



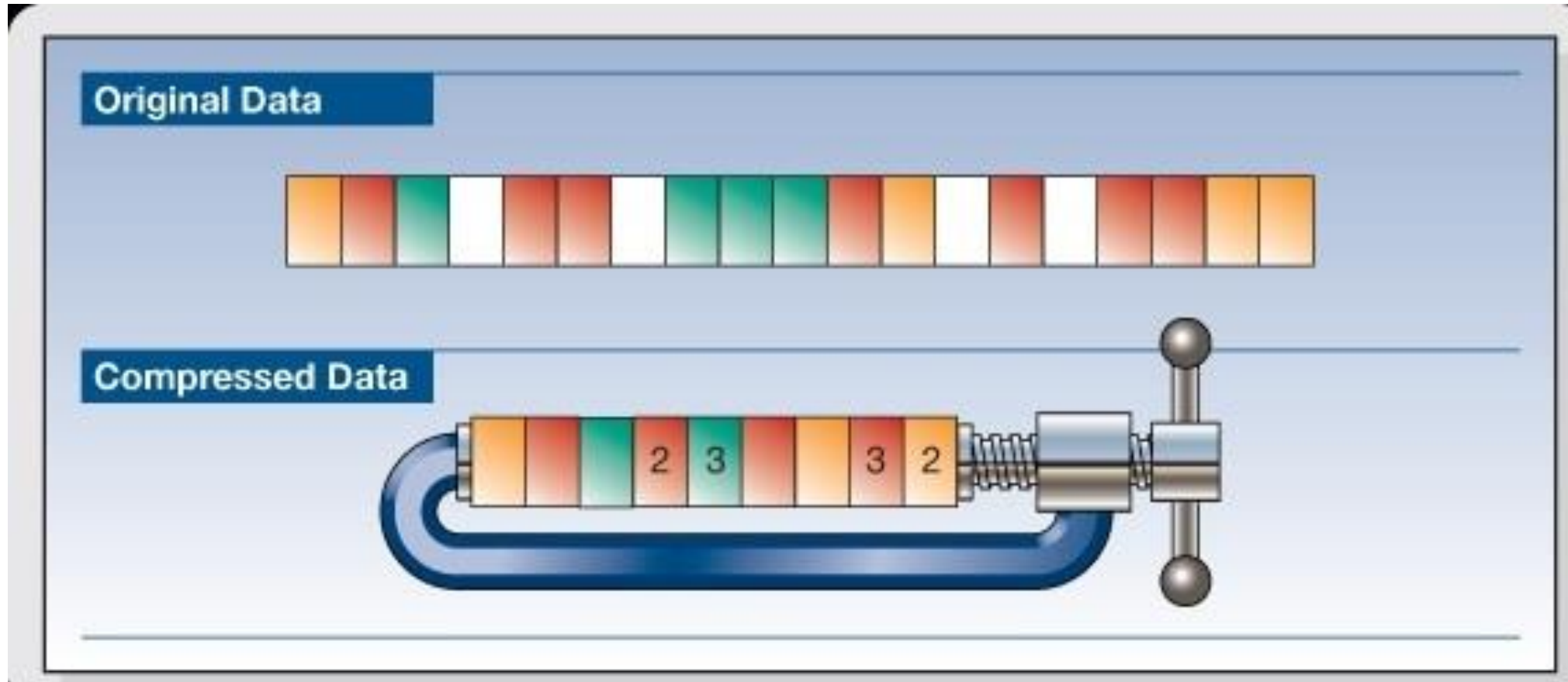
Information Theory

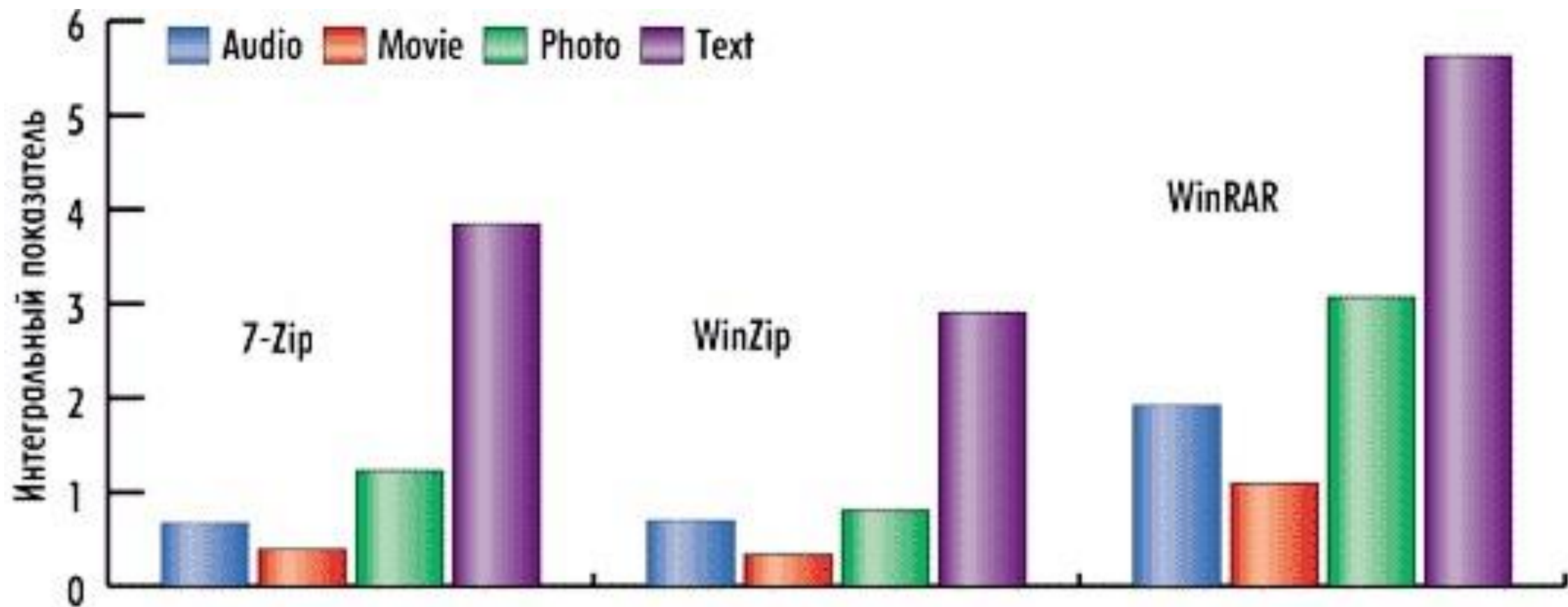
Introduction

YAROSHEVICH
Andrey Olegovich

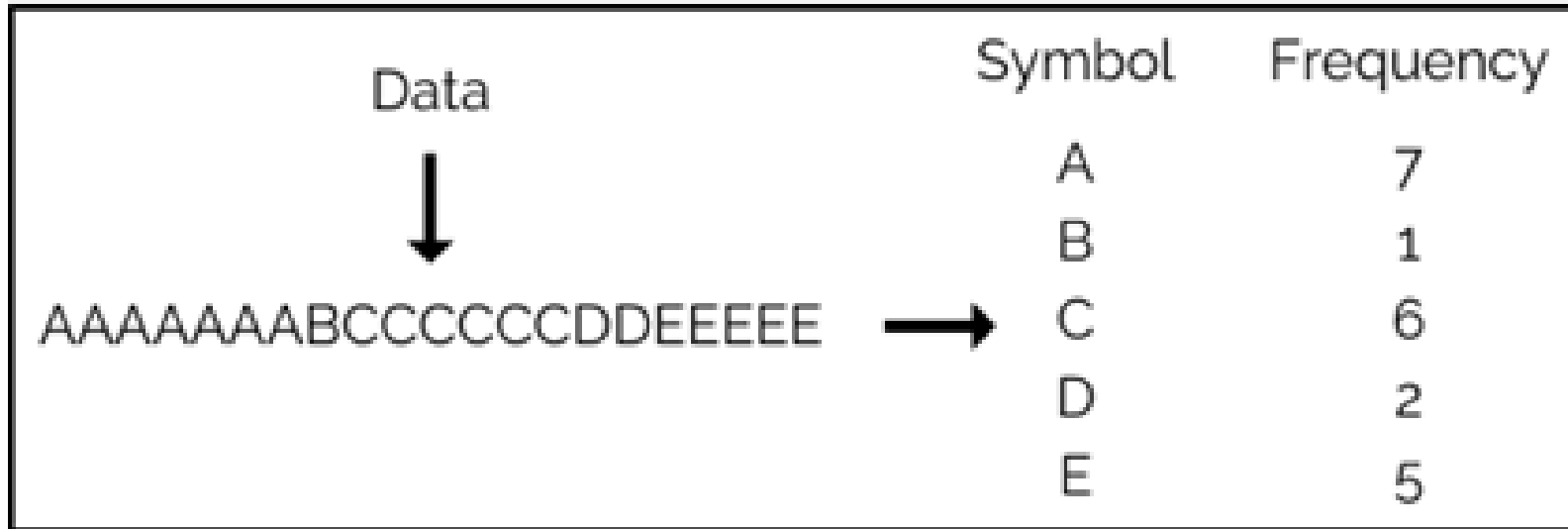
www.YAROSHEVICH.com
ao@YAROSHEVICH.com
+375 29 254-07-92

Comparison of the compression ratio of popular archivers





Before Compression



After Compression

Symbol	Bit Code
A	00
B	111
C	01
D	110
E	10

Entropy
compression ratio

Before Compression = 21×8 bits = 198 bits

After Compression = 7×2 bits + 1×3 bits +
 6×2 bits + 2×3 bits + 5×2 bits = 45 bits

Card cheating



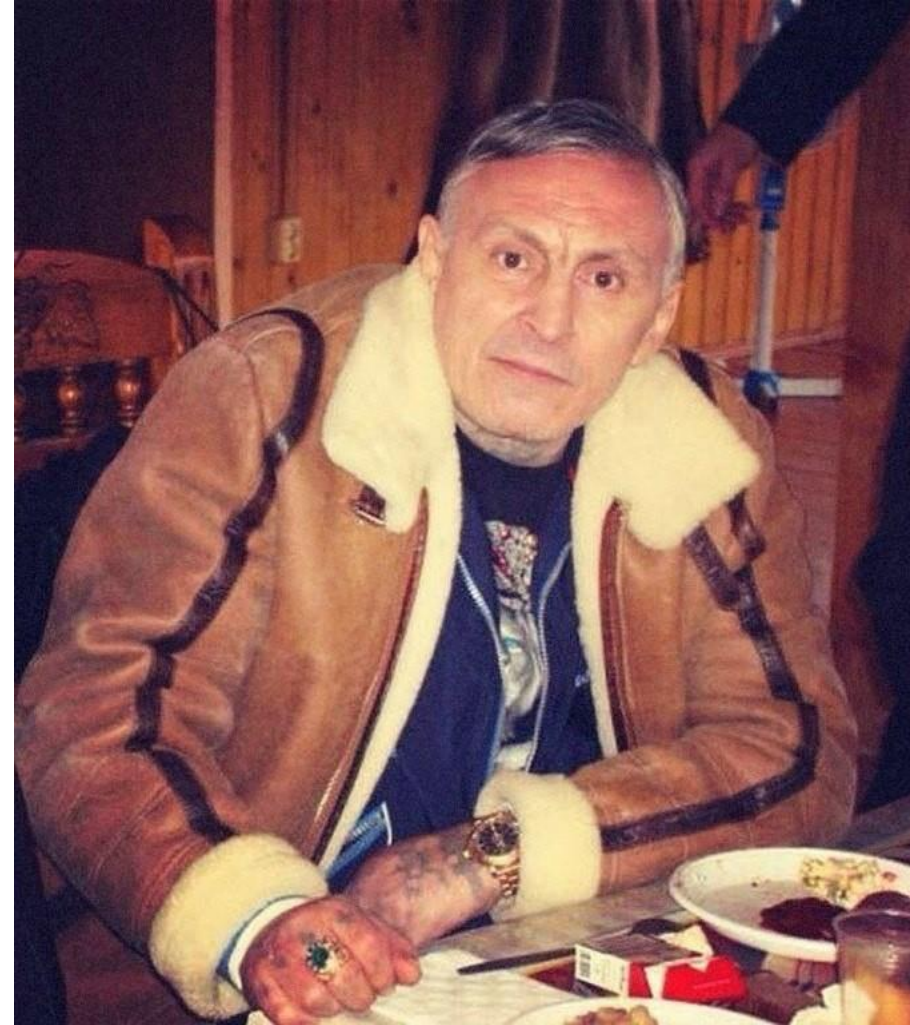
"Card Sharp "
by Caravaggio



Valentin de Boulogne. Shuler.







<https://youtu.be/Ps8ztNaN2ds?t=85>





Poker Hands - Rankings Chart



1 Royal Flush



6 Straight



2 Straight Flush



7 Three of a Kind



3 Four of a Kind



8 Two Pair



4 Full House



9 Pair



5 Flush



10 High Card

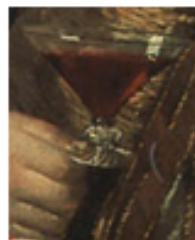
Let's check how deep you understand the problem

Poker.

10 combinations.



How many fingers does it take to pass them?



Or how many kinds of liquor to use?

suit



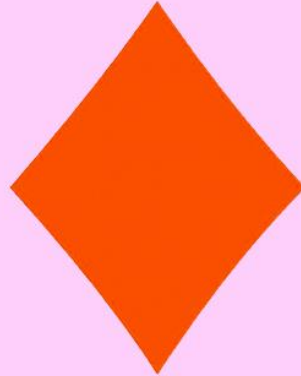
Hearts

Черви



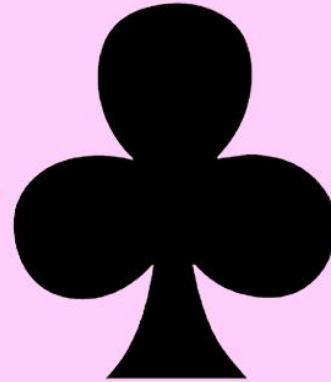
Spades

Пики



Diamonds

Бубны



Clubs

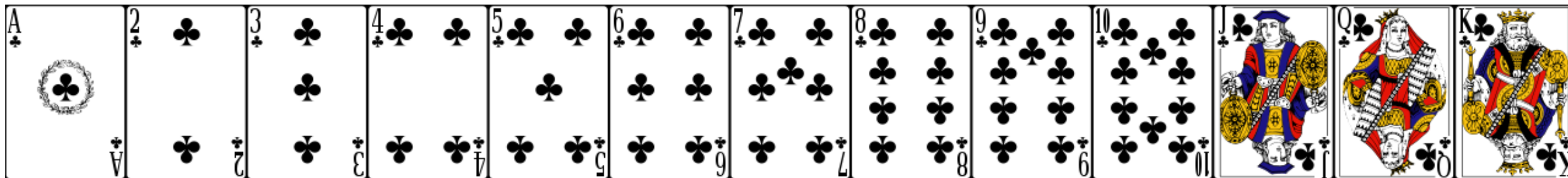
Трефы



How many
fingers to
use?



Or how many kinds
of liquor to use?



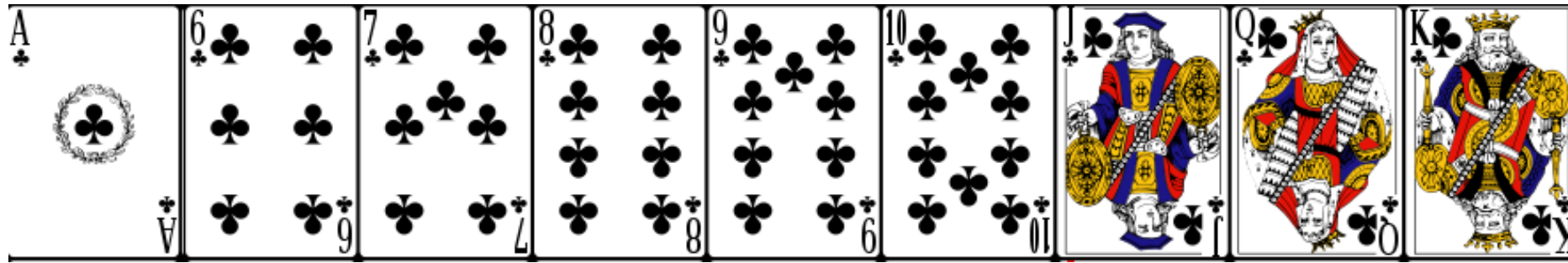
card weight 1-13



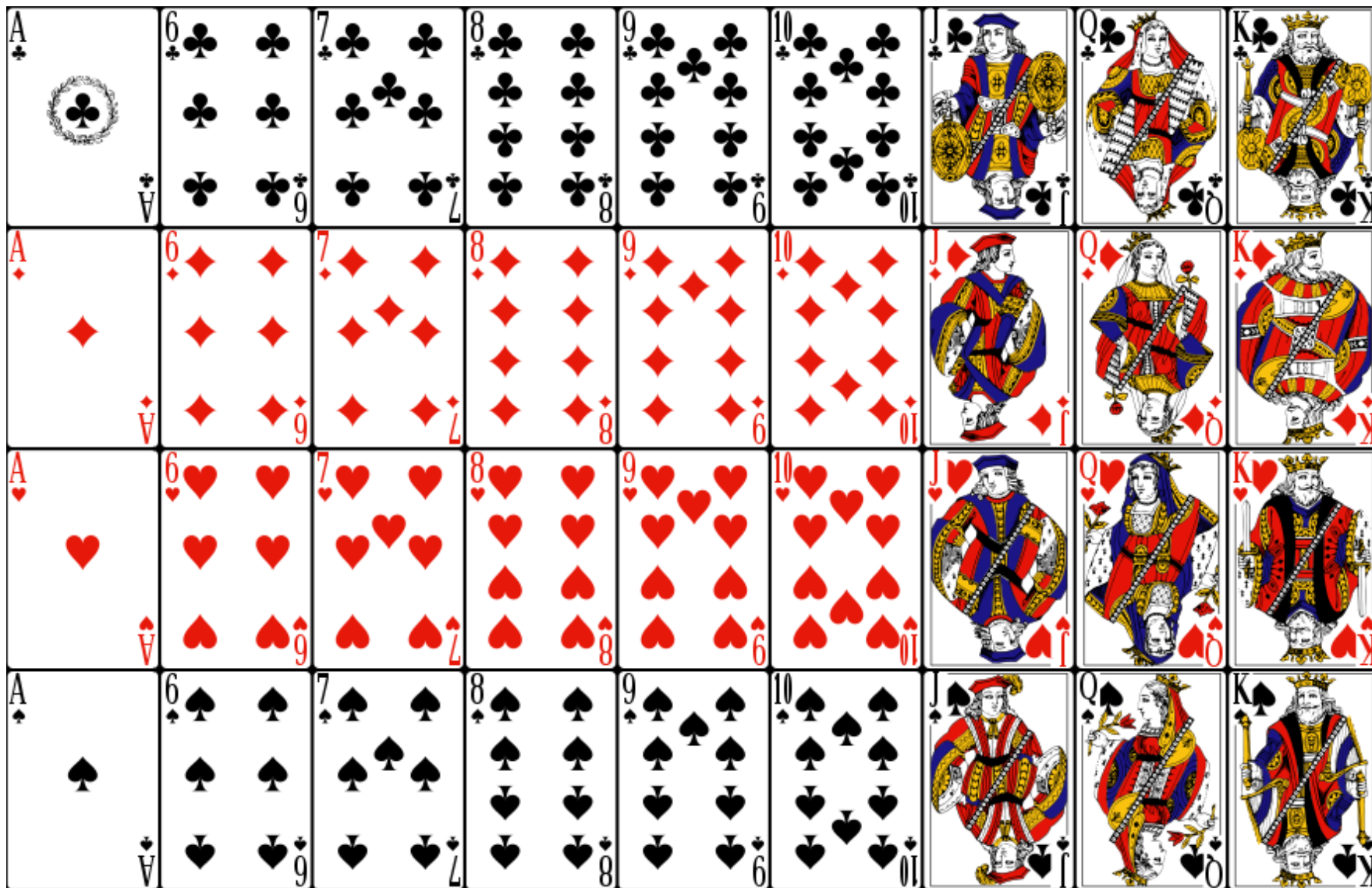
How many fingers does it take to pass it?



Or how many kinds of liquor to use?



card weight 1-9



card from the preference deck $9 \times 4 = 36$



Heads



Tails

DAYS OF THE WEEK

7 days

WEEKDAYS 5 days	Monday	Mon.
	Tuesday	Tue.
	Wednesday	Wed.
	Thursday	Thu.
	Friday	Fri.
WEEKEND 2 days	Saturday	Sat.
	Sunday	Sun.

Exam grade

Оценка по 10-балльной шкале	Оценка по 5-балльной шкале за экзамен	Оценка, используемая в европейском приложении к диплому		Оценка за зачет
10	отлично	A +	Excellent	зачтено
9	отлично	A	Very good	зачтено
8	отлично	A -	Very good	зачтено
7	хорошо	B +	Good	зачтено
6	хорошо	B -	Good	зачтено
5	удовлетворительно	C +	Satisfactory	зачтено
4	удовлетворительно	C -	Satisfactory	зачтено
3	неудовлетворительно	F	Fail	не зачтено
2	неудовлетворительно	F	Fail	не зачтено
1	неудовлетворительно	F	Fail	не зачтено

But a person has only 5 fingers on his hand.

What to do when you need a lot more combinations?

Answer is here

<https://www.youtube.com/watch?v=WTxKEwP01Sc&t=431s>

Deleted scene
from the movie

Inglourious Basterds
by **Tarantino**

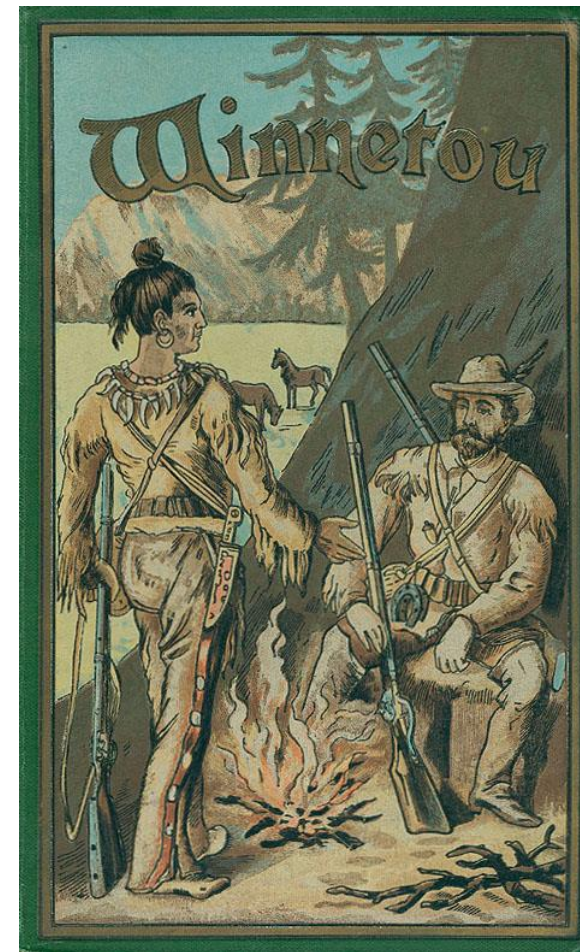


This is children's game usually known "Akinator" <https://en.akinator.com/>.

The first player in this two-player game thinks of something, and by a series of binary questions yes/no, the other player attempts to guess what it is.



- I am from the past?
- Yes
- I am a woman?
- No
- Am I not German? I am American?
- Yes
- If I was married they would call her “squaw”?
- Yes
- I was created by Karl May
- Yes



- I'm **Winnetou - Apache chief**

[film [Winnetou - Apache chief](#) (RU)]

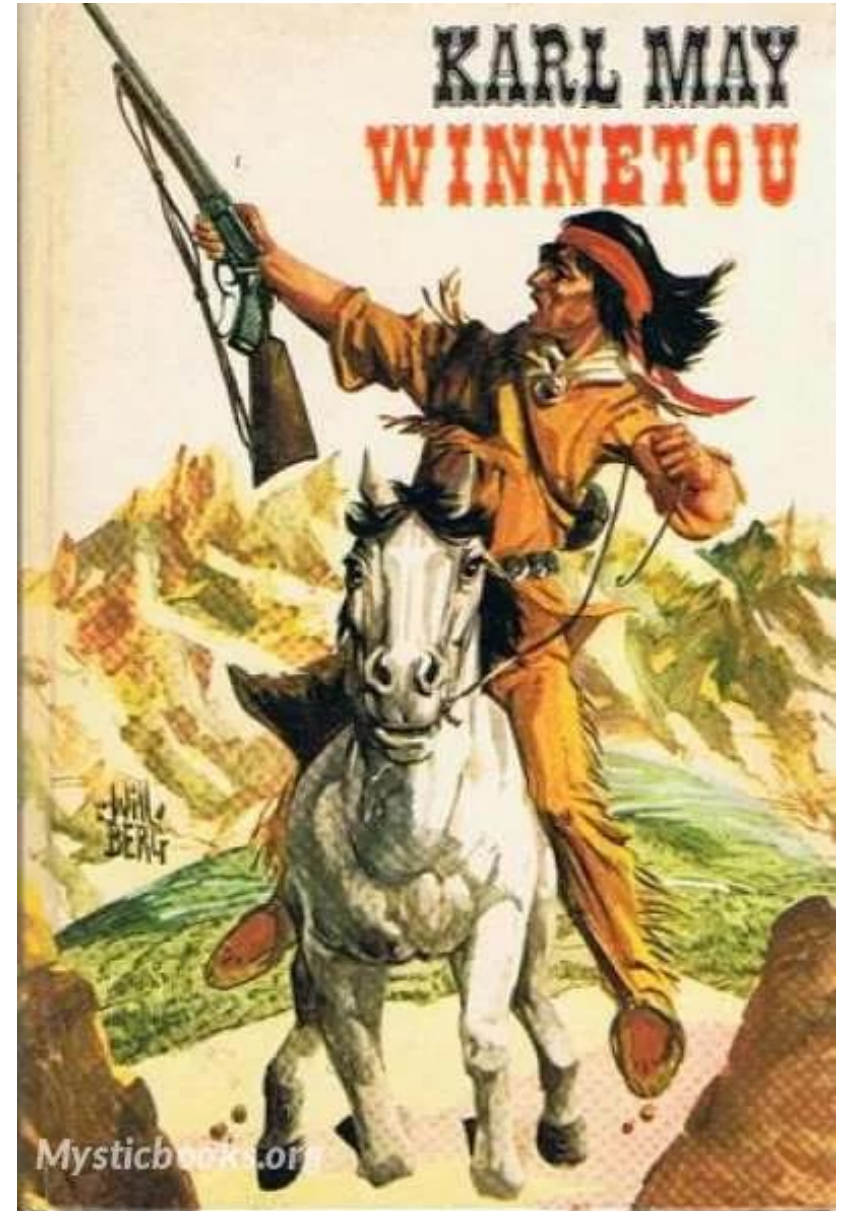
Some questions are better than others questions.

Asking on the first round “is it a Monte-Cristo?” is likely to waste time.

We must find optimal set of questions is to ask.

How to do that?

Let's try to figure it out using the example of «**Winnetou**, the Apache leader».







"Look, he's sending us a text message."

The Indians

- kindled a fire on a rock,
- covered the fire with wet grass,
- and in a certain rhythm stretched a blanket over the smoke





In fact, this is a binary code transmission
(but only at a very low speed).



=====**10110010010110**=====>

How to do it most effectively?

The «direct solution» is the transmission of the letter
(for example **K**)



This is A - No

This is B - No

This is C - No

.....

This is K - Yes

.....

This is Z - No

26 questions.

Too much firewood needed here!

Akinator solution (for **K**)

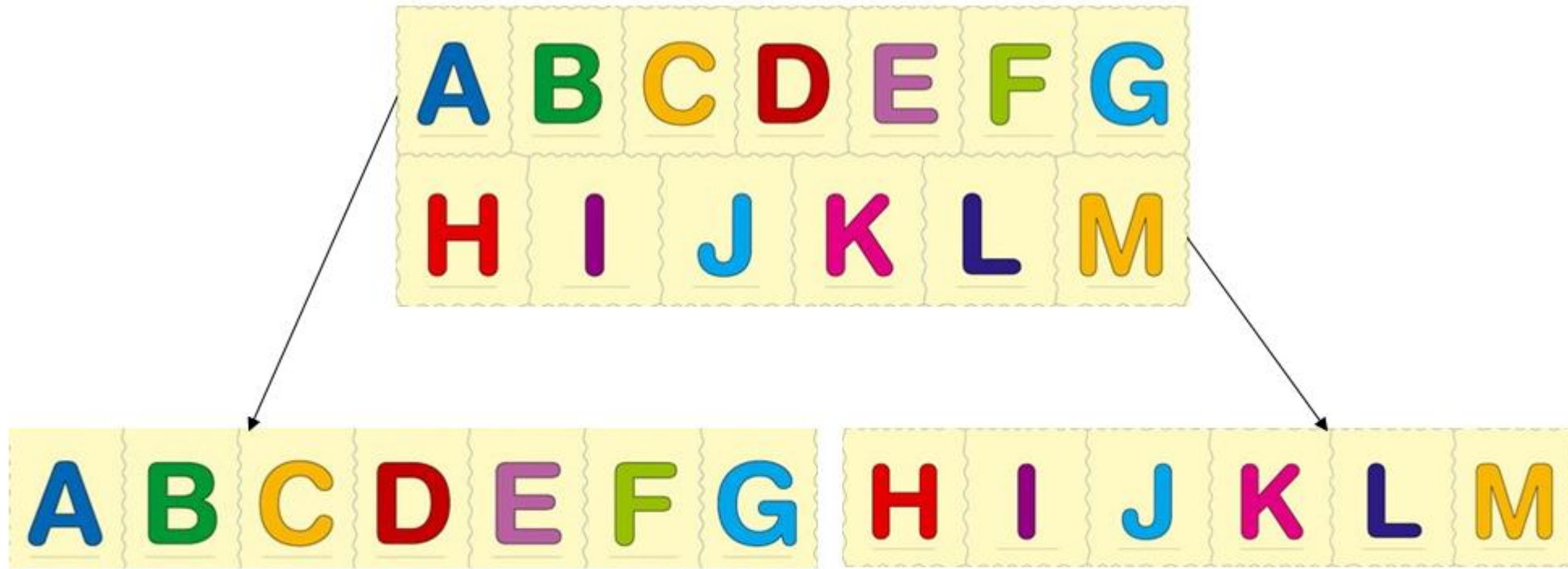
Step 1. We divide the alphabet into 2 equal parts
or approximately 2 equal parts



and fix (ask a question) in which part is our letter.

0

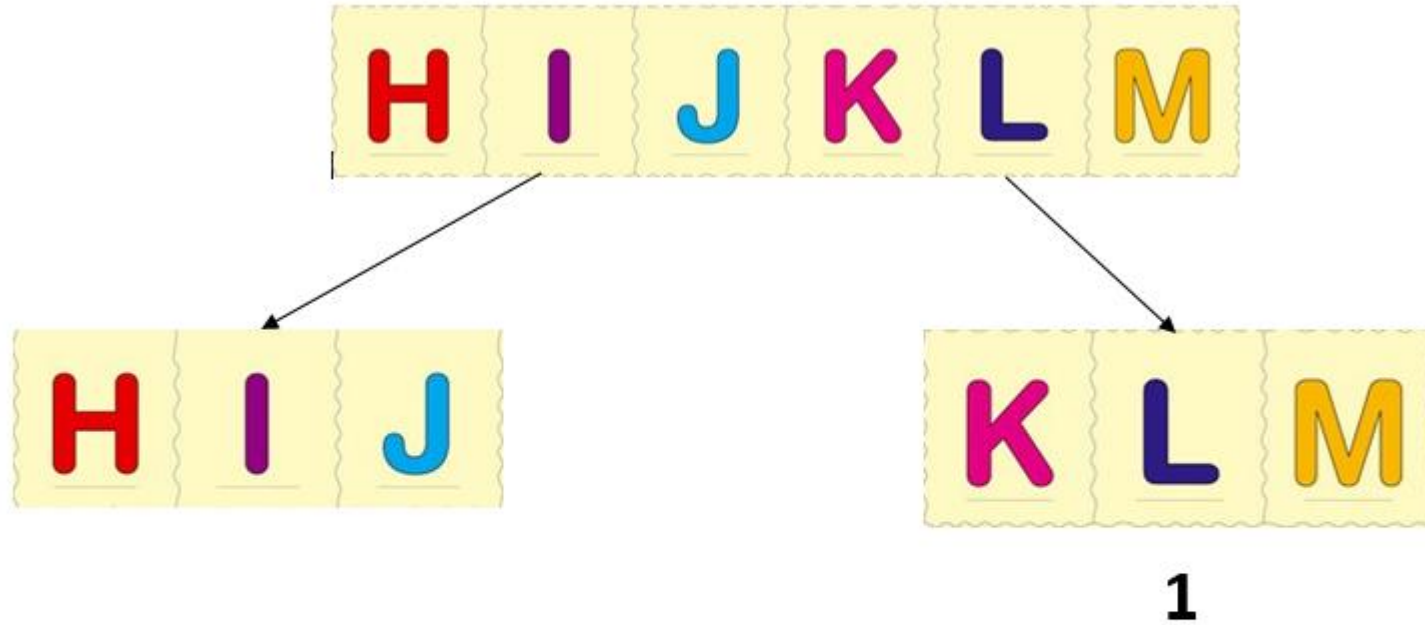
Step 2 Divide the first part of the alphabet into 2 equal parts
or about 2 equal parts



1

and fix (ask a question) in which part is our letter – 1 (1+0=10)

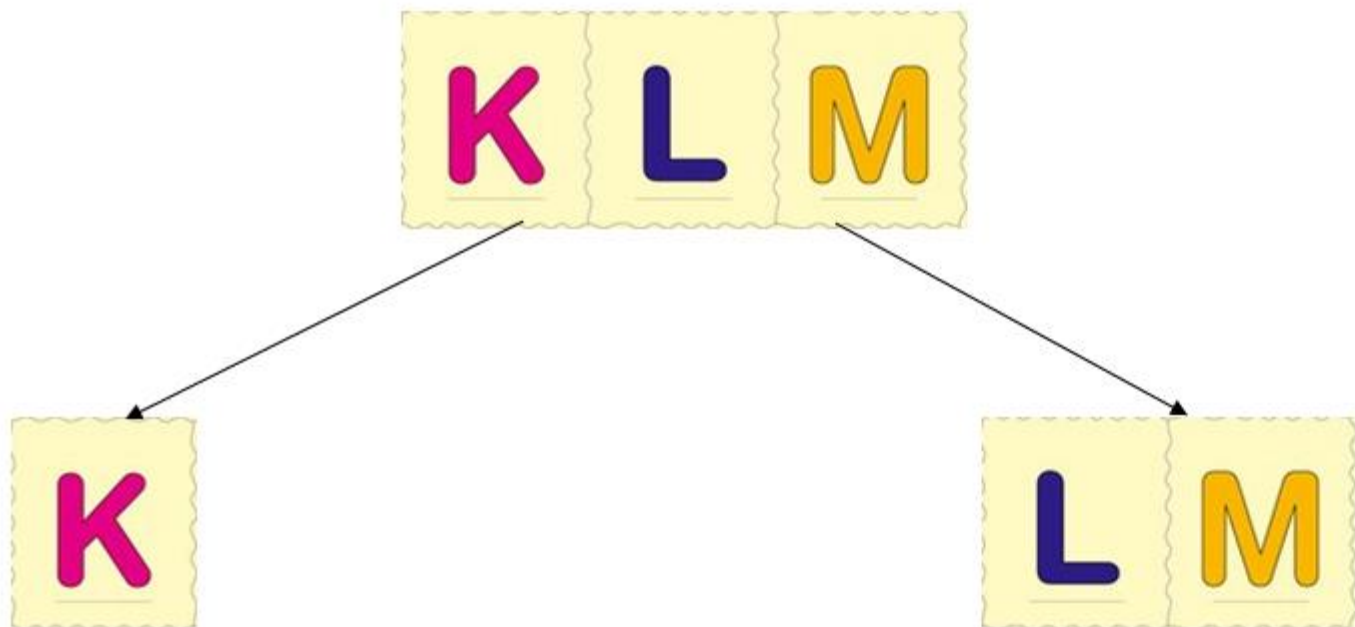
Step 3 Divide the selected part of the alphabet into 2 equal parts



and fix (ask a question) in which part is our letter.

1 ($1+10=110$)

Step 4 Divide the selected part of the alphabet into 2 equal parts



0

and fix (ask a question) in which part is our letter.

0 ($0+110=0110$) **0110** – We get a 4-digit number instead of a 26-digit number.

Direct solution	Akinator solution
<p>This is A - No</p> <p>This is B - No</p> <p>This is C - No</p> <p>.....</p> <p>This is K - Yes</p> <p>.....</p> <p>This is Z - No</p> <p>26 questions</p> <p>00000000001000000000000000000</p>	<p>0 1</p> <p>A B C D E F G N O P Q R S T H I J K L M U V W X Y Z</p> <p>A B C D E F G H I J K L M</p> <p>1</p> <p>H I J K L M</p> <p>1</p> <p>K L M</p> <p>0</p> <p>4 questions</p> <p>0110</p>

Akinator solution (for **L**)

Step 1. We divide the alphabet into 2 equal parts or approximately 2 equal parts

0



1



and fix (ask a question) in which part is our letter.

0

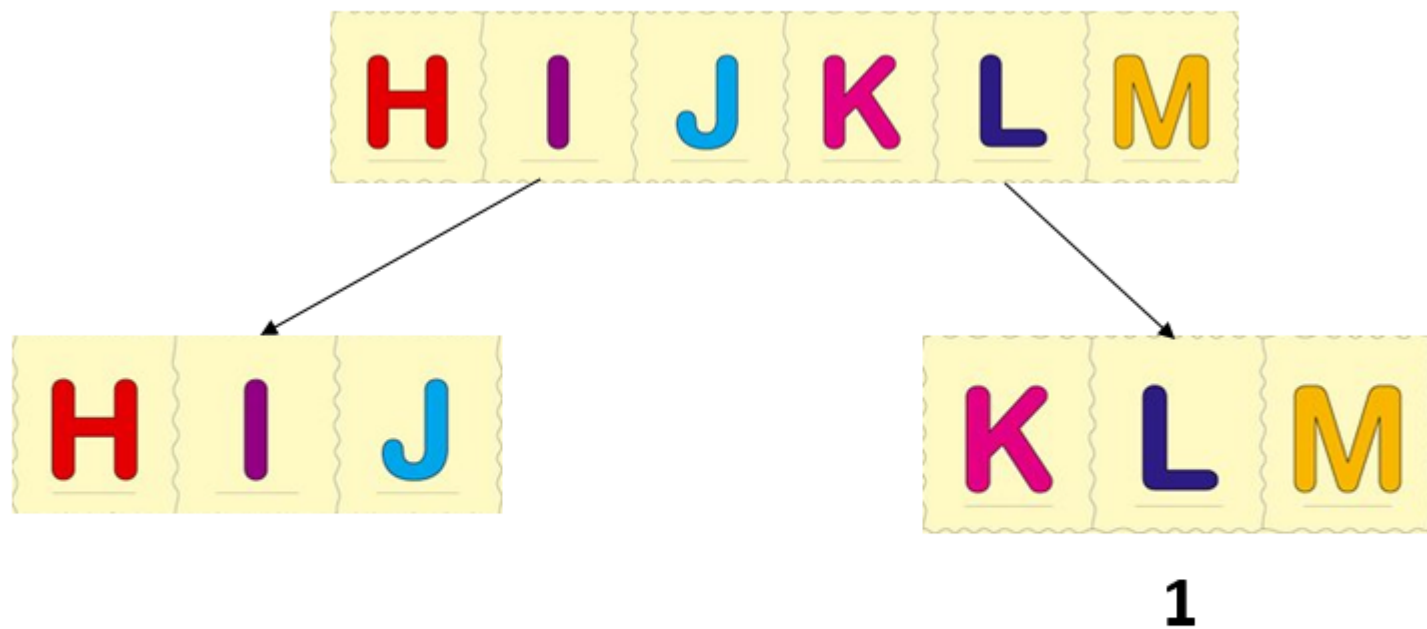
Step 2 Divide the first part of the alphabet into 2 equal parts or about 2 equal parts



1

and fix (ask a question) in which part is our letter – 1 (1+0=10)

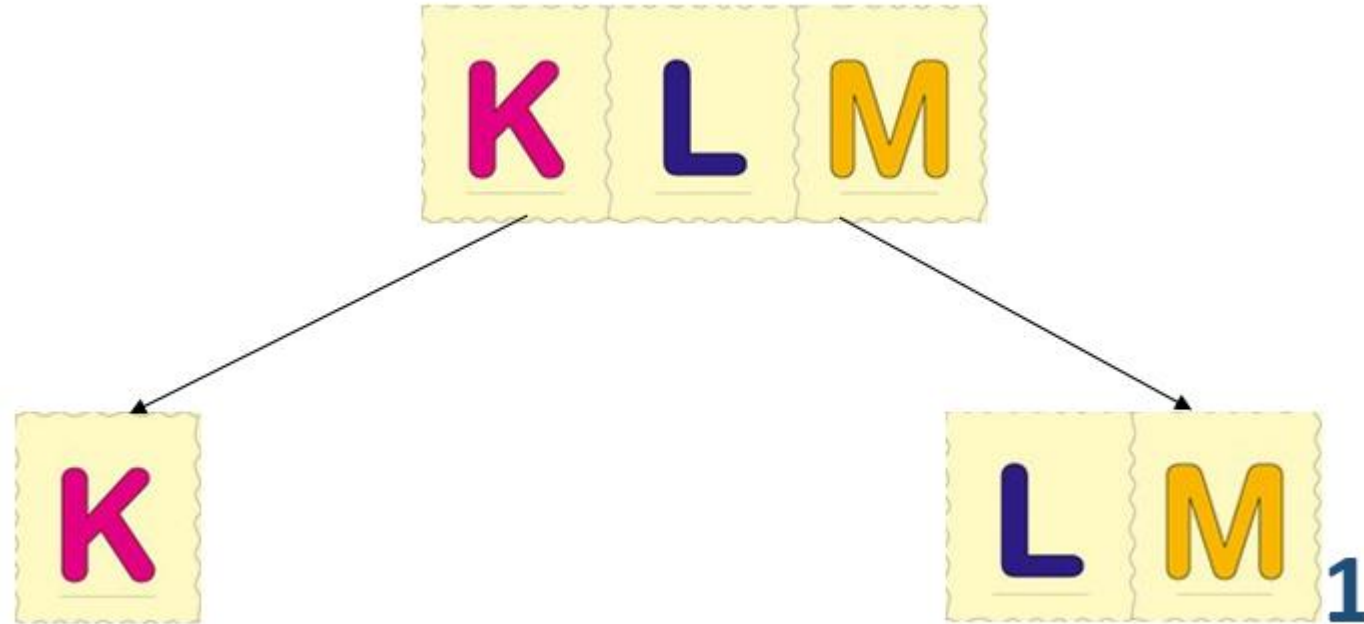
Step 3 Divide the selected part of the alphabet into 2 equal parts



and fix (ask a question) in which part is our letter.

1 (1+10=**110**)

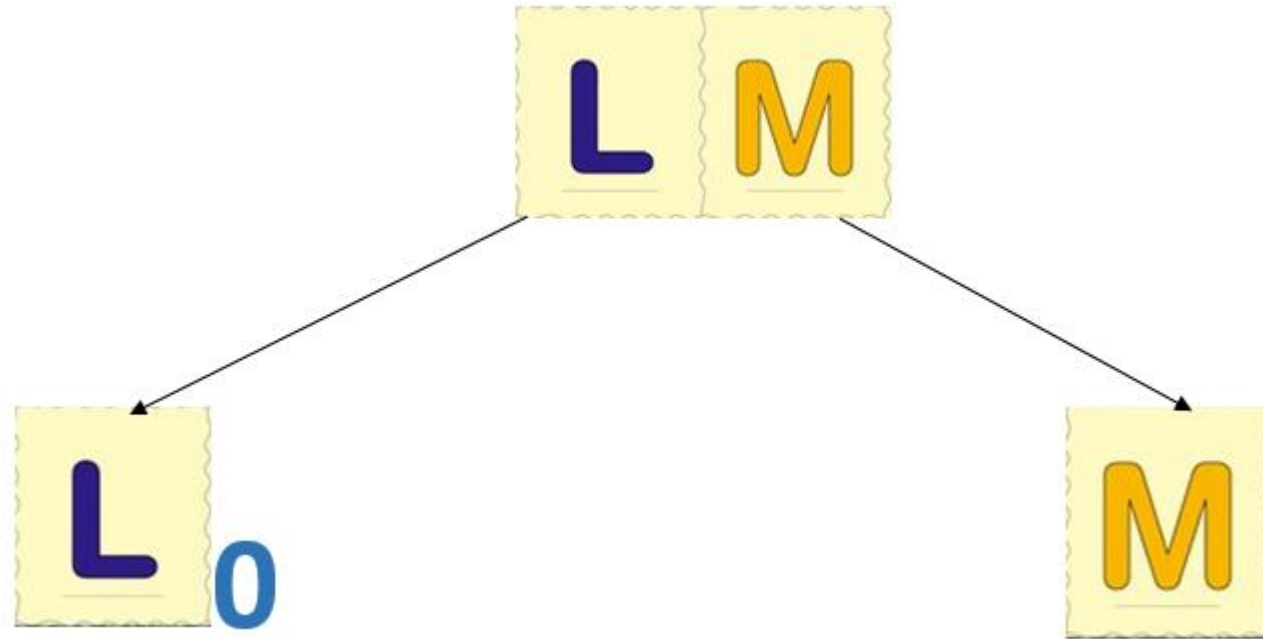
Step 4 Divide the selected part of the alphabet into 2 equal parts



and fix (ask a question) in which part is our letter.

1 (1+110=**1110**)

Step 5 Divide the selected part of the alphabet into 2 equal parts



and fix (ask a question) in which part is our letter.

0 ($0+1110=01110$) **0**1110 – We get a 5-digit number instead of a 26-digit number.

And for M 11110 – We get a 5-digit number instead of a 26-digit number.



0



1



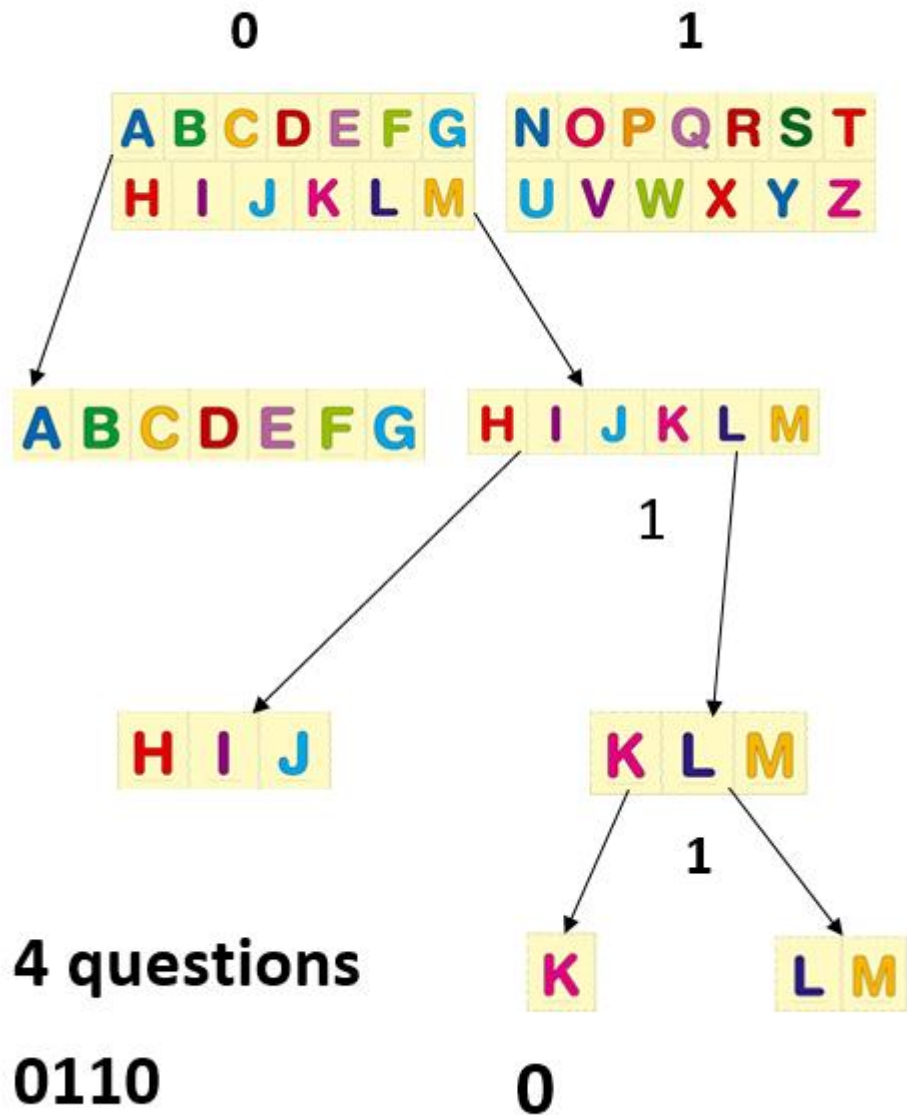
1



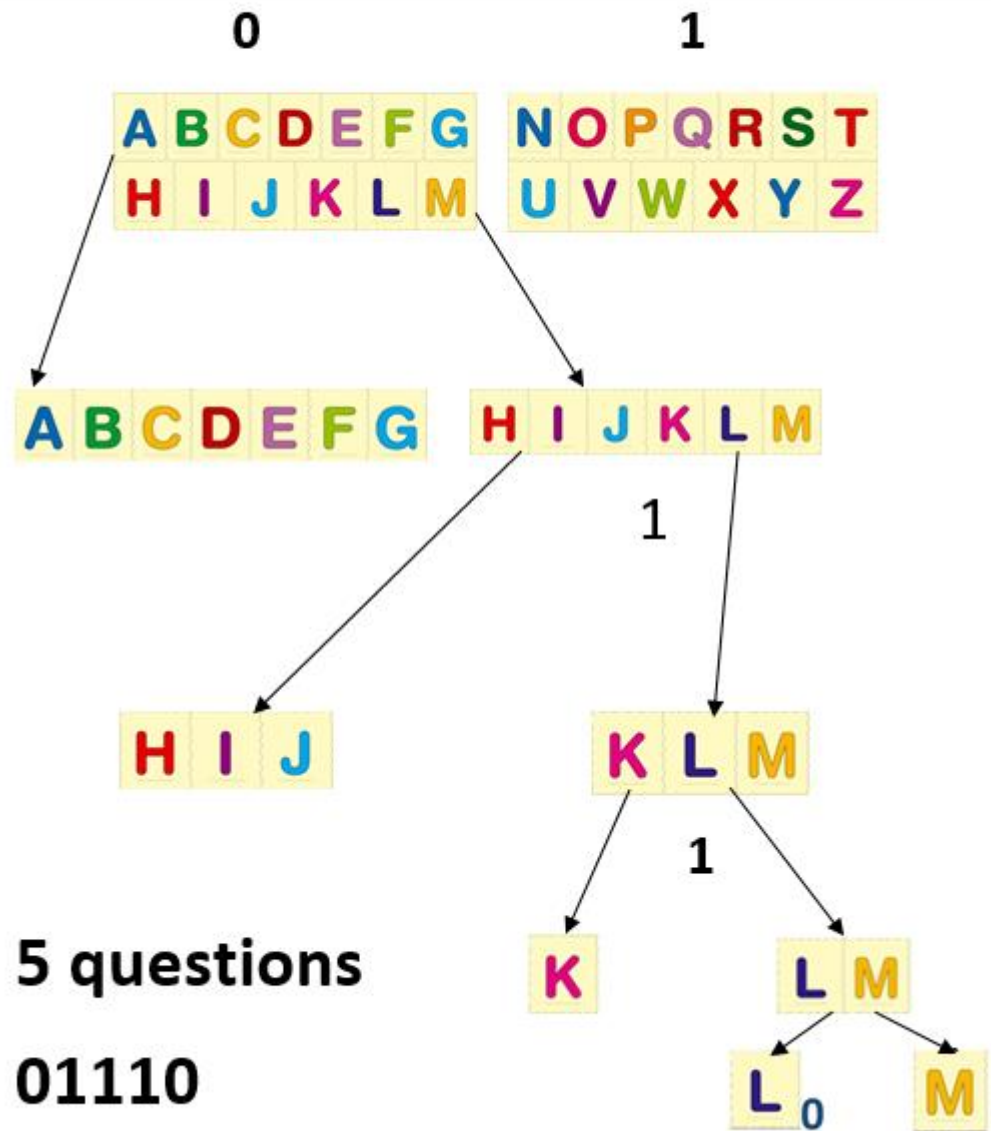
1

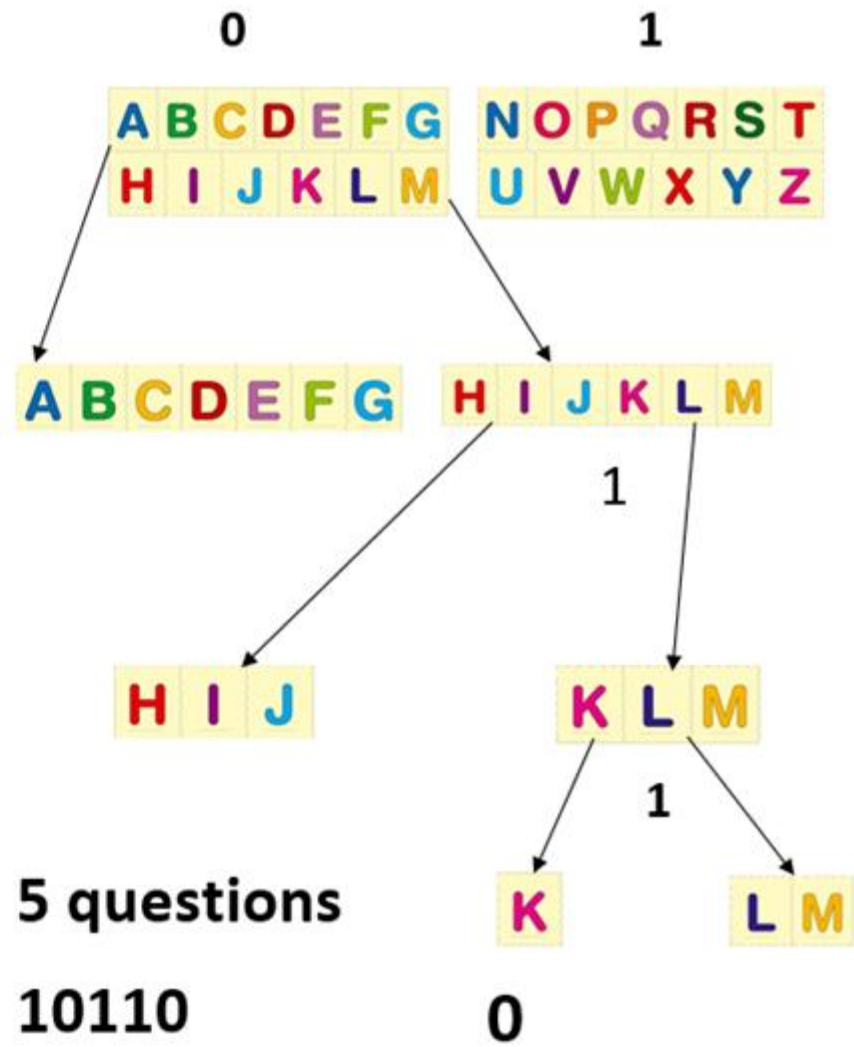
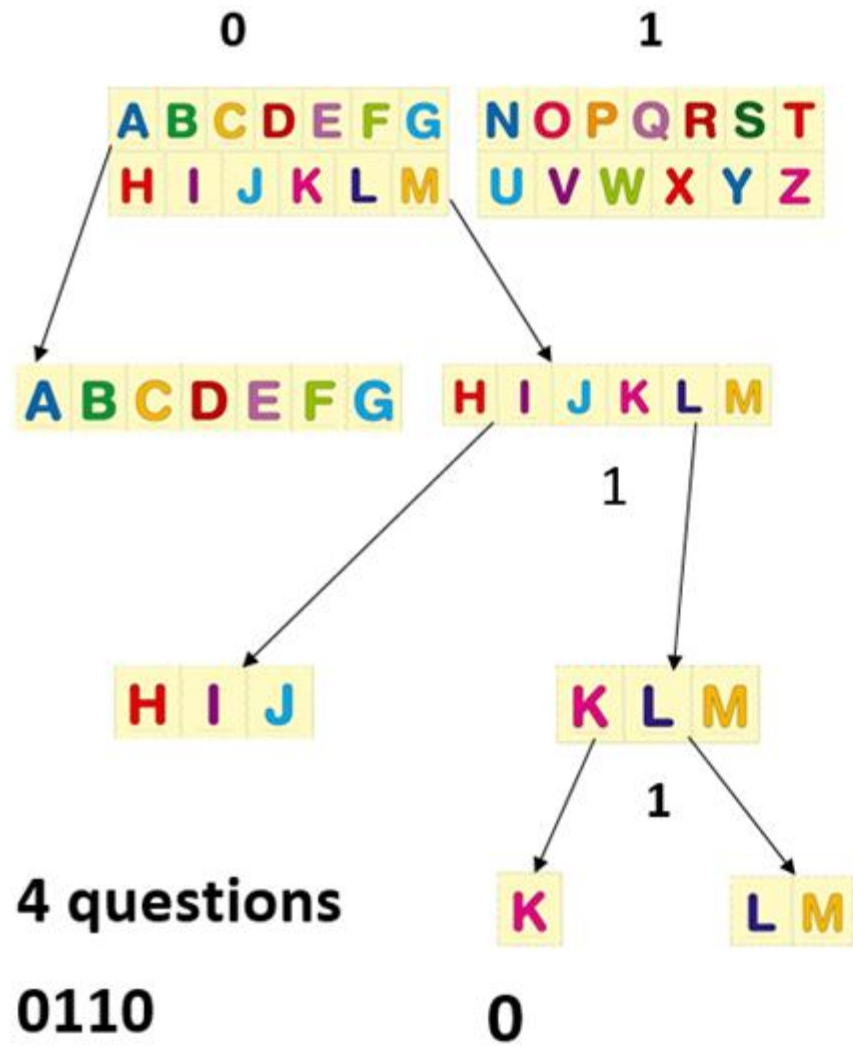


Akinator solution for **K**



Akinator solution for **L (&M)**







11010

For Byelorussian
alphabet we get
5 symbols on
letter

What language has ŭ?

1) Old Greek Πνεύμα – spirit



Esperanto **laŭ** – «according
to»

by

Lazaro Zamenhof

An artificial language that can be learned in 3 months

Адыгейский алфавит на арабской основе

составленный Берсием Умаром в 1855 г.

(Представлен Адыгейскому ИЦЦ А.А.Хатаповым)

آ	ت	خ	ذ	س	ط	غ	ز	ن
آ	ث	خ	ذ	ش	ظ	ف	ل	و
ب	ك	ج	ر	ش	ط	ق	ح	ق
پ	ج	ج	ژ	ص	ع	ك	ح	ه
پ	ح	د	ژ	ص	غ	ژ	م	ي



4->2

2=f(4)

16->4

4=f(16)

26 -> 4,66

4.66=f(26)

32 ->5

5=f(32)

33 -> 5,1

5.1=f(33)

64 -> 6

6=f(64)

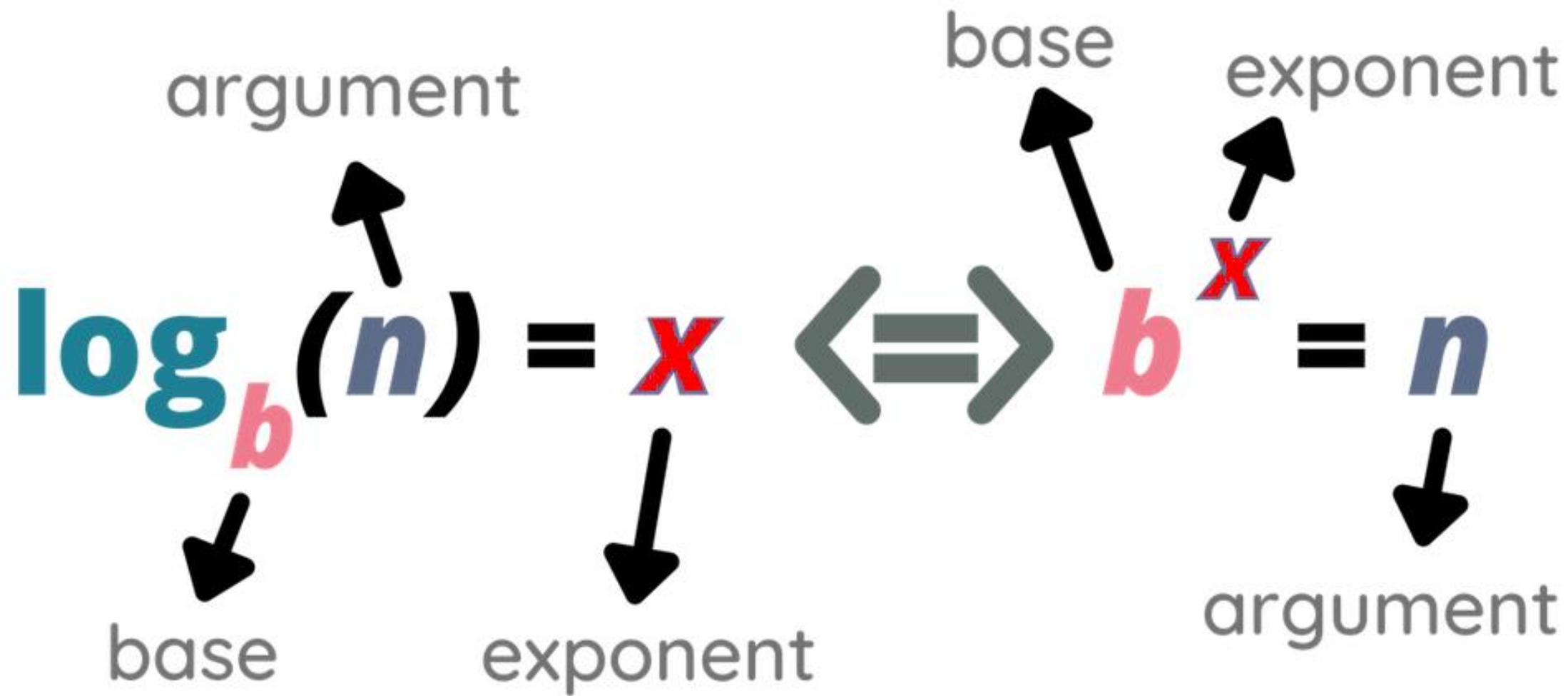
32 -> $f(x)$ -> 5

64 -> $f(x)$ -> 6

$$16 \rightarrow \sqrt[4]{f(x)} \rightarrow 4$$

$$32 \rightarrow \sqrt[5]{f(x)} \rightarrow 5$$

$$64 \rightarrow \sqrt[6]{f(x)} \rightarrow 6$$



$$2^? = 32$$

to what power must 2 be raised to get thirty-two

$$\log_2 32 = 5$$

binary algorithm from thirty-two equals five

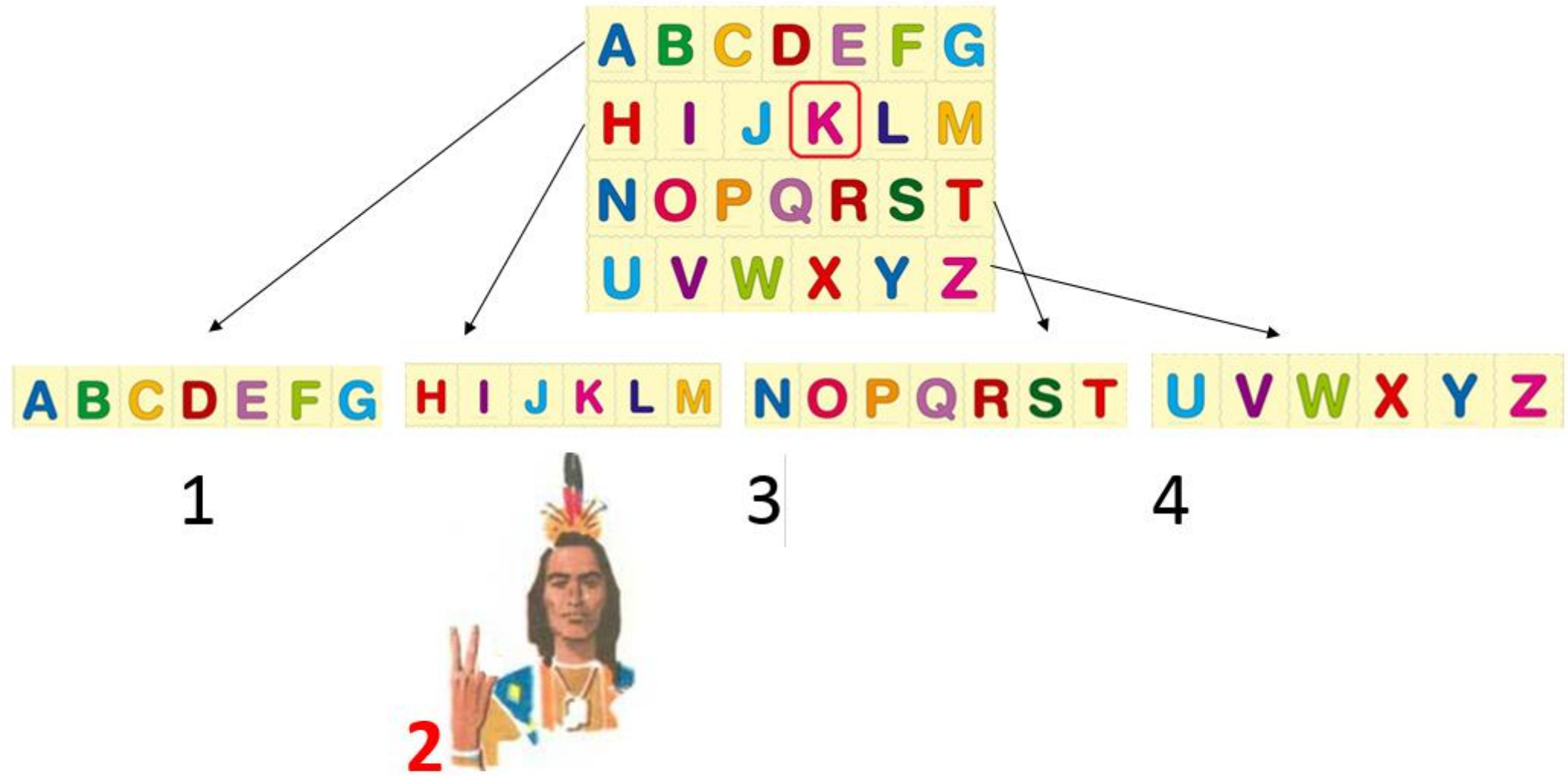
ДЗЯКУЙ

$$5 * 6 = 30$$

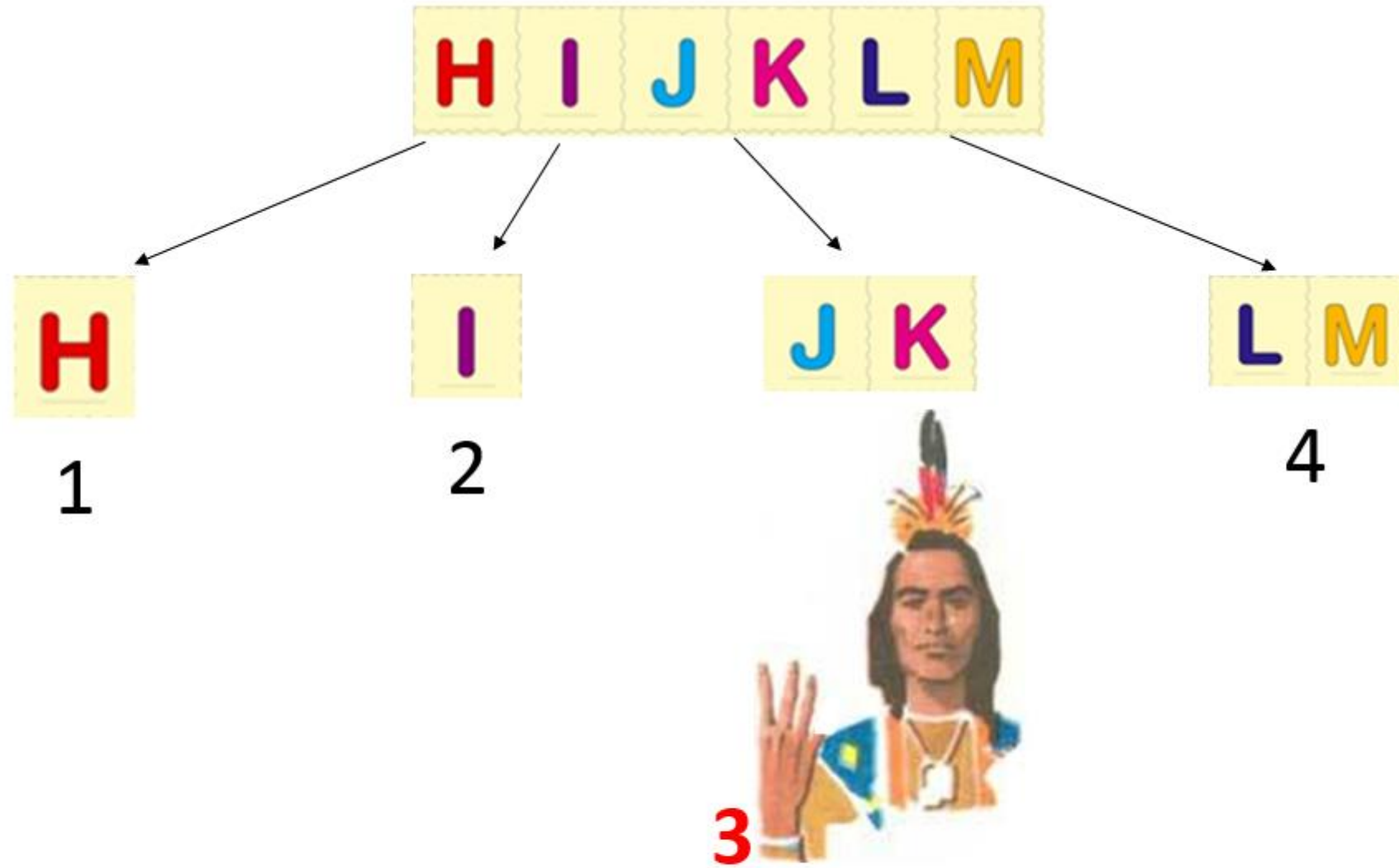
five times six equals thirty



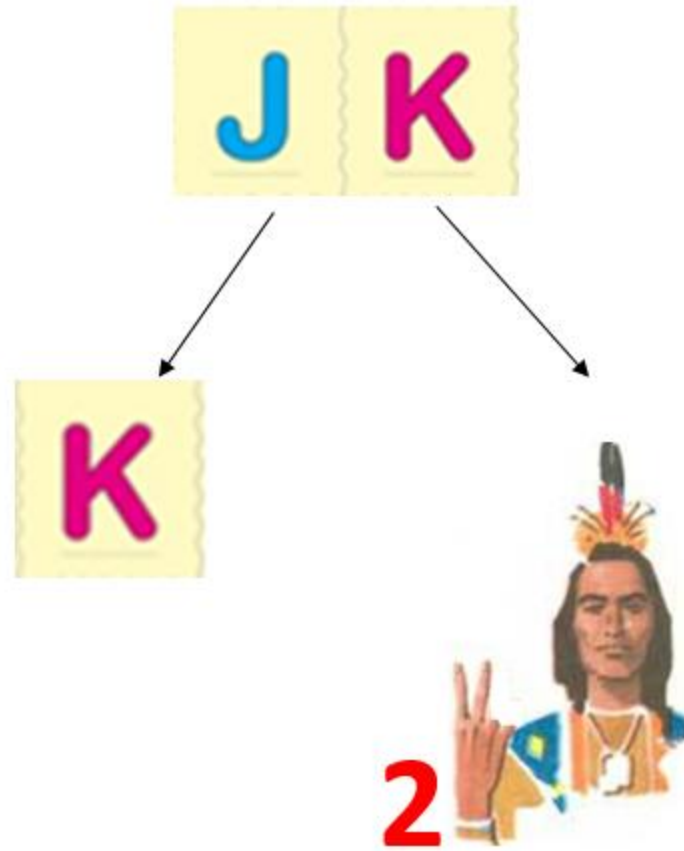
Let's try the method using 4 fingers



Step 1. Divide the selected part of the alphabet into 4 equal parts, and fix (ask a question) in which part is our letter - **2**.

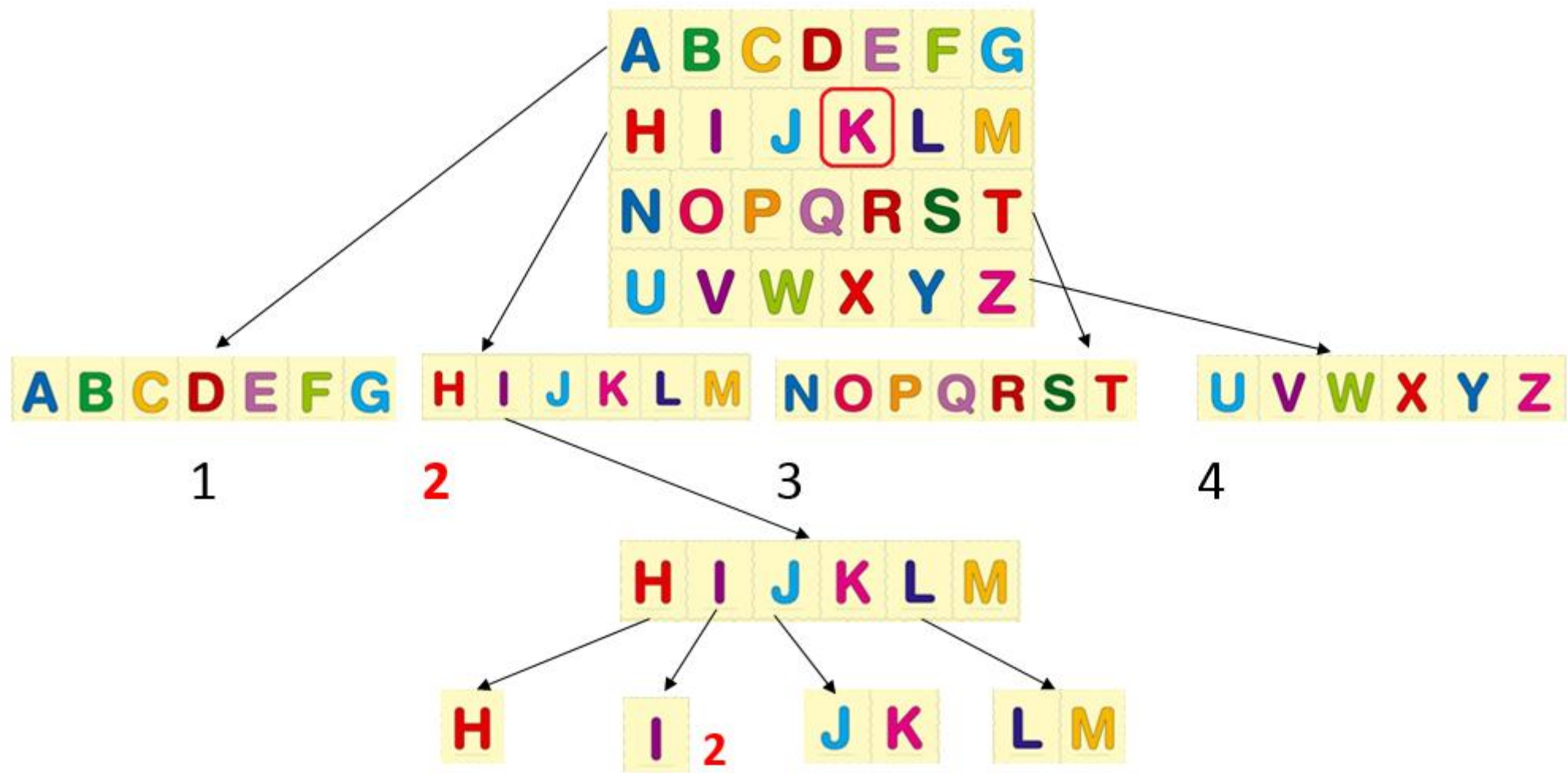


Step 2 Divide the selected part of the alphabet into 4 equal parts, and fix (ask a question) in which part is our letter - **3**.



Step 3 Divide the selected part of the alphabet into 2 equal parts, and fix (ask a question) in which part is our letter - **2**.

Thus - three numbers: two-three-two

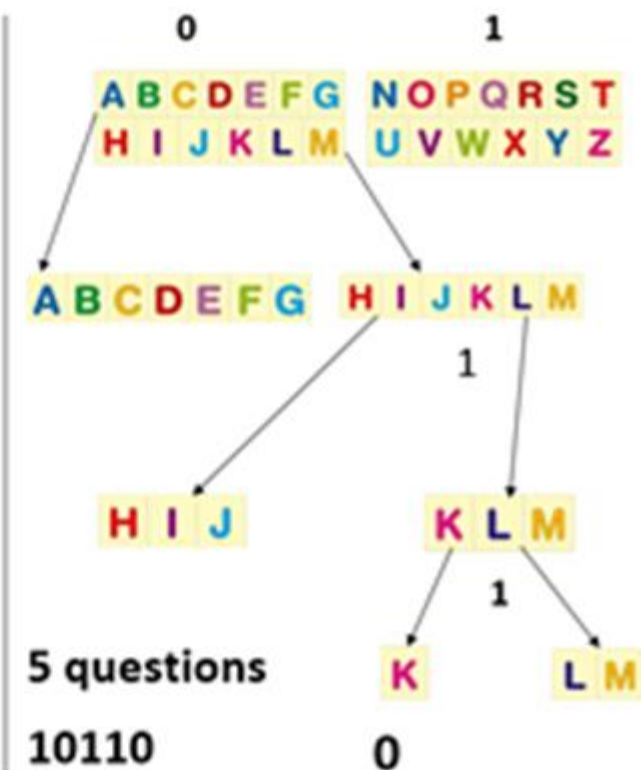
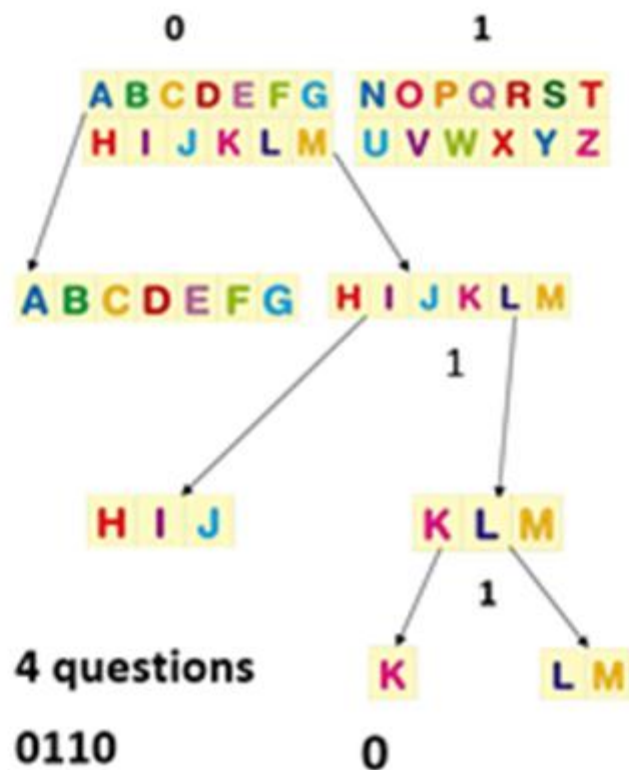


In the case of the letters **H** and **I** - we get two numbers: **2-2** for **I**
 Total Average $(2*2+4*3)/6=2.66$ $\log_4(26)=2.35$

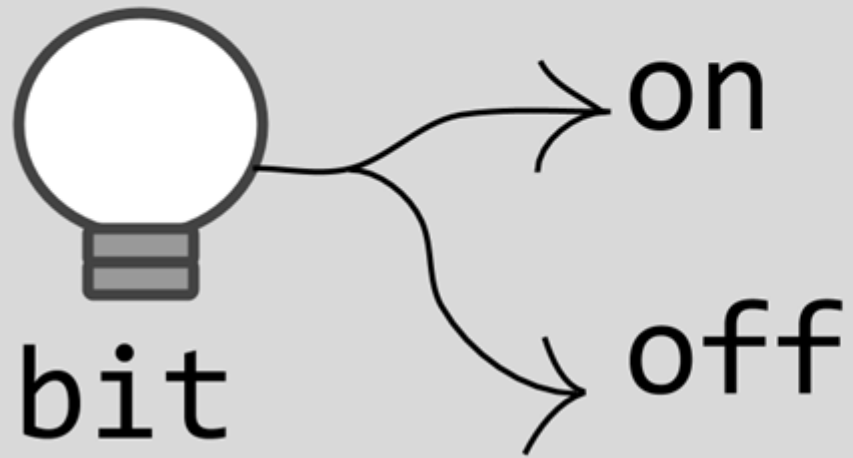
In the case of a binary algorithm
 $(4+5+5)/3=4.6666$
6 symbols per letter

$$\log_2(26)=4.7$$

$$4,7 > 4.666$$



Bit = Binary Digit



Computer Bit



Computer Byte



<http://www.computerhope.com>

Place values

(multiply this number by the 1 or 0 in its place)

128	64	32	16	8	4	2	1
x	x	x	x	x	x	x	x
1	0	1	1	0	1	0	1
=	=	=	=	=	=	=	=

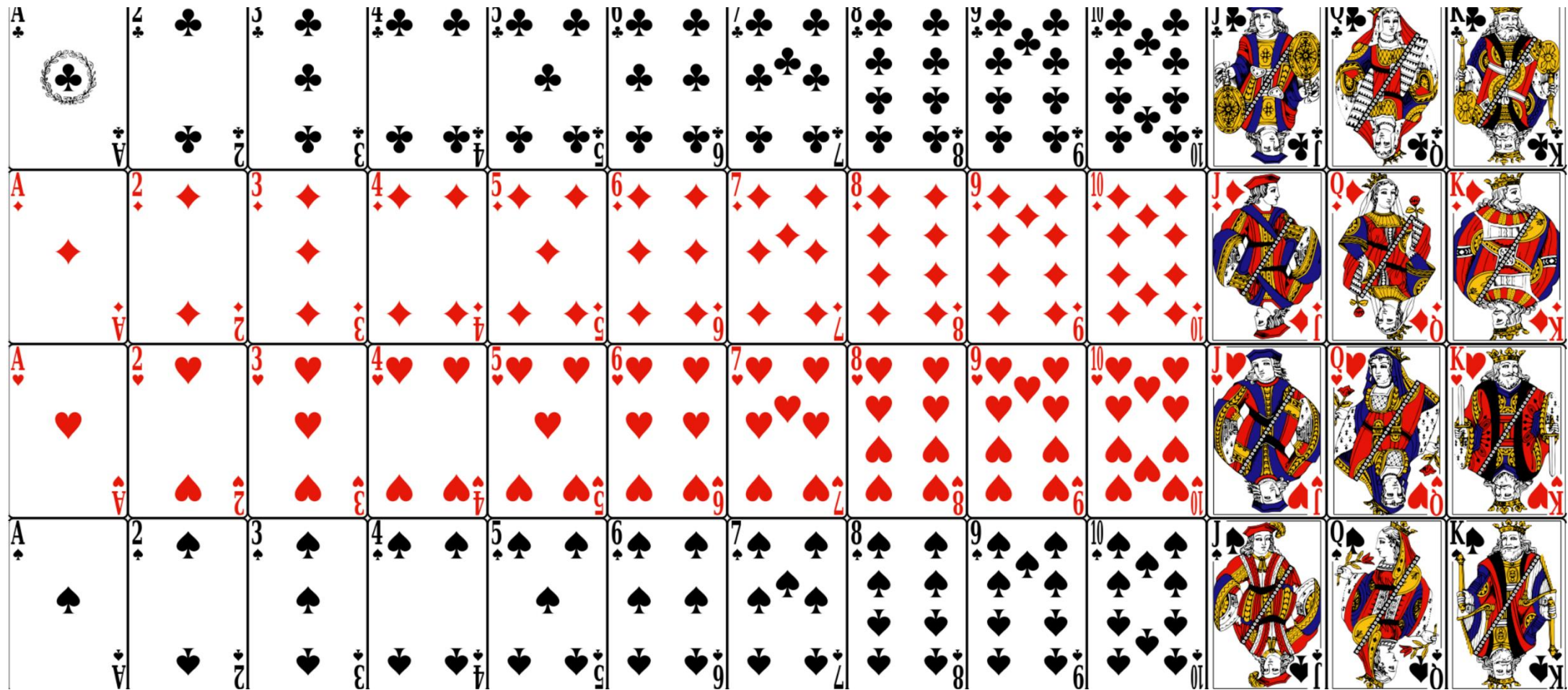
$$128 + 0 + 32 + 16 + 0 + 4 + 0 + 1$$

(add all these together to get the decimal number)

$$= 181$$

Bridge combination

52 cards



$$? = \log_2 52 = 5,7 \text{ per card}$$

Those. The correct answer is guessed in 5-6 questions, an average of **5.7** questions.

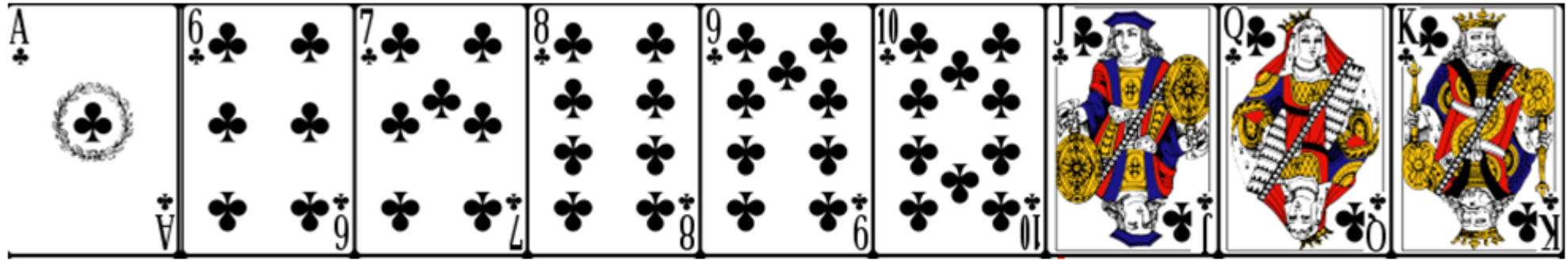


The bridge combination consists of 13 cards,

$$13 * 5,7 = 74.1$$

74.1 questions

The unit of measure is bits.



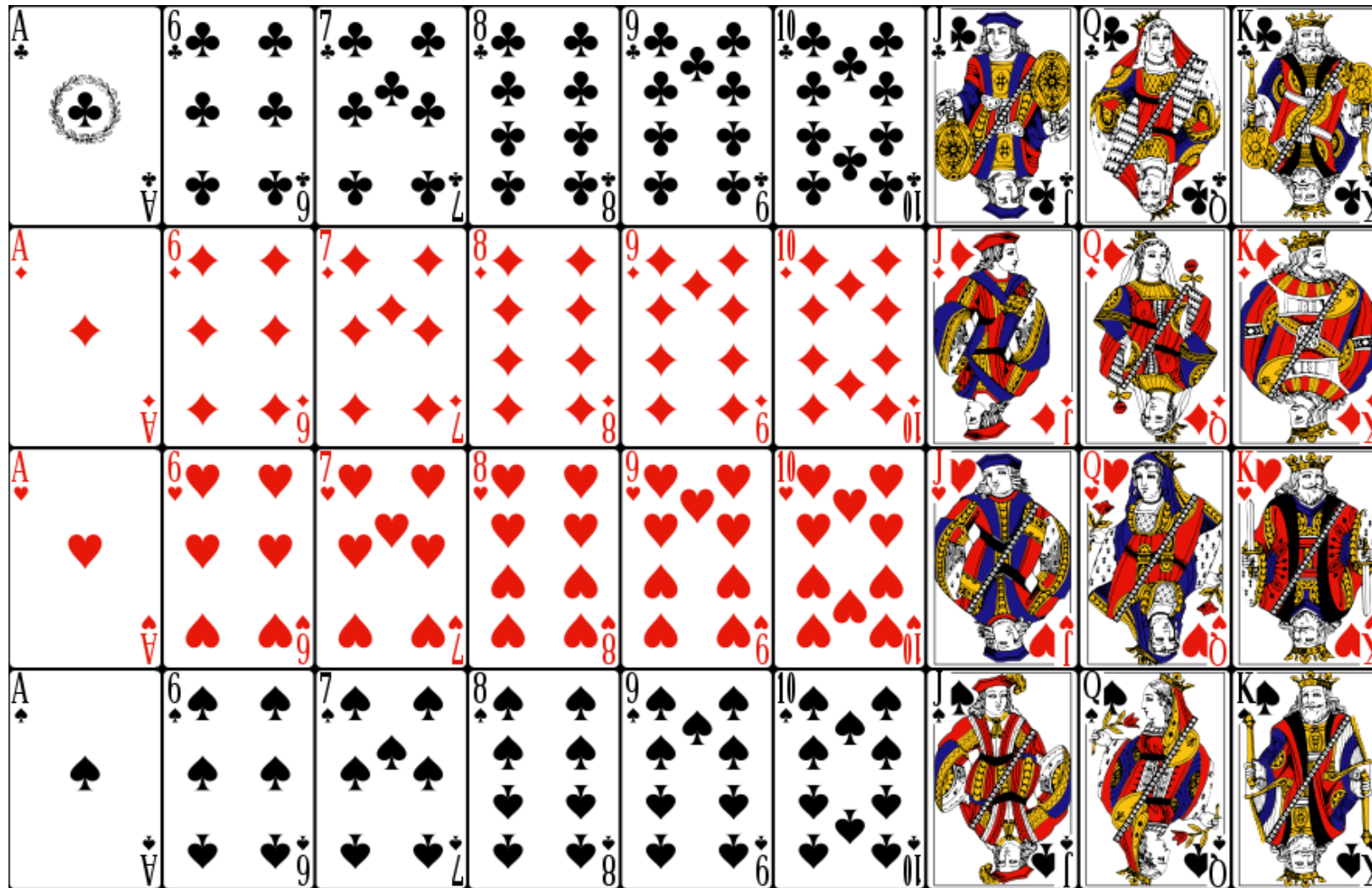
card weight 1-9

$$\log_2 10 = 3.303$$

Total Average $(8*3+2*4)/10=3.2$

3.303=3.2 Not `equal –
but LET IT BE.





Card from
the
preference
deck
 $9 \times 4 = 36$,
how much
informatio
n is there
(in bits)

POKER HAND RANKINGS



1 ROYAL FLUSH



4 FULL HOUSE



2 STRAIGHT FLUSH



5 FLUSH



3 FOUR OF A KIND



6 STRAIGHT



7 THREE OF A KIND



8 TWO PAIR



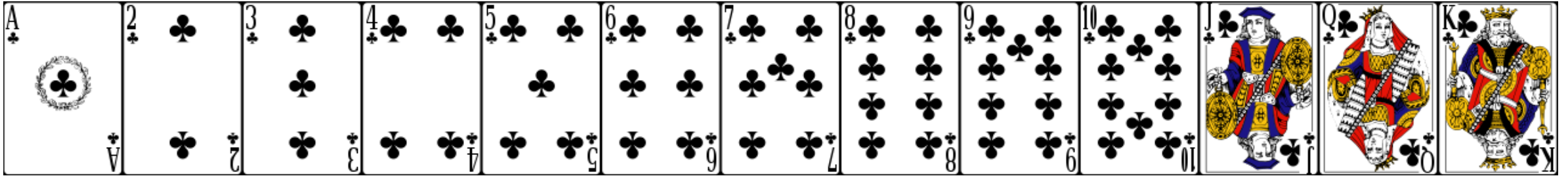
9 PAIR



10 HIGH CARD



How much information is there (in bits)



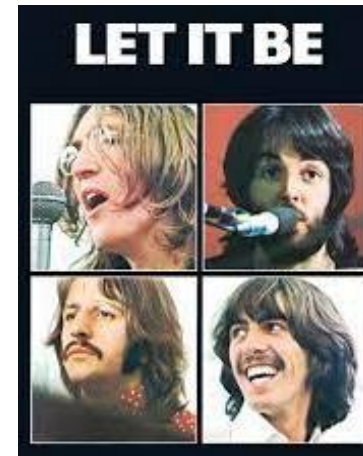
card weight 1-13

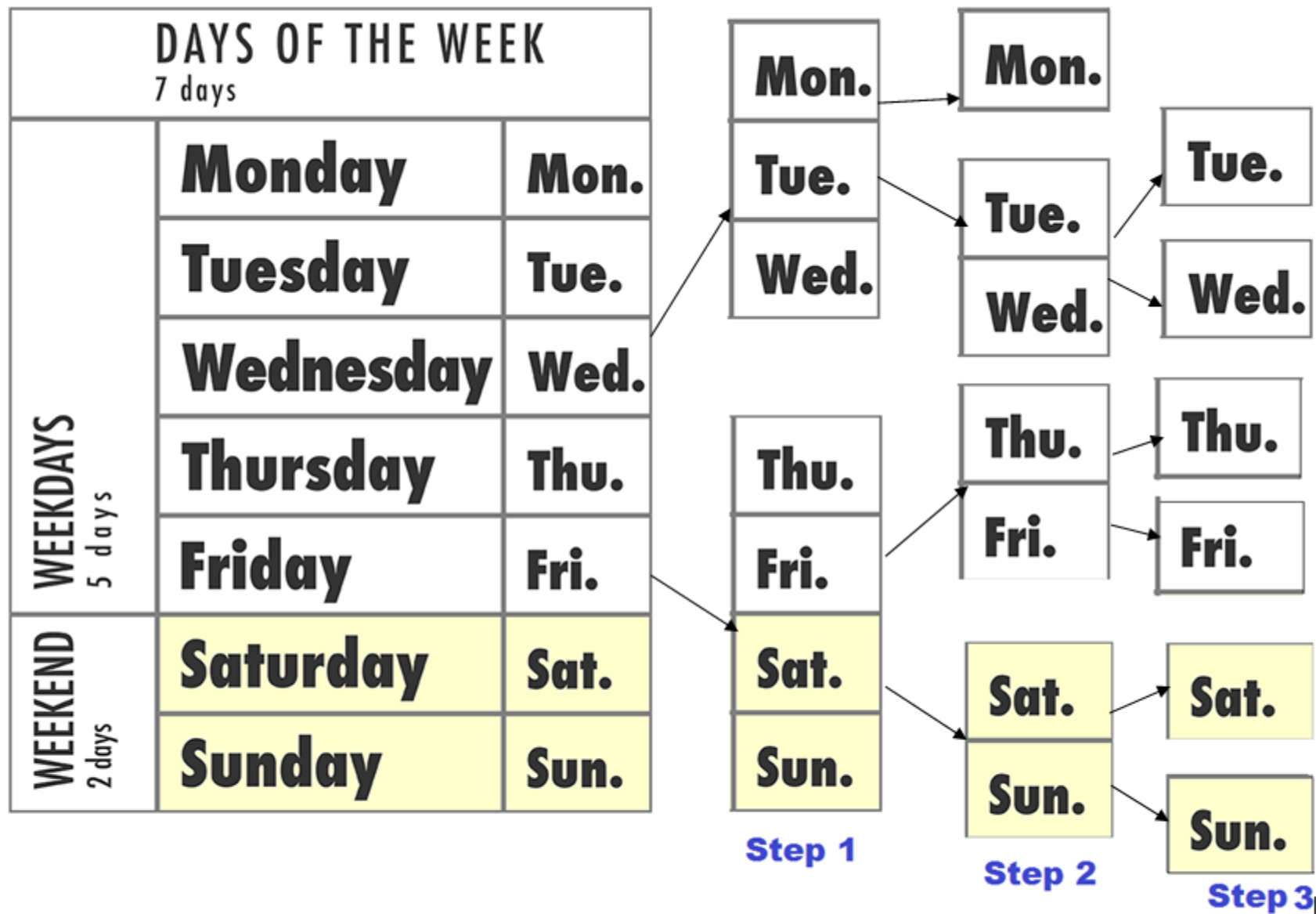
How much information is there (in bits)

$$\log_2 13 = 3.700$$

$$\text{Total Average } (3 \cdot 3 + 10 \cdot 4) / 13 = 3.767$$

$$3.700 = 3.767 \text{ Not equal - but LET IT BE.}$$

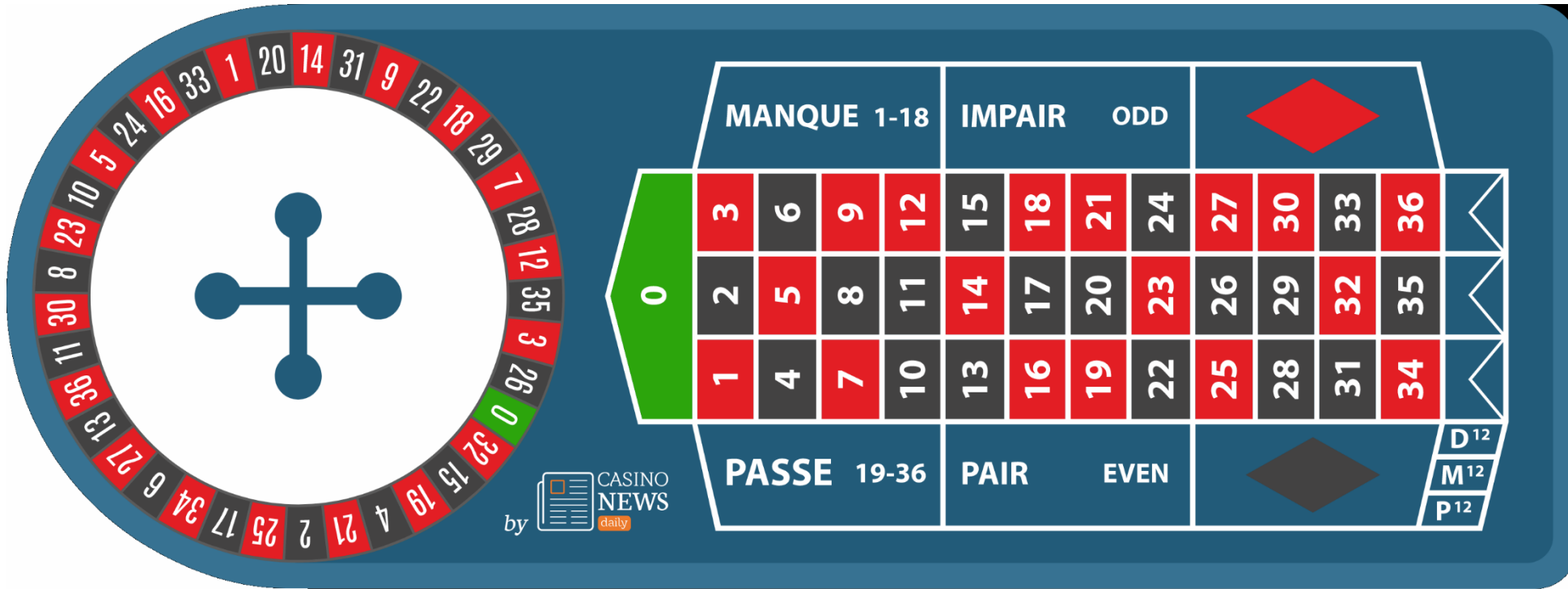




$\log_2 7 = 2,807$ Total Average $(1*2+6*3)/7=2.857!=2,807$ but Let it be






Exam grade

Оценка по 10-балльной шкале	Оценка по 5-балльной шкале за экзамен	Оценка, используемая в европейском приложении к диплому		Оценка за зачет
10	отлично	A +	Excellent	зачтено
9	отлично	A	Very good	зачтено
8	отлично	A -	Very good	зачтено
7	хорошо	B +	Good	зачтено
6	хорошо	B -	Good	зачтено
5	удовлетворительно	C +	Satisfactory	зачтено
4	удовлетворительно	C -	Satisfactory	зачтено
3	неудовлетворительно	F	Fail	не зачтено
2	неудовлетворительно	F	Fail	не зачтено
1	неудовлетворительно	F	Fail	не зачтено



FRENCH ROULETTE

by  CASINO NEWS daily

MANQUE 1-18				IMPAIR ODD								
3	6	9	12	15	18	21	24	27	30	33	36	  
2	5	8	11	14	17	20	23	26	29	32	35	
1	4	7	10	13	16	19	22	25	28	31	34	
PASSE 19-36				PAIR EVEN								
												D ¹²
												M ¹²
												P ¹²

$$74.1 = 13 * 5.7 = 13 * \log_2 52 =$$



$$H = n * \log_2 S = \log_2 S^n =$$



$$\log_2$$

$$74.1 = 13 * 5.7 = 13 * \log_2 52 =$$



$$H = n * \log_2 S = \log_2 S^n =$$

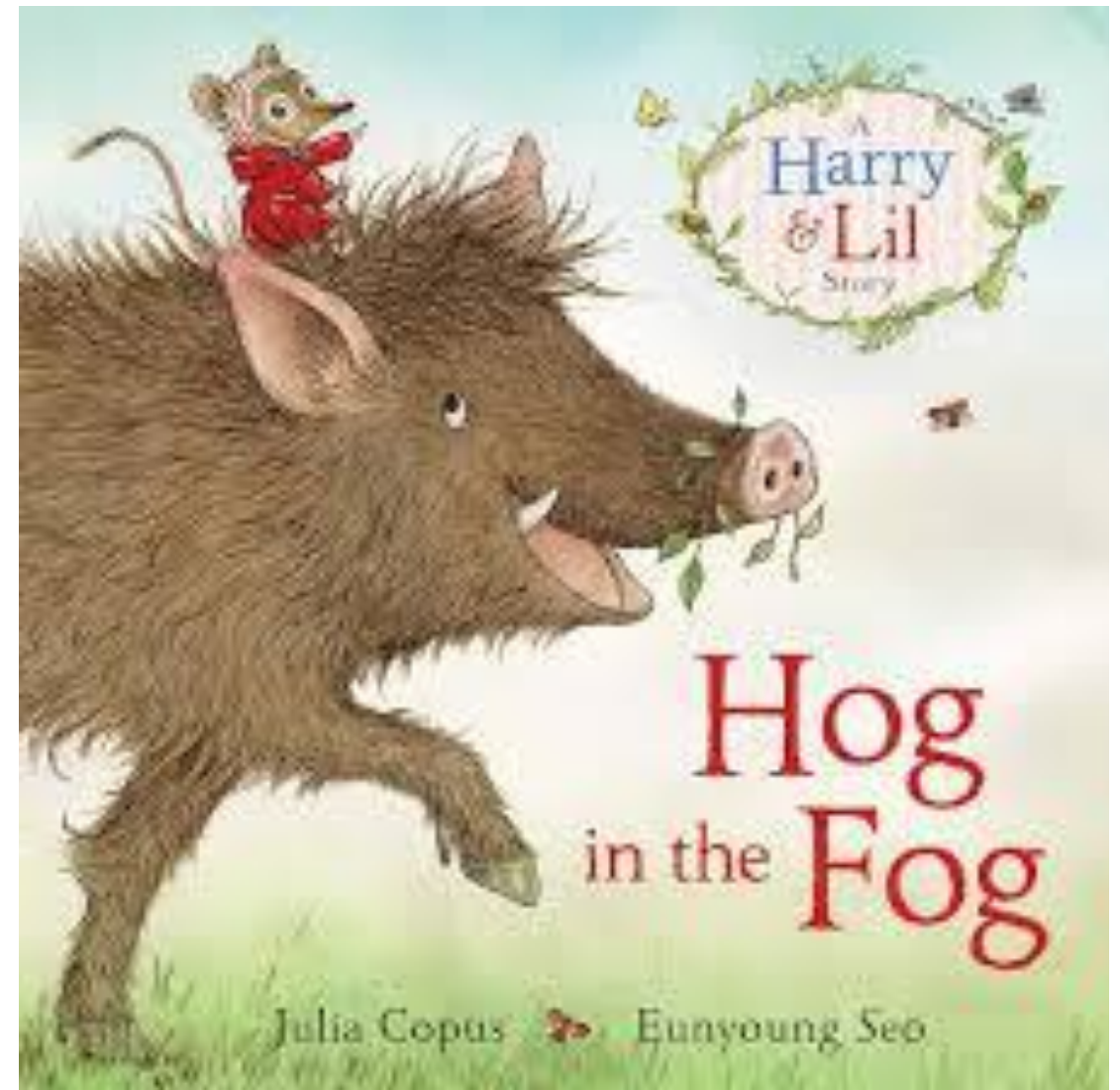


$$H = n * \log_2 S$$



Entropy H_0 amount of information per message symbol







2 gays is **good**

It's easier to make a choice

low entropy

$$\log_2 2 = 1$$



8 gays is **bad**

It's not easier to make a choice

high entropy

$$\log_2 8 = 3$$

Abba - Dancing Queen

<https://www.youtube.com/watch?v=xFrGuyw1V8s>





Picky Bride's problem



The solution of the problem

1. Divide the applicants into 3 parts



Say **NO** to the first





Say **YES** to the second if he is better than the first



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



<https://youtu.be/dUJqf8emQ5I?t=32>

- Nadia is 28 years old, here you will marry not only San Sanych, you will marry a goat.

This is bad advice.





2 gays is **good**

It's easier to make a choice

low entropy

$$\log_2 2 = 1$$



8 gays is **bad**

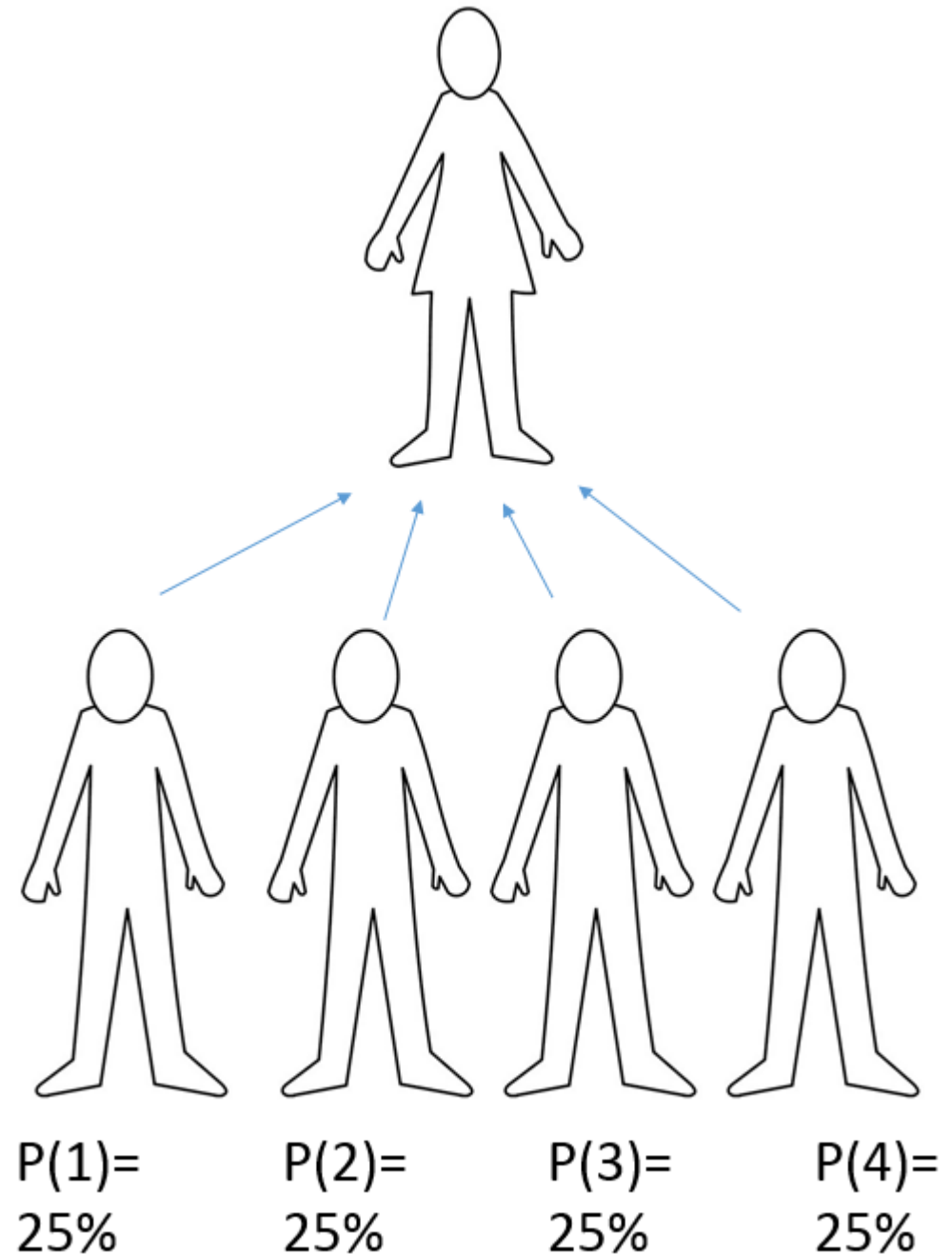
It's not easier to make a choice

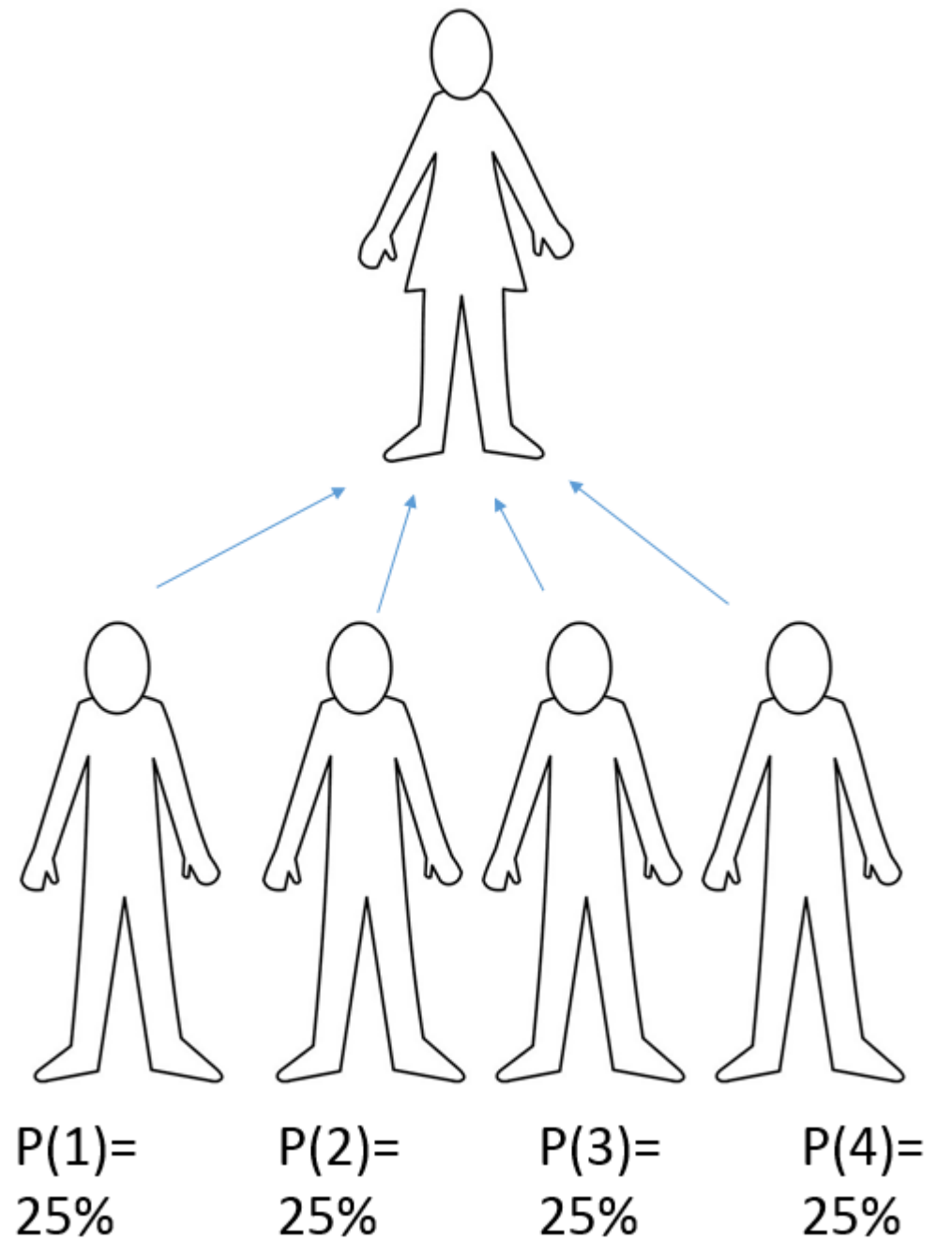
high entropy

$$\log_2 8 = 3$$

The princess is proposed to 4 applicants, but she does not know anything about any of them.

Each groom has a 25% chance of being chosen - isn't it true?







As expected, the suitors of the princess are **billionaires**

50 billion

25 billion

12.5 billion

12.5 billion



Mark
Zuckerberg

Sergey
Brin

Stefan
Quandt

Prince Al
Saud

$P(1)=$

$P(2)=$

$P(3)=$

$P(4)=$

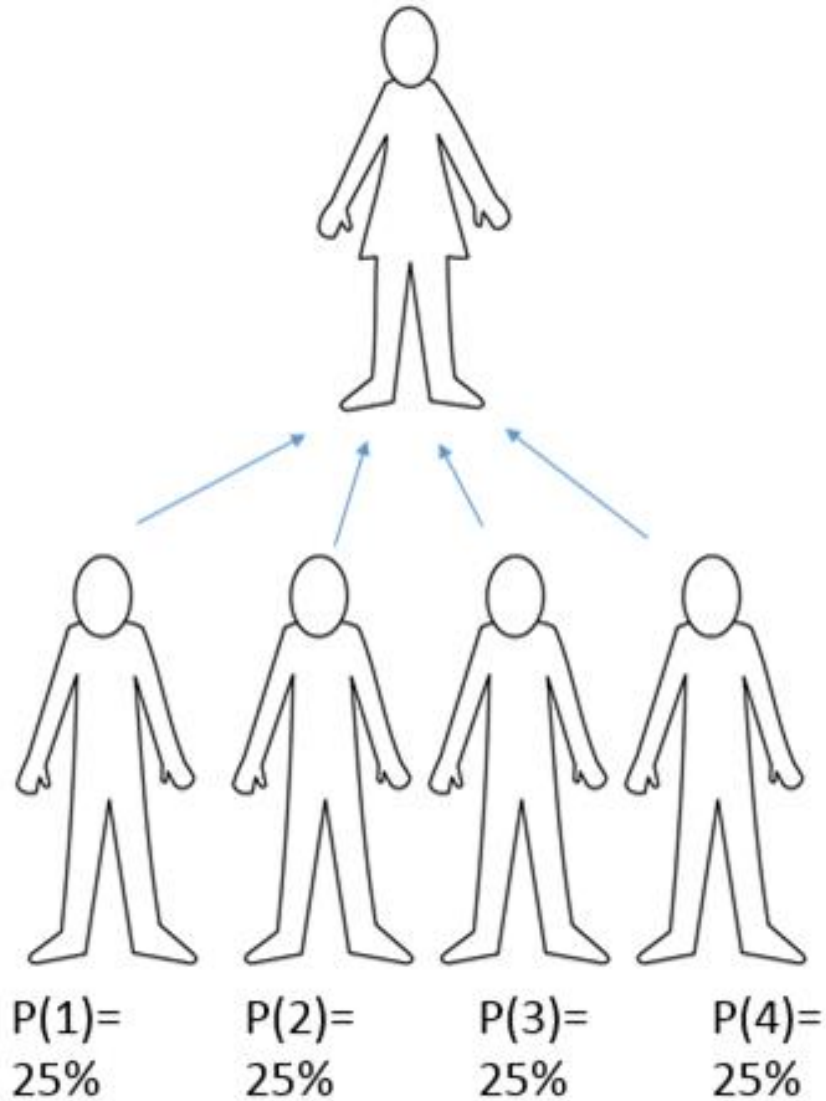
50%

25%

12,5%

12,5%

In which case will the choice be more difficult for the princess?



The World's Billionaires [\[billionaire.com\]](#) by 3 March 2017

50 billion

25 billion

12.5 billion

12.5 billion



Mark
Zuckerberg

Sergey
Brin

Stefan
Quandt

Prince Al
Saud

$P(1)=$
50%

$P(2)=$
25%

$P(3)=$
12,5%

$P(4)=$
12,5%

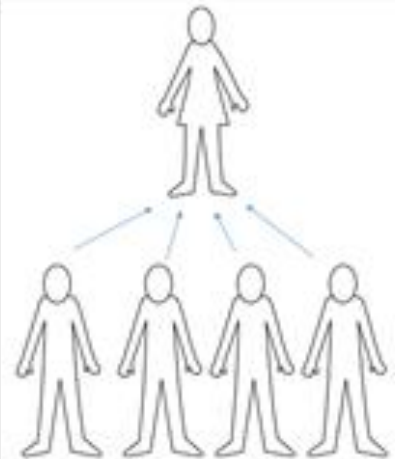


Claude Shannon
1916-2001

Claude Shannon cleverly reformulated the question - if you want to predict the choice of the princess, how many questions and binary answers you need to ask?



In the first case &

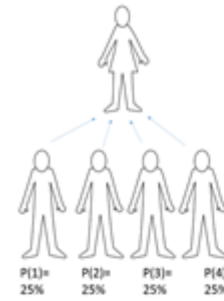


P(1)= 25% P(2)= 25% P(3)= 25% P(4)= 25%

In the second case.

The World's Billionaires by 3 March 2017

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











1 question - the first or second group				
				Question 1
X				
2 question - the first or second group				
				Question 2
	X			

Average number of questions =





$$2 \cdot 0.25 + 2 \cdot 0.25 + 2 \cdot 0,25 + 2 \cdot 0,25$$

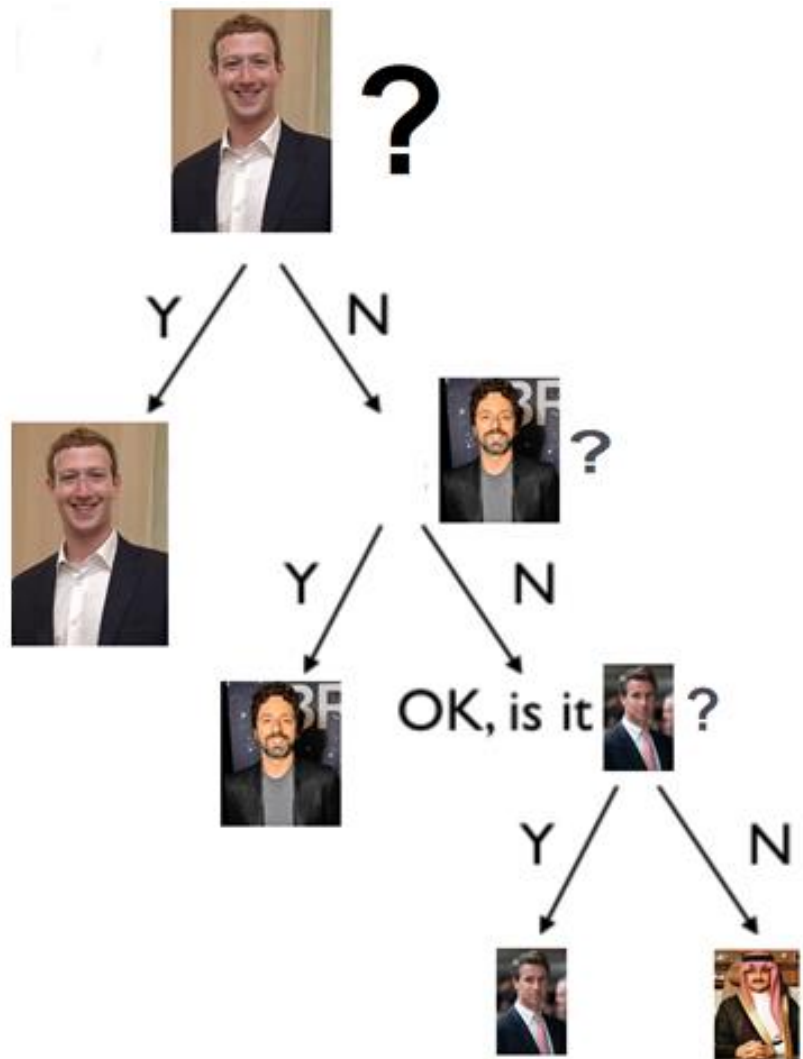
$$= 2$$

Question 1. Is this Zuckerberg?			
 50%			
HET			
Question 2. Is this Sergey Brin?			
 25%			
no			
Question 3. Is this Stefan from BMW?			
 12,5%			12,5%
no		So Prince Saud	

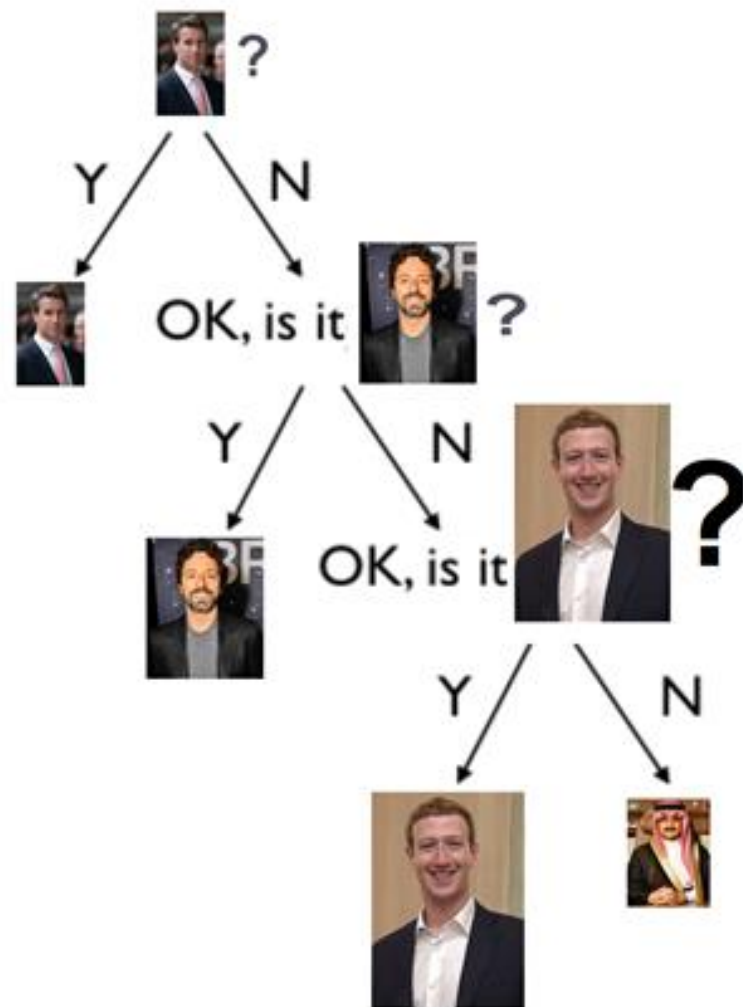
Average number of questions =

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


1 question		$1*0.5$
2 question		$2*0.25$
3 question		$3*0,125$
3 question		$3*0,125$
Average number of questions =		1,75







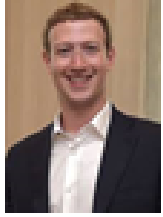



$1 \cdot 0.5 +$	$2 \cdot 0.25 +$	$3 \cdot 0,125 +$	$3 \cdot 0,125$
$= 1,75$			



$1 \cdot 0.125 +$	$2 \cdot 0.25 +$	$3 \cdot 0,5 +$	$3 \cdot 0,125$
$= 2,5$			

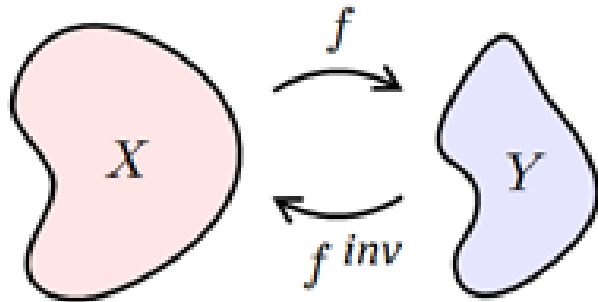
1 question		$f(0.5)=1$
2 question		$f(0.25)=2$
3 question		$f(0,125)=3$
3 question		$f(0,125)=3$
Average number of questions =		1,75

1		$f(0.5)=1$	$f(1/2)=1$	one second turns into one
2		$f(0.25)=2$	$f(1/4)=2$	one fourth becomes two
3		$f(0,125)=3$	$f(1/8)=3$	one eighth becomes three
3		$f(0,125)=3$	$f(1/8)=3$	one eighth becomes three
=		1,75		

1		$f(0.5)=1$	$f(1/2)=1$	$b(2)\rightarrow 1$	two becomes one
2		$f(0.25)=2$	$f(1/4)=2$	$b(4)\rightarrow 2$	4 becomes 2
3		$f(0,125)=3$	$f(1/8)=3$	$b(8)\rightarrow 3$	8 becomes 3
3		$f(0,125)=3$	$f(1/8)=3$	$b(8)\rightarrow 3$	8 becomes 3
		1,75		$b(2)\rightarrow ?$	

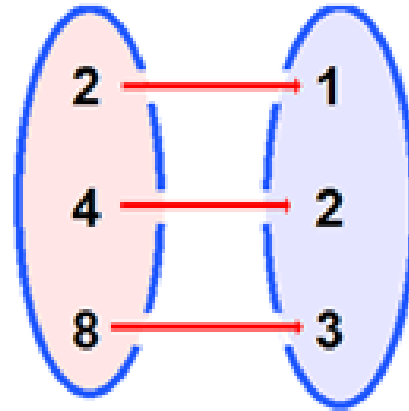
Inverse function

If
 f maps X to Y ,

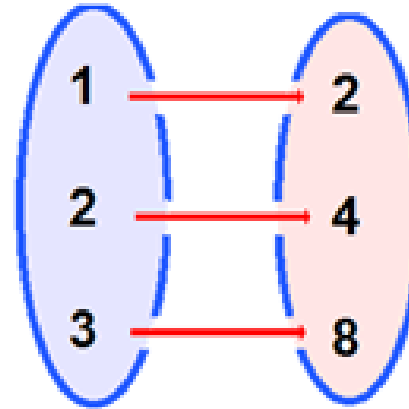


then
 f^{-1} maps Y back to X

b



b^{-1}



Function $f(x)$	Inverse $f^{-1}(y)$
$x + a$	$y - a$
$a - x$	$a - y$
mx	y/m
$1/x$ (i.e. x^{-1})	$1/y$ (i.e. y^{-1})
x^2	(i.e. $y^{1/2}$)
x^3	(i.e. $y^{1/3}$)

Information Theory

A Tutorial Introduction

James V Stone

$$H(X) = -\sum p(x_i) \log_2 p(x_i)$$



Shannon called this
part

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

in a scientific way.





Quantifying Information (Claude Shannon, 1948)

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

Information is measured in bits
(binary digits) number of 0/1's
required to encode choice(s)

$$\log_2 \left(\frac{1}{8} \right) = 3$$

$$\log_2 \left(\frac{1}{4} \right) = 2$$


	Quantifying information			
	$I(x_i) = \log_2\left(\frac{1}{p_i}\right)$			
 40%	$\log_2\left(\frac{1}{0.4}\right)$	= 1,322	$1,322 * 0.4 +$	0.5288+
 20%	$\log_2\left(\frac{1}{0.2}\right)$	= 2,322	$2,322 * 0.2 +$	0.4644+
 20%	$\log_2\left(\frac{1}{0.2}\right)$	= 2,322	$2,322 * 0.2 +$	0.4644+
 20%	$\log_2\left(\frac{1}{0.2}\right)$	= 2,322	$2,322 * 0.2$	0.4644
				1.922

Quantifying information

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



60%

 10%

 10%

 10%

$$\log_2\left(\frac{1}{0.6}\right)$$

$$\log_2\left(\frac{1}{0.1}\right)$$

$$\log_2\left(\frac{1}{0.1}\right)$$

$$\log_2\left(\frac{1}{0.1}\right)$$

$$1,66 * 0.4 +$$

$$3,322 * 0.1 +$$

$$3,322 * 0.1 +$$

$$3,322 * 0.1$$

$$0.664 +$$

$$0.3322 +$$

$$0.3322 +$$

$$0.3322$$

$$1.6606$$

That's what mathematicians did.
They Decided to measure entropy according to
Shannon's formula.

$$\sum_{i=1}^N p(i) * \log_2\left(\frac{1}{p(i)}\right)$$

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

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

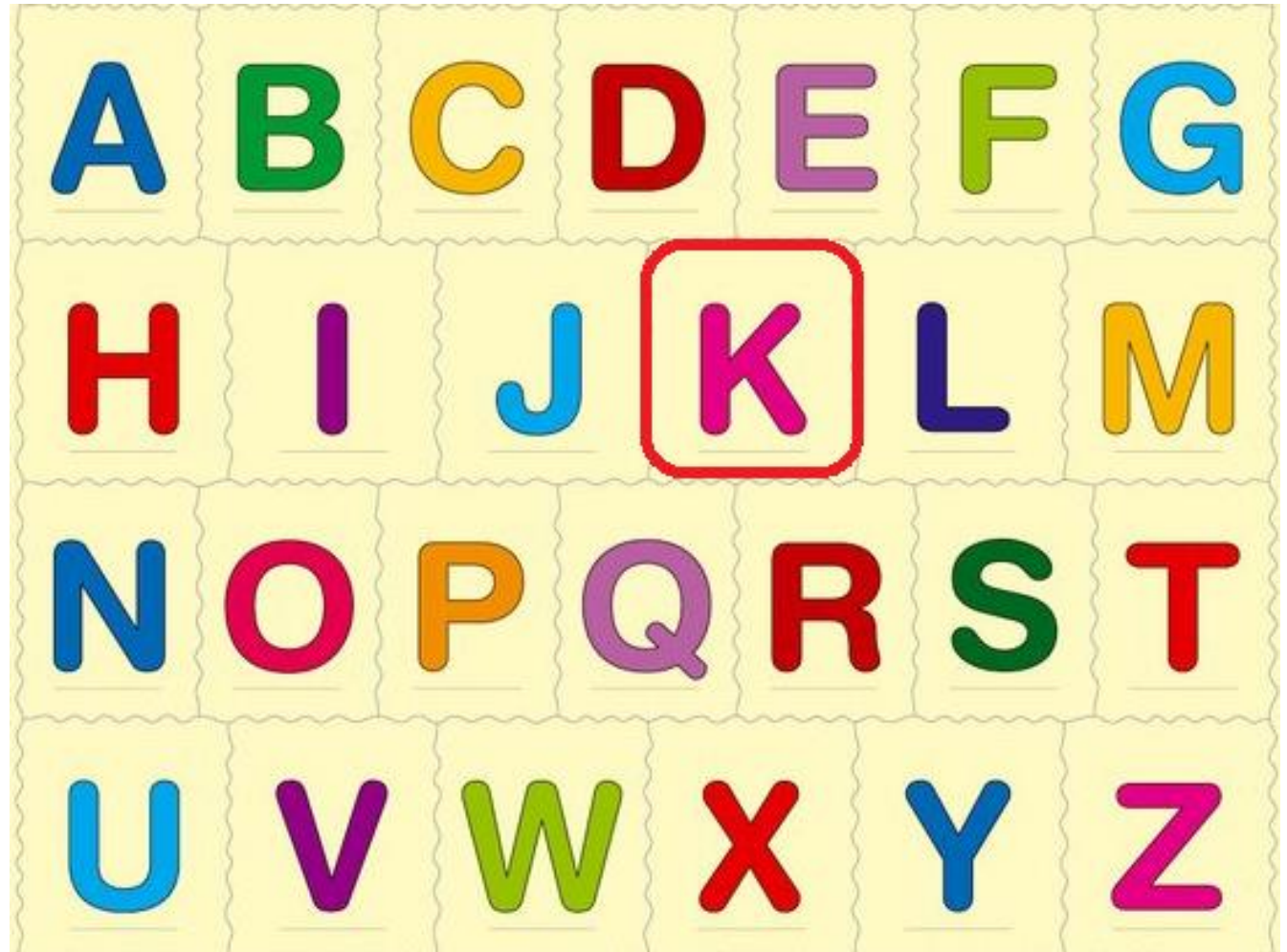
$$H(X) = - \sum_{i=1}^n p(x_i) \log_b p(x_i)$$

Let's repeat

How much
information is
contained in the
Belarusian
alphabet?



How much information is contained in the English alphabet?



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 - - ● ●