+0.1 16.2.26

+0.1 11.2.26

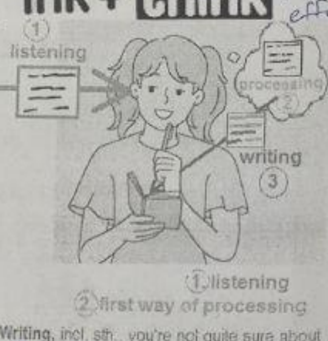
+0, 1

80% chance of rain in Oxford.
This introduces the probabilistic nature of weather predictions.

traditional teaching method

Bad way

A better learning method (more effective)



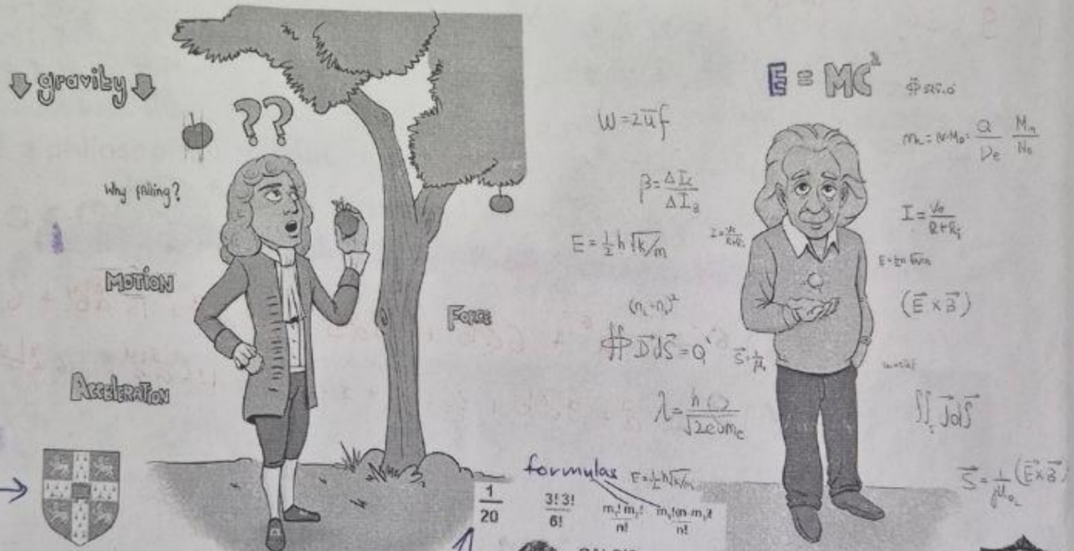
① listening
② first way of processing
③ Writing, incl. sth... you're not quite sure about

Good way

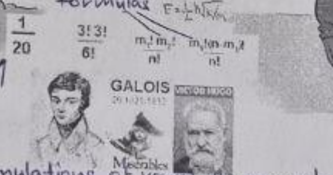
work with concrete

School ⇩gravity⇩ MOTION shift from concrete → abstract ==formalism==> University work with abstract $E = MC^2$ $\oint \vec{D}\, d\vec{S} = Q$ $\iint \vec{J}\, d\vec{S}$

CONCRETE AND ABSTRACT THINKING



formulas



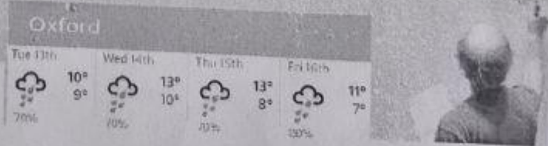
permulations of vowels & consonants in name GALOIS

ISAAC NEWTON

ALBERT EINSTEIN

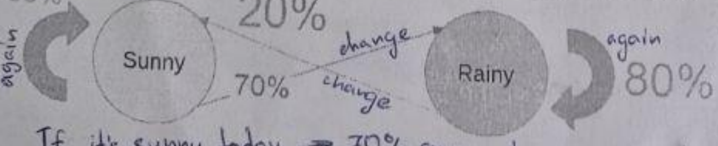
Motivation: 80% chance of rain. Suppose the events A_j ...

Let A_j be the event of rain at 9am ... independently on day j of this term, $1 \le j \le n$



Markoff Chain Probability Model

for Oxford Weather → The probability of tomorrow's weather depends only on today's



If it's sunny today → 70% sunny tomorrow, 20% rain tomorrow;
If it's rain today → 80% rain tomorrow, 20% sunny tomorrow.

Pascal's triangle is used to quickly find numbers associated with combinations & probabilities in repeated trials (games, coin)

binomial coefficients (how many ways to select k successes from n trials)

$\binom{n}{k}$ (C_k^n)

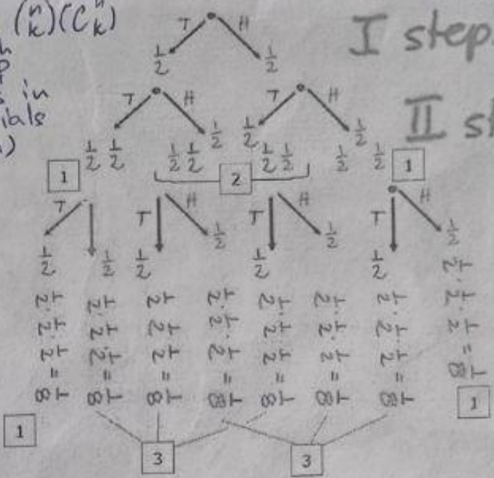
Pascal's triangle

1	1
1 1	2
1 2 1	4
1 3 3 1	8
1 4 6 4 1	16
1 5 10 10 5 1	32
1 6 15 20 15 6 1	64
1 7 21 35 35 21 7 1	128

1 8 28 56 70 56 28 8 1

1 9 36 84 126 126 84 36 9 1

$\boxed{1|2|1}$ – $C_2^0 = 1$, $C_2^1 = 2$, $C_2^2 = 1$



I step

II steps

III steps

Newton's Binomial

– expation of $(a+b)^n$ with coefficients C_k^n, which are taken from Pascal's triangle

$$(a+b)^0 = 1$$
$$(a+b)^1 = a+b$$
$$(a+b)^2 = a^2+2ab+b^2$$
$$(a+b)^3 = a^3+3a^2b+3ab^2+b^3$$
$$(a+b)^4 = a^4+4a^3b+6a^2b^2+4ab^3+b^4$$
$$(a+b)^5 = a^5+5a^4b+10a^3b^2+10a^2b^3+5ab^4+b^5$$
$$(a+b)^6 = a^6+6a^5b+15a^4b^2+20a^3b^3+15a^2b^4+6ab^5+b^6$$
$$(a+b)^9 = a^9+9a^8b+36a^7b^2+84a^6b^3+126a^5b^4+126a^4b^5+84a^3b^6+36a^2b^7+9ab^8+b^9$$

$$\frac{1}{20} = \frac{\overset{\text{Сочи. Лис.}}{3!\,3!}}{\underset{\text{бук.}}{6!}} = \frac{m_1!(m-m_1)!}{m!}$$

+0,1

Events & Probabilities

$\omega \in \Omega$ - sample space

For example:

a) $\Omega = \{H, T\}$ - a coin

b) $\Omega = \{(i;j)\}\ i, j \in [1,2,3,4,5,6]$ - a dice

A subset of Ω is called an event.

1) coin comes up tail $A = \{T\}$

2) we observe of total of 9

$A = \{(3,6); (4;5); (5,4), (6,3)\}$ - events

If $\omega \in \Omega$ is the outcome, we say A occur if $\omega \in A$.

Complement of A

$\bar{A} = A^c$ occurs if A does not occur.

Union: $A \cup B$ occurs if A or B occurs (both)

Intersection: $A \cap B$ occurs if both A and B occurs.

Set differennes: $A \setminus B = \underset{A \text{ occur}}{A} \cap \underset{B \text{ doesn't occur}}{B^c}$

Disjoint: A and B disjoint if $A \cap B = \emptyset$

A and B cannot occur together

We assign a probability $P(A)$ of each A

Simplest case:

Ω is finite and all outcomes

Then $P(A) = \frac{|A|}{|\Omega|}$ Bout a) $\left.\begin{matrix}|\Omega| = 2 \\ |A| = 1\end{matrix}\right\} P(A) = \frac{1}{2}$

b) $\left.\begin{matrix}|\Omega| = 36 \\ |A| = 4\end{matrix}\right\} \Rightarrow P(A) = \frac{4}{36} = \frac{1}{9}$

Elementary combinatorics

Elementary cambinatorics distingwisable objects

Arragements of distingwisable objects

Suppoose we have n distingwisable objects

How many ways are there to order them?

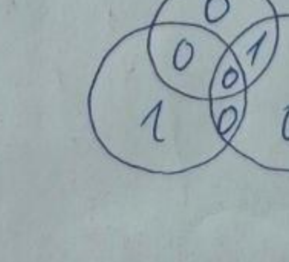
$n = 3$ $\quad n!$

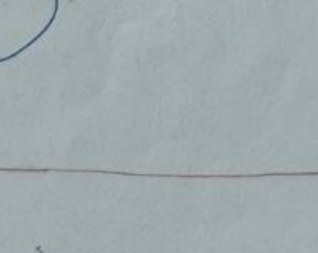
$3!\ 6$

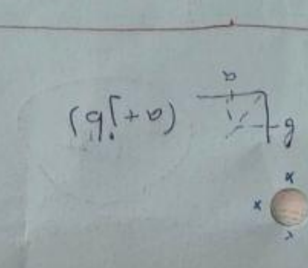
C_m^n - binomial coefficient $= \frac{m!}{n!(m-n)!}$

${}^nC_m - \binom{n}{m}$

Смотреть вот так







+0.1

$e^{j\pi} = -1$ $\sqrt{-1}$ $(a+jb)$

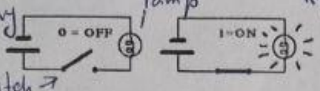
Boolean vs Aristotelian logic:
Aristotelion - is logic of propositions about classes (syllogisms)
Boolean - is the operations AND, OR and NOT on true/false

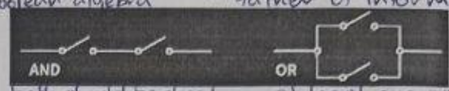
Massachusetts Institute of Technology (MIT)

Lecture by Pr. Bob Gallagher
Boole (1815-1864) & **Shannon** (1916-2001)

↑ Boolean algebra ↑ "father of information theory"

Shannon — (B. logic as the basis of switch)





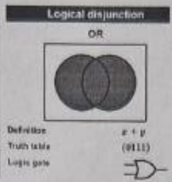
both should be closed

at least one must be closed

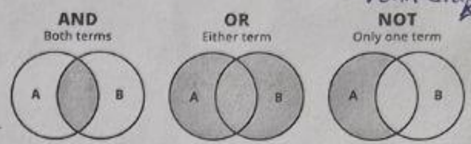
Logical addition (disjunction)

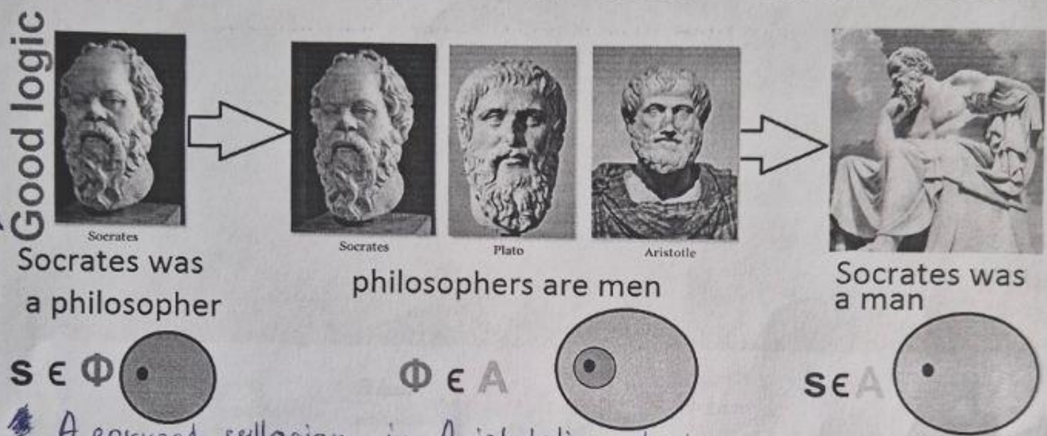
A	B	F=A∨B
0	0	0
0	1	1
1	0	1
1	1	1

A	B	A ∨ B
True	True	True
True	False	True
False	True	True
False	False	False

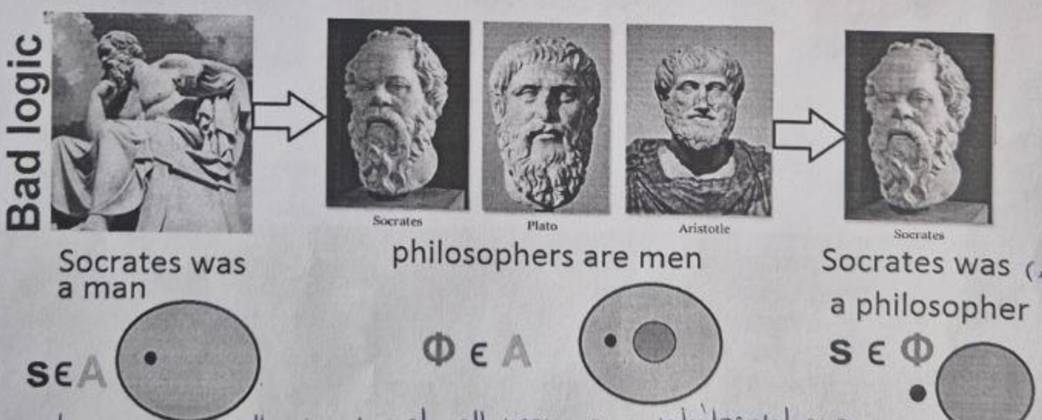


BOOLEAN LOGIC – Venn diagrams





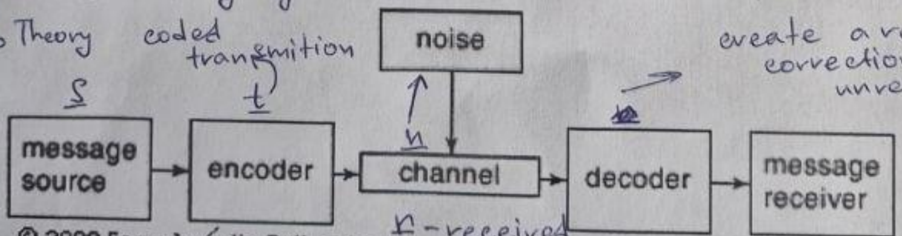
A correct syllogism in Aristotelian logic.



(error)

Incorrect syllogism: not all men are philosophers

This is what classical Info Theory about...



s

coded transmition t

n

r - received

create a relaiable correction over and unreliable (noisy) channel.

infers s and n

(Shannon communication model)

+ error correction

Resume of Lecture by Pr. Bob Gallagher from MIT

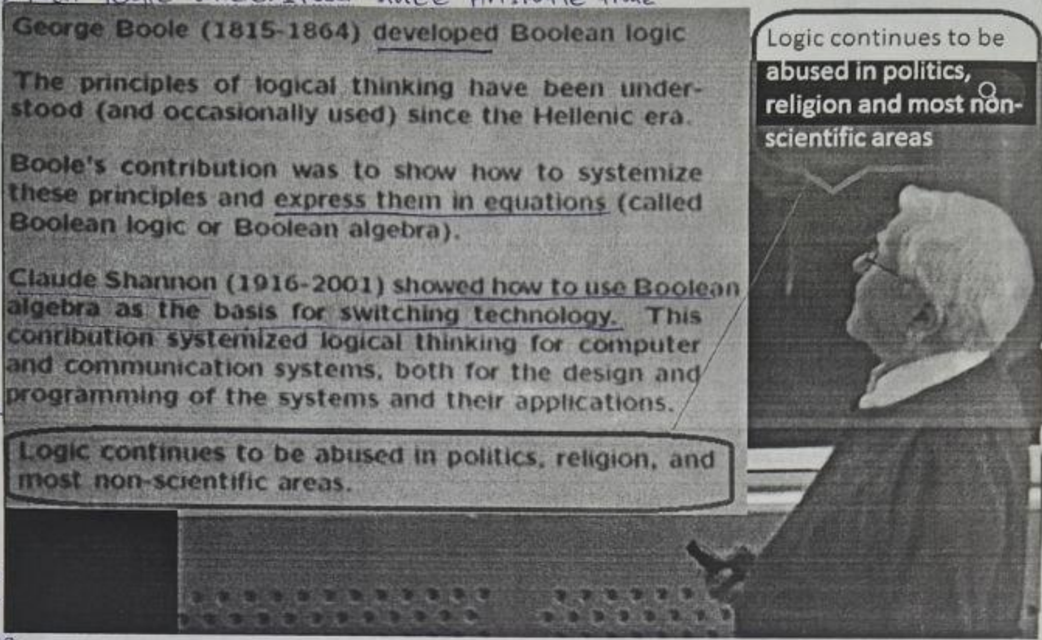
Massachusetts Institute of Technology (MIT)

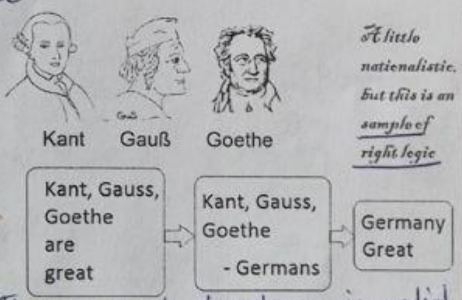
→ But logic understood since Aristotle time

Boole created the mathematical of logic and Shannon connected logic with electrical engineering

Created Information theory, his majior work

People often use logical forms incorrectly (draw false conclusions) and try to abuse the normal one.





The logical structure is valid and follows a clear, correct reasoning path

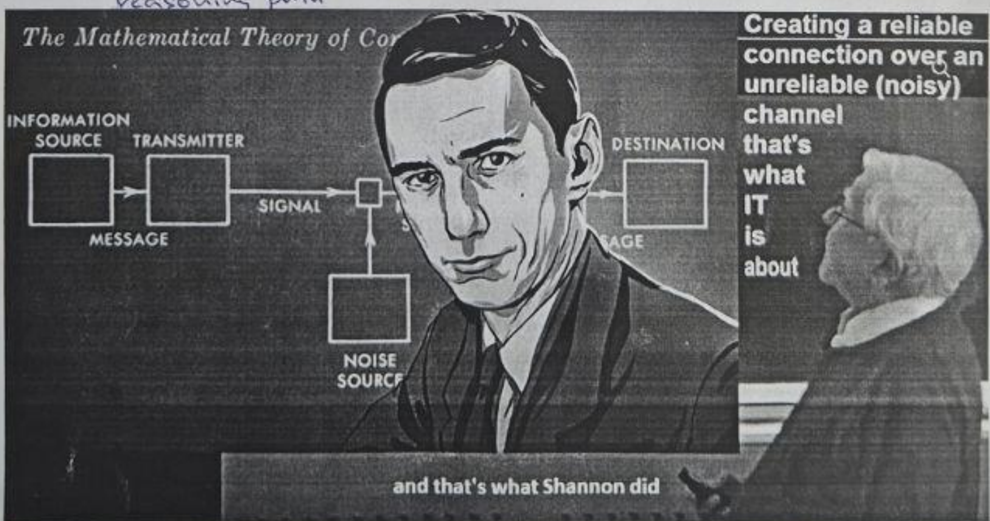
Bad logic (abuse of logic)

An individual incorrectly generalized personel attributes on group identity.

A diagram of a communication system

Data => facts

Info => organized data



Main Idea

Hedgehog in the fog ⇒ Information theory.

If you encode information correctly, then even throught a noisy channel you can transmit data almost without errors

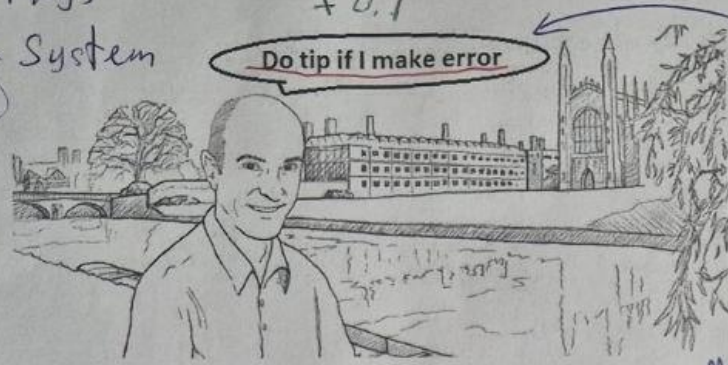
+0.1 +0.1

$\boxed{9000} \pm \boxed{30}$ ω (1)

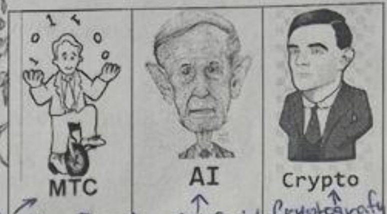
Solution – Phys System

+0.1

+ 0.1

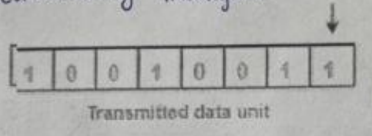


Sir **Dr. D. MacKay,**
University of Cambridge
(22 April 1967 – 14 April 2016)

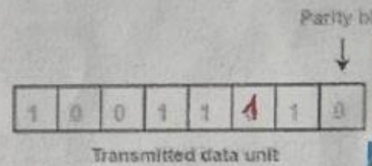


Machine Translation and coding – Artificial inteligence – Cryptografy

"I believe in clean energy,
but I also believe in mathematics"

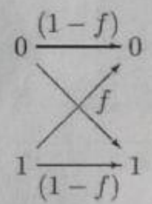


Parity bit = extra bit that makes the sum of 1s even (odd) to detect single-bit errors



1 0 1 0 . . . . 1 1

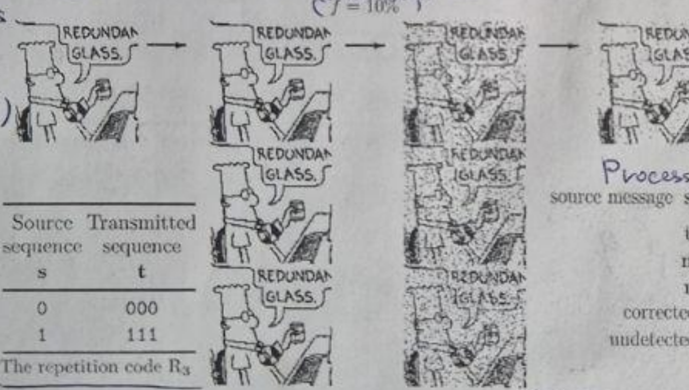
Redundant data helps ensure accuracy and reliability in the presence of errors.



f=10% – probability of error

massage source (s) —ENCODER→ each bit is repeated (t) coded transmition —CHANNEL (with noize) f = 10%→ (r) massage reciever —DECODER→ $\hat{s}$

more channels better (more copies ↓ less data loss)



+0.1

+0.1

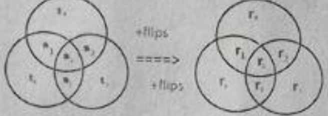
Source sequence s	Transmitted sequence t
0	000
1	111

The repetition code R_3

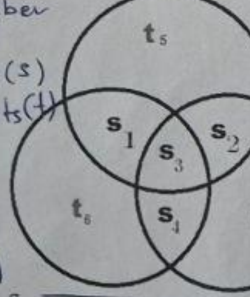
Process of error correction:

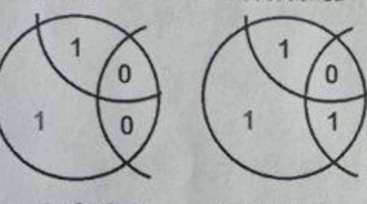
source message	s	0	0	1	0	1	1	0
	t	000	000	111	000	111	111	000
	n	000	001	000	000	101	000	000
	r	000	001	111	000	010	111	000
corrected errors			*					
undetected errors						*		

7.4. Hamming code. $\left(\frac{4}{\Sigma} \to \frac{7}{t}\right)$

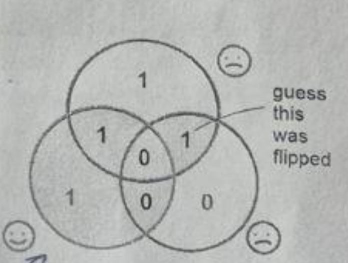


7 - total number of bits (t, s)
4 - data bits (s)
3 - parity bits (t)





1+1+0+0=0 1+1+0+1=1



Ex.:
$n = 2^m - 1, m = 3$
(m - parity bits)
$n = 2^m - 1 = 2^3 - 1 = 7$

represents different sets of parity bits. The intersections - the data bit was protected ~~with~~ by parity bit.

$(1)\ \sigma^2 = npq$ кол-во исходов, вероятность

10000 90% 10%

how the overlaps helps in pointing the errors.

$p_1 = 0,6$ $I_1 = \log \frac{1}{0,6} = 0,736966$

$p_2 = 0,2$ $I_2 = \log_2 \frac{1}{0,2} = 2,32193$

$p_3 = 0,1$ $I_3 = \log_2 \frac{1}{0,1} = 3,32193$

$p_4 = 0,1$ $I_4 = \log_2 \frac{1}{0,1} = 3,32193$

$0,6 \cdot 0,736966 + 0,2 \cdot 2,32193 + 0,1 \cdot 3,32193 + 0,1 \cdot 3,32193 = 1,5709516$

An example of a stratage for selecting the best candidate through sequential comparisons.

Say NO to the first

Say YES to the second if it is better than the first

Say NO to the third only if it is worse than all the others

The decision-making process involves comparing these candidats into groups to determain the best.

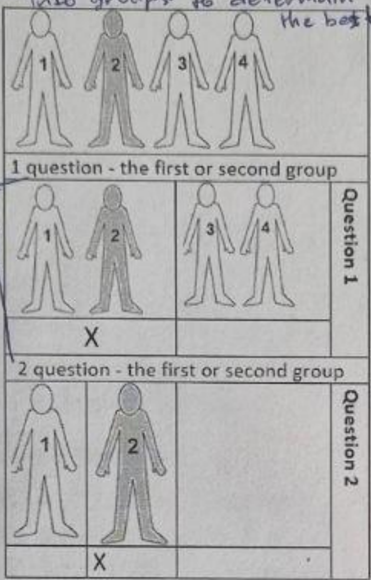
Entropy

?. (p) of all candidats equal

(yes/no)

1. □ □ □ □ □ □

2. □ □



$1 \cdot 0.6$

$2 \cdot 0.2 = 0.4$

$3 \cdot 0.1 = 0.3$

$3 \cdot 0.1 = 0.3$

1.6

Average number of questions =

2*0.25+ 2*0.25+ 2*0,25+ 2*0,25 =2

explains how to determine the average num of questions needed to indentify a specific person from a set of candidats based on their (p) diffrent

Average number of questions =

1*0.5+	2*0.25+	3*0,125+	3*0,125

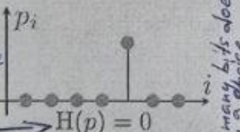
Question 1. Is this Zuckerberg?	 50%	1*0.5
Question 2. Is this Sergey Brin?	 25%	2*0.25
Question 3. Is this Stefan from BMW?	 12,5%	3*0,125
So Prince Saud	 12,5%	3*0,125
Average number of questions =1,75		

Shannon's entropy formula $H(p)$

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

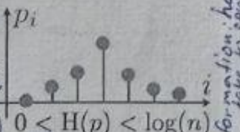
$$\sum_{i=1}^{n} p(i) \log_2 \frac{1}{p(i)}$$

Zero entropy: there is almost no uncertanty; we know in advance that the same outcome will always occur.



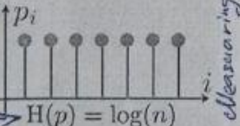
$H(p) = 0$

Several outcomes have probabilities but not equal. There is uncertanty, but it's not max, so the entropy is between 0 and log(n)



$0 \leq H(p) < \log(n)$

All outcomes are equaly likely → the max uncertainty (entropy): we can't guess the outcome at all



$H(p) = \log(n)$

Measuring information: how many bits does it take to convey a choice.

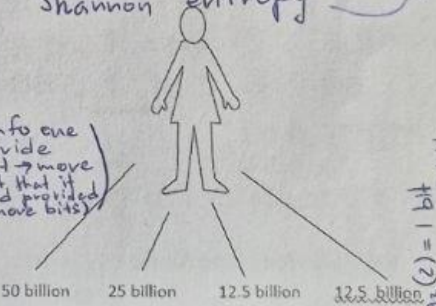
Quantifying information (how much info one outcome provide (rarer event → more info the fact that it occurred provided (more bits))

$$I(x_i) = \log_2 \left(\frac{1}{p_i} \right)$$

number of bits required to encode choice (the average num of bits needed to report the result (entropy of source))

$$\sum_{i=1}^{n} p(x_i) \, I(x_i)$$

Shannon entropy



Mark Zukerberg	Sergey Brin	Stefan Quandt	Prince Al Saud
P(1)= 50%	P(2)= 25%	P(3)= 12,5%	P(4)= 12,5%

$I(1) = \log_2 \left(\frac{1}{0,5}\right) = \log_2(2) = 1$ bit

$I(2) = \log_2 \left(\frac{1}{0,25}\right) = \log_2(4) = 2$ bit

$I(3) = I(4) = \log_2 \left(\frac{1}{0,125}\right) = \log_2(8) = 3$ bit

$H(x) = \sum_{i=1}^{4} I(x_i) = 0,5 \cdot 1 + 0,25 \cdot 2 + 0,125 \cdot 3 + 0,125 \cdot 3 = 1,75$ bits

$p_1 = 0.6$ $I_1 = \log_2 \frac{1}{0,8} = \log_2 1,25 = 0,321928$

~~$p_2 = 0.2$~~ ~~$I_3 = \log_2 \frac{1}{0,1} = \log_2 10 = 3,32153$~~

$p_3 = 0,1$

$p_4 = 0,1$ $I_4 = \log_2 \frac{1}{0,1} = \log_2 10 = 3,32193$

3.32193

+0.3

$0,8 \cdot 0,321928 + 0,1 \cdot 3,32193 + 0,1 \cdot 3,32153 = 0,9219284 \approx 0,92$

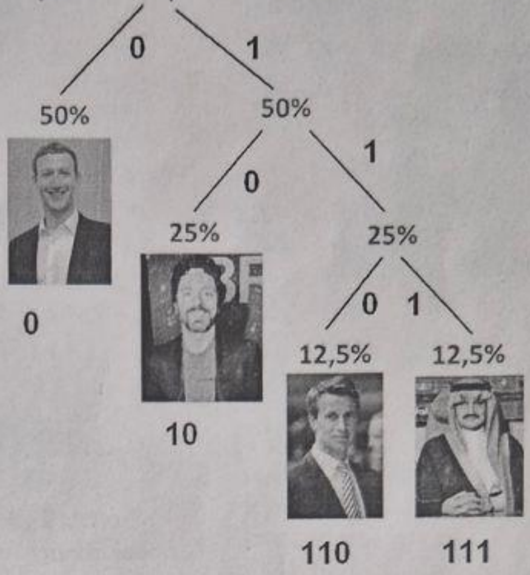
$H(x) = \sum_{i=1}^{n} p(x_i) \log_2 \frac{1}{p(x_i)}$ - среднее кол-во инфо на 1 исход.

$p(x_i)$ - вероятность того, что произойдет исход x_i

$\frac{1}{p(x_i)}$ - "обратная вероятность

$\log_2 \frac{1}{p(x_i)} = I(x)$ = кол-во инфо в одном конкретном исходе (в битах $\log_2$)

Huffman Tree:
Yes/No question tree: we give a short code (0) to a frequent person, and longer ones (10, 110, 111) to use fewer bits on average fore rare people.
(Show the process of creating a Huffman coding Tree!



First-order approximation (symbols independent but with frequencies of Belarusian txt). ↑ we calculate the probabilities of individual letters in the massage

Мама мыла ра

М - 3	— 30%	1-3 М
а - 4	— 40%	4-7 а
ы -1	— 10%	8 -ы
л - 1	— 10%	9 -л
р - 1	— 10%	10 -р
10		

лла $\boxed{\text{маммa}}$ р

1st-order approx.

Second-order approximation (digram (2-symbols) structure as in Belarusian) → letter pair frequences for more efficient coding.

Мама мыла ра

Ма - 2	22%	1-2	ма
ам - 2	22%	3-4	ам
мы - 1	11%	5	мы
ыл - 1	11%	6	ыл
ла - 1	11%	7	ла
ар - 1	11%	8	ар
ра - 1	11%	9	ра
9			

Ni - position in the text

0. 4 6 7 3 1 9 1 6 7 3 5

ам ыл ла ам ма ра ма ыл ла ам мы

мылла рама

2nd-order approx.

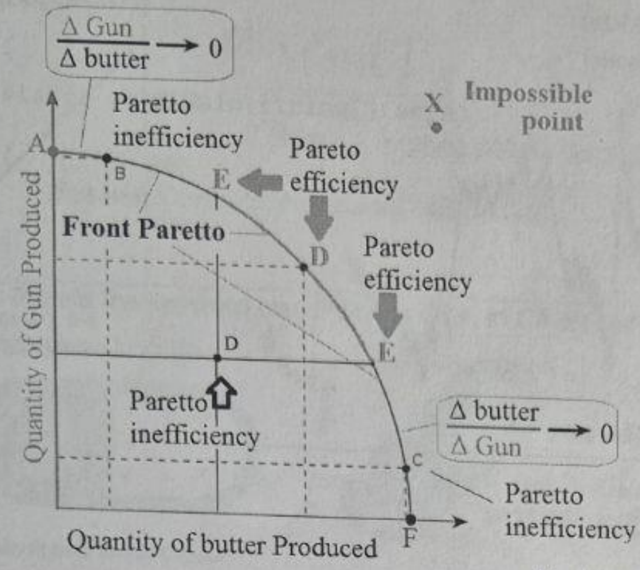
mamma mia

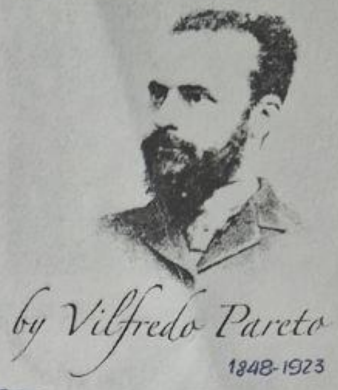
	0	
m - 40%	0-25 %	0:a__a__ma
A - 30%	26 -50 %	1st: _imaiaamm__
␣ - 20 %	51 - 75 %	
i - 10%	76 - 100%	

1st	2nd		
		1-12 %	
0-40%	ma - 12,5%		mimiammaa_mimammia
41-70%	am - 25%	13-38%	
71-90%	mm - 12,5%	39 - 51%	
91-100%	mi - 25%	52% - 75%	
	ia - 12,5 ..		
	a_ - 25% . . .		

Pareto curve: at the frontier, one resource cannot be improved without worsening another; inside, resources are used inefficiently.

Pareto efficiency ~~[illegible]~~ - Is a state where resources are allocated in the most efficient manner.



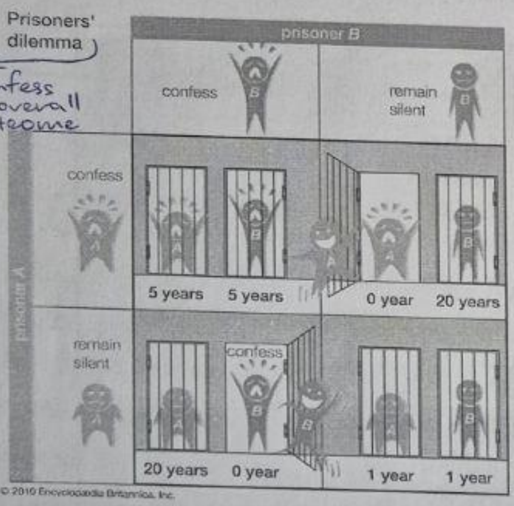


by Vilfredo Pareto
1848-1923

The orange sector E-D-E is the most Pareto efficient - since an increase in one indicator leads to a decrease in another.

Nash equilibrium - a concept of game theory where no player can benifit by changing their strategy while the other players keep their unchanged.

Individual choice to confess worsens the overall outcome



** => Nash equilibrium

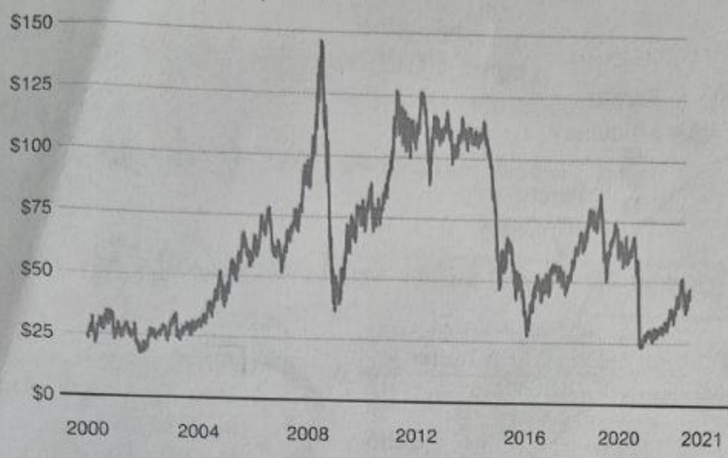
$H_2(x)$ / $H_1(x)$	Player 2	Recognition;	Non-recognition;
Player 1		1	2
Recognition;	1	* -5 / -5 *	-20 / 0
Non-recognition;	2	0 / -20	-1 / -1

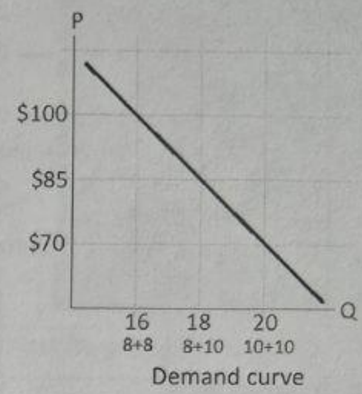
-1-1
Pareto Optimality

Every mafioso has 2 options: remain silent or confess.
1) If both remain silent, they are considered "man of respect" and receive a more lenient punishment.
2) If one confesses and the other remains silent the one who confess recieves a prison sentence with freedom and is labeled as "Penito". Who remais silent - receives the full punishment 20y.

Oil price hits 18-year low

Brent crude, US dollars per barrel



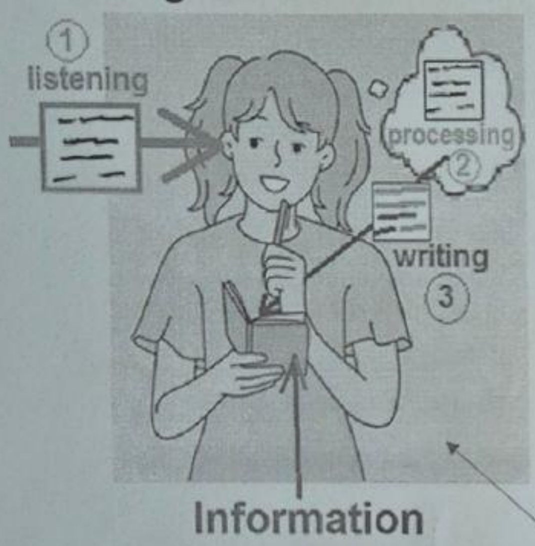


Pareto

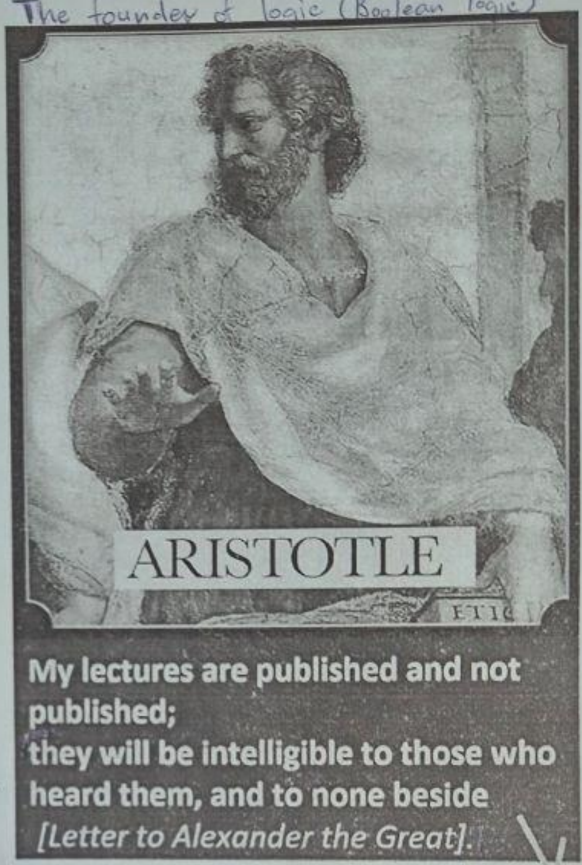
Barrel	1. $8 \cdot 10^6$ barrel/day	2. $10 \cdot 10^6$ barrel/day
1. $8 \cdot 10^6$ barrel	$800 millions per day; $100 / barrel; $800 millions per day	$850 millions per day; $85 / barrel; $680
2. $10 \cdot 10^6$ barrel/day	$680 millions per day; $85 / barrel; $850 millions per day	$700 millions per day; $70 / barrel; $700 millions per day

Information Theory

Hearding



The founder of logic (Boolean logic)



Little information is like a hadgekog in the fog: hight uncertanty.

$H=K\,\mathrm{lb}(N)$ - entropy Shannon - the average number of bits nedded to describe a massage.

$$\sum_x P(x) \cdot \log\left(\frac{1}{P(x)}\right)$$