C Y B E R N E T I C S

Information in communication:

Shannon Information Theory

Object of Information Theory

To provide a mathematical approach to the acquisition, coding and communication of information

Questions addressed by Information Theory

- How can the amount of information be measured?
- How can information be efficiently encoded in a symbolic language?
- How much information can be transmitted per unit of time?
- How can noise be characterized and what is the influence of noise on the communication?

Relation to other domains

Mathematics	Inequalities
Probability theory	Limit theorem
Markov theory	Influence of context
Statistics	Hypothesis of minimal information
Communication theory	Limits of communication; universal distribution; MDL
Control theory	Regulation of processes
Technology	Data transmission, storage, and handling; AI; Complexity; Regulation
Physics	Thermodynamic entropy; reversible computation
Physiology	Neural, hormonal, genetic, immunological information exchange
Psychology	Cognitive psychology
Pedagogy	Learning and teaching
Linguistics	Language analysis; semiotics
Cryptography	Secret codes
Economics	Game theory and decision theory; feedback
Esthetics	Repetition and variation; theory of styles
Philosophy	Philosophy of science, epistemology: induction, Occam's razor
Logic	Deduction, induction, optimal conclusions from incomplete information
Biology	Self organizing systems; origin of cooperativity

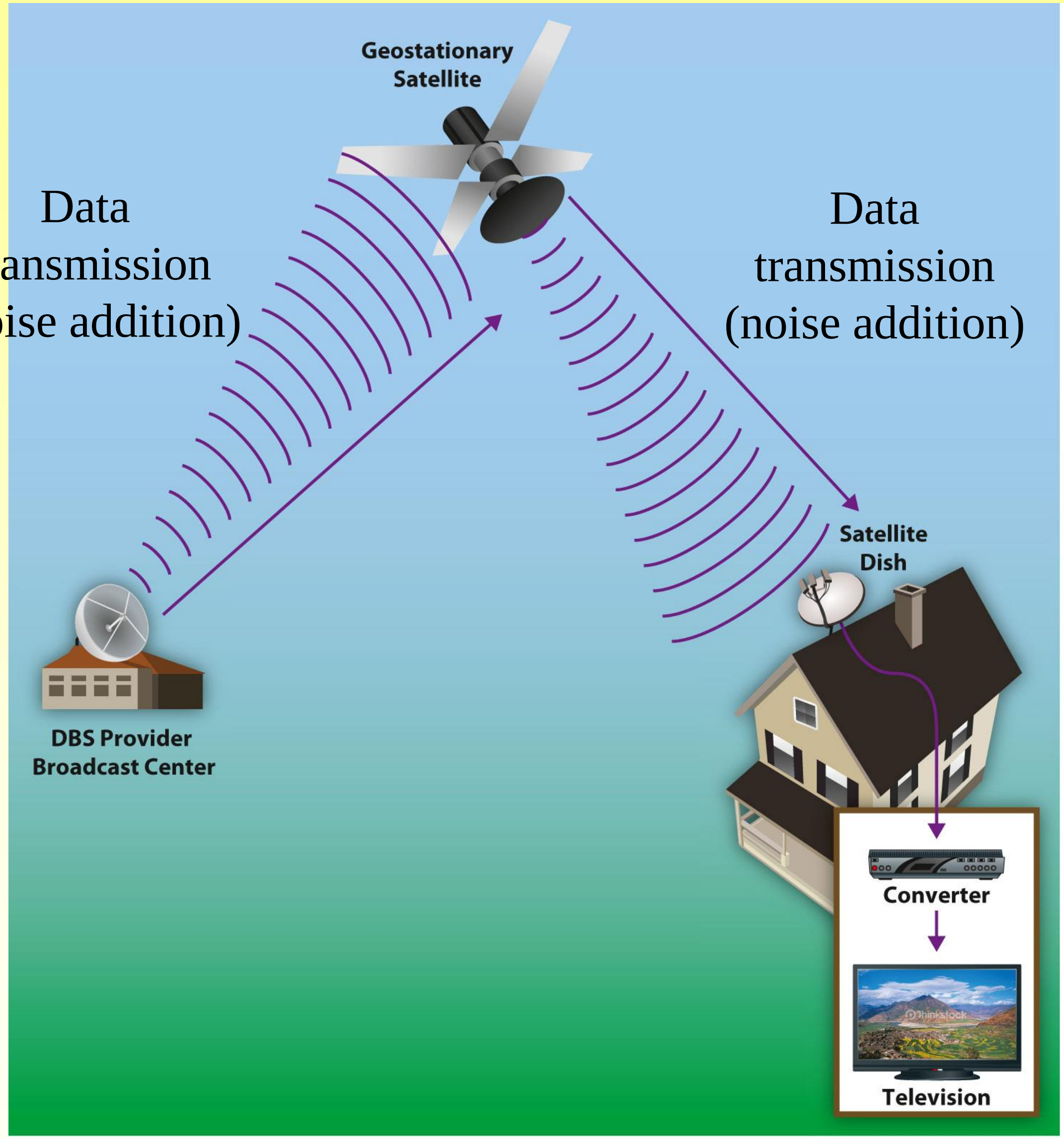
What is information?

- Is an abstract concept
- Is measure of how much ignorance can be removed
- Can be quantified

What is information?

Data codification
(Compression)

Data
transmission
(noise addition)

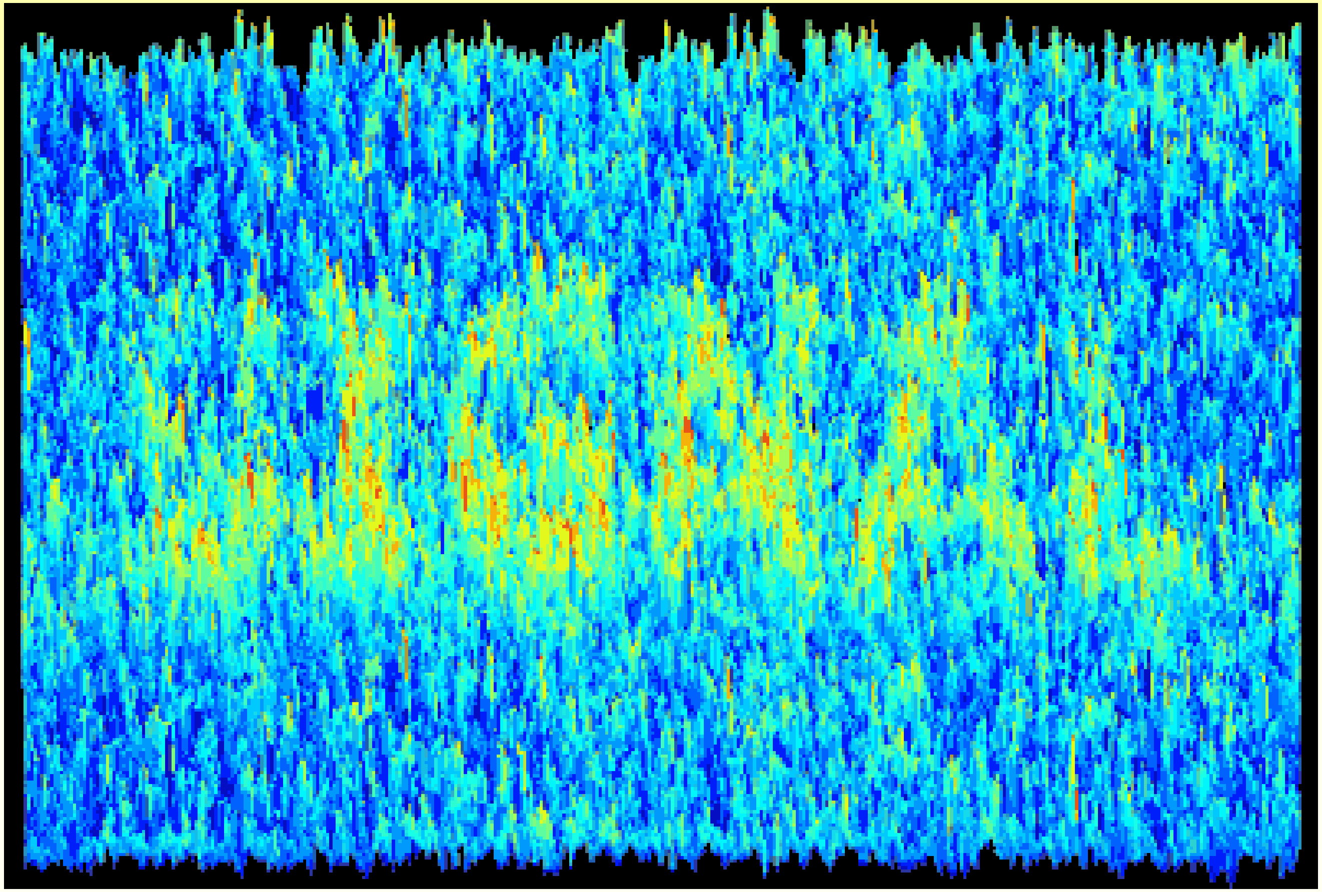


Data
transmission
(noise addition)

Data de-codification
(De-compression)
Noise removal

What is information?

Information processing requires distinguishing useful signals from noise: concept of *signal-to-noise ratio*



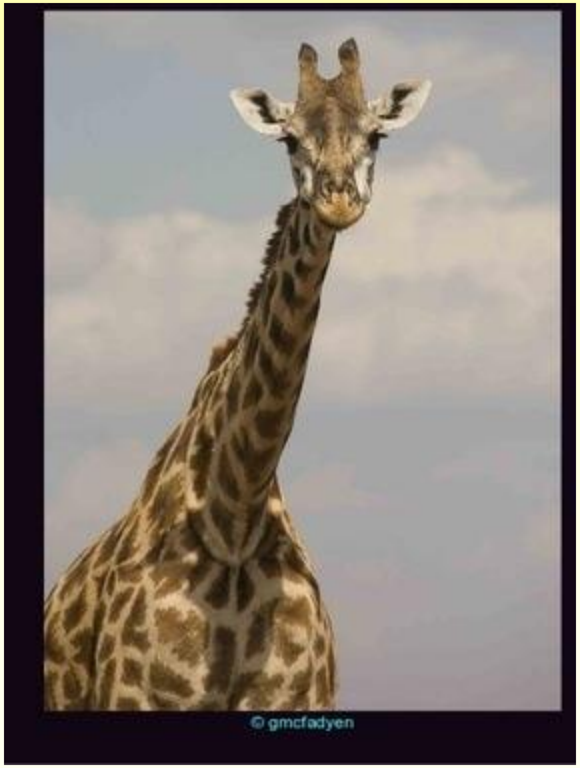
What is information?

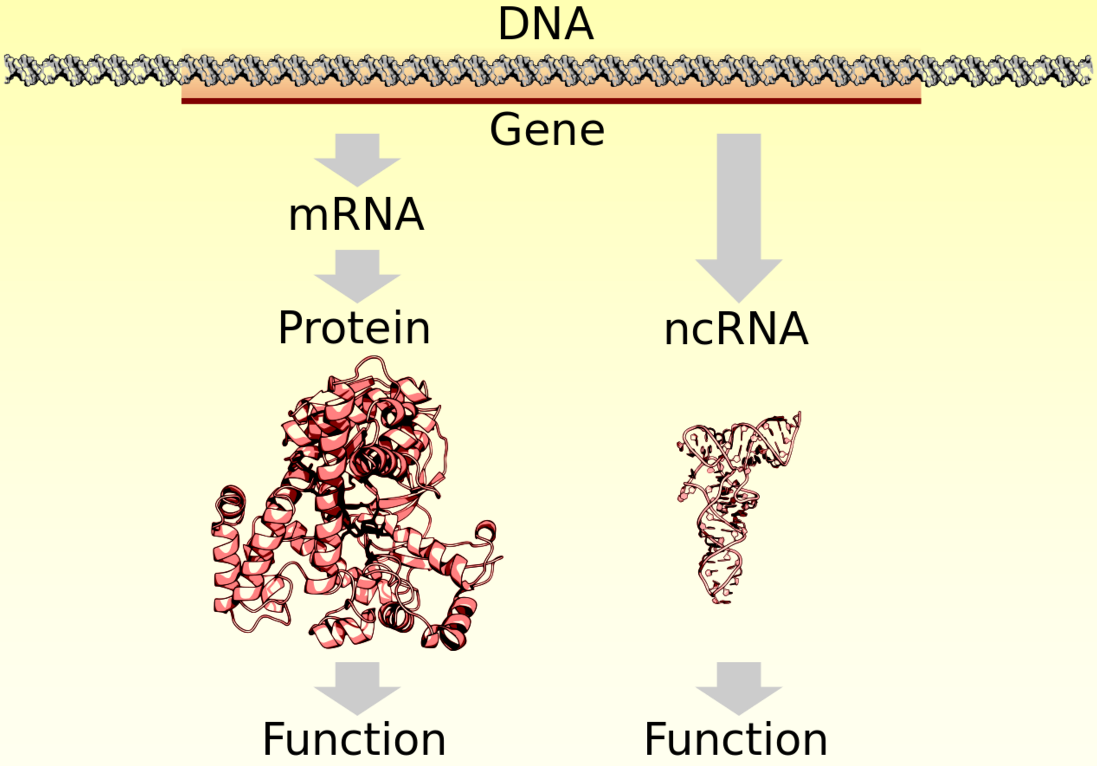
Sensory systems in Biology: The *efficent coding hypothesis*

Evolution of sense organs and brain are driven by the need to minimize the energy expended in acquiring information

What is information?

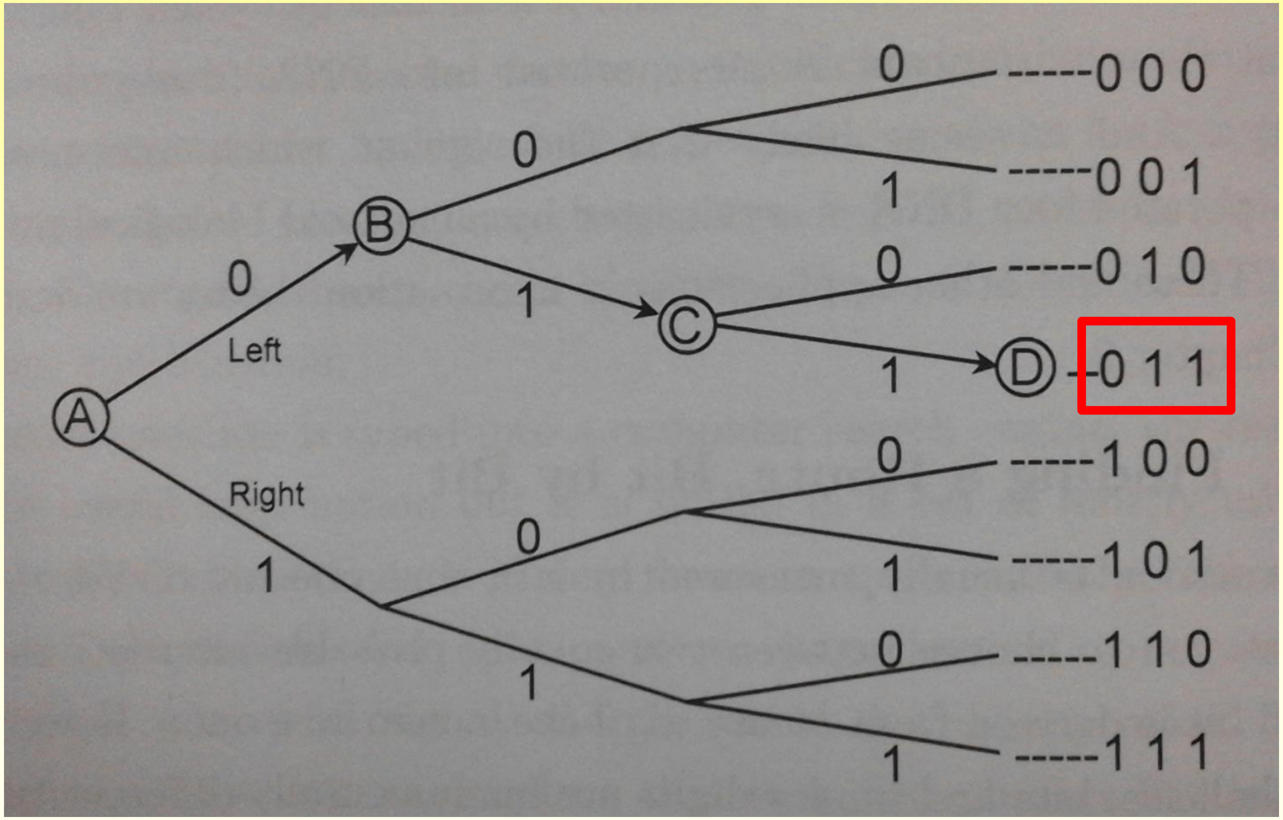
Natural selection (adaptation) implies encoding environmental information in the genetic code (DNA)





What is information?

Finding a route



One bit is the amount of information required to choose between **two equally probable alternatives**

What is information?

Finding a route

1 bit $\rightarrow$ 2 alternatives

2 bit $\rightarrow$ 4 alternatives

3 bit $\rightarrow$ 8 alternatives

.

.

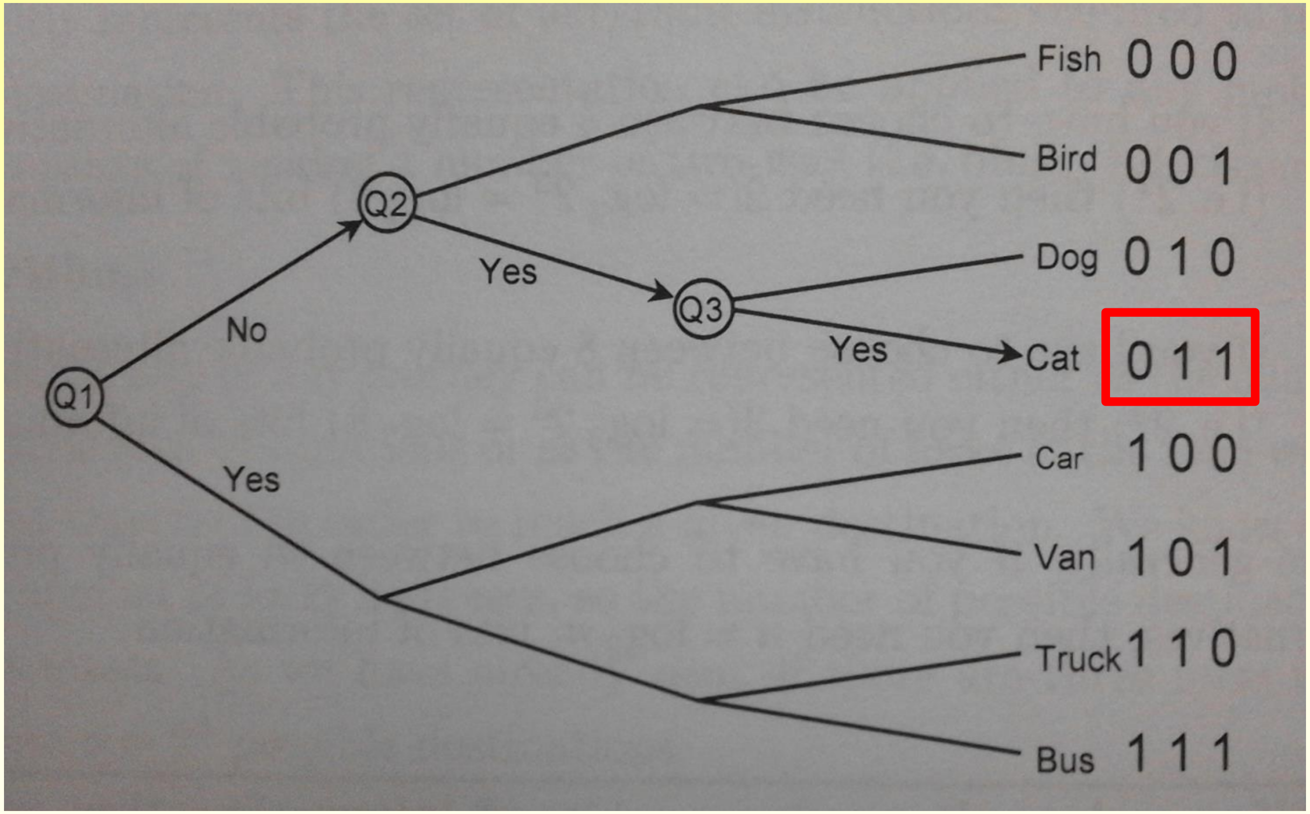
.

n bit $\rightarrow$ 2^n alternatives

To choose from m equiprobable alternatives we need $n = log_2 m$ bits

What is information?

Finding an answer



n questions that can be answered by ‘yes’ or ‘no’ can be used to choose from 2^n words: if $n = 20 \rightarrow$ *1,048,576 words*

What is information?

Finding an answer

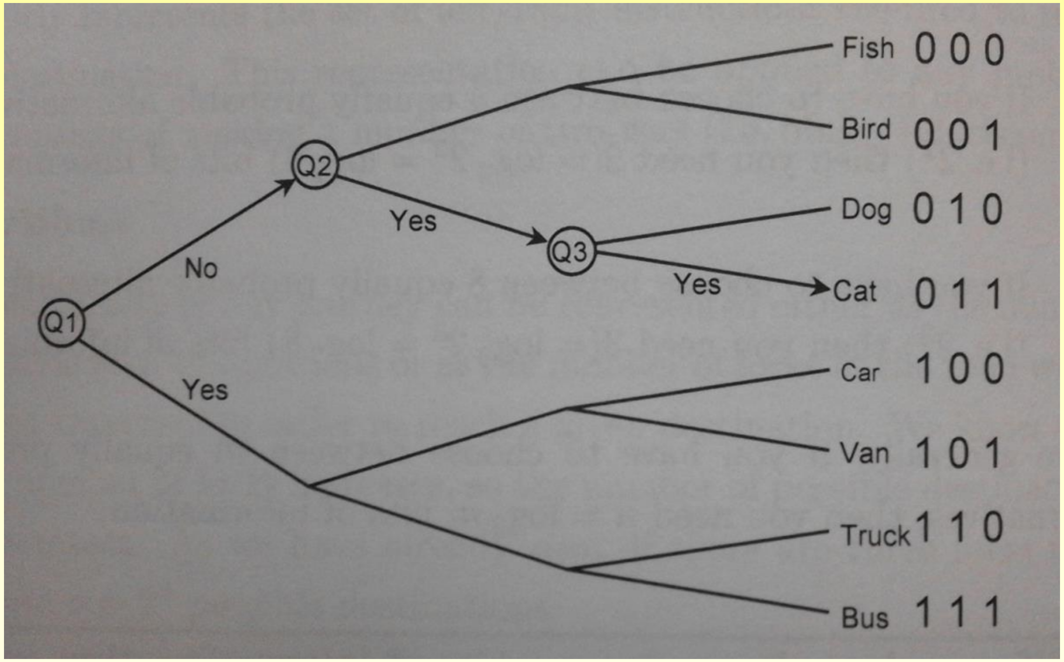
In a "game" of choices, the most useful questions are:

- those which can halve the number of choices left

- those for which you have no idea about the answer

What is information?

Difference between *<u>bi</u>*nary digi*<u>t</u>* and bit of information

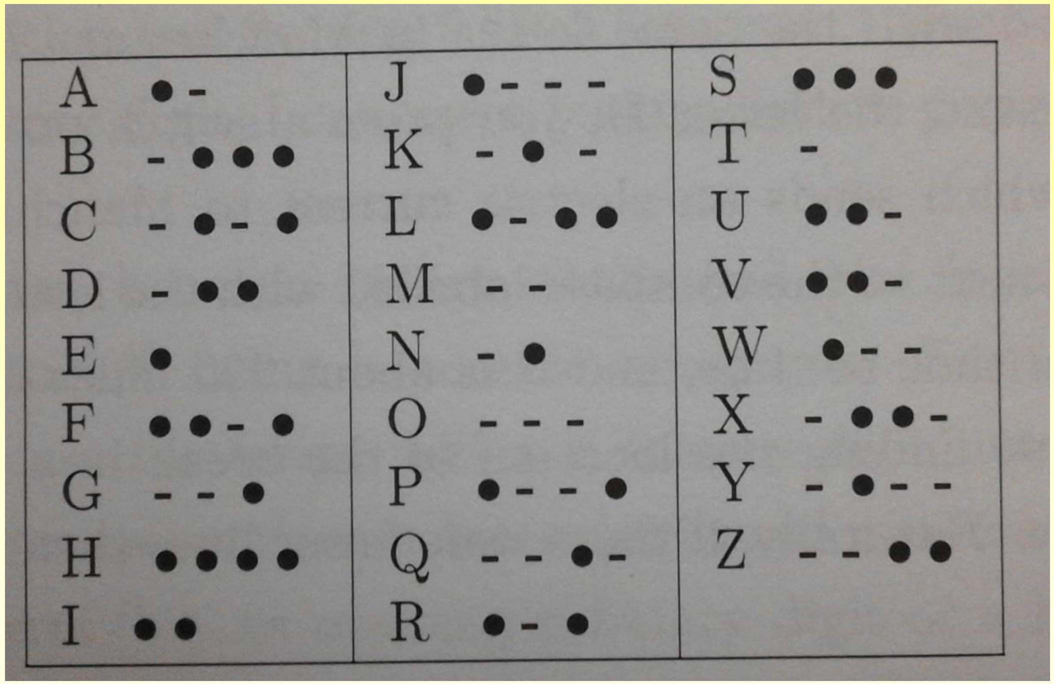


- If you already know the answer to the question: 1 binary digit given = 0 bit of information

- If you have no idea about the answer: 1 binary digit given = 1 bit of information

What is information?

Information coding – Telegraphy, Morse code



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

26 letters are encoded in 26 *codewords*

Frequent letters → shorter codewords

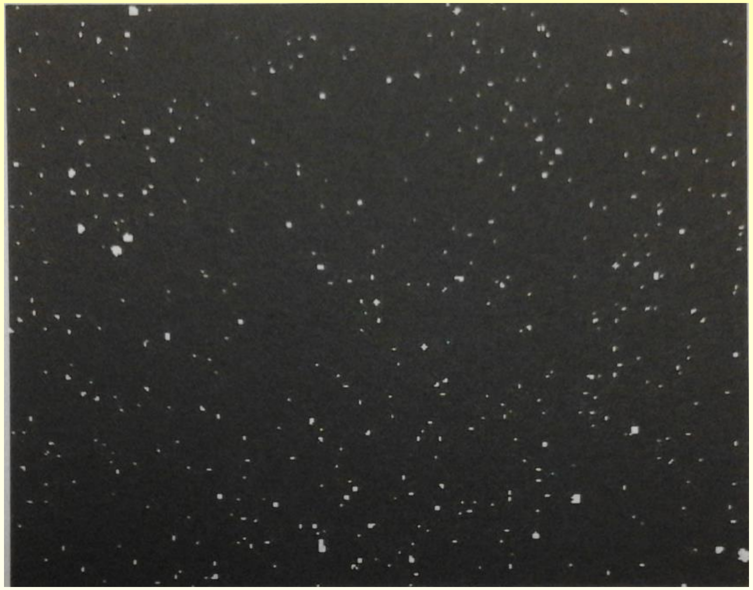
Infrequent letters → longer codewords

→ Optimization of transmission!

Fundamental question of information theory

How can we tell if a communication channel is being used as efficiently as possible?

Black and white image of stars
(100 x 100 pixels)



Coding method 1:

100 x 100 binary digits
black → 0
white → 1

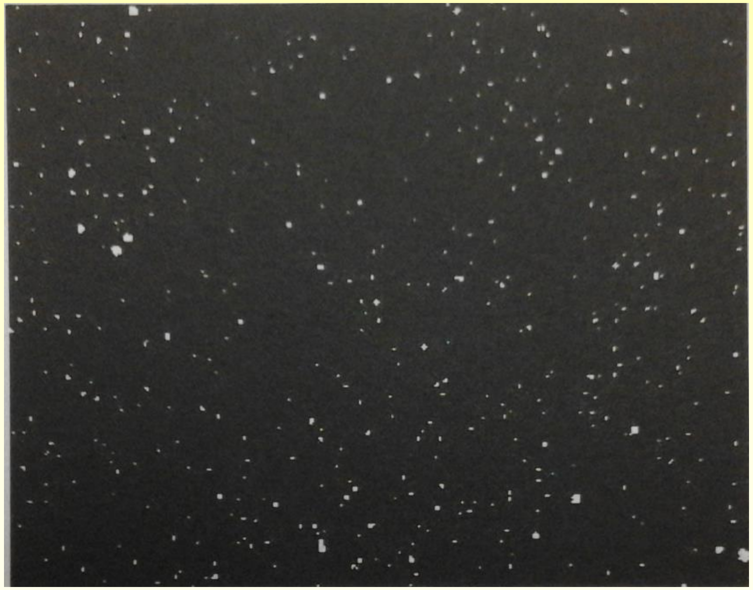
10,000 bits

Very inefficient coding
for this image!

Fundamental question of information theory

How can we tell if a communication channel is being used as efficiently as possible?

Black and white image of stars
(100 x 100 pixels)



Coding method 2:

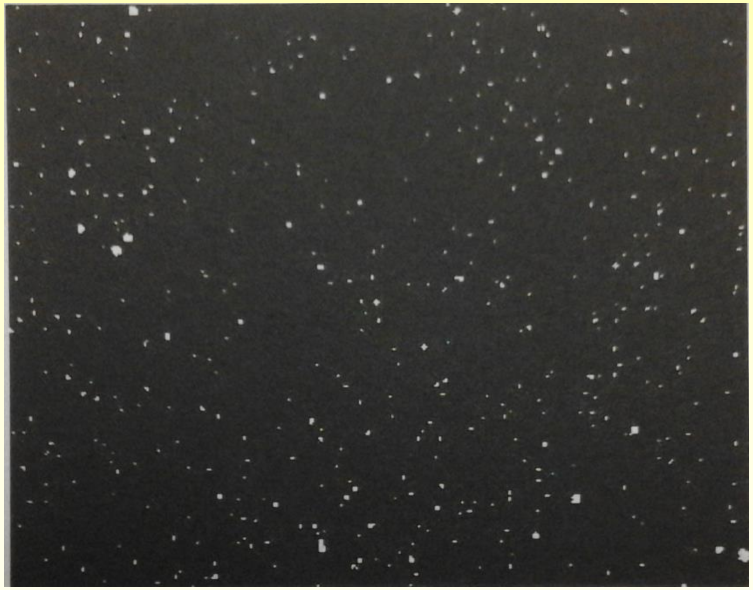
Send location of white pixels

Code: [(19,13),(22,30),...]

Fundamental question of information theory

How can we tell if a communication channel is being used as efficiently as possible?

Black and white image of stars
(100 x 100 pixels)



Coding method 3:

Concatenate the rows

Send the position of the white pixels

Code: [13,9,...]

The choice of the best method will depend of the type of image

Fundamental question of information theory

How can we tell if a communication channel is being used as efficiently as possible?

Black and white image
(100 x 100 pixels)

Coding method 3:

Concatenate the rows

Send the position of the white pixels

Too many consecutive white pixels!

Better to send the positions where the color changes

Fundamental question of information theory

How can we tell if a communication channel is being used as efficiently as possible?

Grey-level image
(100 x 100 pixels)

Color coded in 256 grey levels

$log_2 256 = 8$ bits/pixel

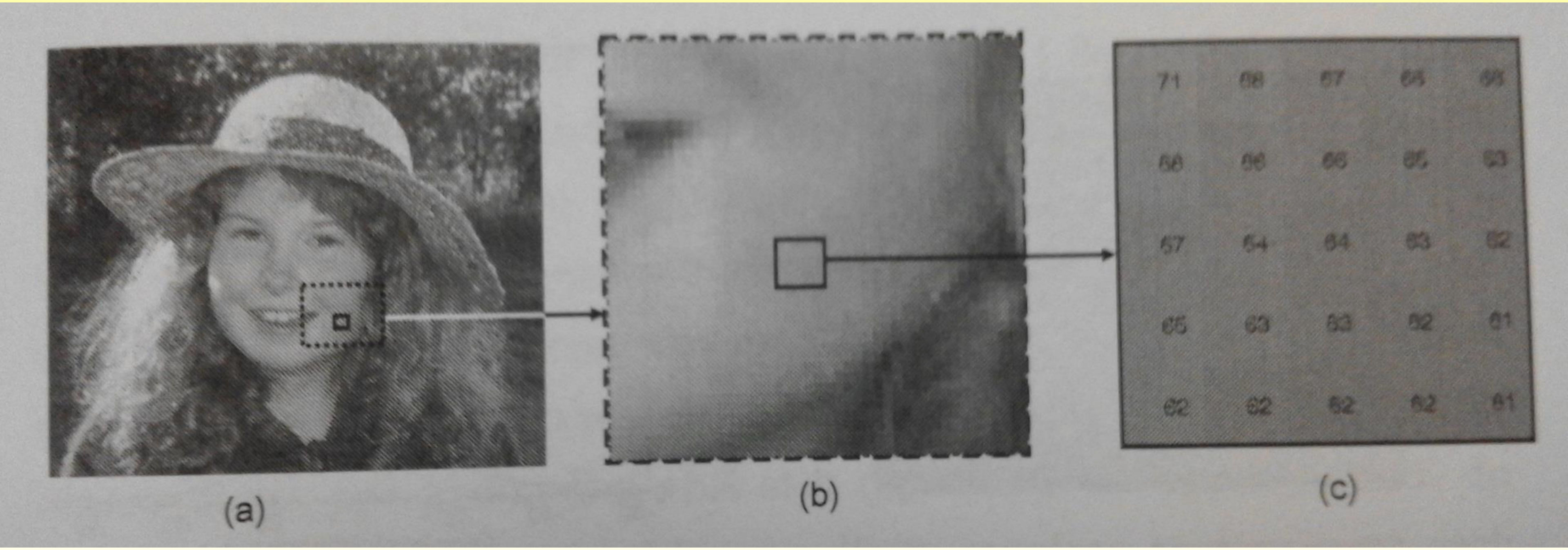
80,000 bits per image!

Can we make it better?

Notice the *redundancy* in the image

Fundamental question of information theory

How can we tell if a communication channel is being used as efficiently as possible?

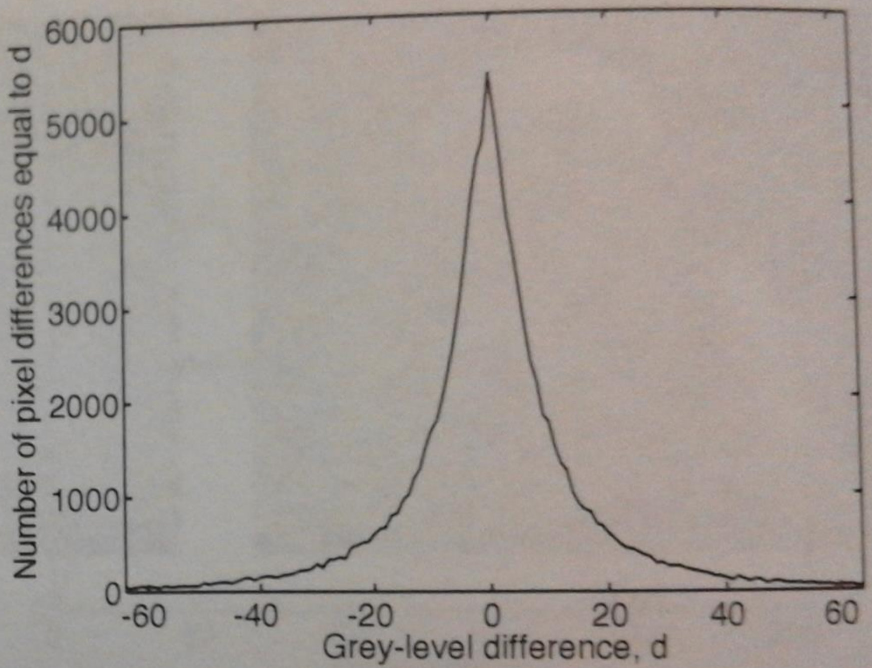


Redundancy in the color levels:
Most grey levels in contiguous pixels are not independent

Fundamental question of information theory

How can we tell if a communication channel is being used as efficiently as possible?

Image reconstructed from the differences in grey level of contiguous pixels

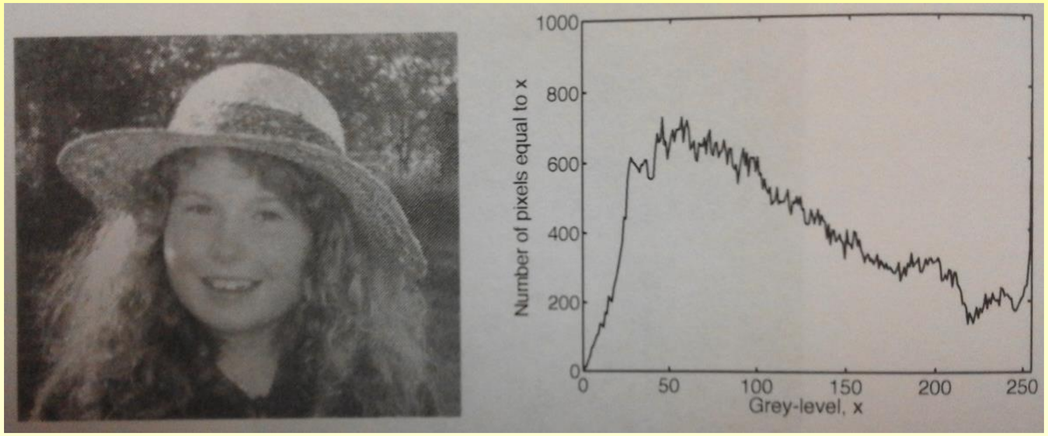


Most values of grey level differences are in a narrow interval

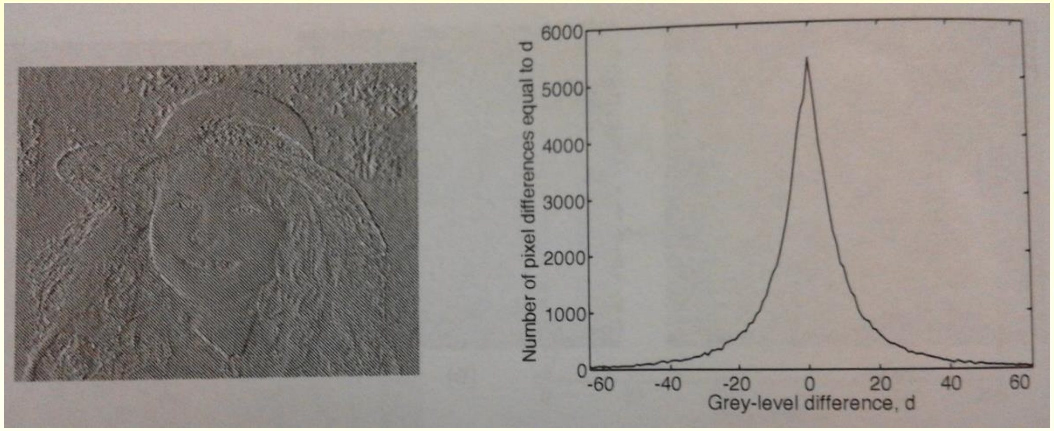
Most information can be encoded in 127 values → 7 bits/pixel

Fundamental question of information theory

How can we tell if a communication channel is being used as efficiently as possible?



$log_2 256 = 8$ bits/pixel



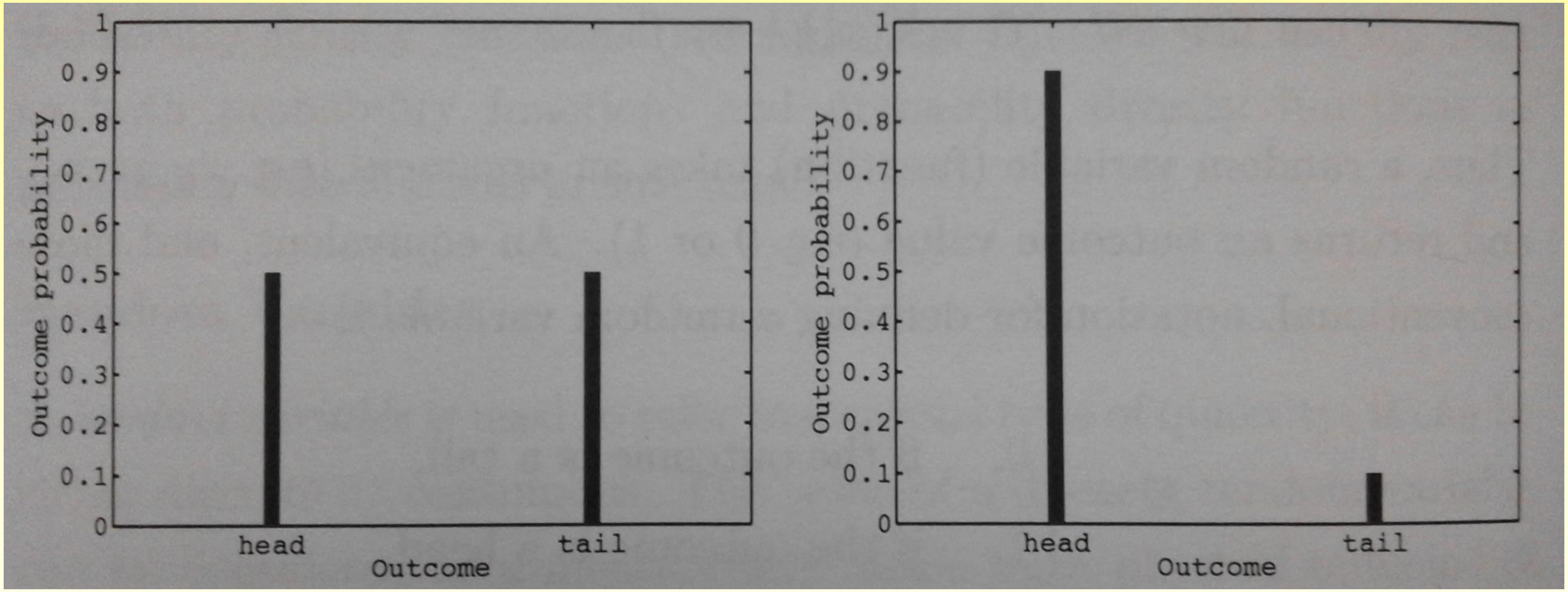
$log_2 127 = 7$ bits/pixel

How much actual <u>information</u> does each pixel contain?

How to measure information?

Flipping coins:

Unbiased coin (50-50) Biased coin (90-10)



Unexpected → Informative Expected → Not informative

Information should be inversely proportional to the expectancy: $h(x) \sim 1/p(x)$

How to measure information?

Mathematical properties of Shannon Information:

- **Continuity**: continuous function of the probability of possible outcomes

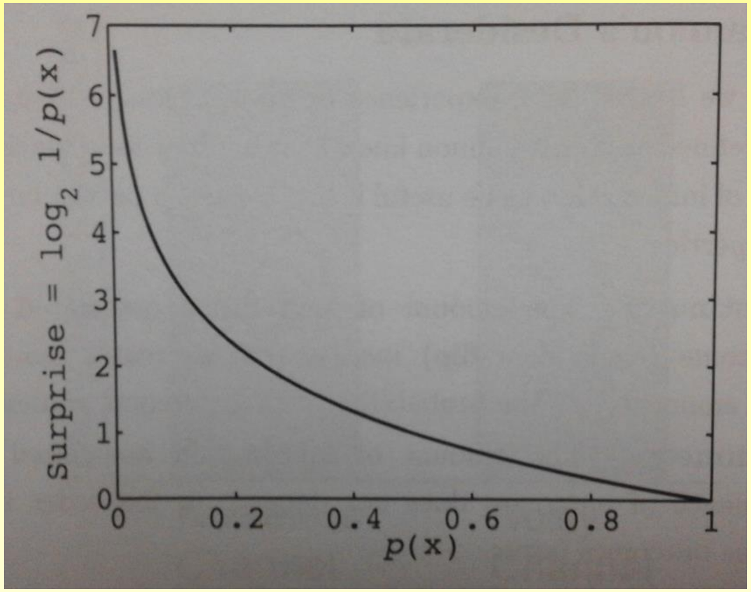
- **Additive**: the information associated with a set of outcomes is obtained by adding the information of individual outcomes

- **Symmetry**: the information associated with a sequence of outcomes does not depend on the order in which those outcomes occur

- **Maximal value**: information is maximal for outcomes that occur with equal probability

How to measure information?

Properties of Shannon Information



$h(x) \sim 1/p(x) \rightarrow h(x) = log_2(1/p(x))$ bits
$= -log_2(p(x))$ bits

But we are actually interested in the average information contained in a set of possible values

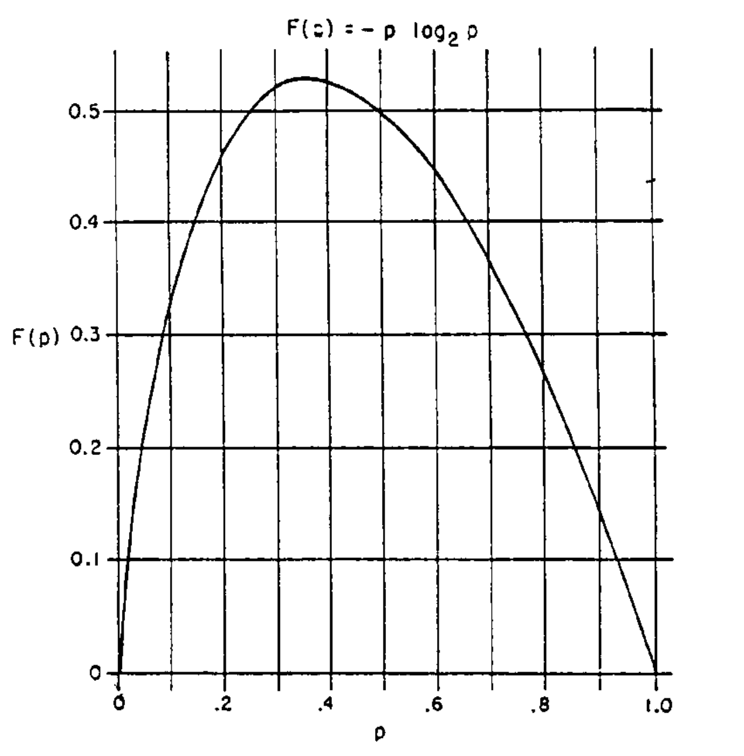
$$H(X) = \sum_{i=1}^{n} p(x_i) log_2 \frac{1}{p(x_i)}$$

Shannon Entropy

How to measure information?

Properties of Shannon Entropy:

$$H(X) = \sum_{i=1}^{n} p(x_i) log_2 \frac{1}{p(x_i)}$$



Entropy is always be larger than or equal to zero

An event with small probability has small contribution to total uncertainty

The entropy of an experiment is only zero if one of the probabilities equals 1

For the case of equally probable events, the Shannon entropy reduces to Hartley's entropy

The uncertainty is largest for events with equal probability: $p(x_i) = 1/n$; $H_{max} = \log n$

Entropy can be interpreted as the average surprise value of the different outcomes

Entropy

Information theory:

$$\boldsymbol{H(X) = \sum_{i=1}^{n} p(x_i) log_2 \frac{1}{p(x_i)}}$$ (Shannon)

in case of equal probabilities:

$H = \log n$ (Hartley)

Uncertainty is related to number of possibilities

Thermodynamics: $S = k \log W$ (Boltzmann)

(S = entropy, W number of possible microscopic states, k Boltzmann constant)

Entropy is related to number of different possible states

In thermodynamics as well as in information theory:

Entropy is related to disorder, uncertainty, number of possible states

Entropy

Uncertainty of experiment with 2 possible outcomes as function of p

Unbiased coin

$H = -0.5\log_2 0.5 - 0.5\log_2 0.5$

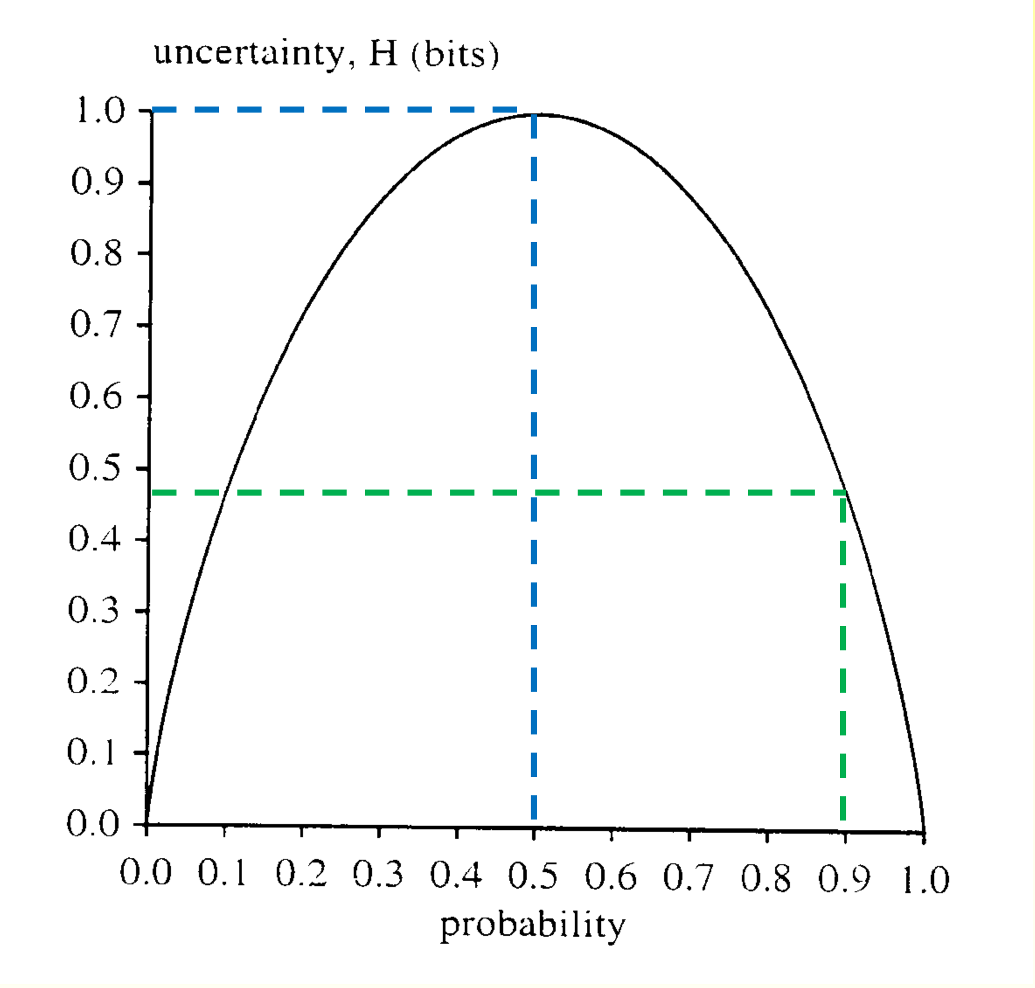
$H = 1$ bit

Biased coin

$H = -0.9\log_2 0.9 - 0.1\log_2 0.1$

$H = 0.469$ bit

The biased coin is like an unbiased coin with $2^{0.469} = 1.38$ sides



Entropy

For equi-probable outcomes the Shannon entropy of a variable X is the logarithm of the number *m* of outcomes of X

$$H(X) = \log_2 m \text{ bits}$$

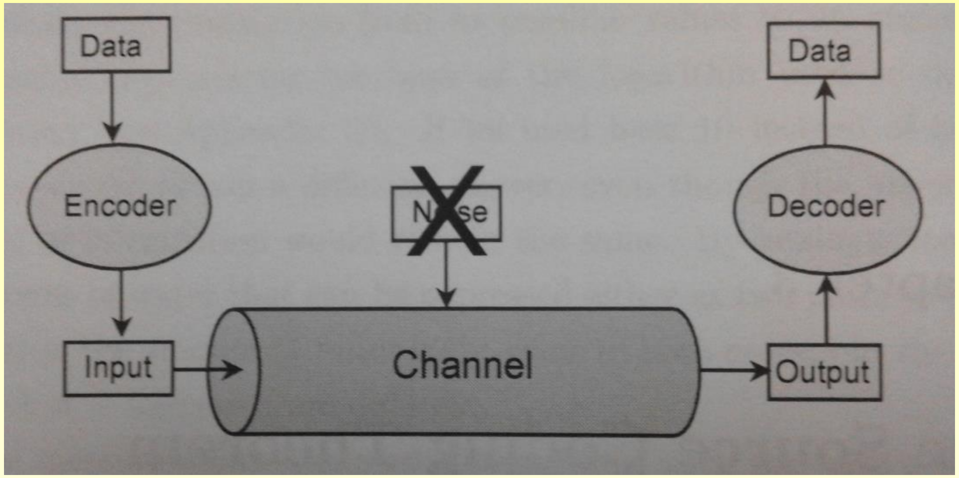
A variable with an entropy of $H(X)$ provides enough Shannon information to choose between $m = 2^{H(X)}$ equi-probable outcomes

The average uncertainty of a variable X is given by its entropy $H(X)$

If we are told the value X, the amount of information given is $H(X)$

Fundamental question of information theory

How can we tell if a communication channel is being used as efficiently as possible?



Most natural signals contain information in a diluted form
Example: contiguous “pixels” tend to have similar values

For efficient communication (coding):
- Inputs should be transformed to signals with independent values
- The transformed signal should have a distribution optimized for the particular channel

Shannon Source Coding Theorem

Channel capacity: C = # of information bits communicated per second

But! The amount of bits that can be encoded in binary digits depend on the coding method

The **Shannon Source Coding Theorem** established that *for every channel there is a* *<u>coding method</u>* *for which it is possible to transmit at an average of* $C/H - \varepsilon$ *symbols per second, where* ε *is arbitrarily small.*

Shannon Source Coding Theorem

Channel capacity: C = # of information bits communicated per second

But! The amount of bits that can be encoded in binary digits depend on the coding method

The full capacity of a channel is utilized if the source is encoded in such a way that each transmitted binary digit represents an average of one bit of information

Shannon Source Coding Theorem

Examples of different performance of a code

8-sided die

Symbol	Codeword
$s_1 = 1$	$x_1 = 000$
$s_2 = 2$	$x_2 = 001$
$s_3 = 3$	$x_3 = 010$
$s_4 = 4$	$x_4 = 011$
$s_5 = 5$	$x_5 = 100$
$s_6 = 6$	$x_6 = 101$
$s_7 = 7$	$x_7 = 110$
$s_8 = 8$	$x_8 = 111$

H = $\log_2 8$ = 3 bits/symbol

6-sided die

Symbol	Codeword
$s_1 = 1$	$x_1 = 000$
$s_2 = 2$	$x_2 = 001$
$s_3 = 3$	$x_3 = 010$
$s_4 = 4$	$x_4 = 011$
$s_5 = 5$	$x_5 = 100$
$s_6 = 6$	$x_6 = 101$

H = $\log_2 6$ = 2.58 bits/symbol

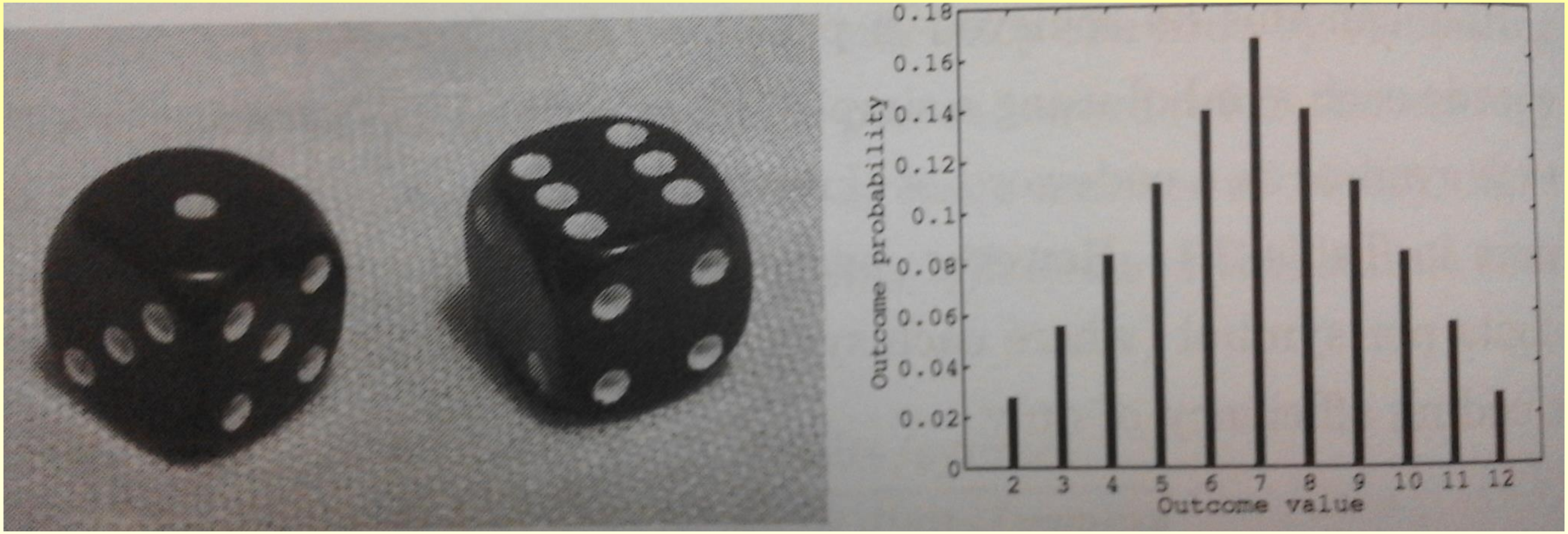
Lenght = 3 binary digit/symbol

Coding efficiency
H/Lenght = 1 bits/binary digit

Coding efficiency
H/Length = 0.86 bits/binary digit

Data compression

Throw of 2 6-sided dice



H = 3.27 bits/symbol

3 binary digits are not enough to code all outputs

4 binary digits are too many and give a coding efficiency
H/Lenght = 0.818 bits/binary digit

Data compression

Throw of 2 6-sided dice

Symbol	Sum	Dice	Freq	p	h	Code x
s_1	2	1:1	1	0.03	5.17	10000
s_2	3	1:2, 2:1	2	0.06	4.17	0110
s_3	4	1:3, 3:1, 2:2	3	0.08	3.59	1001
s_4	5	2:3, 3:2, 1:4, 4:1	4	0.11	3.17	001
s_5	6	2:4, 4:2, 1:5, 5:1, 3:3	5	0.14	2.85	101
s_6	7	3:4, 4:3, 2:5, 5:2, 1:6, 6:1	6	0.17	2.59	111
s_7	8	3:5, 5:3, 2:6, 6:2, 4:4	5	0.14	2.85	110
s_8	9	3:6, 6:3, 4:5, 5:4	4	0.11	3.17	010
s_9	10	4:6, 6:4, 5:5	3	0.08	3.59	000
s_{10}	11	5:6, 6:5	2	0.06	4.17	0111
s_{11}	12	6:6	1	0.03	5.17	10001

$\langle L \rangle$ = Sum $p(x_i)L(x_i)$ = 3.31

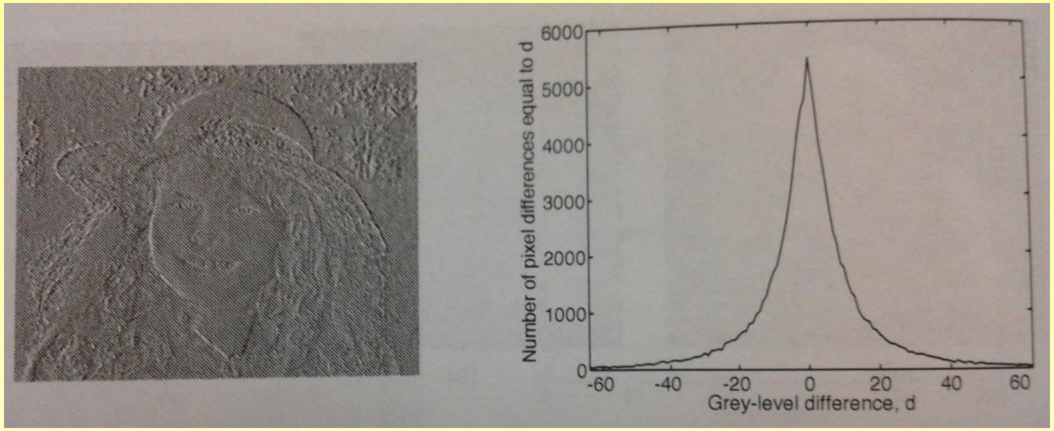
Then the coding efficiency: H/L = 3.27/3.31 bits/binary digit

= 0.99 bits/binary digit

Optimal prefix code: Huffman code
Principle of Huffman code

Source Character	$P(a_i)$	$P(a_i')$	$P(a_i'')$	$P(a_i''')$	Code Word
a_1	0.3	0.3	0.45	1 0.55	11
a_2	0.25	0.25	1 0.3	0 0.45	10
a_3	0.25	1 0.25	0 0.25		01
a_4	1 0.1	0 0.2			001
a_5	0 0.1				000
E.g. a_4	1	0	-	0	= 001

Huffman code for grey-level images



$log_2 127 = 7$ bits/pixel

If we use the histogram to feed the Huffman coding we obtain:

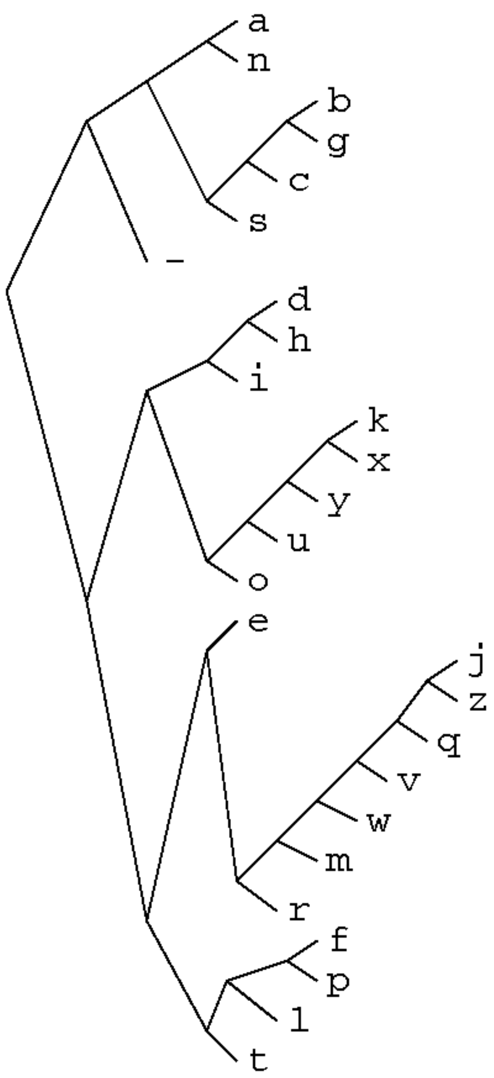
$L = \text{Sum } p(x_i)L(x_i) = 5.97$

Then the coding efficiency: H/L = 5.94/5.97 bits/binary digit

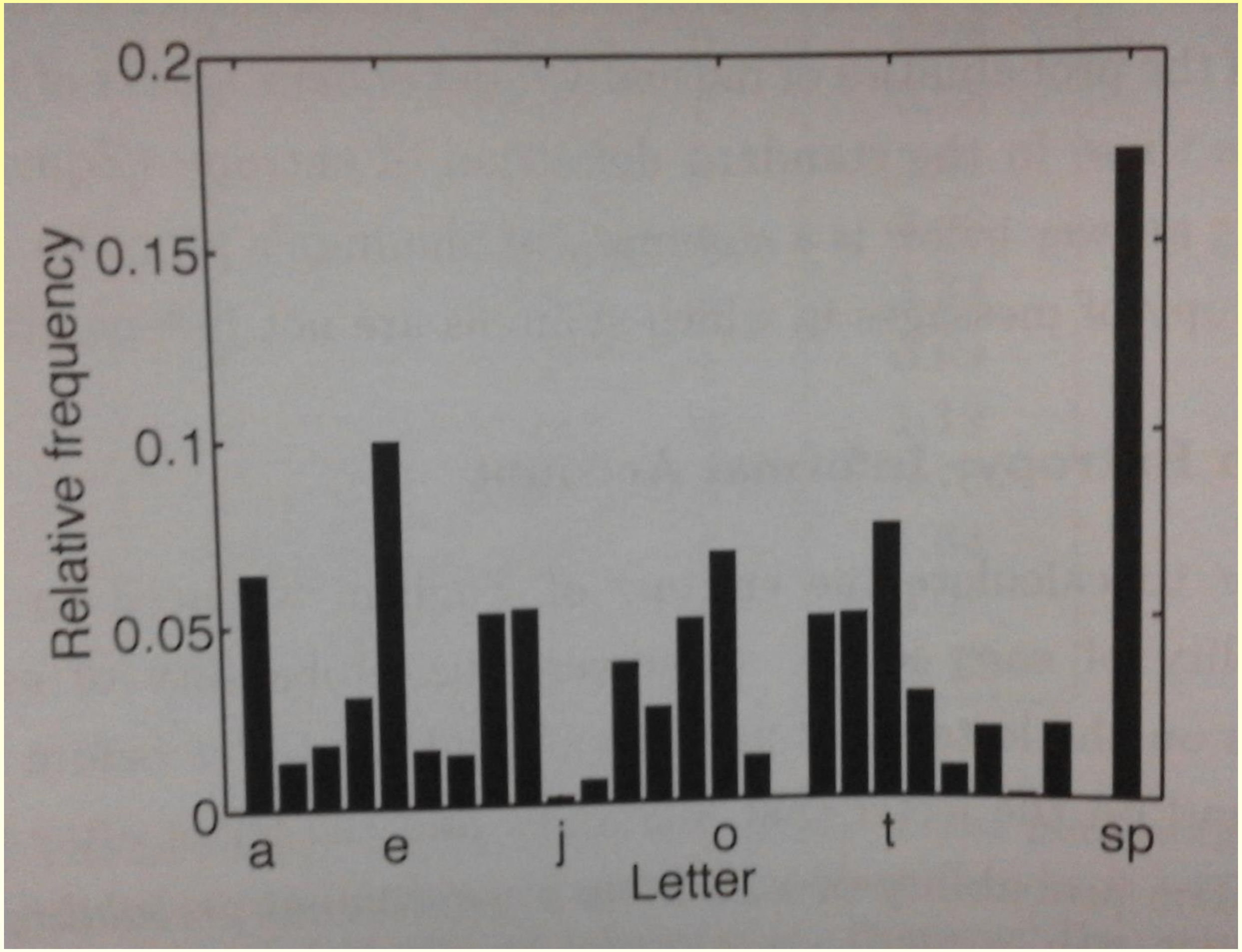
= 0.995 bits/binary digit

Huffman code for English

a_i	p_i	$\log_2 \frac{1}{p_i}$	l_i	$c(a_i)$
a	0.0575	4.1	4	0000
b	0.0128	6.3	6	001000
c	0.0263	5.2	5	00101
d	0.0285	5.1	5	10000
e	0.0913	3.5	4	1100
f	0.0173	5.9	6	111000
g	0.0133	6.2	6	001001
h	0.0313	5.0	5	10001
i	0.0599	4.1	4	1001
j	0.0006	10.7	10	1101000000
k	0.0084	6.9	7	1010000
l	0.0335	4.9	5	11101
m	0.0235	5.4	6	110101
n	0.0596	4.1	4	0001
o	0.0689	3.9	4	1011
p	0.0192	5.7	6	111001
q	0.0008	10.3	9	110100001
r	0.0508	4.3	5	11011
s	0.0567	4.1	4	0011
t	0.0706	3.8	4	1111
u	0.0334	4.9	5	10101
v	0.0069	7.2	8	11010001
w	0.0119	6.4	7	1101001
x	0.0073	7.1	7	1010001
y	0.0164	5.9	6	101001
z	0.0007	10.4	10	1101000001
–	0.1928	2.4	2	01



Huffman code for English



Entropy of written English: H_0

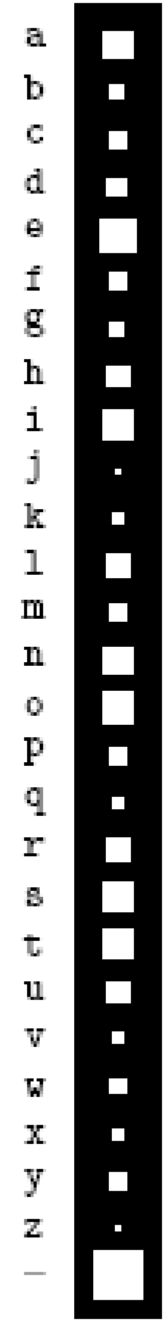
$$H_0 = \log 27 = 4.75 \text{ bit}$$

XFOML RHKHJFFJUJ ZLPWCFWCKCYJ
FFJEYVKCQSGHYD QPAAM KBZAACIBZLHJQD

English

Probabilities: p(x)

i	a_i	p_i
1	a	0.0575
2	b	0.0128
3	c	0.0263
4	d	0.0285
5	e	0.0913
6	f	0.0173
7	g	0.0133
8	h	0.0313
9	i	0.0599
10	j	0.0006
11	k	0.0084
12	l	0.0335
13	m	0.0235
14	n	0.0596
15	o	0.0689
16	p	0.0192
17	q	0.0008
18	r	0.0508
19	s	0.0567
20	t	0.0706
21	u	0.0334
22	v	0.0069
23	w	0.0119
24	x	0.0073
25	y	0.0164
26	z	0.0007
27	–	0.1928



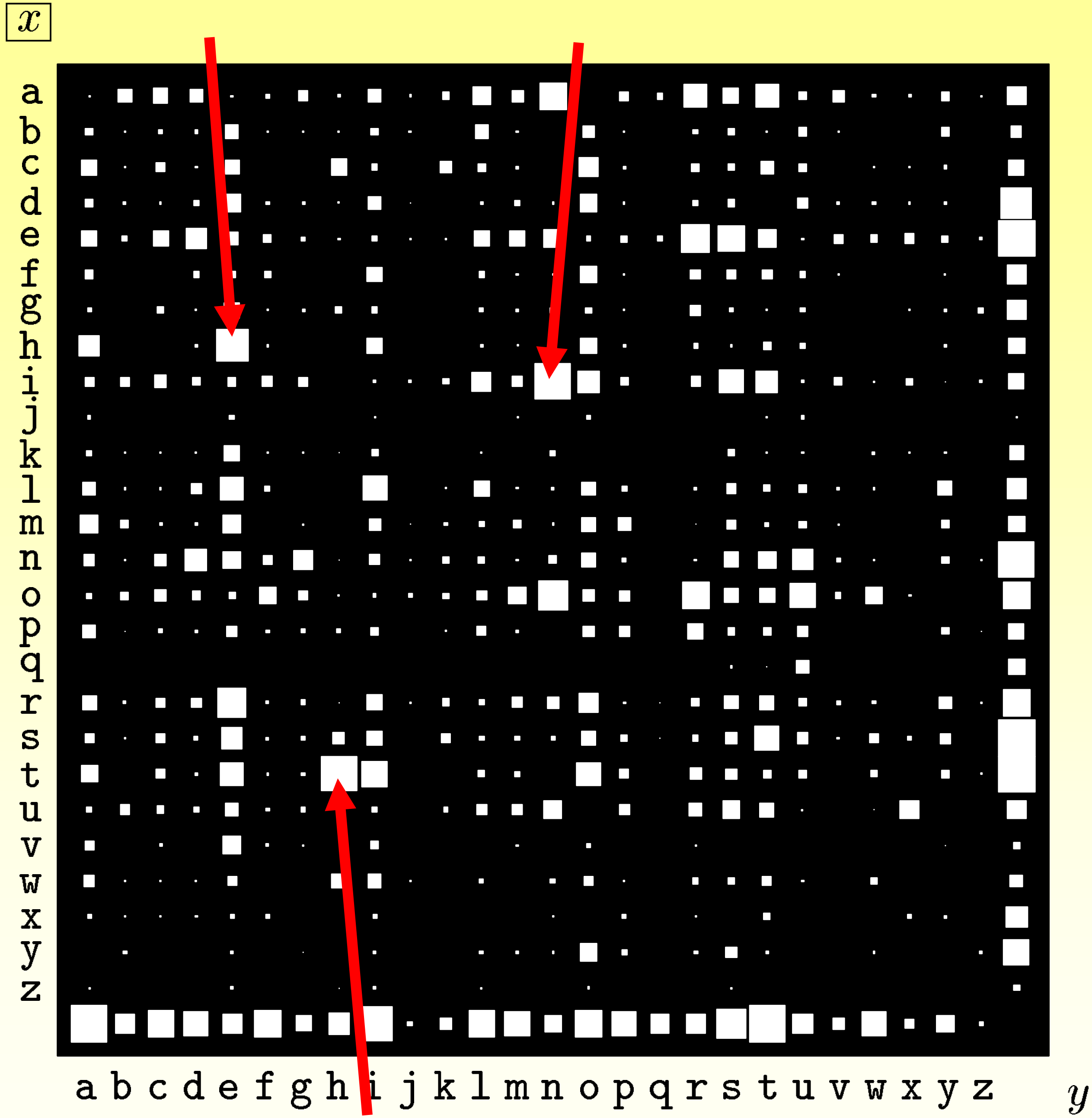
Entropy of written English: H_1

$$H_1 \equiv H(A_1) = 4.03 \text{ bit}$$

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

English

Bigram probabilities: p(xy)



Entropy of written English: H_2

$$H_2 \equiv H(A_2|A_1) = H(A_1 A_2) - H(A_1)$$

Entropy of written English: H_2

$$H_2 \equiv H(A_2|A_1) = H(A_1A_2) - H(A_1)$$

ON IE ANTSOUTINYS ARE T INCORE ST BE S DEAMY ACHIN D ILONASIVE TOCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTSIBE.

Entropy of written English: H_3

$$H_3 \equiv H(A_3|A_1A_2) = H(A_1A_2A_3) - H(A_1A_2)$$

IS NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTION A OF CRE

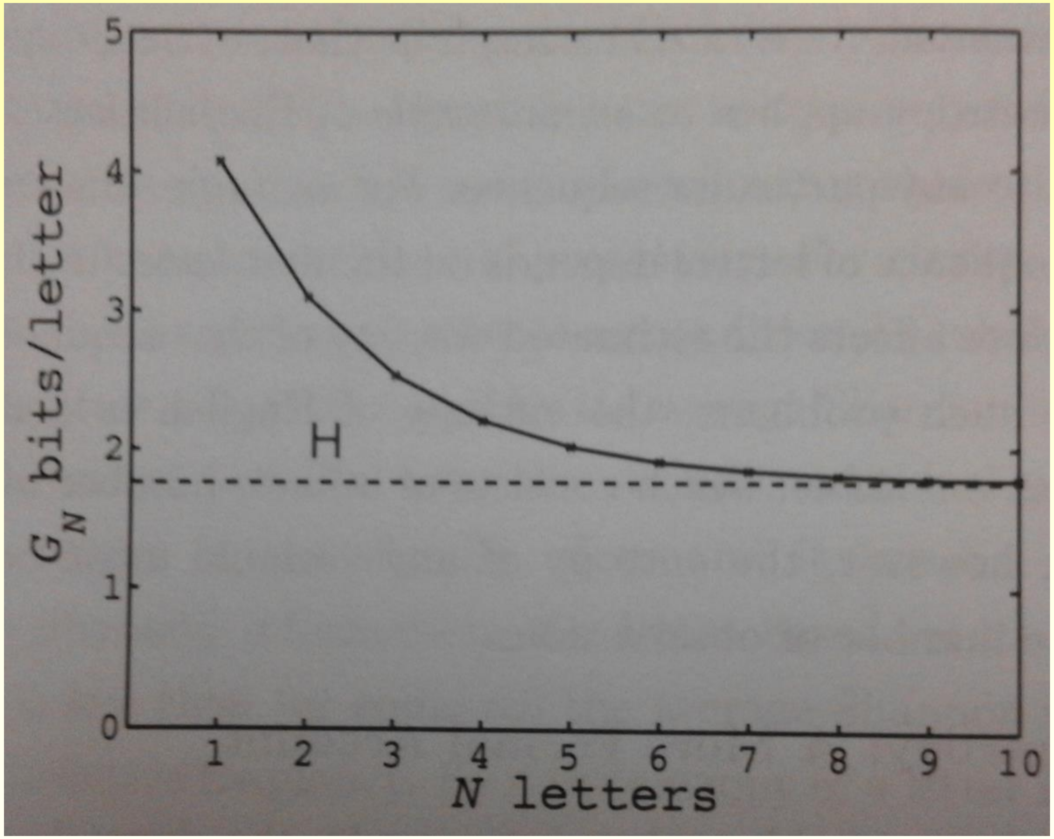
Entropy of written English: H_4

$$H_4 \equiv H(A_4|A_1A_2A_3) = H(A_1A_2A_3A_4) - H(A_1A_2A_3)$$

THE GENERATED KBOB PROVIDUAL BETTER TRAND THE DISPLAYED CODE. ABOVERY UPONDULTS WELL THE CODERST IN THESTICAL IT DO HOCK BOTHE MERG. (INSTATES CONS ERATION. NEVER ANY OF PUBLE AND TO THEORY. EVENTIAL CALLEGAND TO LEAST BENERATED IN WITH PIES AS IS WITH THE)

Entropy of written English: H_N and H_∞

$$H_N \equiv H(A_N|A_1A_2...A_{N-1}) = H(A_1A_2...A_{N-1}A_N) - H(A_1A_2...A_{N-1})$$



Uncertainties in English

H_0	H_1	H_2	H_3	H_5	H_8	H_∞
4.75	4.03	3.32	3.10	2.16	1.86	1.33 bit

A character in English texts thus contains not much more than 1 bit of information!

The redundancy $R \simeq 72\%$

Th_r_ _s _nly _n_ w_y t_ f_ll _n th_ v_w_ls _n th_s s_nt_nc_.

Example of redundancy in Dutch

Vlgones een oznrdeeok van een Eglnese uvinretsiet mkaat het neit uit in wlkee vloogdre de ltteers in een wrood saatn, zlonag de ersete en de latsate ltteer maar op de jiutse patals saatn. De rset van de ltteers mgoen wllikueirg gpletaast wdoren, je knut gwoeon lzeen wat er saatt. Dit kmot odmat we neit ekle ltteer arpat lzeen maar het wrood als gheeel. Wdooren als “arngchstseeuw", waar in nrolemale omdigstasnheden ahct mdeeklikners elakar oplogevn, lgigen eits meoilijker.

(Volgens een onderzoek van een Engelse universiteit maakt het niet uit in welke volgorde de letters in een woord staan, zolang de eerste en de laatste letter maar op de juiste plaats staan. De rest van de letters mogen willekeurig geplaatst worden, je kunt gewoon lezen wat er staat. Dit komt omdat we niet elke letter apart lezen maar het woord als geheel. Woorden als “angstschreeuw", waar in normale omstandigheden acht medeklinkers elkaar opvolgen, liggen iets moeilijker.

Uncertainties and redundancy of different languages

In bits	Samoa	English	Old Russian	Hebrew
H_0	4.085	4.754	5.169	4.4
H_1	3.402	4.086	4.548	4.0
H_2	2.684	3.013	3.435	
H_3	1.330	1.330	1.330	
R	0.674	0.72	0.743	