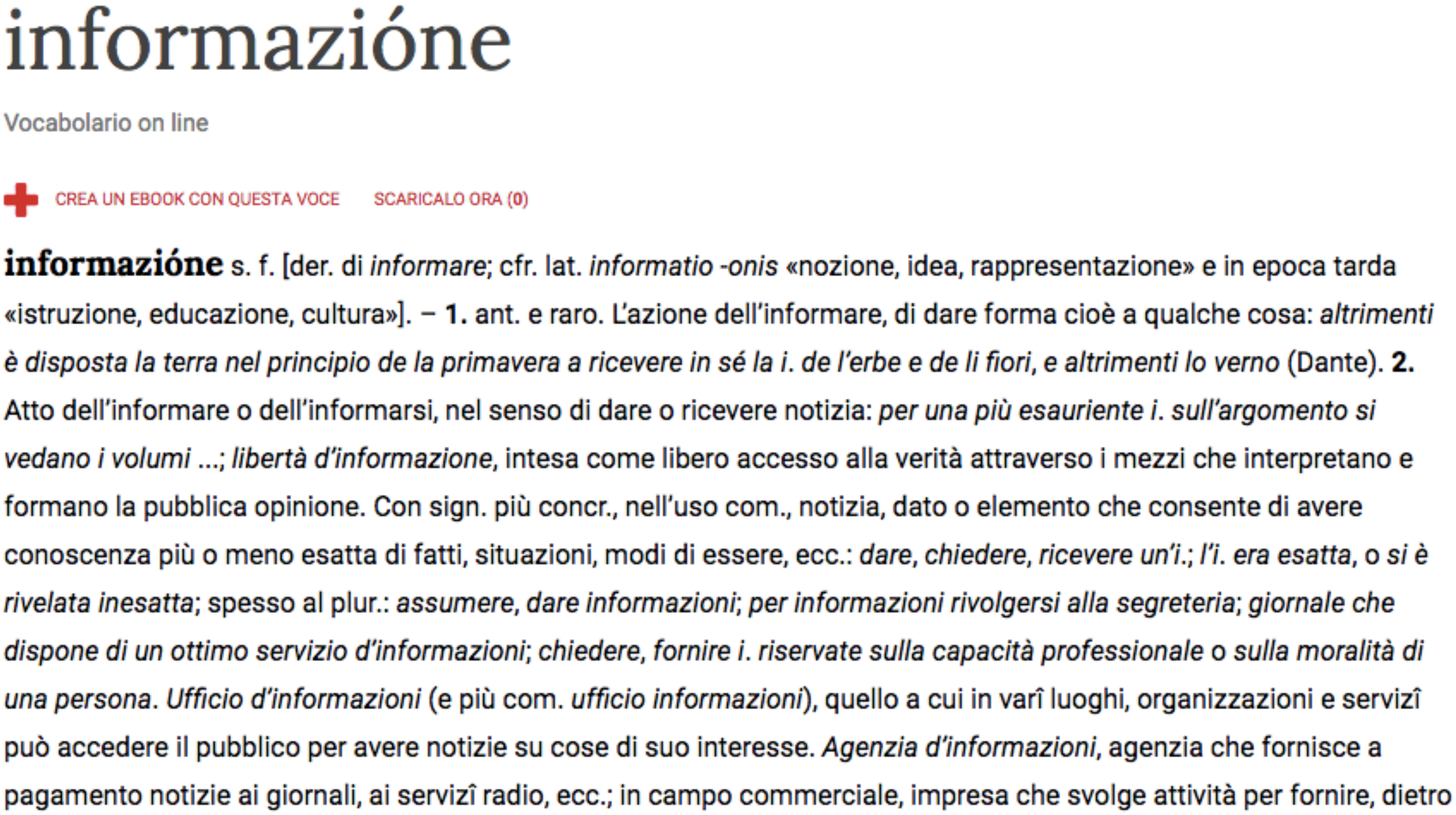
Information theory I

Fisica dell’Energia - a.a. 2020/2021

What is Information

What is Information



informazióne

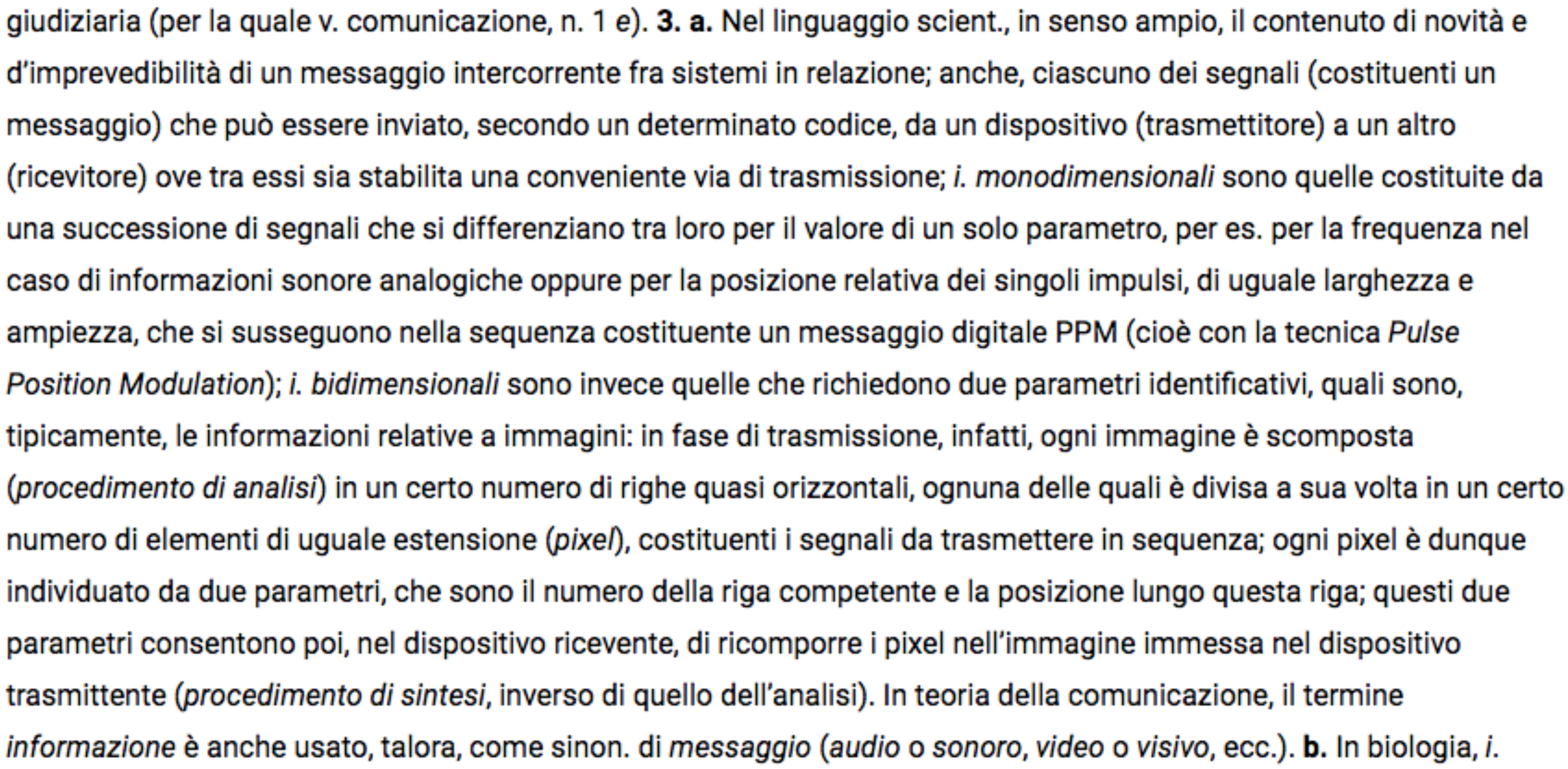
Vocabolario on line

+ CREA UN EBOOK CON QUESTA VOCE   SCARICALO ORA (**0**)

informazióne s. f. [der. di *informare*; cfr. lat. *informatio -onis* «nozione, idea, rappresentazione» e in epoca tarda «istruzione, educazione, cultura»]. – **1.** ant. e raro. L’azione dell’informare, di dare forma cioè a qualche cosa: *altrimenti è disposta la terra nel principio de la primavera a ricevere in sé la i. de l’erbe e de li fiori, e altrimenti lo verno* (Dante). **2.** Atto dell’informare o dell’informarsi, nel senso di dare o ricevere notizia: *per una più esauriente i. sull’argomento si vedano i volumi ...*; *libertà d’informazione*, intesa come libero accesso alla verità attraverso i mezzi che interpretano e formano la pubblica opinione. Con sign. più concr., nell’uso com., notizia, dato o elemento che consente di avere conoscenza più o meno esatta di fatti, situazioni, modi di essere, ecc.: *dare, chiedere, ricevere un’i.*; *l’i. era esatta*, o *si è rivelata inesatta*; spesso al plur.: *assumere, dare informazioni*; *per informazioni rivolgersi alla segreteria*; *giornale che dispone di un ottimo servizio d’informazioni*; *chiedere, fornire i. riservate sulla capacità professionale* o *sulla moralità di una persona*. *Ufficio d’informazioni* (e più com. *ufficio informazioni*), quello a cui in varî luoghi, organizzazioni e servizî può accedere il pubblico per avere notizie su cose di suo interesse. *Agenzia d’informazioni*, agenzia che fornisce a pagamento notizie ai giornali, ai servizî radio, ecc.; in campo commerciale, impresa che svolge attività per fornire, dietro

http://www.treccani.it/vocabolario/informazione/

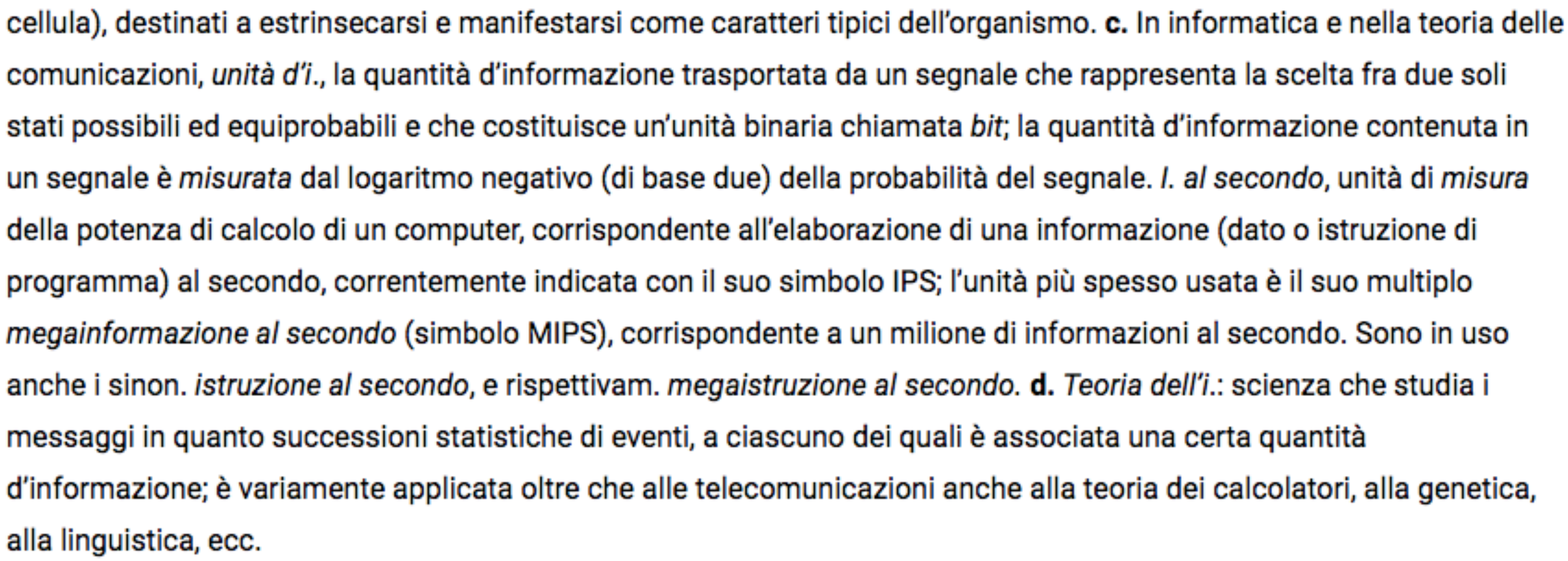
What is Information



giudiziaria (per la quale v. comunicazione, n. 1 *e*). **3. a.** Nel linguaggio scient., in senso ampio, il contenuto di novità e d'imprevedibilità di un messaggio intercorrente fra sistemi in relazione; anche, ciascuno dei segnali (costituenti un messaggio) che può essere inviato, secondo un determinato codice, da un dispositivo (trasmettitore) a un altro (ricevitore) ove tra essi sia stabilita una conveniente via di trasmissione; *i. monodimensionali* sono quelle costituite da una successione di segnali che si differenziano tra loro per il valore di un solo parametro, per es. per la frequenza nel caso di informazioni sonore analogiche oppure per la posizione relativa dei singoli impulsi, di uguale larghezza e ampiezza, che si susseguono nella sequenza costituente un messaggio digitale PPM (cioè con la tecnica *Pulse Position Modulation*); *i. bidimensionali* sono invece quelle che richiedono due parametri identificativi, quali sono, tipicamente, le informazioni relative a immagini: in fase di trasmissione, infatti, ogni immagine è scomposta (*procedimento di analisi*) in un certo numero di righe quasi orizzontali, ognuna delle quali è divisa a sua volta in un certo numero di elementi di uguale estensione (*pixel*), costituenti i segnali da trasmettere in sequenza; ogni pixel è dunque individuato da due parametri, che sono il numero della riga competente e la posizione lungo questa riga; questi due parametri consentono poi, nel dispositivo ricevente, di ricomporre i pixel nell'immagine immessa nel dispositivo trasmittente (*procedimento di sintesi*, inverso di quello dell'analisi). In teoria della comunicazione, il termine *informazione* è anche usato, talora, come sinon. di *messaggio* (*audio* o *sonoro*, *video* o *visivo*, ecc.). **b.** In biologia, *i.*

http://www.treccani.it/vocabolario/informazione/

What is Information



cellula), destinati a estrinsecarsi e manifestarsi come caratteri tipici dell'organismo. **c.** In informatica e nella teoria delle comunicazioni, *unità d'i.*, la quantità d'informazione trasportata da un segnale che rappresenta la scelta fra due soli stati possibili ed equiprobabili e che costituisce un'unità binaria chiamata *bit*; la quantità d'informazione contenuta in un segnale è *misurata* dal logaritmo negativo (di base due) della probabilità del segnale. *I. al secondo*, unità di *misura* della potenza di calcolo di un computer, corrispondente all'elaborazione di una informazione (dato o istruzione di programma) al secondo, correntemente indicata con il suo simbolo IPS; l'unità più spesso usata è il suo multiplo *megainformazione al secondo* (simbolo MIPS), corrispondente a un milione di informazioni al secondo. Sono in uso anche i sinon. *istruzione al secondo*, e rispettivam. *megaistruzione al secondo.* **d.** *Teoria dell'i.*: scienza che studia i messaggi in quanto successioni statistiche di eventi, a ciascuno dei quali è associata una certa quantità d'informazione; è variamente applicata oltre che alle telecomunicazioni anche alla teoria dei calcolatori, alla genetica, alla linguistica, ecc.

http://www.treccani.it/vocabolario/informazione/

Measuring information

6 flips of a coin

6 flips of a coin

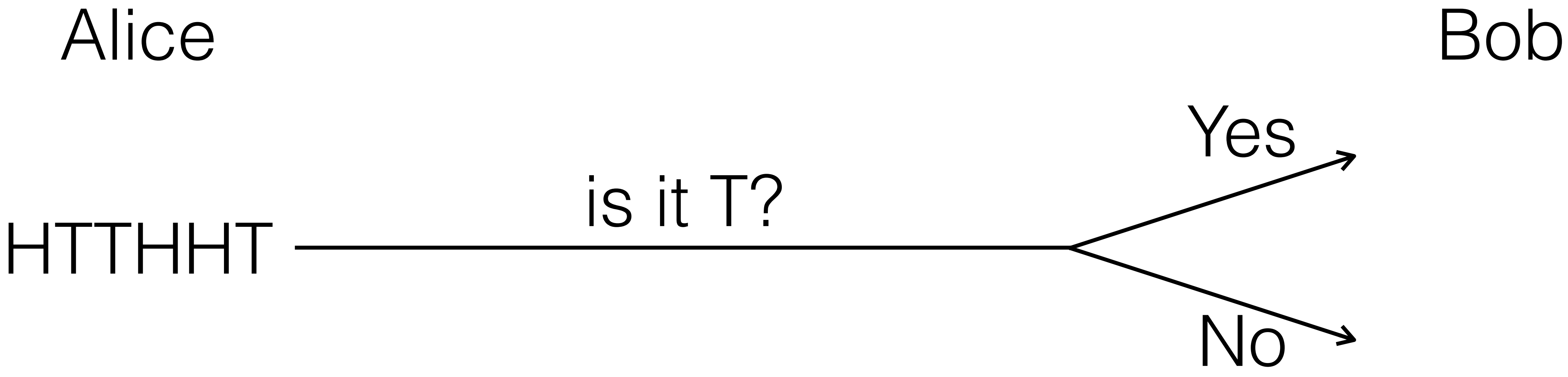
Alice

Bob

as a string

HTTHHT → HTTHHT

Reduction to YES or NO answers



Alice

Bob

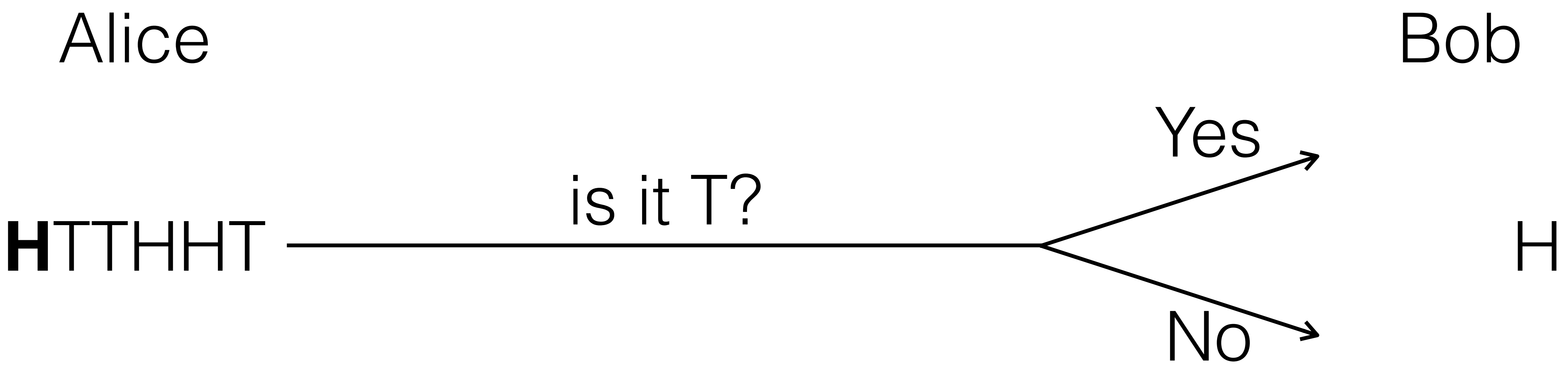
HTTHHT

is it T?

Yes

No

Reduction to YES or NO answers



Alice

Bob

HTTHHT

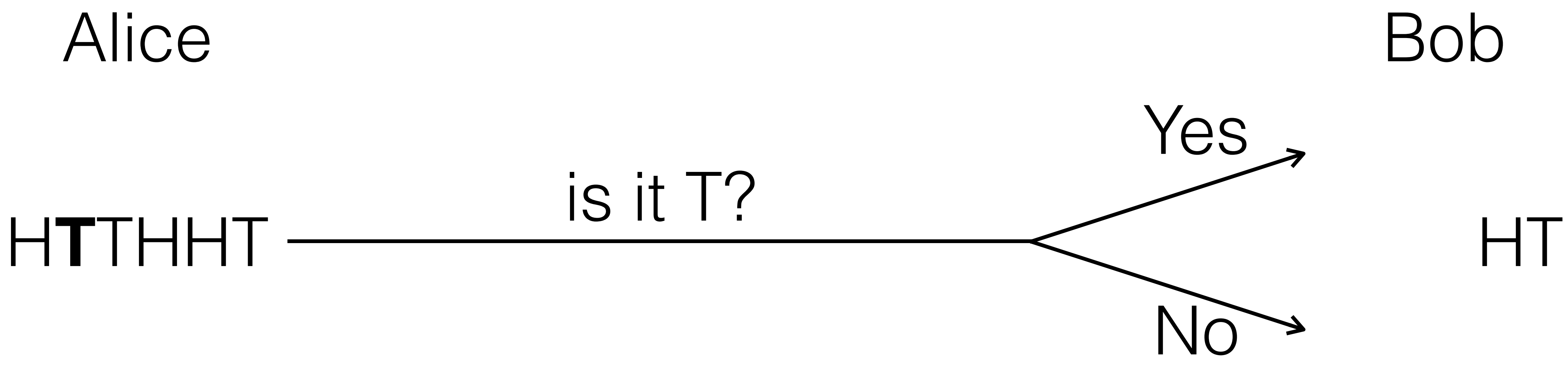
is it T?

Yes

No

H

Reduction to YES or NO answers



Alice

Bob

H**T**THHT

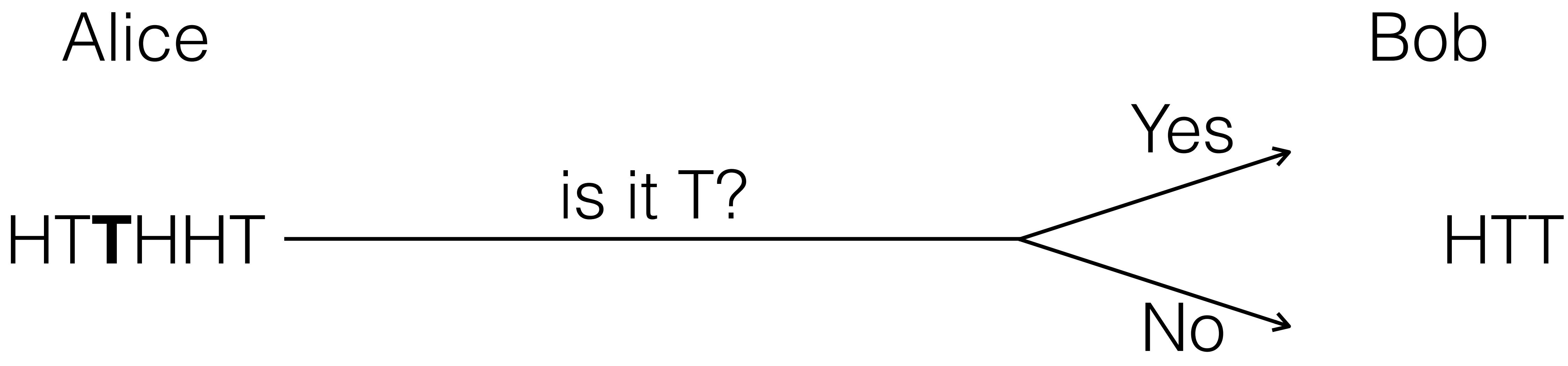
is it T?

Yes

No

HT

Reduction to YES or NO answers



Alice

Bob

Yes

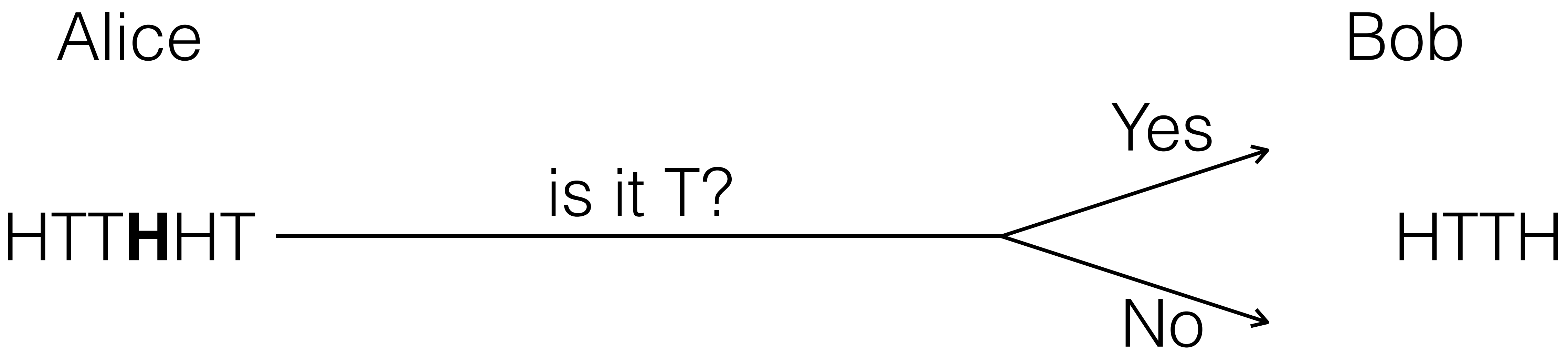
is it T?

HT**T**HHT

HTT

No

Reduction to YES or NO answers



Alice

Bob

Yes

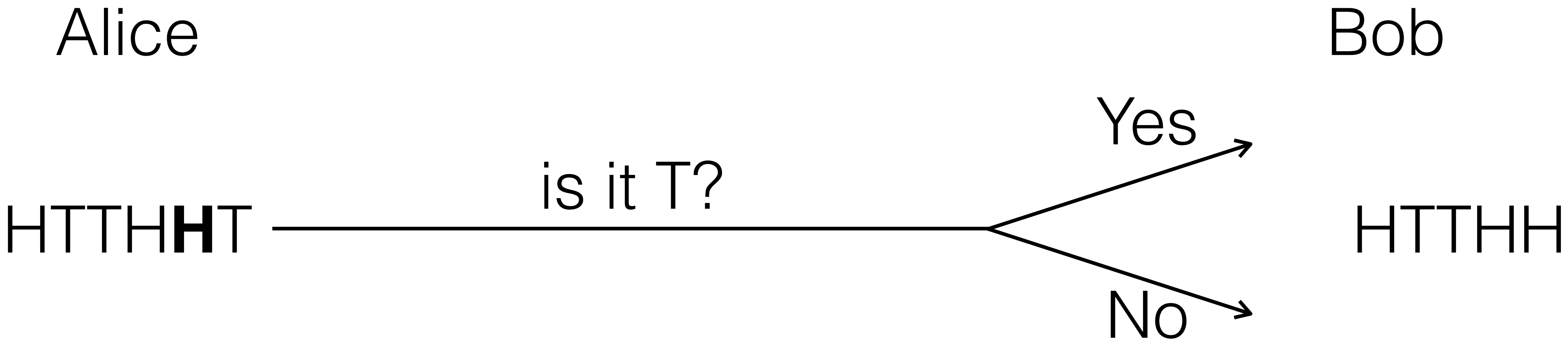
is it T?

HTT**H**HT

HTTH

No

Reduction to YES or NO answers



Alice

Bob

HTTH**H**T

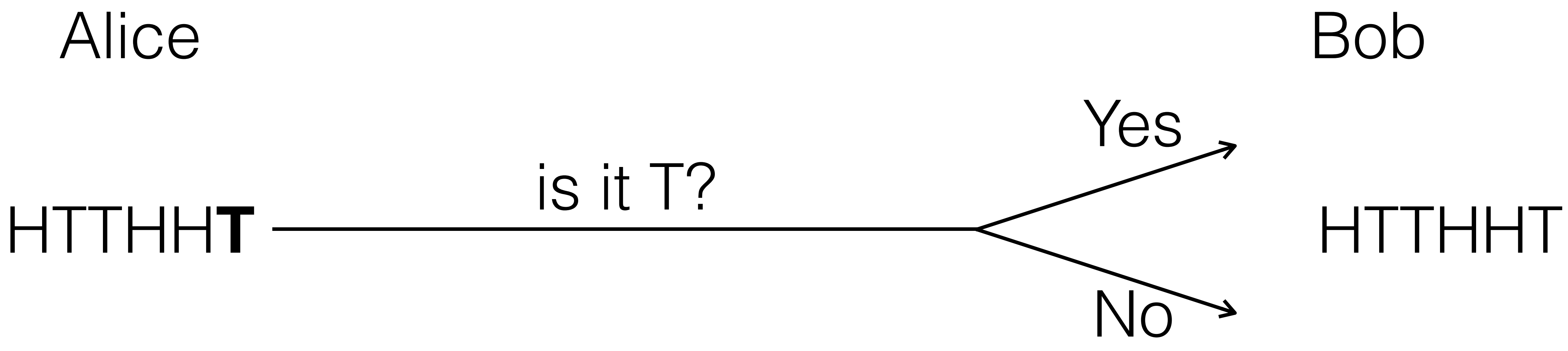
is it T?

Yes

No

HTTHH

Reduction to YES or NO answers



Alice

Bob

HTTHH**T**

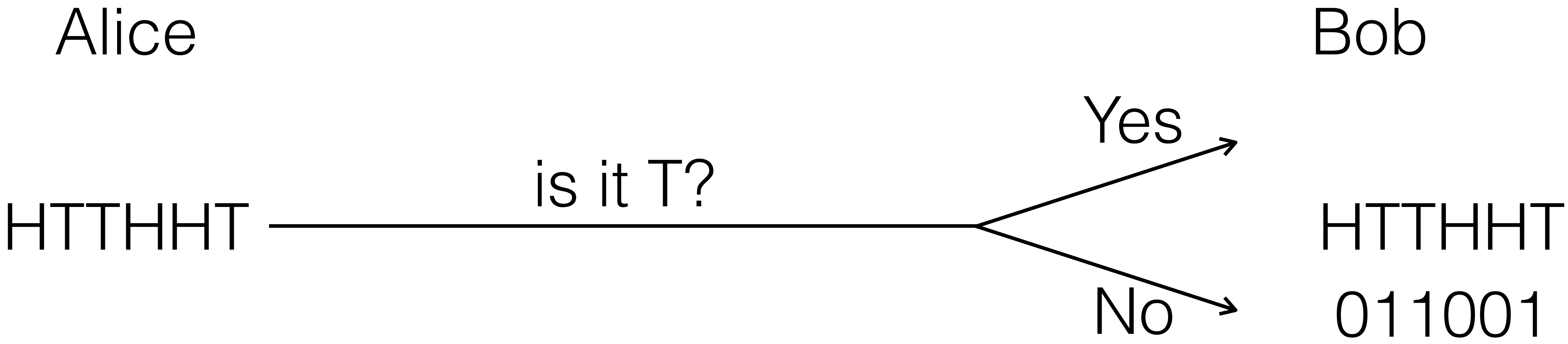
is it T?

Yes

No

HTTHHT

Reduction to YES or NO answers



Alice

Bob

HTTHHT

is it T?

Yes

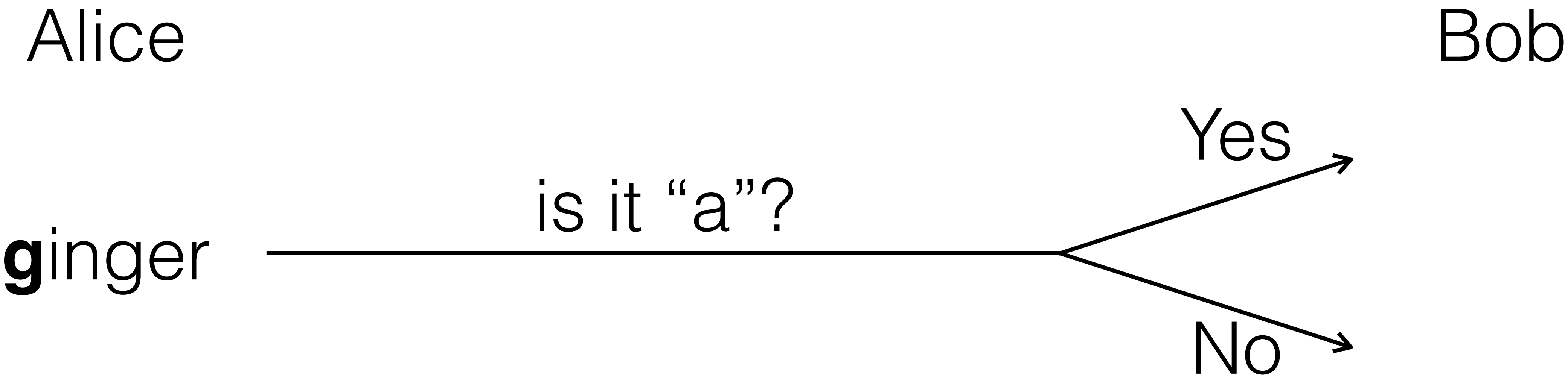
No

HTTHHT

011001

Transmission of 6 symbols requires 6 questions (bits) word composed by 6 characters

Reduction to YES or NO answers



Alice

Bob

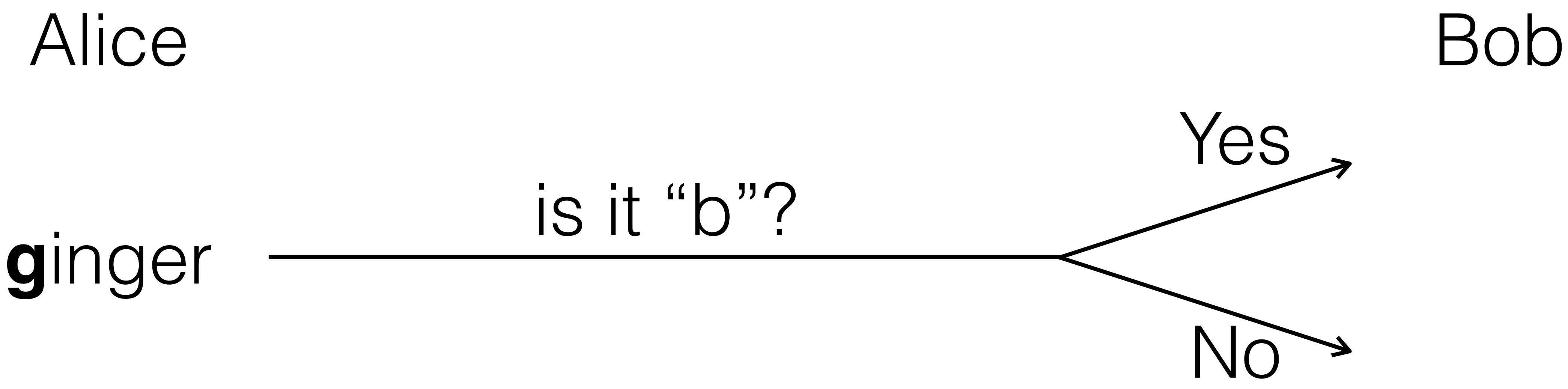
ginger

is it “a”?

Yes

No

Reduction to YES or NO answers



Alice

Bob

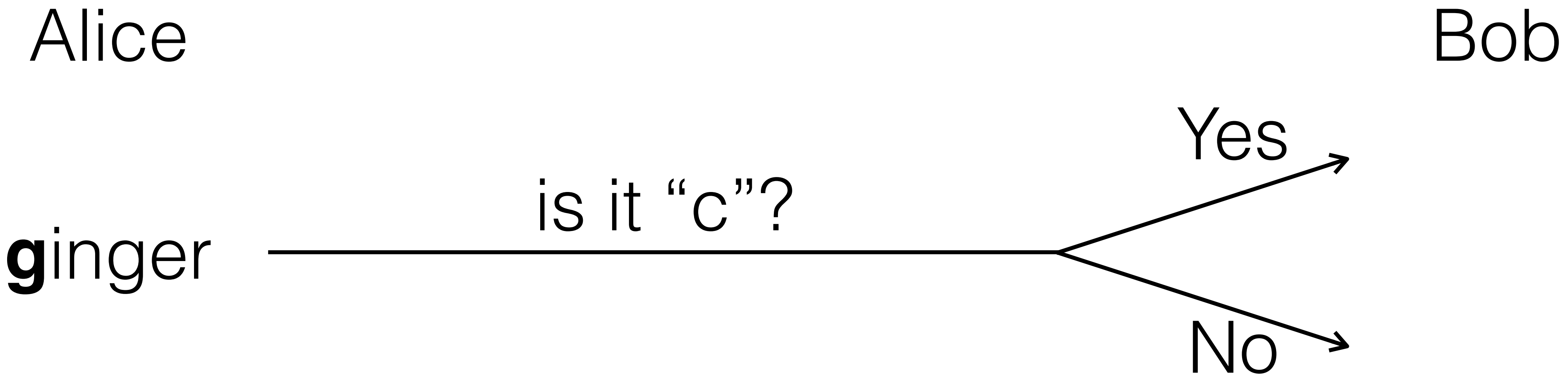
ginger

is it “b”?

Yes

No

Reduction to YES or NO answers



Alice

Bob

ginger

is it “c”?

Yes

No

Maximum of 26 questions, 13 on average (if characters outcome are i.i.d.)

Inefficient!

Reduction to YES or NO answers

ABCDEF**G**HIJKLMNOPQRSTUVWXYZ

is it lesser than “N”?

ABCDEF**G**HIJKLM~~NOPQRSTUVWXYZ~~

is it lesser than “F”?

~~ABCDE~~F**G**HIJKLM~~NOPQRSTUVWXYZ~~

is it lesser than “J”?

~~ABCDE~~F**G**HI~~JKLMNOPQRSTUVWXYZ~~

is it lesser than “H”?

~~ABCDE~~F**G**~~HIJKLMNOPQRSTUVWXYZ~~

after 5 questions we correctly individuate the character

Minimum number of questions

- $2^{\# \text{ questions}} = 26$ (for english alphabet)

- # questions = $\log_2(26) = 4.7$ expected number of questions

- for a word composed by 6 character 6*4.7 = 28.2 questions needed

Reduction to YES or NO answers

- Rationale: reduce at each iteration the size off the set of one half
- Build a decision tree where the leafs of the tree are the available symbols
- Maximum number of questions equal to the height of the tree

Telegraphy

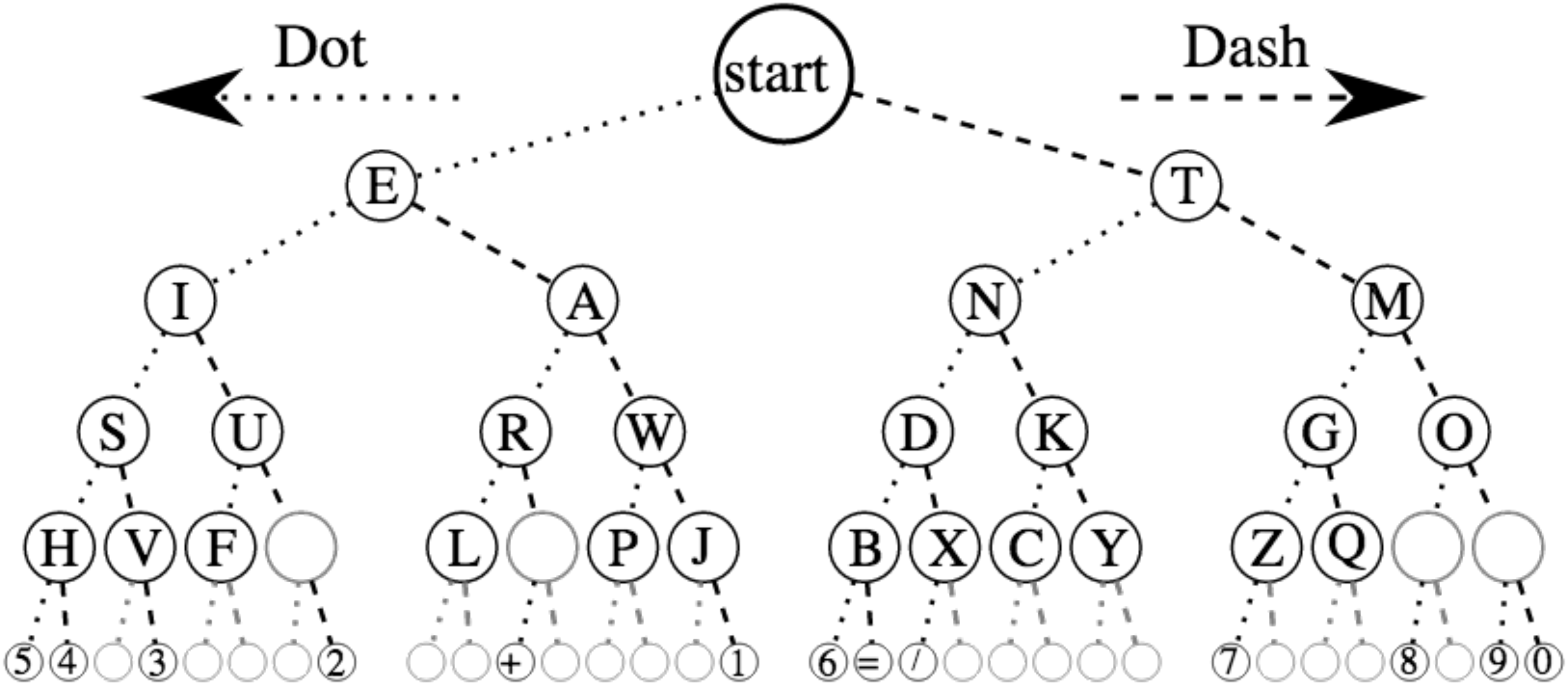
- **Telegraphy** (from Greek: **tele** "at a distance", and **graphein** "to write")

- **Long distance** transmission of textual/symbolic messages

- Method used for encoding the message be known to both sender and receiver

- Even **e-mail** is an example of telegraphy

Morse code



Measuring information

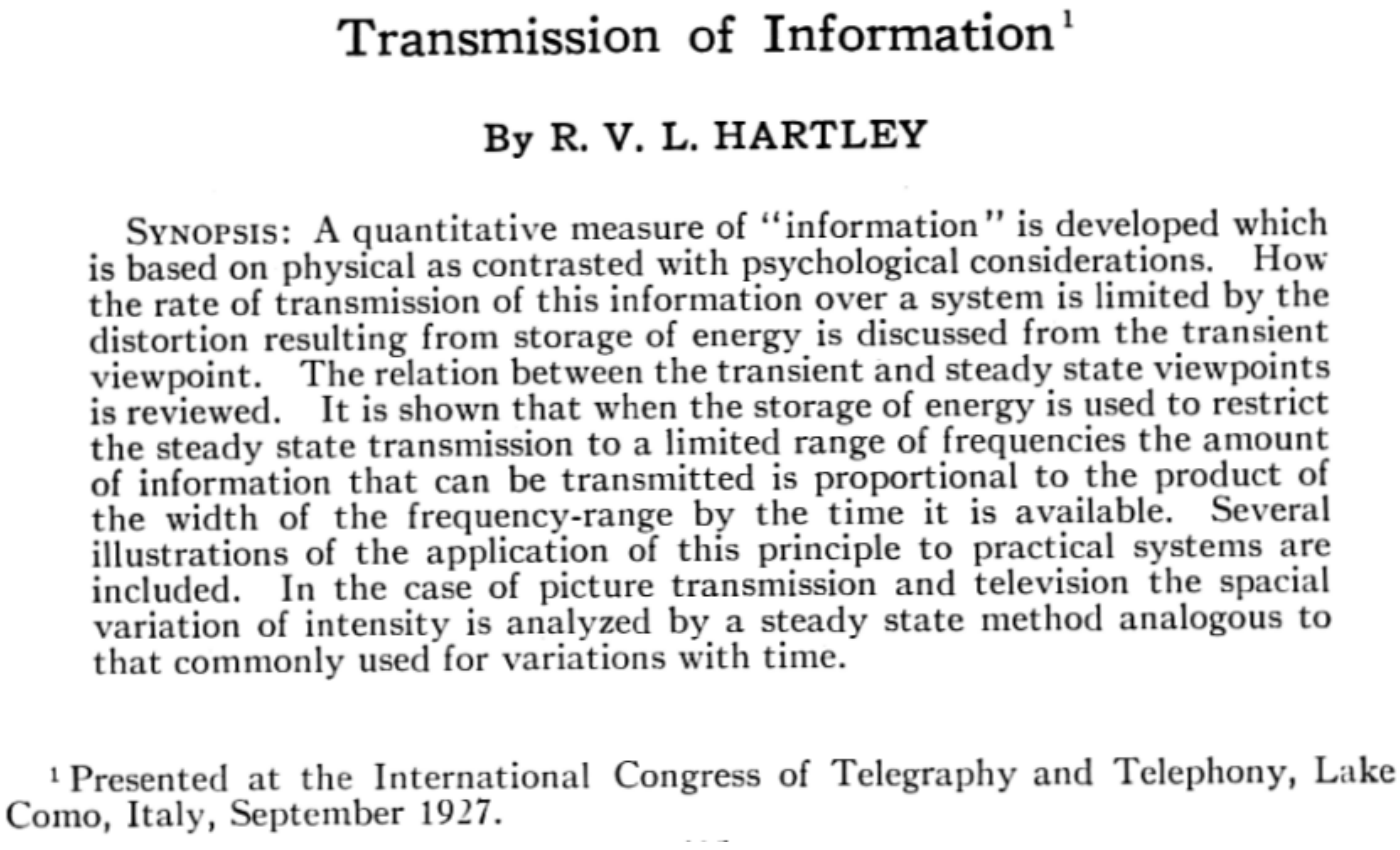
- ***s*****:** symbols (binary, decimal, …)
- ***n*****:** message length
- ***s^n*****:** possible messages

- The problem is to estimate the quantity of information relative to a message

Ralph Hartley

- R. Hartley was an electronics researcher
- Contributed to the **foundations of information theory**
- The **hartley**, a unit of information equal to one decimal digit, is named after him

Ralph Hartley



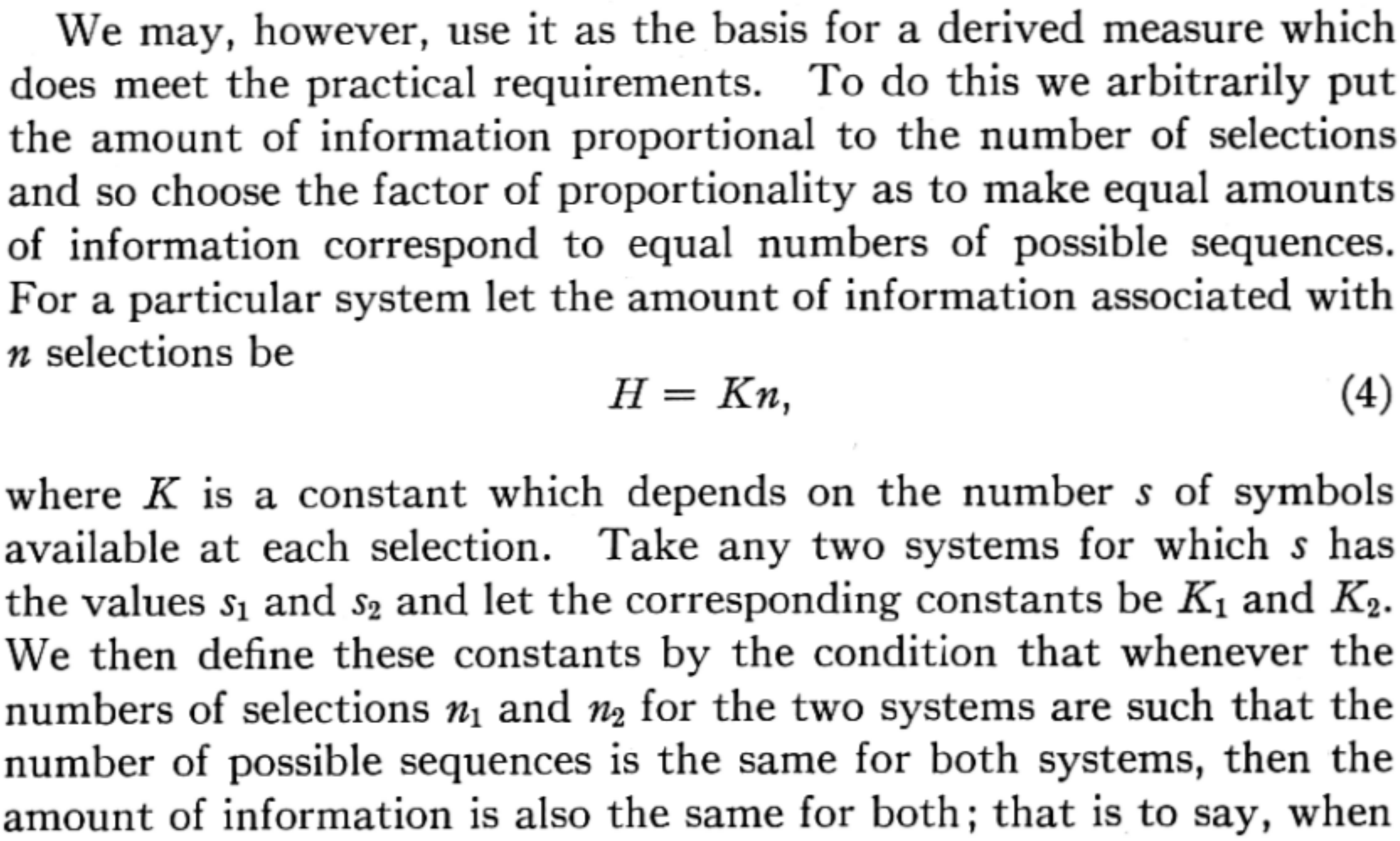
Transmission of Information[1]

By R. V. L. HARTLEY

SYNOPSIS: A quantitative measure of "information" is developed which is based on physical as contrasted with psychological considerations. How the rate of transmission of this information over a system is limited by the distortion resulting from storage of energy is discussed from the transient viewpoint. The relation between the transient and steady state viewpoints is reviewed. It is shown that when the storage of energy is used to restrict the steady state transmission to a limited range of frequencies the amount of information that can be transmitted is proportional to the product of the width of the frequency-range by the time it is available. Several illustrations of the application of this principle to practical systems are included. In the case of picture transmission and television the spacial variation of intensity is analyzed by a steady state method analogous to that commonly used for variations with time.

[1] Presented at the International Congress of Telegraphy and Telephony, Lake Como, Italy, September 1927.

Ralph Hartley

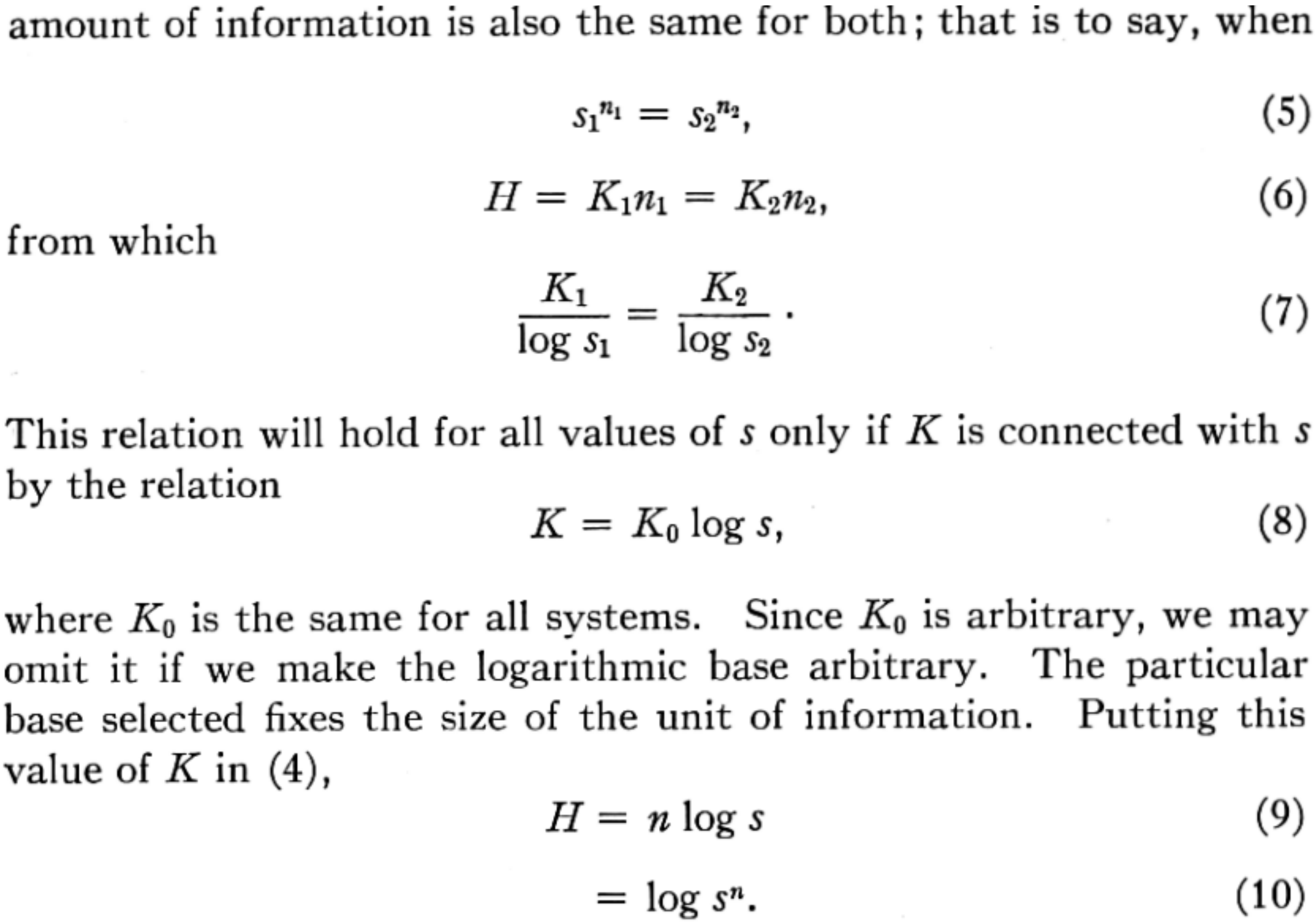


We may, however, use it as the basis for a derived measure which does meet the practical requirements. To do this we arbitrarily put the amount of information proportional to the number of selections and so choose the factor of proportionality as to make equal amounts of information correspond to equal numbers of possible sequences. For a particular system let the amount of information associated with n selections be

$$H = Kn, \tag{4}$$

where K is a constant which depends on the number s of symbols available at each selection. Take any two systems for which s has the values s_1 and s_2 and let the corresponding constants be K_1 and K_2. We then define these constants by the condition that whenever the numbers of selections n_1 and n_2 for the two systems are such that the number of possible sequences is the same for both systems, then the amount of information is also the same for both; that is to say, when

Ralph Hartley



amount of information is also the same for both; that is to say, when

$$s_1^{n_1} = s_2^{n_2}, \tag{5}$$

$$H = K_1 n_1 = K_2 n_2, \tag{6}$$

from which

$$\frac{K_1}{\log s_1} = \frac{K_2}{\log s_2} \,. \tag{7}$$

This relation will hold for all values of s only if K is connected with s by the relation

$$K = K_0 \log s, \tag{8}$$

where K_0 is the same for all systems. Since K_0 is arbitrary, we may omit it if we make the logarithmic base arbitrary. The particular base selected fixes the size of the unit of information. Putting this value of K in (4),

$$H = n \log s \tag{9}$$

$$= \log s^n. \tag{10}$$

A mathematical theory of communication by Claude Shannon

Information source

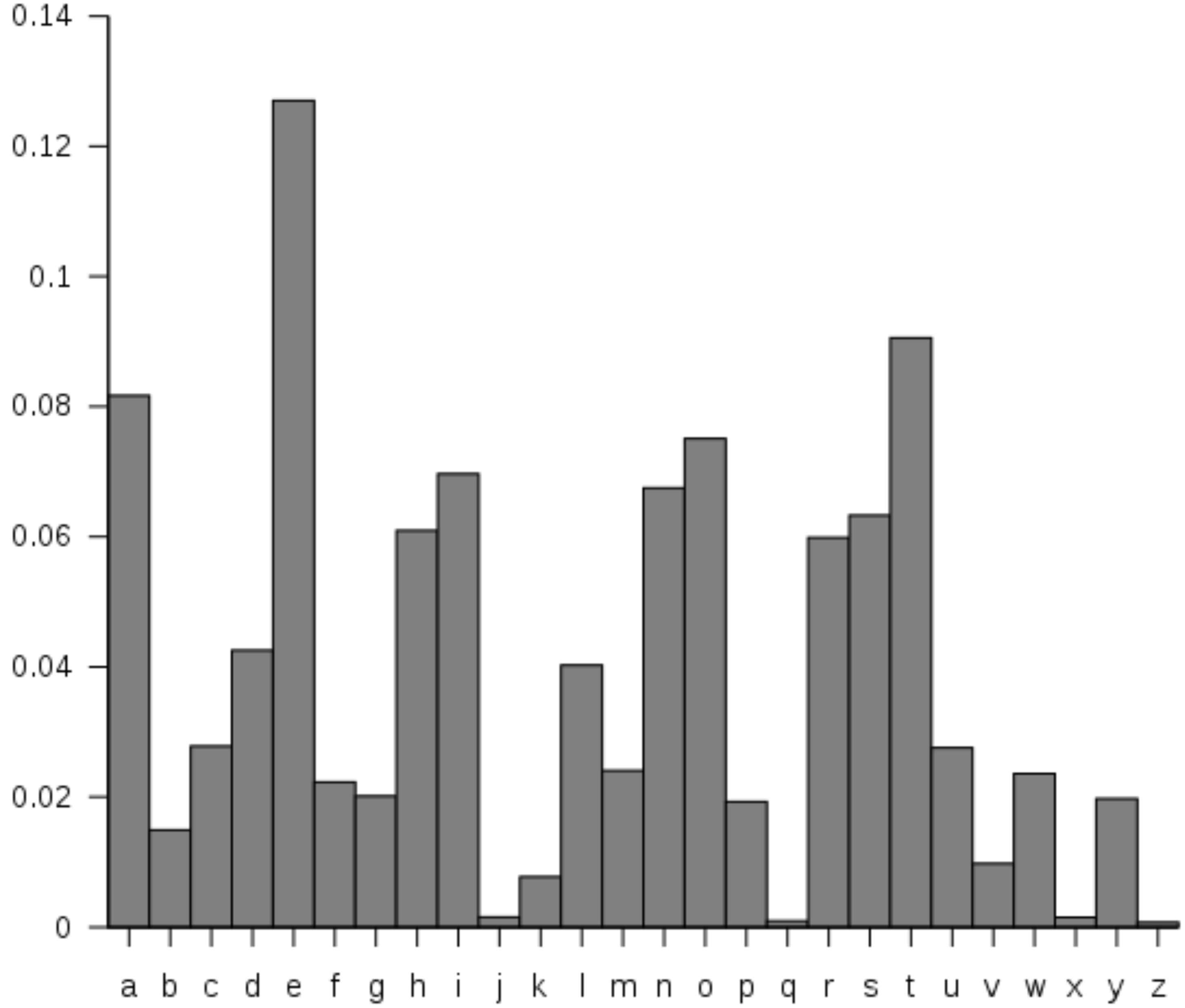
- How is an **information source** to be described mathematically?

- How much **information** in bits per second is **produced** in a given source?

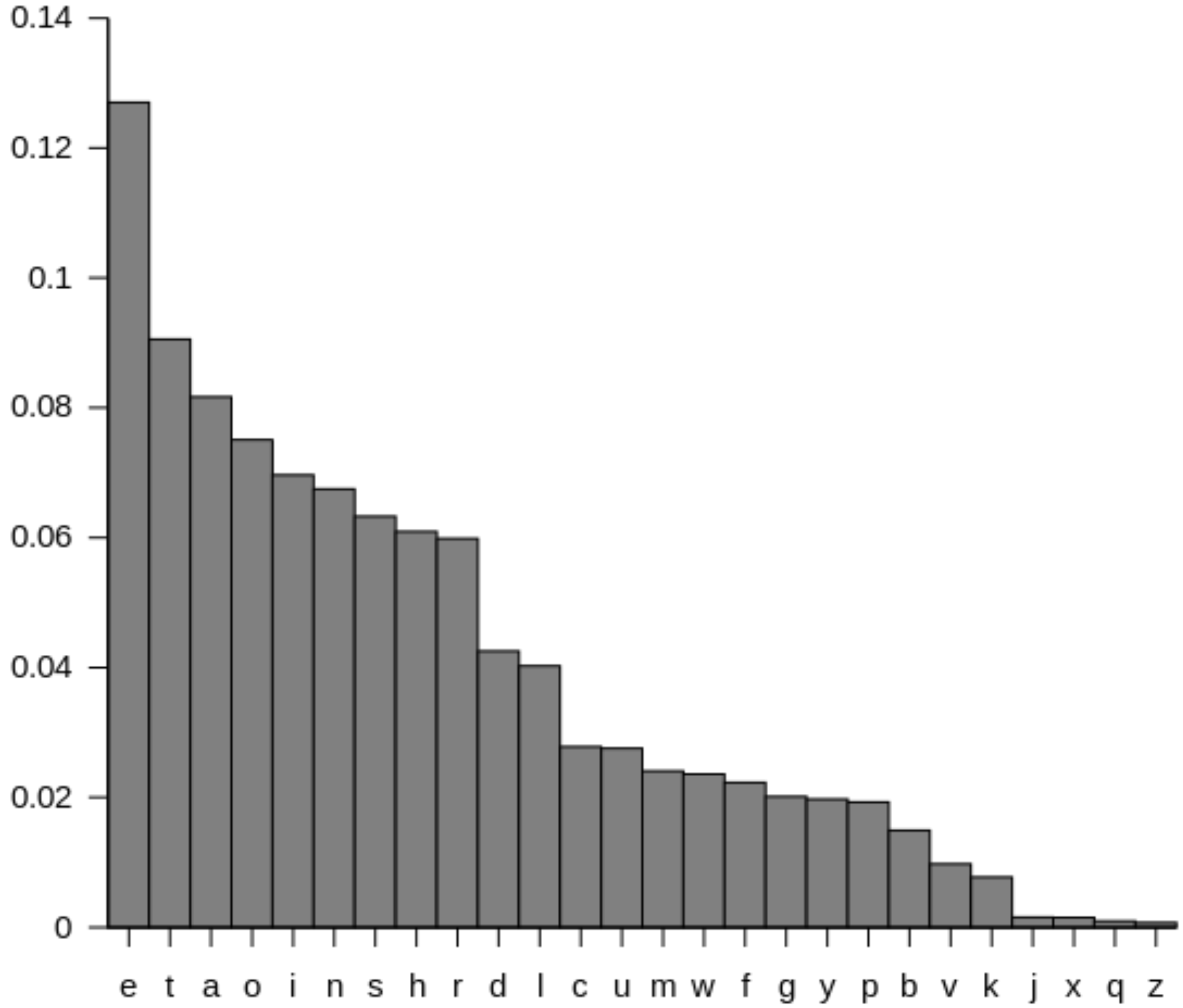
How is an information source to be described mathematically?

In telegraphy, for example, the messages to be transmitted consist of sequences of letters. These sequences, however, are **not completely random**. In general, they form sentences and have the **statistical structure** of, say, English. The letter E occurs more frequently than Q, the sequence TH more frequently than XP, etc. The existence of this structure allows one to make a saving in time (or channel capacity) by properly encoding the message sequences into signal sequences.

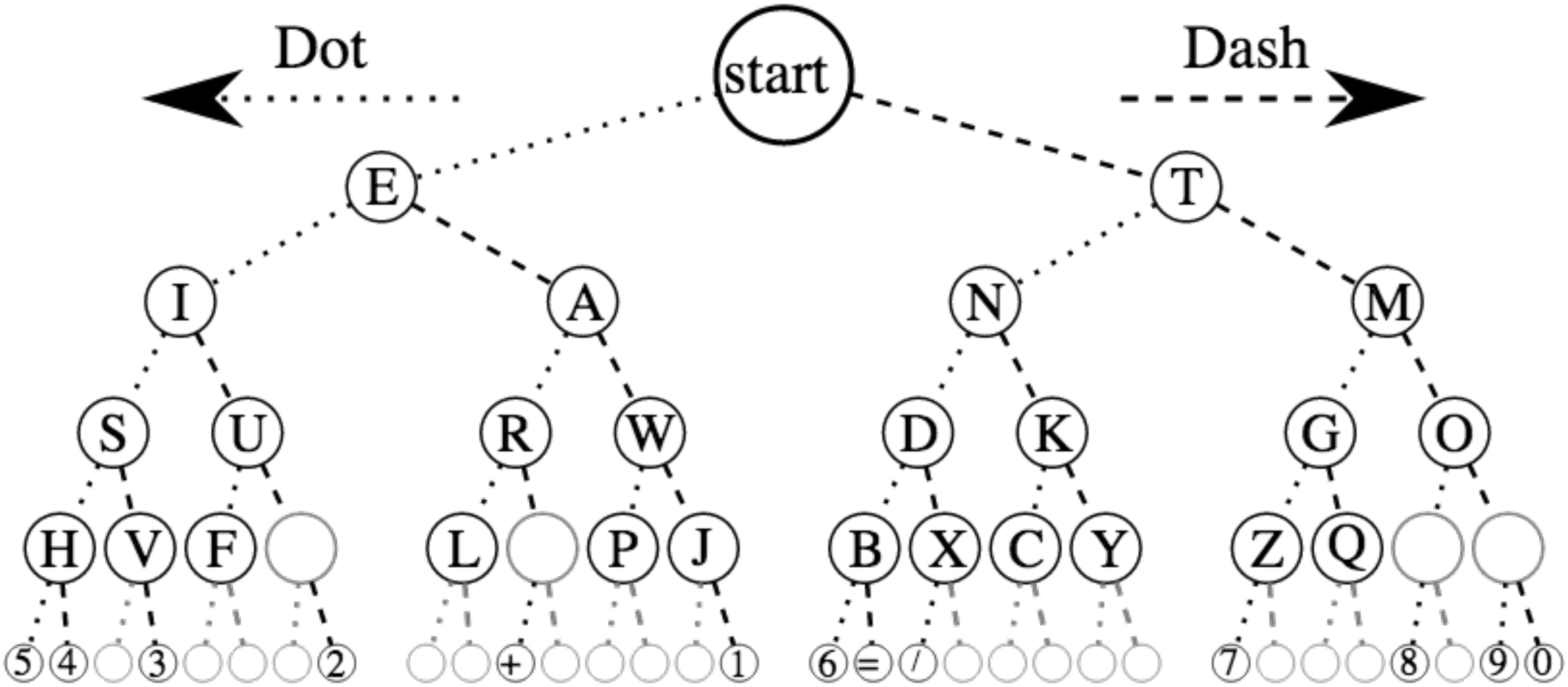
How is an information source to be described mathematically?



How is an information source to be described mathematically?



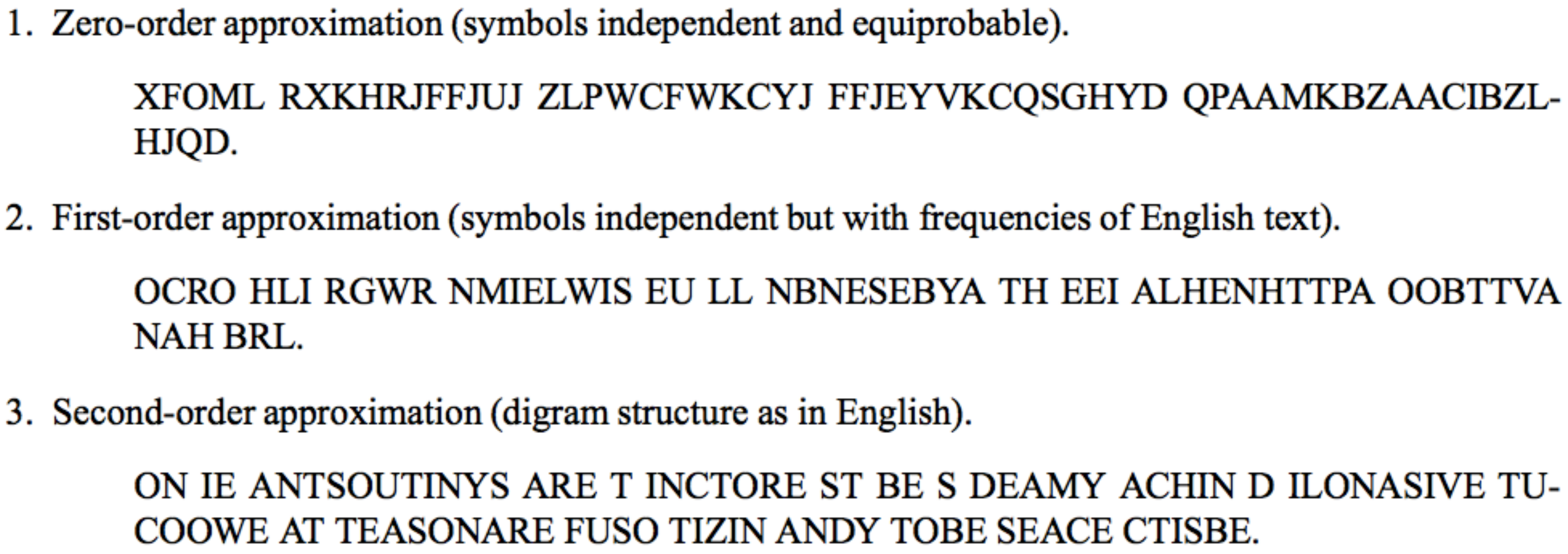
Morse code



How is an information source to be described mathematically?

We can think of a **discrete source** as generating the message, **symbol by symbol**. It will choose successive symbols according to certain probabilities depending, in general, on preceding choices as well as the particular symbols in question. A physical system, or a mathematical model of a system which produces such a sequence of symbols governed by a set of probabilities, is known as a **stochastic process**.

The series of approximation to English



1. Zero-order approximation (symbols independent and equiprobable).

   XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD.

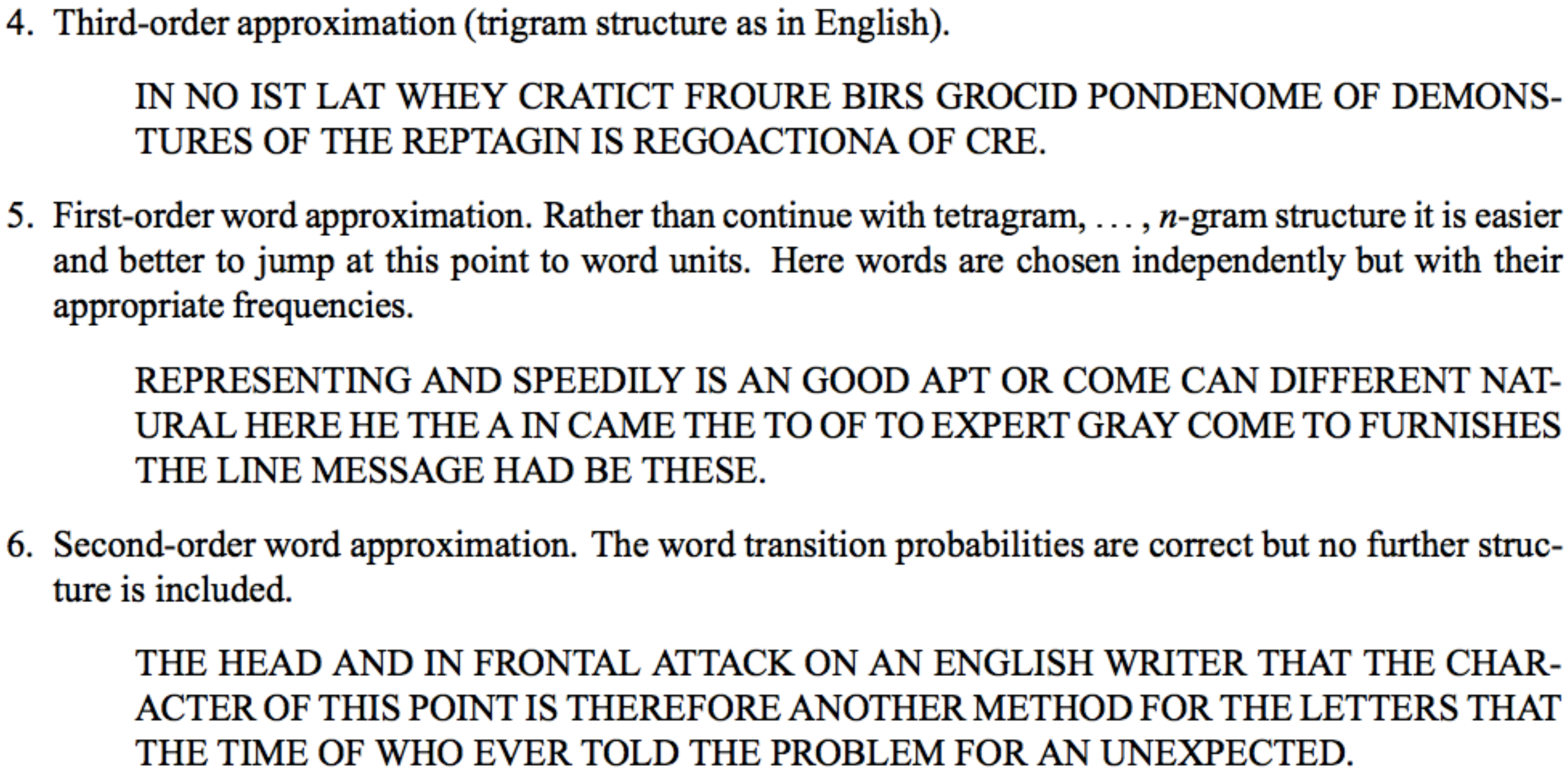
2. First-order approximation (symbols independent but with frequencies of English text).

   OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL.

3. Second-order approximation (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.

The series of approximation to English



4. Third-order approximation (trigram structure as in English).

   IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE.

5. First-order word approximation. Rather than continue with tetragram, . . . , *n*-gram structure it is easier and better to jump at this point to word units. Here words are chosen independently but with their appropriate frequencies.

   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.

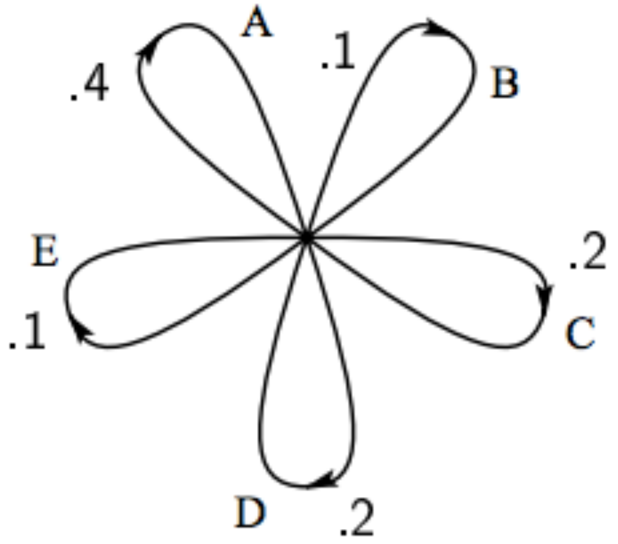
6. Second-order word approximation. The word transition probabilities are correct but no further structure is included.

   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.

Stochastic process which generates a sequences of symbols

Using the same five letters (ABCDE) let the probabilities be .4, .1, .2, .2, .1, respectively, with successive choices independent. A typical message from this source is then:

- AAACDCBDCEAADADACEDA
- E A D C A B E D A D D C E C A A A A A D

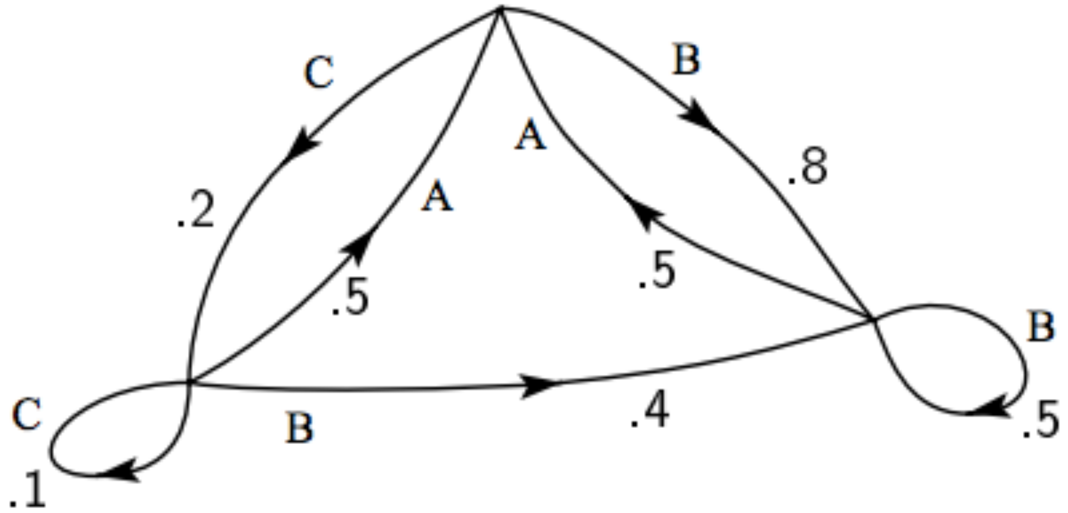


Stochastic process which generates a sequences of symbols

A more complicated structure is obtained if successive symbols are **not chosen independently** but their **probabilities depend on preceding letters**.

In the simplest case of this type a choice depends only on the preceding letter and not on ones before that.

The statistical structure can then be described by a set of **transition probabilities** $p_i(j)$, the probability that letter *i* is followed by letter *j*



Choice, Uncertainty and Entropy

- We have represented a discrete information source as a **Markov process**. Can we define a quantity which will measure, in some sense, how much **information** is "produced" by such a process, or better, at what rate information is produced?

- Suppose we have a set of possible events whose probabilities of occurrence are $p_1 ; p_2 ;\ldots; p_n$. These probabilities are known but that is all we know concerning which event will occur. Can we find a measure of how much "**choice**" is involved in the selection of the event or of how **uncertain** we are of the outcome?

Choice, Uncertainty and Entropy

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

where the constant K merely amounts to a choice of a unit of measure

Choice, Uncertainty and Entropy

“My greatest concern was what to call it. I thought of calling it ‘information’, but the word was overly used, so I decided to call it ‘uncertainty’. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, ‘You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage.”

Example: tossing a fair coin

- Total number of possible outcomes: N = 2
- probability heads: p(H) = 0.5
- probability tails: p(T) = 0.5
- Shannon's entropy: H = -(1/2)log(1/2) - (1/2)log(1/2) = -(1/2)(-1) - (1/2)(-1) = 1/2+1/2 = **1 bit**
- 1 bit of **information gained**
- 1 bit of **uncertainty reduced**

Example: tossing a double head coin

- Total number of possible outcomes: N = 1
- probability heads: p(H) = 1
- Shannon's entropy: H = -(1)log(1) = -(1)(0) = **0 bits**
- 0 bits of **information gained**
- 0 bits of **uncertainty reduced**

Shannon entropy characteristics

- **Continuity:** the measure should be continuous, so that changing the values of the probabilities by a very small amount should only change the entropy by a small amount.

- **Symmetry**: the measure should be unchanged if the outcomes are re-ordered

$$H_n\left(p_1, p_2, \ldots\right) = H_n\left(p_2, p_1, \ldots\right)$$

Shannon entropy characteristics

- **Additivity:** the amount of entropy should be independent of how the process is regarded as being divided into parts

$$H_n(p_1, p_2) = H_n(p_1) + H_n(p_2)$$

if p_1 and p_2 are independent

Shannon entropy characteristics

- **Maximum:** the measure should be maximal if all the outcomes are equally likely (uncertainty is highest when all possible events are equiprobable)

$$H_n(p_1, \ldots, p_n) \leq H_n\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) = \log_b(n)$$

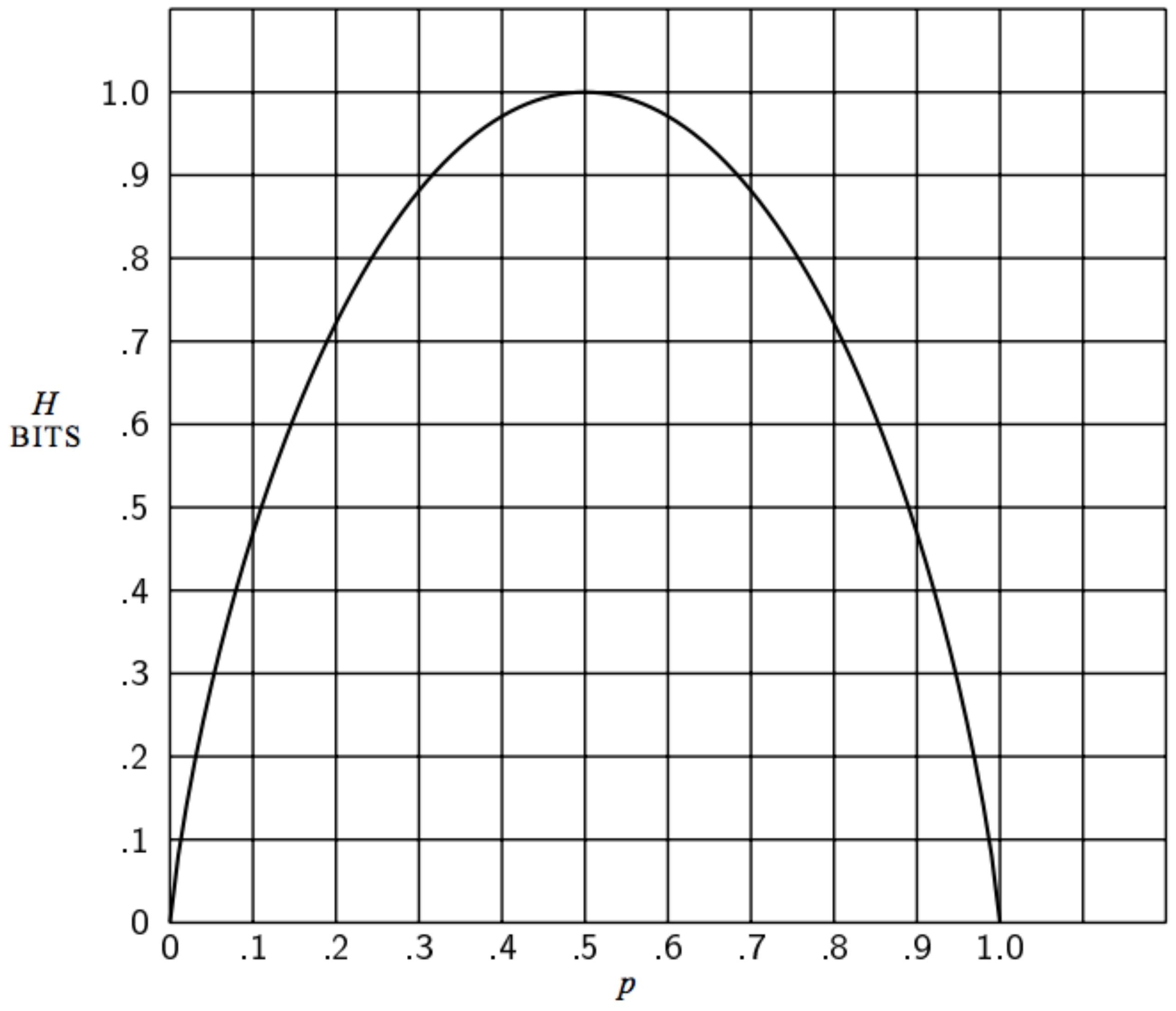
For equiprobable events the entropy should increase with the number of outcomes

$$H_n\bigg(\underbrace{\frac{1}{n}, \ldots, \frac{1}{n}}_{n}\bigg) = \log_b(n) < \log_b(n+1) = H_{n+1}\bigg(\underbrace{\frac{1}{n+1}, \ldots, \frac{1}{n+1}}_{n+1}\bigg)$$

Entropy in the case of two possibilities

- Entropy in the case of two possibilities with probabilities p and q = 1 - p

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



Choice, Uncertainty and Entropy

- Let's suppose that all symbols are **equiprobable** and **independent** with probability p_i=1/q (q symbols)

$$H = -K \sum_{i=1}^{q} p_i \log p_i$$

the entropy of a message can be written as

$$\begin{aligned} H &= -KN \sum_{i=1}^{q} \left(\frac{1}{q}\right) \log \left(\frac{1}{q}\right) \\ &= -KNq \left(\frac{1}{q}\right) \log \left(\frac{1}{q}\right) \\ &= KN \log(q) \end{aligned}$$

Choice, Uncertainty and Entropy

$$\begin{aligned} H &= -KN\sum_{i=1}^{q}\left(\frac{1}{q}\right)\log\left(\frac{1}{q}\right) \\ &= -KNq\left(\frac{1}{q}\right)\log\left(\frac{1}{q}\right) \\ &= KN\log(q) \end{aligned}$$

- If the number of symbols is equal to 2 (binary system) and assuming K=1

$$H = KN\log(q) = N\log(2) = N$$

the entropy of the message coincide with its length

Shannon entropy

- Quantitative measure of **information**
- Quantitative information of “**surprise**”

Unit measure

- Depending on the base (*b*) of the logarithm (and constant K) the **unit measure** of information changes:
    - *b* = 2 -> **bit**
    - *b* = e -> **nat**
    - *b* = 10 -> **dit** (or hartley)

Estimating the information entropy of the written English text

- 27 characters (26 letters + space). Hence N = 27

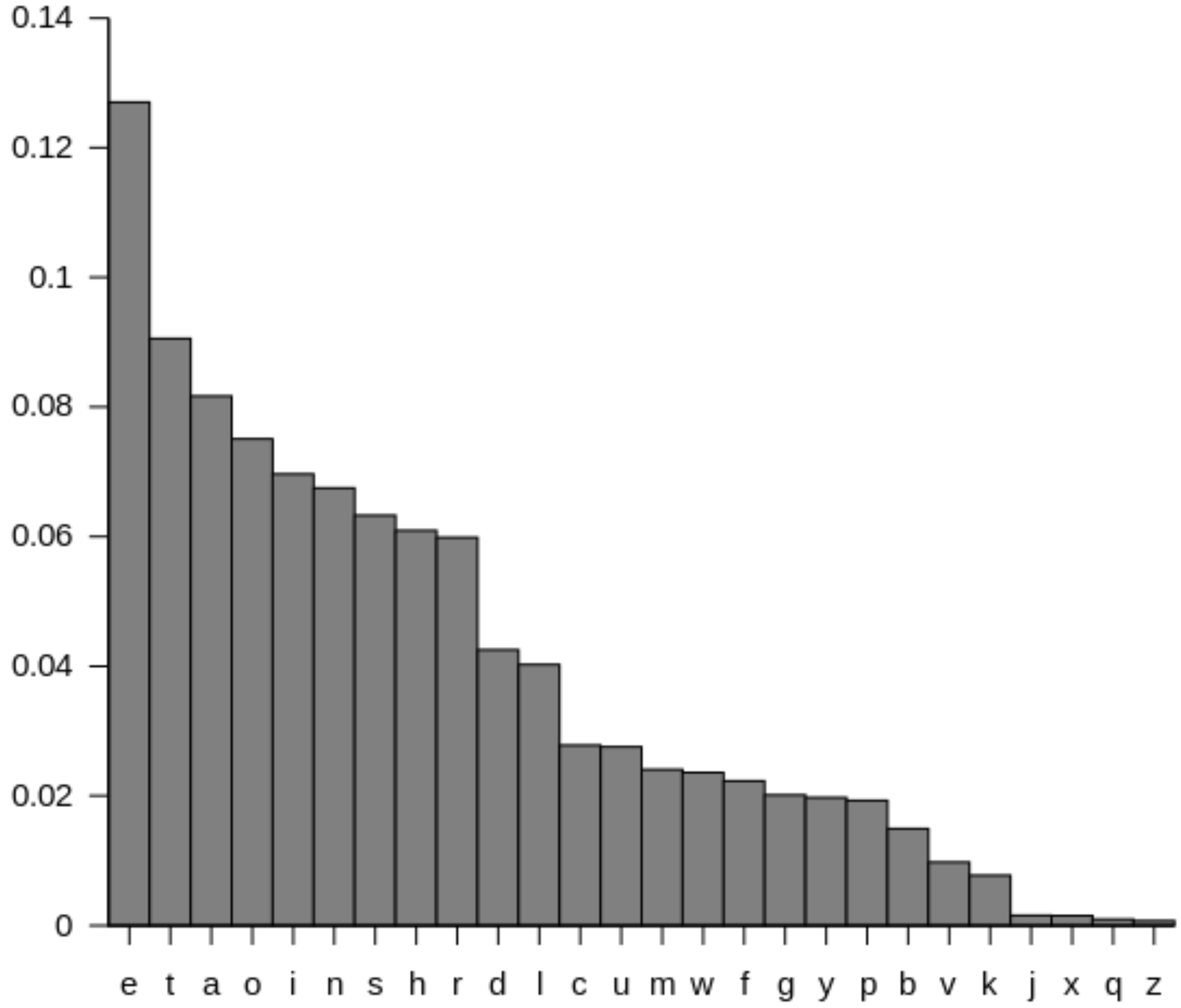
- We assume all character equally probable: p(each char) = 1/27

- The information entropy per character is therefore: H = -27(1/27)log(1/27) = log(27) = **4.75 bits**

Estimating the information entropy of the written English text

E highest frequency

Z lowest frequency



Redundancy in written english text

- Redundancy: nothing more than the number of constraints imposed on the text of the English language.

- For example, the letter Q is always followed by U, and we also have rules such as "I before E except after C", and so on.

Estimating the information entropy of the written English text

- According to Shannon, considering redundancy and contextually the entropy per character of english text was estimated to be:

H between 0.6 and 1.3 bits

H ≈ 1 bit

Estimating the information entropy of the written English text

- Equally probable characters

$$\mathbf{H = 4.75\ bits}$$

- Considering redundancy, contextually etc..

$$\mathbf{H \approx 1\ bit}$$

Redundancy

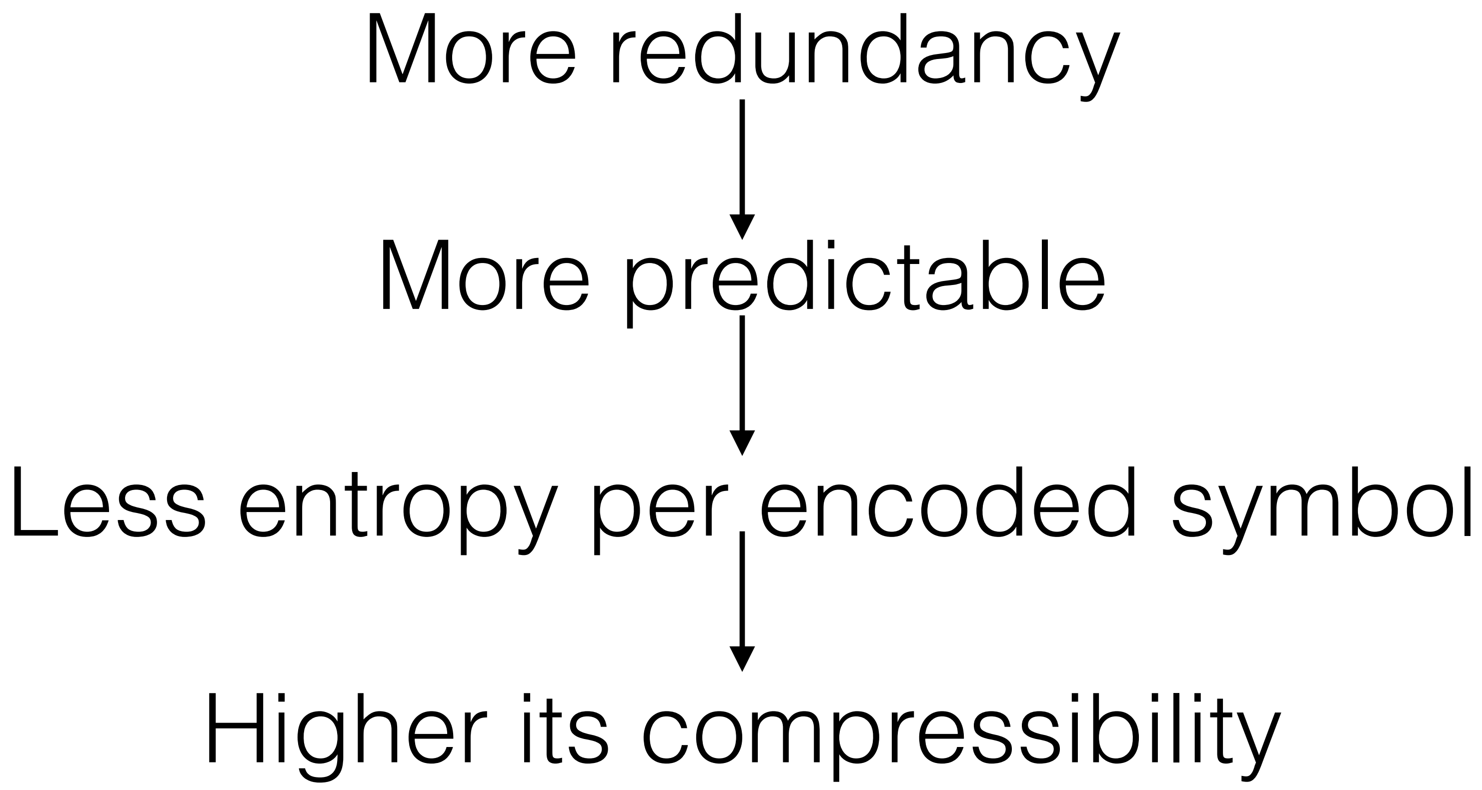
- Redundancy = number of bits to encode a message - number of bits of Shannon's information

- Redundancy in a message is a measure of the compressibility of the message

Loss-less data compression

- To reduce the number of bits used to encode a message by identifying and eliminating statistical redundancy.
- The exact original data can be reconstructed from compressed data

Redundancy



More redundancy

More predictable

Less entropy per encoded symbol

Higher its compressibility

Compress data

- Extract redundancy from the message
- Encoding the same amount of Shannon’s information by using less bits
- More Shannon’s information per encoded symbol
- Total Shannon’s information preserved
- Compressed message less predictable

ZIP example

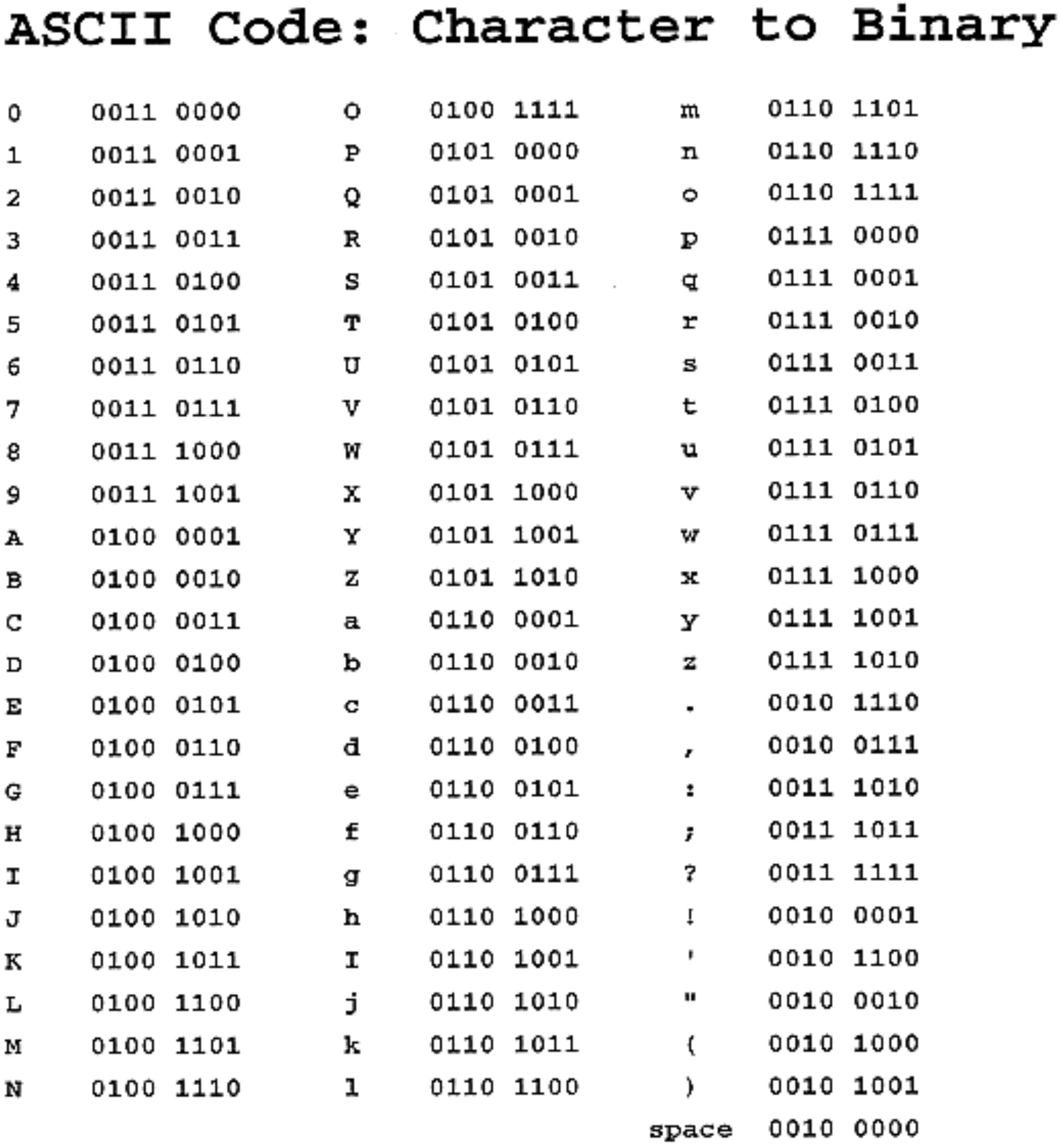
Shannon’s Entropy

- Shannon’s entropy represents a **lower limit** for lossless data compression: the minimum amount of bits that can be used to encode a message without loss

- A lossless data compression scheme cannot compress messages, on average, to have more than one bit of Shannon’s information per bit of encoded message

Example of coding: ASCII Code

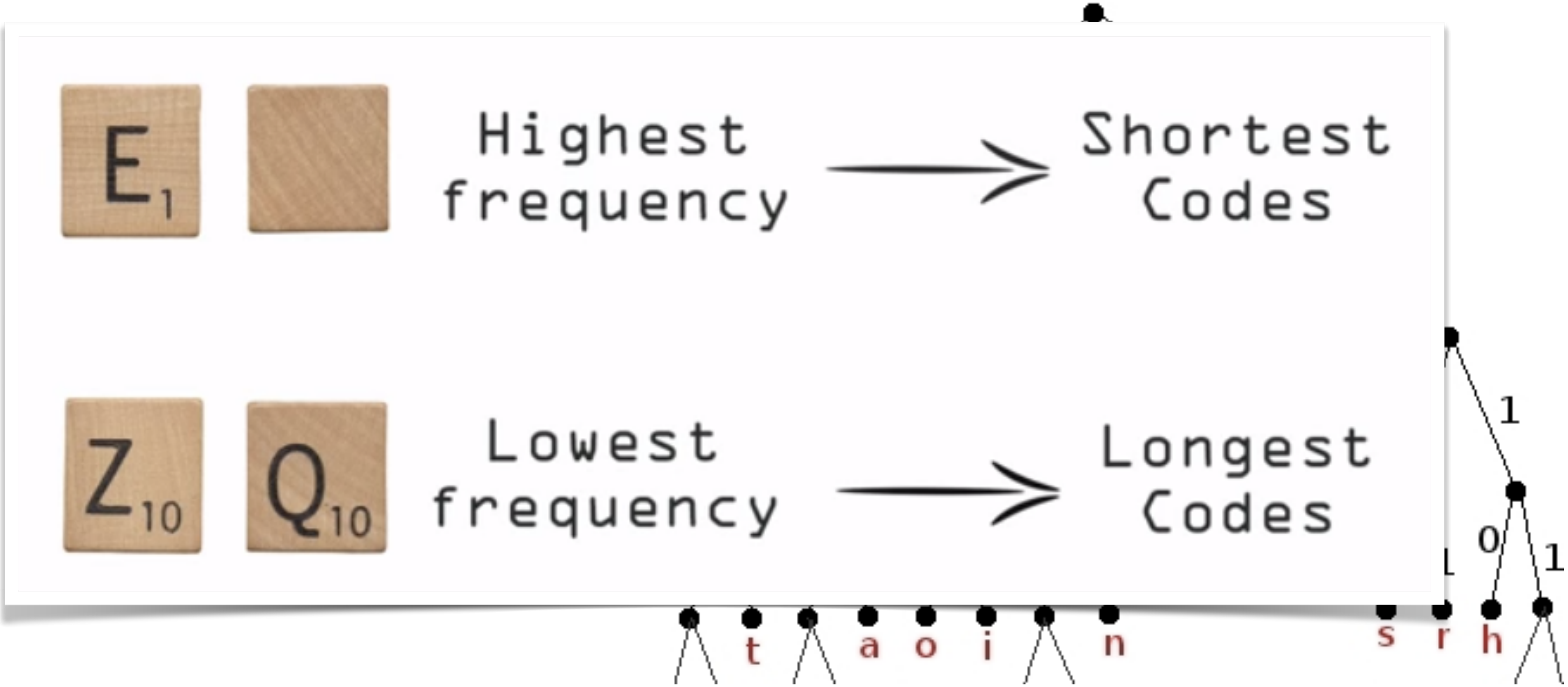
- 1 byte (8 bits) per character
- Very inefficient
- Theoretical optimum: 1 bit per character
- In theory there exist a code 8 times more efficient than ASCII code



ASCII Code: Character to Binary

0	0011 0000	O	0100 1111	m	0110 1101
1	0011 0001	P	0101 0000	n	0110 1110
2	0011 0010	Q	0101 0001	o	0110 1111
3	0011 0011	R	0101 0010	p	0111 0000
4	0011 0100	S	0101 0011	q	0111 0001
5	0011 0101	T	0101 0100	r	0111 0010
6	0011 0110	U	0101 0101	s	0111 0011
7	0011 0111	V	0101 0110	t	0111 0100
8	0011 1000	W	0101 0111	u	0111 0101
9	0011 1001	X	0101 1000	v	0111 0110
A	0100 0001	Y	0101 1001	w	0111 0111
B	0100 0010	Z	0101 1010	x	0111 1000
C	0100 0011	a	0110 0001	y	0111 1001
D	0100 0100	b	0110 0010	z	0111 1010
E	0100 0101	c	0110 0011	.	0010 1110
F	0100 0110	d	0110 0100	,	0010 0111
G	0100 0111	e	0110 0101	:	0011 1010
H	0100 1000	f	0110 0110	;	0011 1011
I	0100 1001	g	0110 0111	?	0011 1111
J	0100 1010	h	0110 1000	!	0010 0001
K	0100 1011	I	0110 1001	'	0010 1100
L	0100 1100	j	0110 1010	"	0010 0010
M	0100 1101	k	0110 1011	(	0010 1000
N	0100 1110	l	0110 1100	)	0010 1001
				space	0010 0000

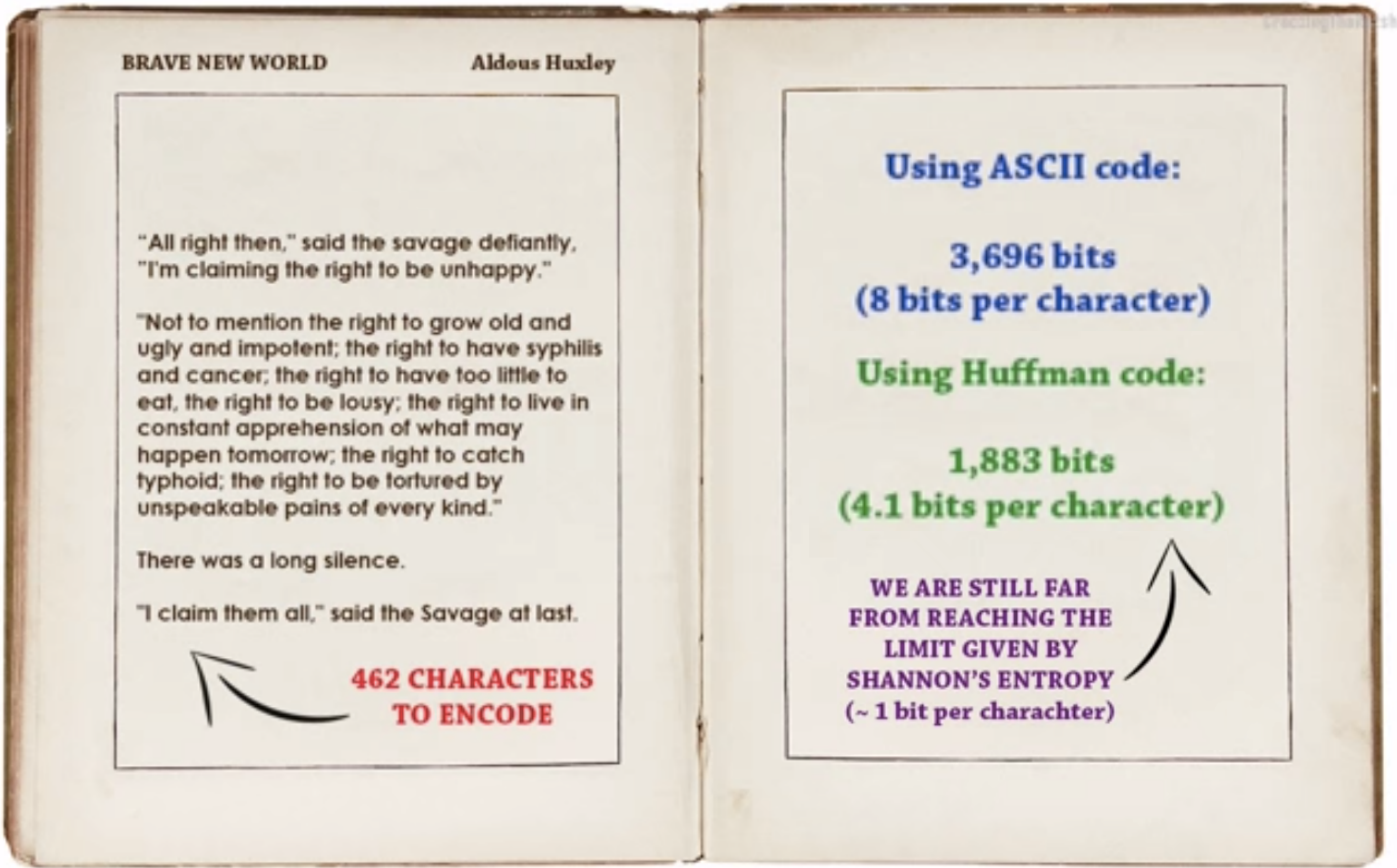
Huffman code



Highest frequency → Shortest Codes

Lowest frequency → Longest Codes

Huffman code



BRAVE NEW WORLD Aldous Huxley

"All right then," said the savage defiantly. "I'm claiming the right to be unhappy."

"Not to mention the right to grow old and ugly and impotent; the right to have syphilis and cancer; the right to have too little to eat, the right to be lousy; the right to live in constant apprehension of what may happen tomorrow; the right to catch typhoid; the right to be tortured by unspeakable pains of every kind."

There was a long silence.

"I claim them all," said the Savage at last.

462 CHARACTERS TO ENCODE

Using ASCII code:

3,696 bits
(8 bits per character)

Using Huffman code:

1,883 bits
(4.1 bits per character)

WE ARE STILL FAR FROM REACHING THE LIMIT GIVEN BY SHANNON'S ENTROPY (~ 1 bit per charachter)

Shannon’s entropy

- Shannon's entropy is a measure of uncertainty, of unpredictability, and also a measure of information content, of potential information gain.

- Shannon’s entropy can also represent a lower limit for lossless data compression: the minimum amount of bits that can be used to encode a message without loss.

Shannon’s entropy

- Also note that with this definition, more information content has nothing to do with its quality. So in this sense, a larger amount of Shannon's entropy does not necessarily imply a better quality of its content

- Encoding bits and Shannon’s bits have different meanings

Entropy:
an application

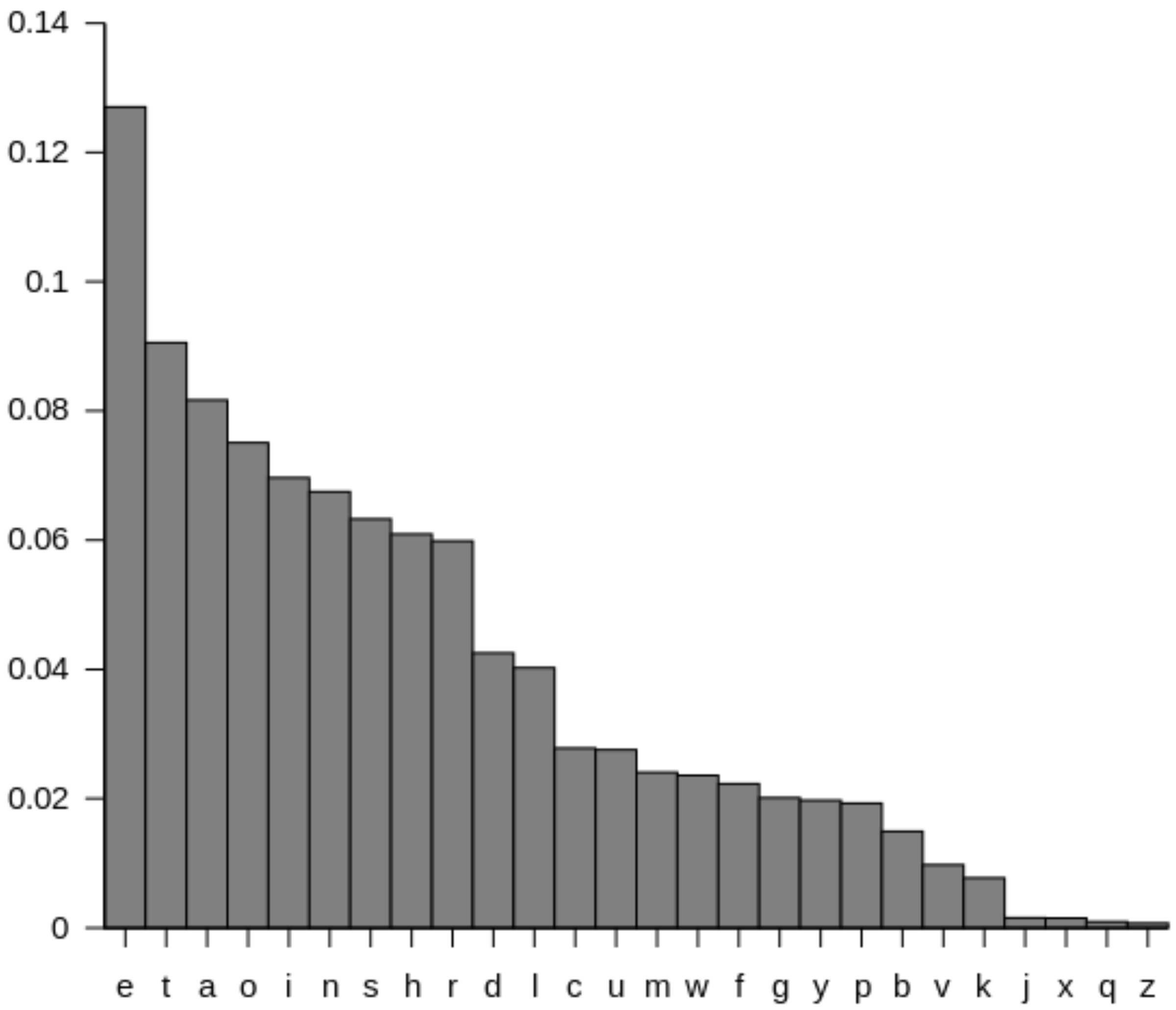
Vigenère cipher

- 1587: Vigenère Cipher
- Polyalphabetic: one to many relationship
- Example
    - Encrypt: lamp
    - Keyword: ubc
    - Ciphertext: fboj

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

Tabula recta

Vigenère cipher

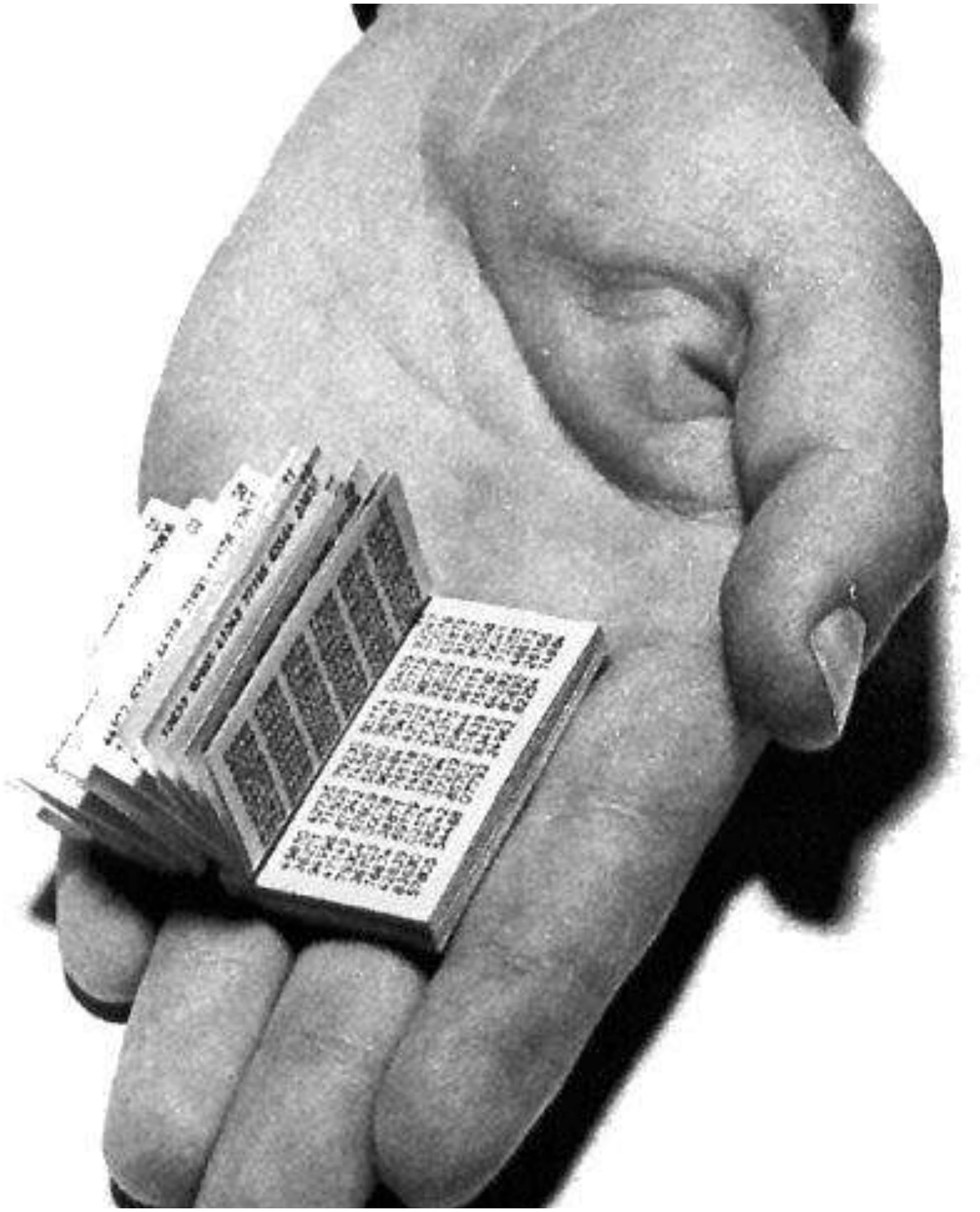


key	ABCDAB CD ABCDA BCD ABCDABCDABCD
plaintext	**CRYPTO** IS SHORT FOR **CRYPTO**GRAPHY
ciphertext	**CSASTP** KV SIQUT GQU **CSASTP**IUAQJB

Redundancy, Entropy and Security

- Shannon gave the mathematical description of a **perfect secrecy** based on the maximum entropy of a message

- **Perfect secrecy** cannot be cracked if used correctly



One Time Pad

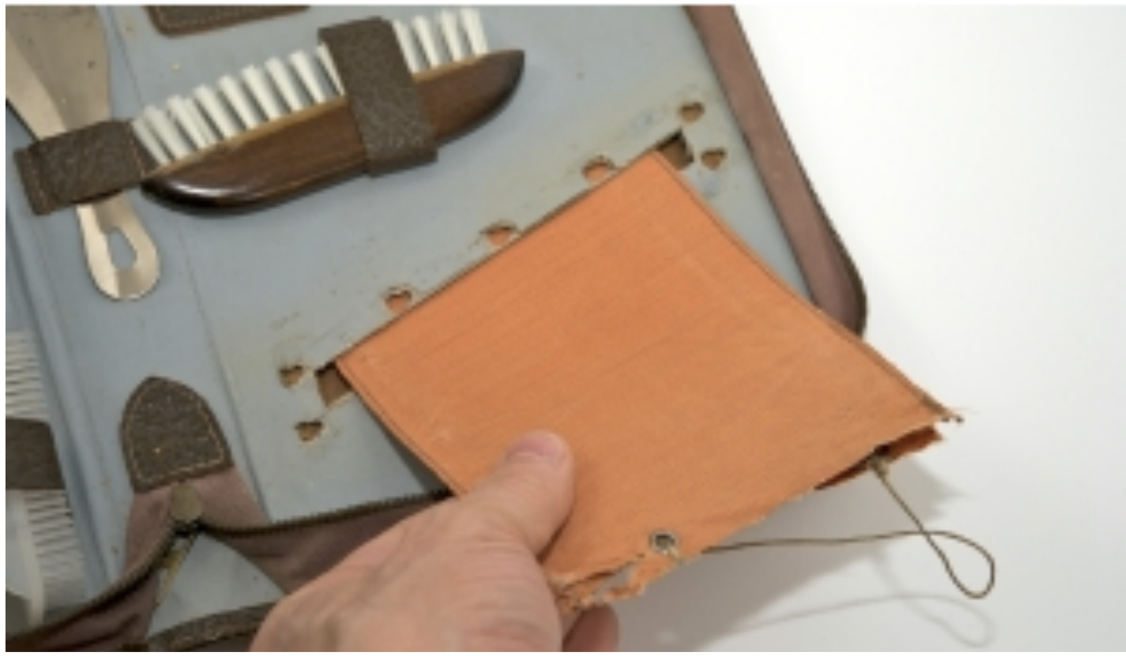
The unbreakable code

One Time Pad: requirements

- The encryption-key has at least the same length as the message
- The OTP should consist of truly random numbers
- Precisely two copies of the OTP should exist
- The OTP should only be used once
- Both copies of the OTP are destroyed immediately after use

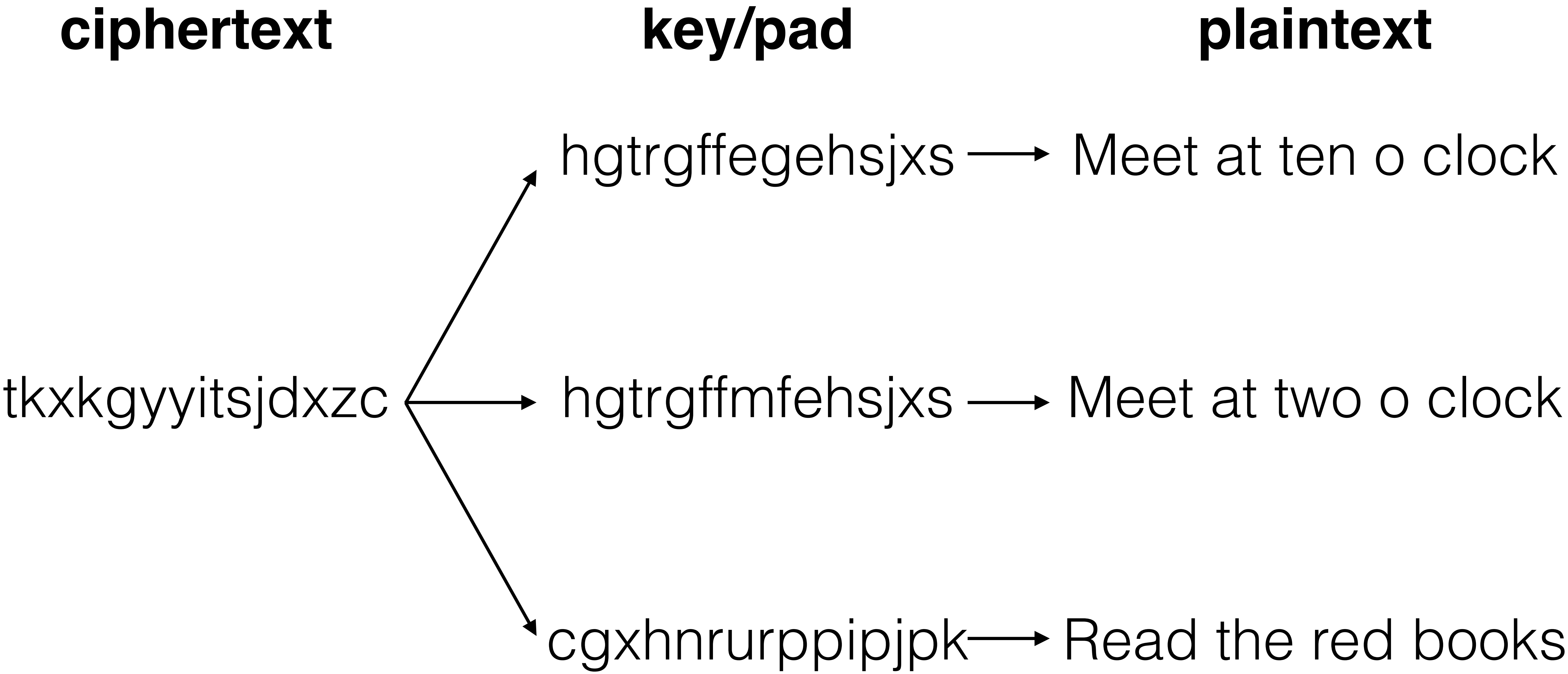
One Time Pad: distribution

- The major problem with OTPs however, is their distribution. A unique set of OTP booklets needs to be issued and distributed to each individual spy or agent abroad. As the OTP was destroyed immediately after use, sufficient and timely supply of new OTPs had to be guaranteed



The key is either a secret message!

One Time Pad



ciphertext **key/pad** **plaintext**

tkxkgyyitsjdxzc

hgtrgffegehsjxs → Meet at ten o clock

hgtrgffmfehsjxs → Meet at two o clock

cgxhnrurppipjpk → Read the red books