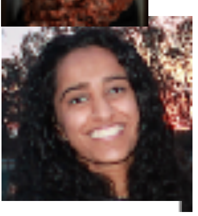
- lectures (Tue, Thu, 1:30-2:50pm, Building 320, room 105)
- HW (5pm Fridays, submitted on Gradescope, accessed via course website)
- recitations (Tuesdays 4pm, 1-1.5 hours)
- midterm (Friday February 16th, 5-7pm, reach out if you have conflicts)
- final (Tuesday, March 19th, 3:30-6:30pm)

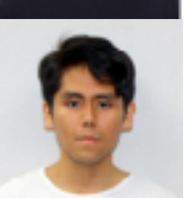
re the lectures and material

- prereq: probability, conditional probability, expectation, etc.
- you'll be held 'accountable' only to material covered in in-person lectures and HWs
- course website rich with additional resources, including course notes and videos of additional lectures from previous years, lectures and material from EE274, books, etc.
- parts of these will be referred to for further reading/viewing

staff





- Instructor: Tsachy Weissman, OH Thursdays 3-4pm or by appointment

- TA: Basil N. Saeed

- TA: Noah A. Huffman

- CSs (course supporters): Lara Arikan, Divija Hasteer, Jiwon Jeong, Cesar Lema (office hours will start next week)

- more details including emails, office hours, etc. on the course website (main resource): https://web.stanford.edu/class/ee276/

- Gradescope and Piazza for the course accessible via website

questions?

example I: lossless compression of a ternary source

$$\text{Source is } U_1, U_2, \cdots \overset{\text{i.i.d}}{\sim} U \in \mathcal{U} = \{A, B, C\}$$

$$P(U = A) = 0.7, \qquad P(U = B) = 0.15,\ P(U = C) = 0.15$$

how can/should we represent the source succinctly with bits?

first code suggestion:

$$A \to \text{`0'}$$
$$B \to \text{`10'}$$
$$C \to \text{`11'}$$

Let $\bar{L}$ denote the average number of bits per symbol. For the coding above,

$$\bar{L} = 0.7 \times 1 + 0.15 \times 2 + 0.15 \times 2 = 1.3 \text{ bits/symbol}$$

note how easily we can decode, e.g.:

$$0011010011101011$$

(thanks to the “prefix condition” satisfied by this code)

second code suggestion:

pair	probability	Code word	Num. Bits Used
AA	0.49	0	1
AB	0.105	100	3
AC	0.105	111	3
BA	0.105	101	3
CA	0.105	1100	4
BB	0.0225	110100	6
BC	0.0225	110101	6
CB	0.0225	110110	6
CC	0.0225	110111	6

$$\bar{L} = \frac{1}{2}(0.49 \times 1 + 0.105 \times 3 \times 3 + 0.105 \times 4 + 0.0225 \times 6 \times 4)$$
$$= 1.1975 \text{ bits/symbol}$$

we'll see:

source "entropy":

$$H(U) = \sum_{u \in \mathcal{U}} p(u) \log_2 \frac{1}{p(u)} \simeq 1.1829$$

"converse" result: for any compressor $H(U) \leq \bar{L}$

"direct" result: for any eps>0 there exists a compressor satisfying

$$\bar{L} \leq H(U) + \epsilon$$

example ii: binary source and channel

Source: $\mathbb{U} = \{U_1, U_2, ...\}$ where $Pr[U_i = 0] = Pr[U_i = 1] = \frac{1}{2}$. The U_i's are i.i.d.

Channel: The channel flips each bit given to it with probability $q < \frac{1}{2}$. We define the channel input to be $\mathbb{X} = \{X_i\}$, the channel noise to be $\mathbb{W} = \{W_i\}$ and the channel output to be $\mathbb{Y} = \{Y_i\}$ such that:

$$W_i \sim Ber(q)$$
$$Y_i = X_i \oplus_2 W_i$$

The W_i are i.i.d. and the X_i are functions of the input source sequence $\mathbb{U}$.

Probability of error per source bit: P_e

encoding scheme 1:

trivial encoding: $$X_i = U_i$$

yields: $$P_e = q$$

the ***rate*** of this scheme is 1 information bits/channel use

Encoding Scheme 2: We can repeat each source bit three times:

$$\mathbb{U} = 0\ 1\ 1\ 0 \ldots$$

$$\mathbb{X} = 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 0\ 0\ 0 \ldots$$

$$P_e = 3q^2(1-q) + q^3 < q$$

the rate of this scheme is 1/3 information bits/channel use can repeat K times (repetition coding)

as K grows we'll get:

arbitrarily small Pe

at the cost of vanishing rate

Shannon 1948: $\exists\, R > 0$ and schemes with rate $\geq R$ satisfying $P_e \to 0$

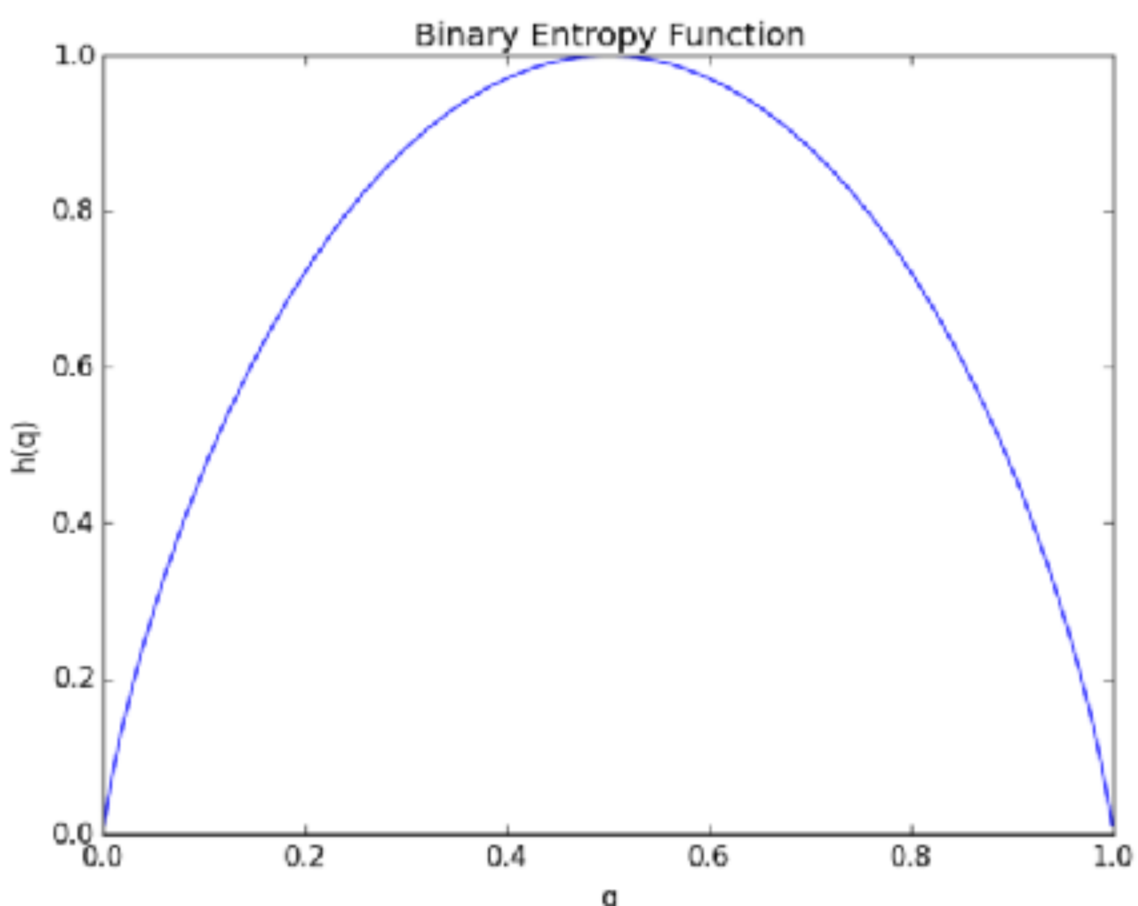
C = "Channel Capacity" = largest such R

in our example:

$$C(q) = 1 - h(q)$$

$$h(q) \triangleq H(Ber(q)) = q \log \frac{1}{q} + (1-q) \log \frac{1}{1-q}$$

The figure below plots $h(q)$ for $q \in [0, 1]$.

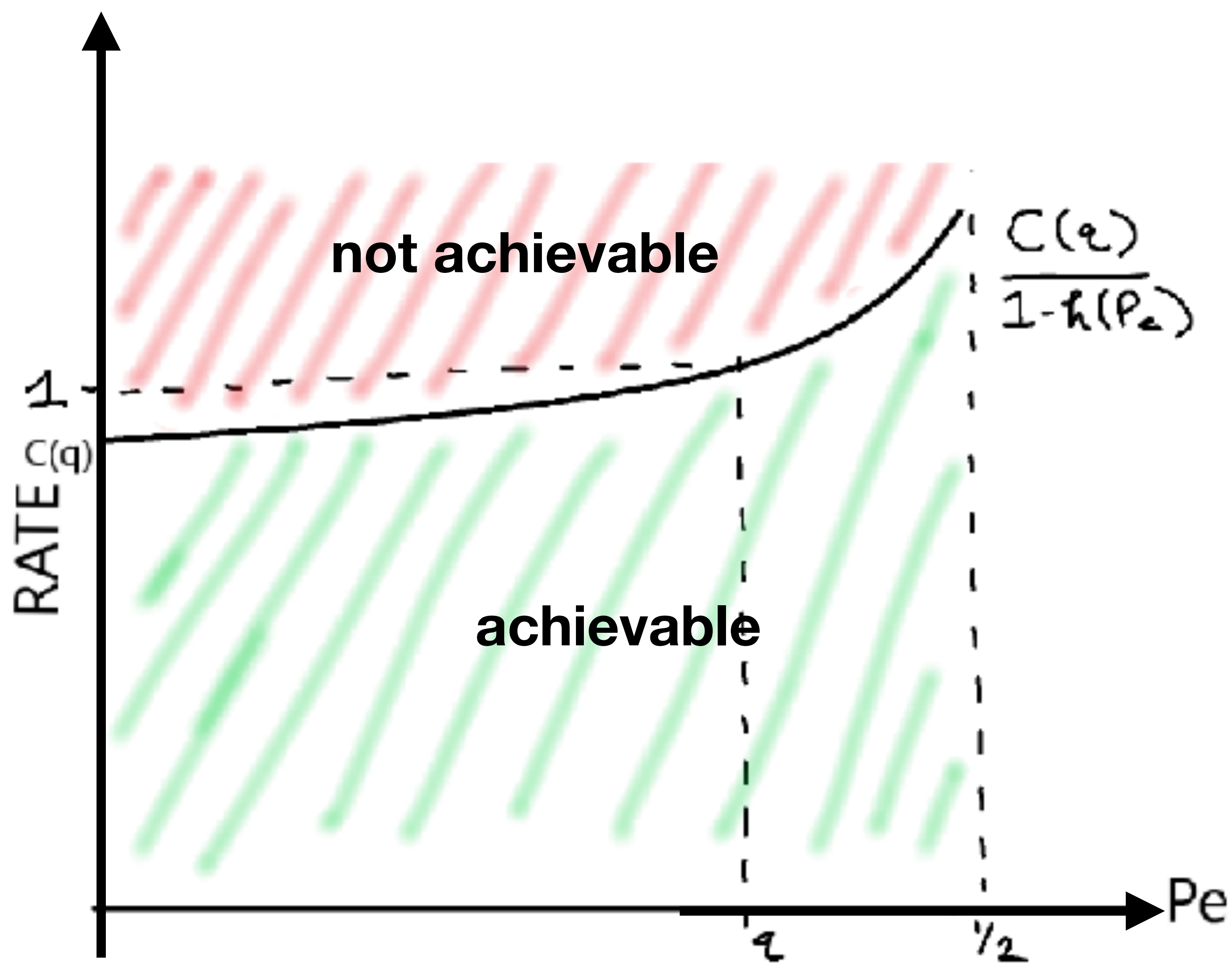


Here too we'll see:

a "converse" part:
no scheme can communicate reliably at a rate above C(q)

a "direct" part:
for any rate below C(q), there exist schemes that can communicate reliably at that rate

we'll also see that, if you're ok with Pe>0:



compression of a Gaussian source

communication via an additive white Gaussian noise channel