ILLC Project Course in Information Theory

Crash course

13 January – 17 January 2014
12:00 to 14:00

Student presentations

27 January – 31 January 2014
12:00 to 14:00

Location

ILLC, room F1.15,
Science Park 107, Amsterdam

Materials

informationtheory.weebly.com

Contact

Mathias Winther Madsen
mathias.winther@gmail.com

Monday

Probability theory
Uncertainty and coding

Tuesday

The weak law of large numbers
The source coding theorem

Wednesday

Random processes
Arithmetic coding

Thursday

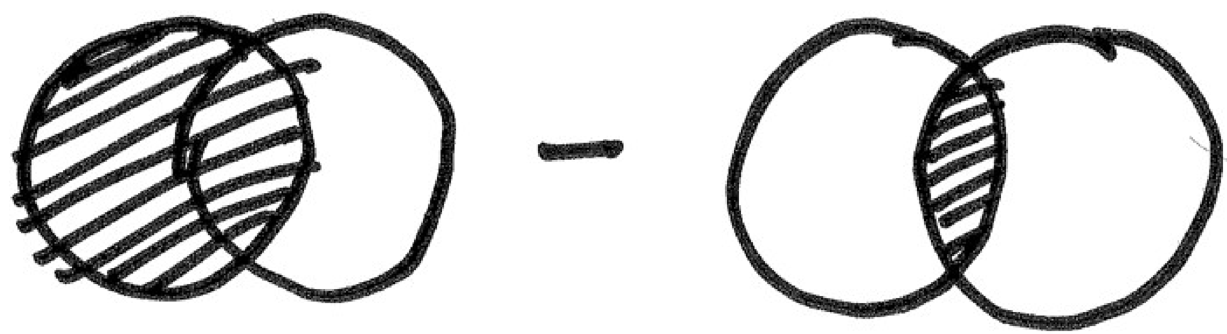
Divergence
Kelly Gambling

Friday

Kolmogorov Complexity
The limits of statistics

Basic idea:

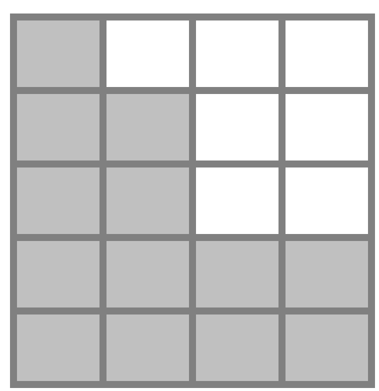
Amount of Information = Prior uncertainty − Posterior uncertainty



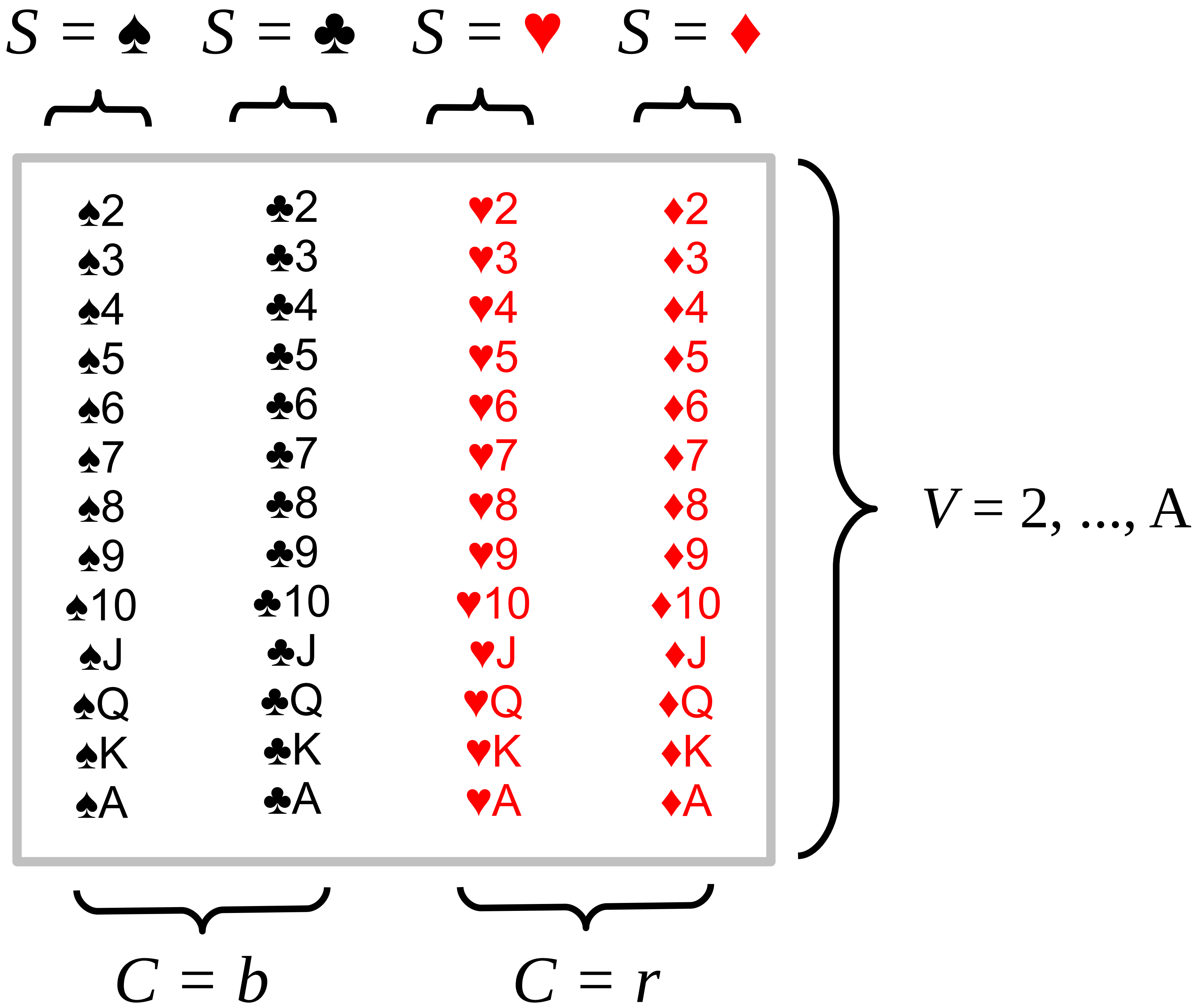
$$I(X;\varphi) \quad = \quad H(X) \quad - \quad H(X|\varphi)$$

First suggestion:

$$\text{Uncertainty} == \log\left(\begin{array}{c}\textbf{Number of}\\ \textbf{remaining}\\ \textbf{possibilities}\end{array}\right)$$

H () == $\log 7$

R. V. L. Hartley: “Transmission of Information,”
Bell System Technical Journal, 1928



$S = ♠$
$S = ♣$
$S = ♥$
$S = ♦$
♠2 ♠3 ♠4 ♠5 ♠6 ♠7 ♠8 ♠9 ♠10 ♠J ♠Q ♠K ♠A
♣2 ♣3 ♣4 ♣5 ♣6 ♣7 ♣8 ♣9 ♣10 ♣J ♣Q ♣K ♣A
♥2 ♥3 ♥4 ♥5 ♥6 ♥7 ♥8 ♥9 ♥10 ♥J ♥Q ♥K ♥A
♦2 ♦3 ♦4 ♦5 ♦6 ♦7 ♦8 ♦9 ♦10 ♦J ♦Q ♦K ♦A
$V = 2, \ldots, \mathrm{A}$
$C = b$
$C = r$

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

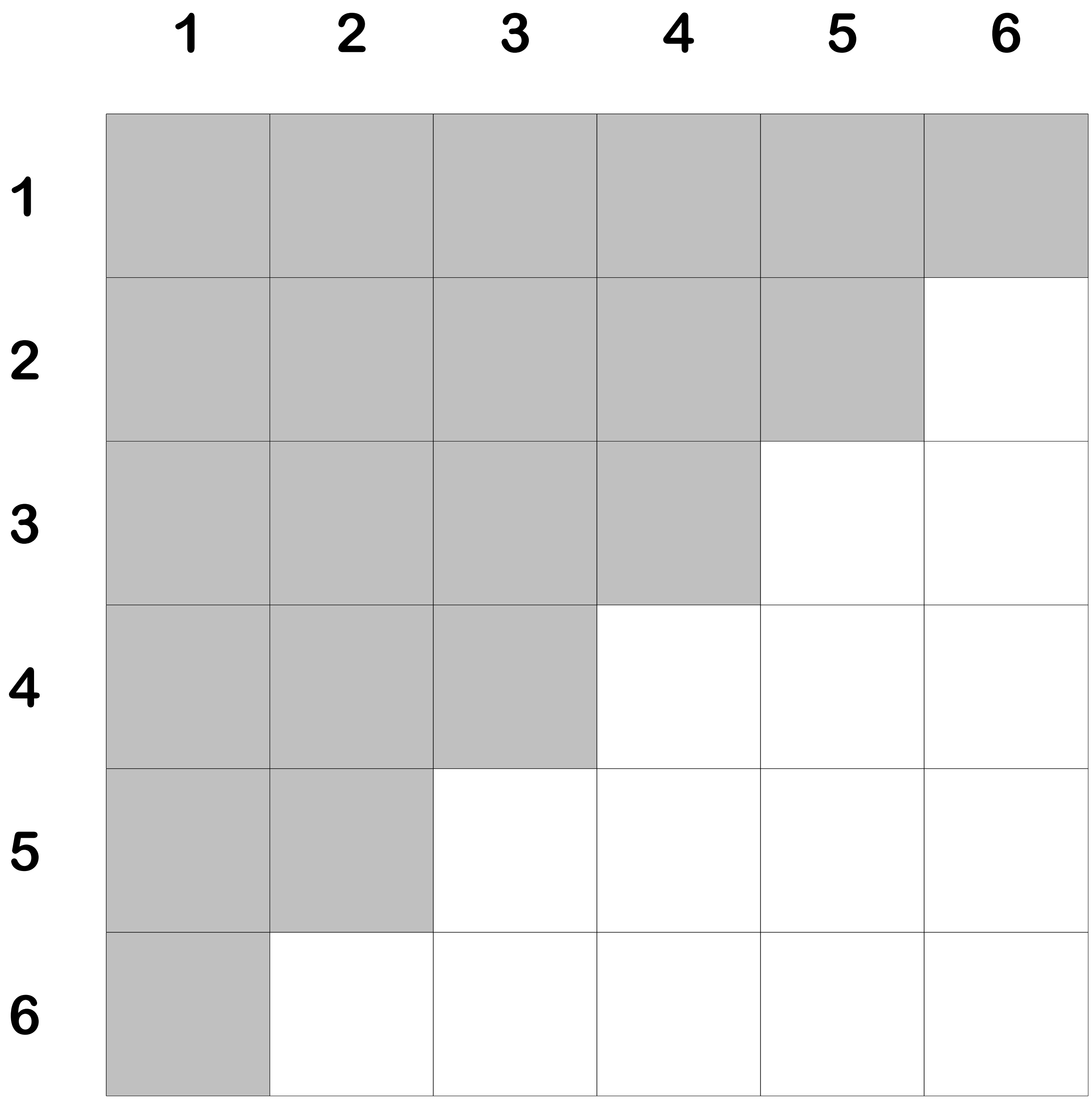
$H(X)$? $H(Y)$? $H(X \quad Y)$?

	1	2	3	4	5	6
1	000 000	000 001	000 010	000 011	000 100	000 101
2	000 110	000 111	001 000	001 001	001 010	001 011
3	001 100	001 101	001 110	001 111	010 000	010 001
4	010 010	010 011	010 100	010 101	010 110	010 111
5	011 000	011 001	011 010	011 011	011 100	011 101
6	011 110	011 111	100 000	100 001	100 010	100 011

$$H(\mathrm{X} \quad \mathrm{Y}) == 5.17$$

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

$$H(X \quad Y \mid X + Y > 7)?$$

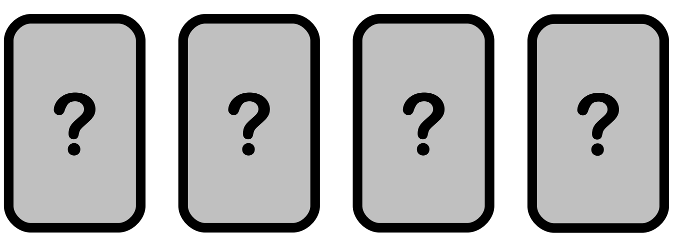


$$H(X \quad Y | X + Y > 7)?$$

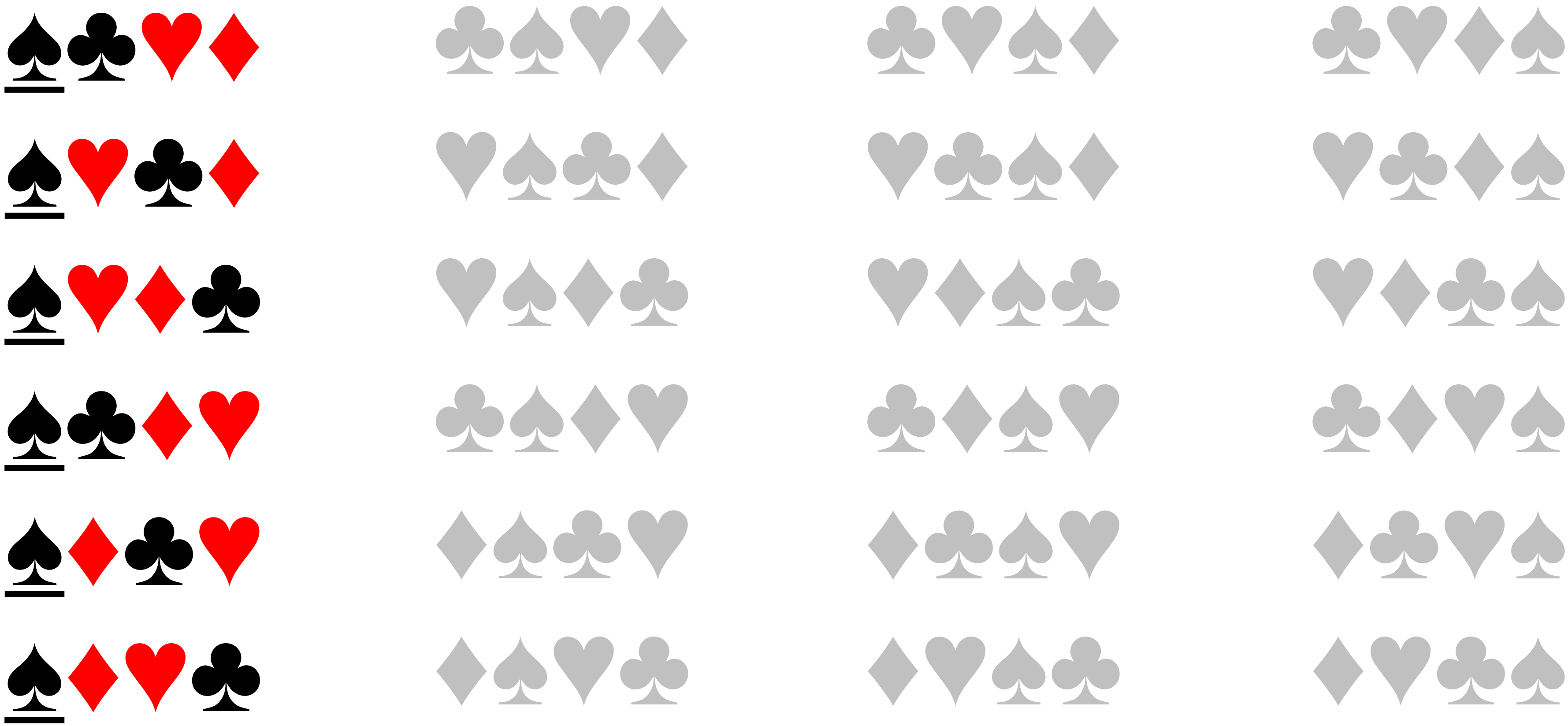
	1	2	3	4	5	6
1						
2						00 00
3					00 01	00 10
4				00 11	01 00	01 01
5			01 10	01 11	10 00	10 01
6		10 10	10 11	11 00	11 01	11 10

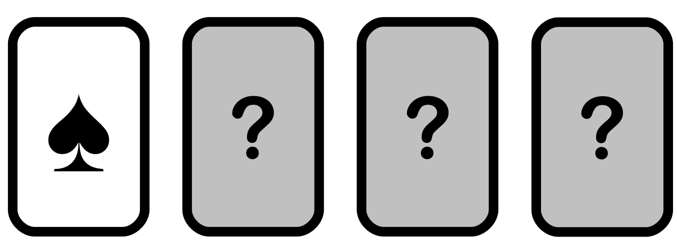
$$H(X \quad Y | X + Y > 7) \;==\; 3.9$$

♠♣♥♦	♣♠♥♦	♣♥♠♦	♣♥♦♠
♠♥♣♦	♥♠♣♦	♥♣♠♦	♥♣♦♠
♠♥♦♣	♥♠♦♣	♥♦♠♣	♥♦♣♠
♠♣♦♥	♣♠♦♥	♣♦♠♥	♣♦♥♠
♠♦♣♥	♦♠♣♥	♦♣♠♥	♦♣♥♠
♠♦♥♣	♦♠♥♣	♦♥♠♣	♦♥♣♠

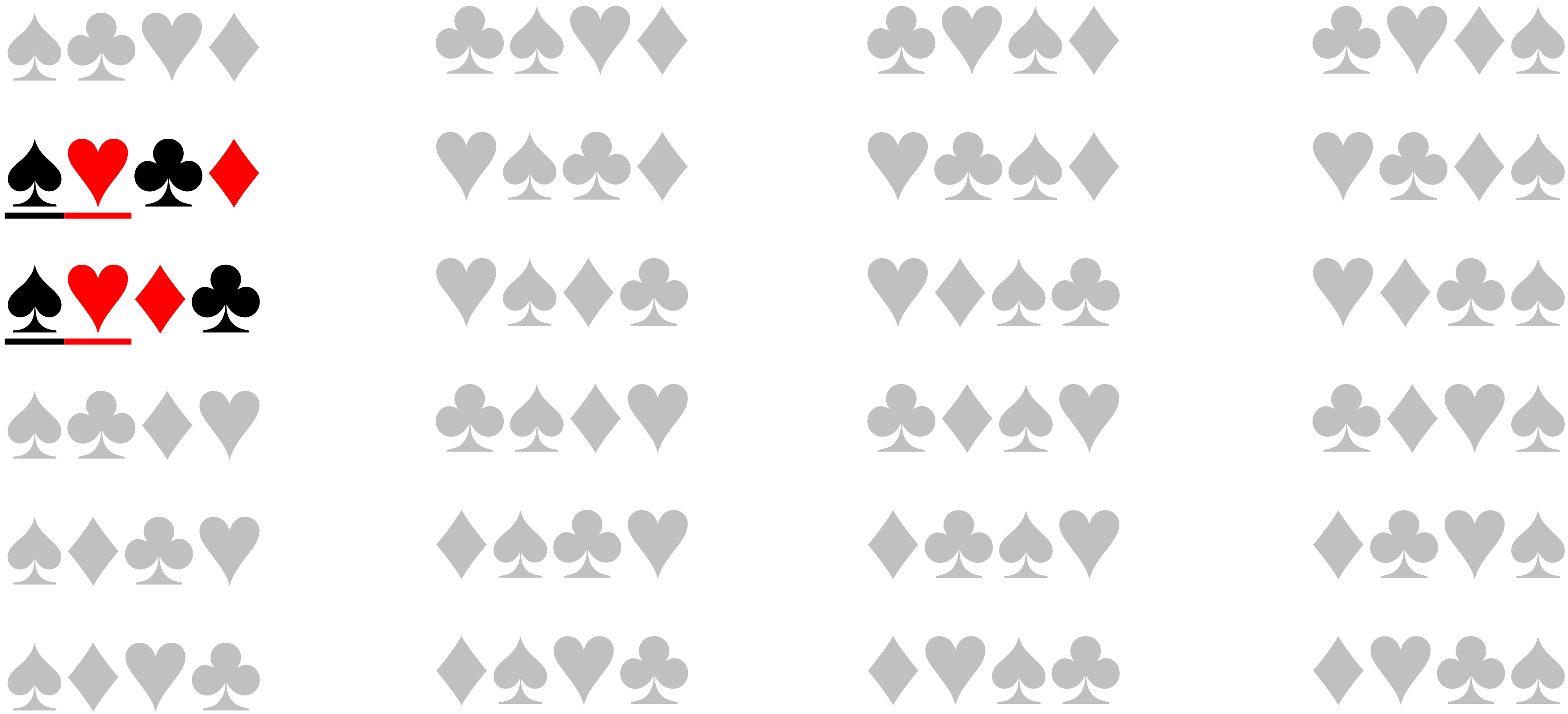


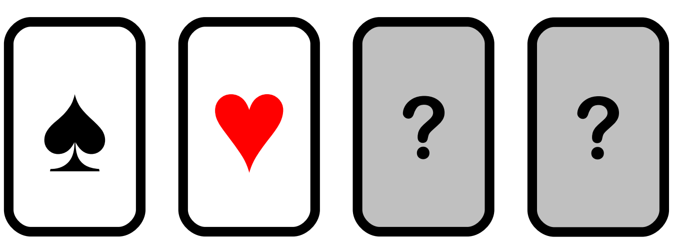
$$H == \log 24 == 4.58$$



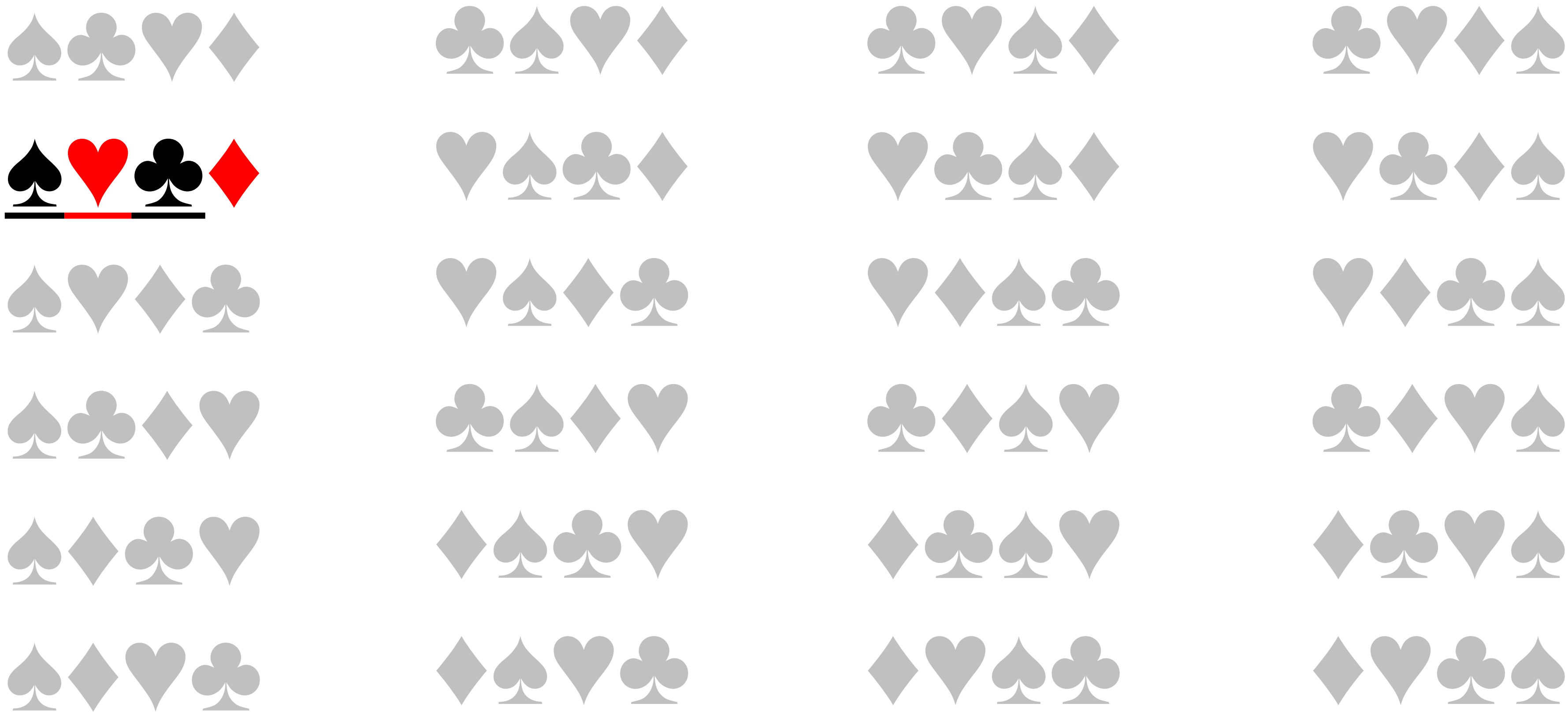


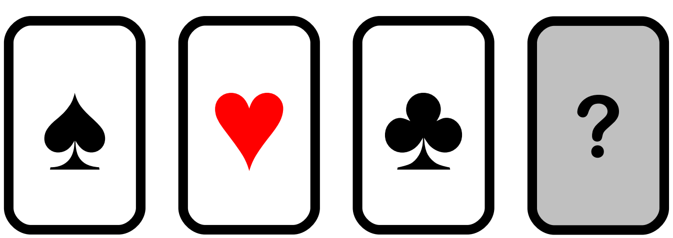
$$H \;==\; \log 6 \;==\; 2.58$$



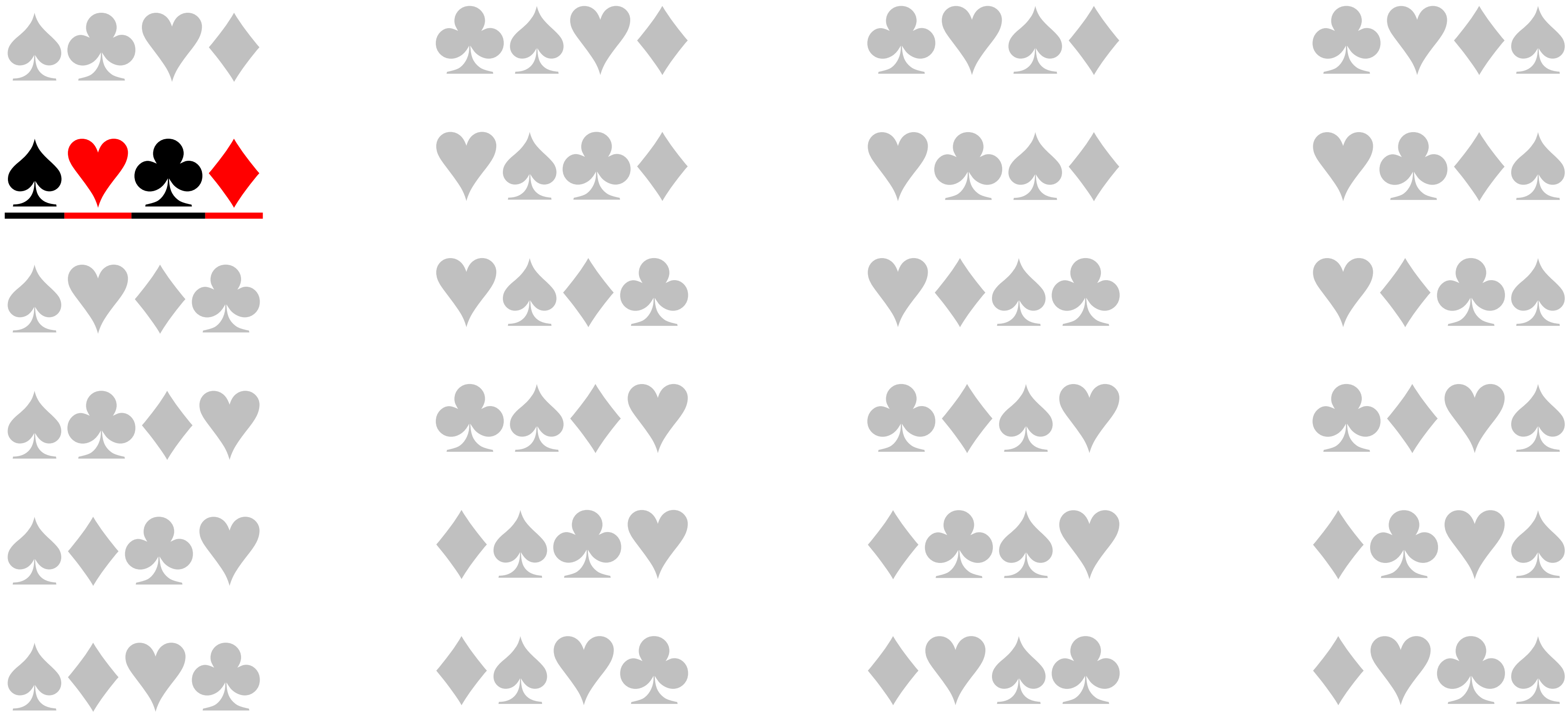


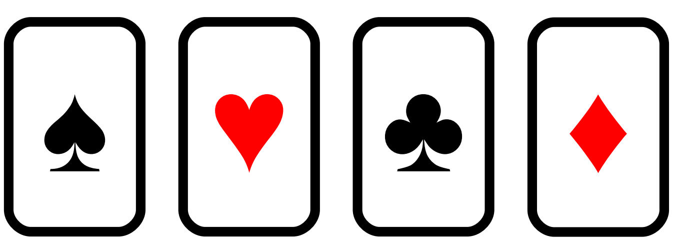
$$H \;==\; \log 2 \;==\; 1$$





$$H == \log 1 == 0$$





$$H == \log 1 == 0$$

Nice properties of the Hartley measure:

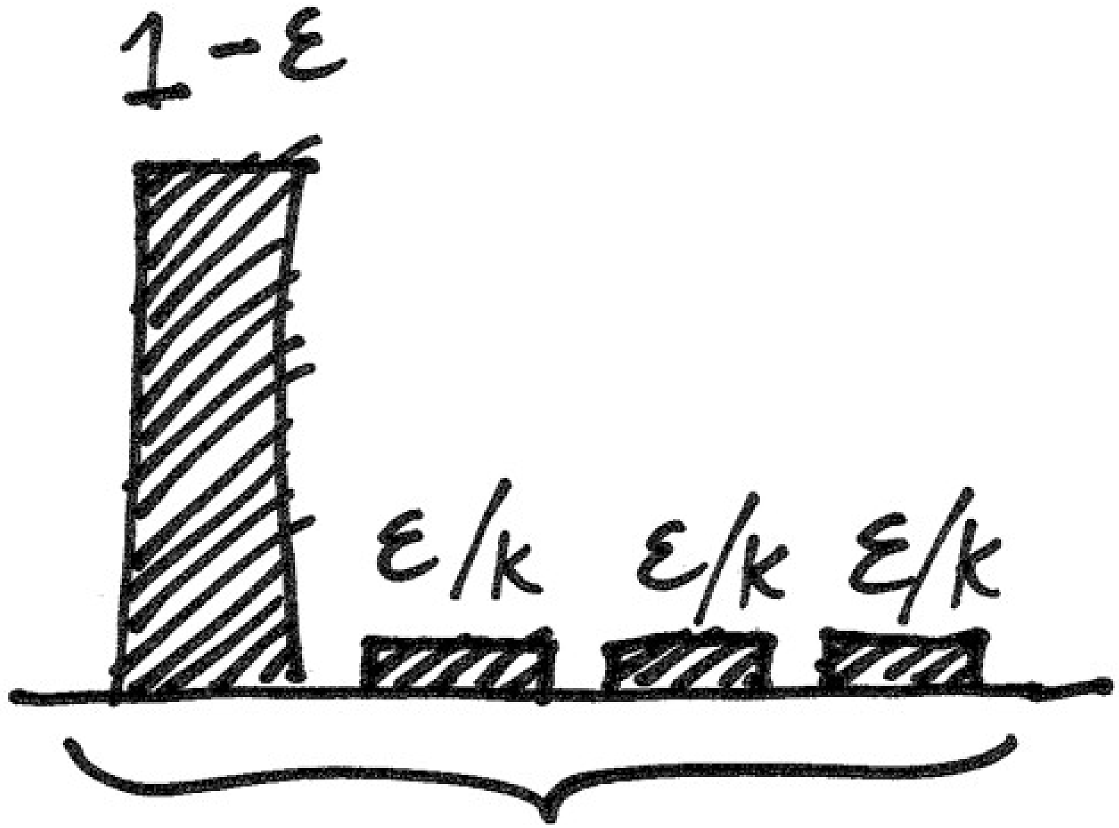
$$H(X) \quad\quad 0.$$

For logically independent X and Y,

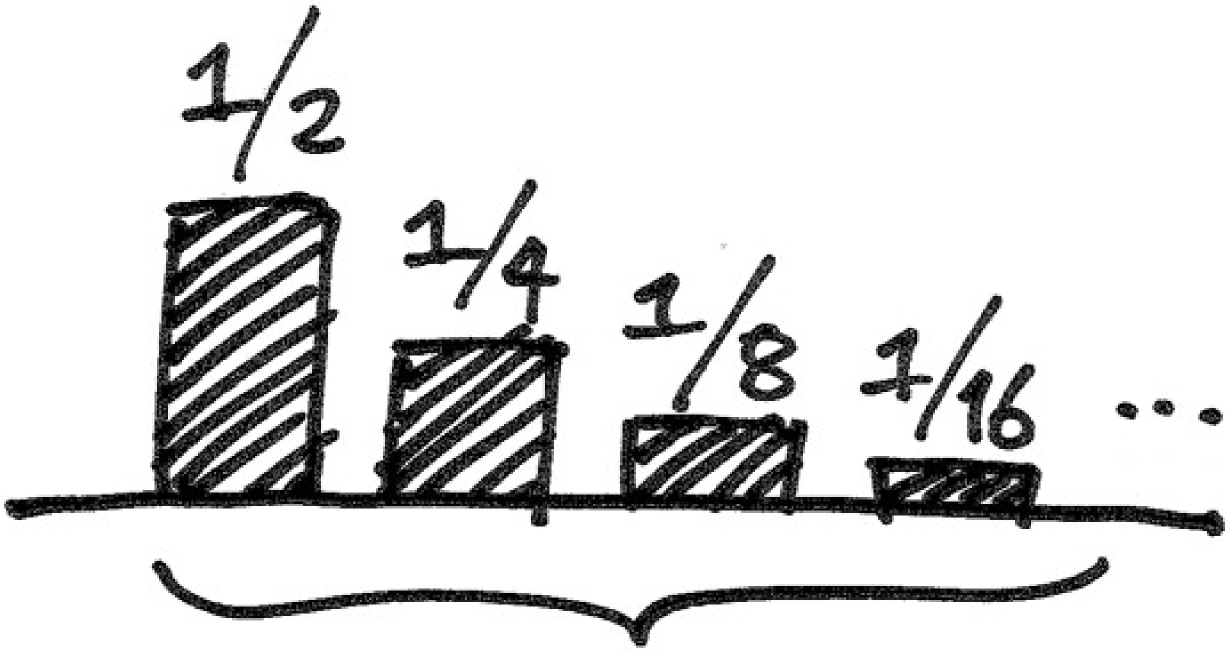
$$H(X \quad Y) \;==\; H(X) + H(Y).$$

For any X and Y,

$$\mathrm{H}(X) \quad\quad \mathrm{H}(X \mid Y = y).$$



$$H \ == \ \log(k + 1) \ ?$$



H == log(   ) ?!?

φ	$\Pr(\varphi)$	$I(X;\varphi)$
$X = 0$	p	$-\log(p)$
$X = 1$	$1 - p$	$-\log(1 - p)$

$$H(X) = I(X;X) = \mathrm{E}\left[I(X;X=x) \right]$$

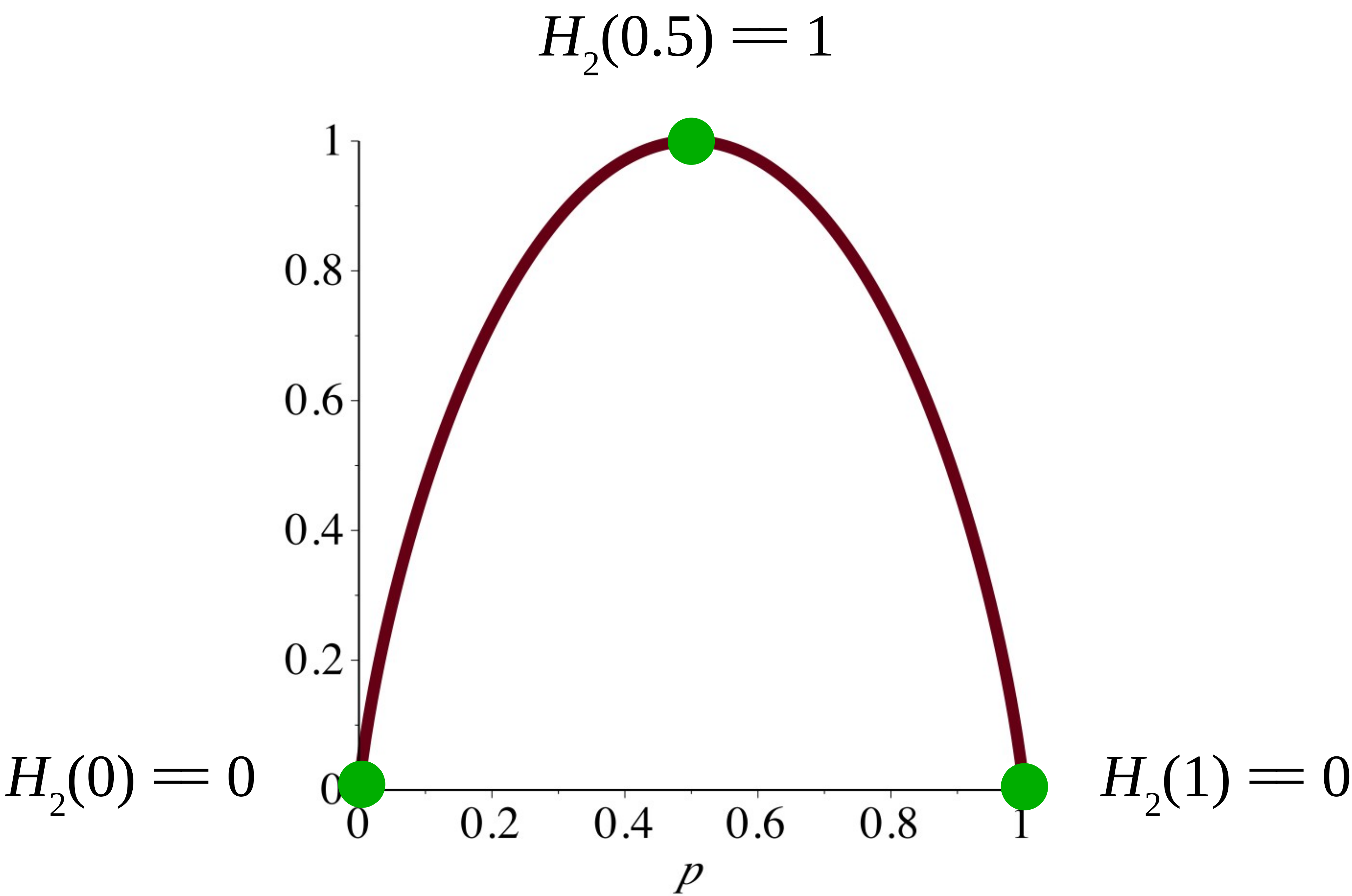
Claude Shannon: "A Mathematical Theory of Communication," *The Bell System Technical Journal*, 1948.

φ	$\Pr(\varphi)$	$I(X;\varphi)$
$X = 0$	p	$-\log(p)$
$X = 1$	$1 - p$	$-\log(1 - p)$

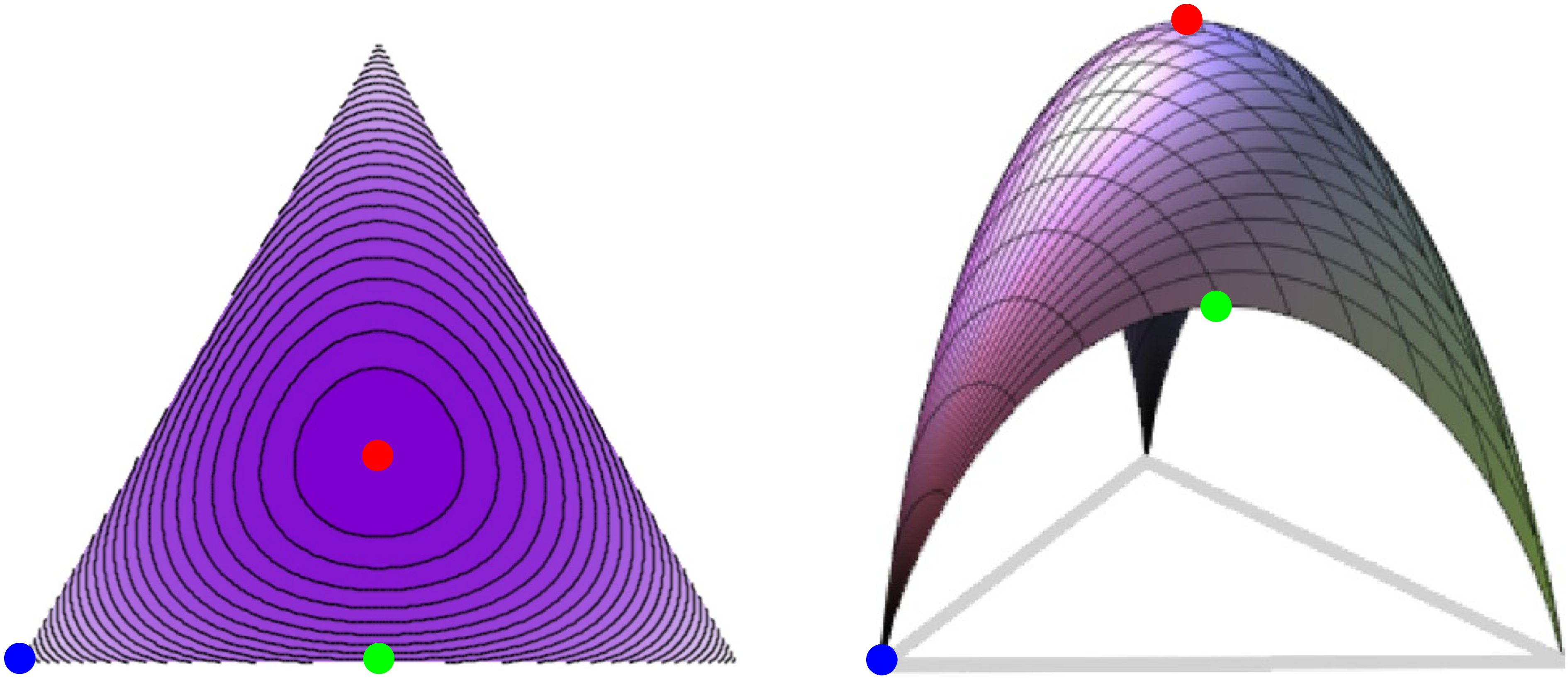
$$H(X) = \mathrm{E}\left[\log \frac{1}{\Pr(X{=}x)}\right]$$

Claude Shannon: "A Mathematical Theory of Communication," *The Bell System Technical Journal*, 1948.

Categorical distribution with 2 outcomes

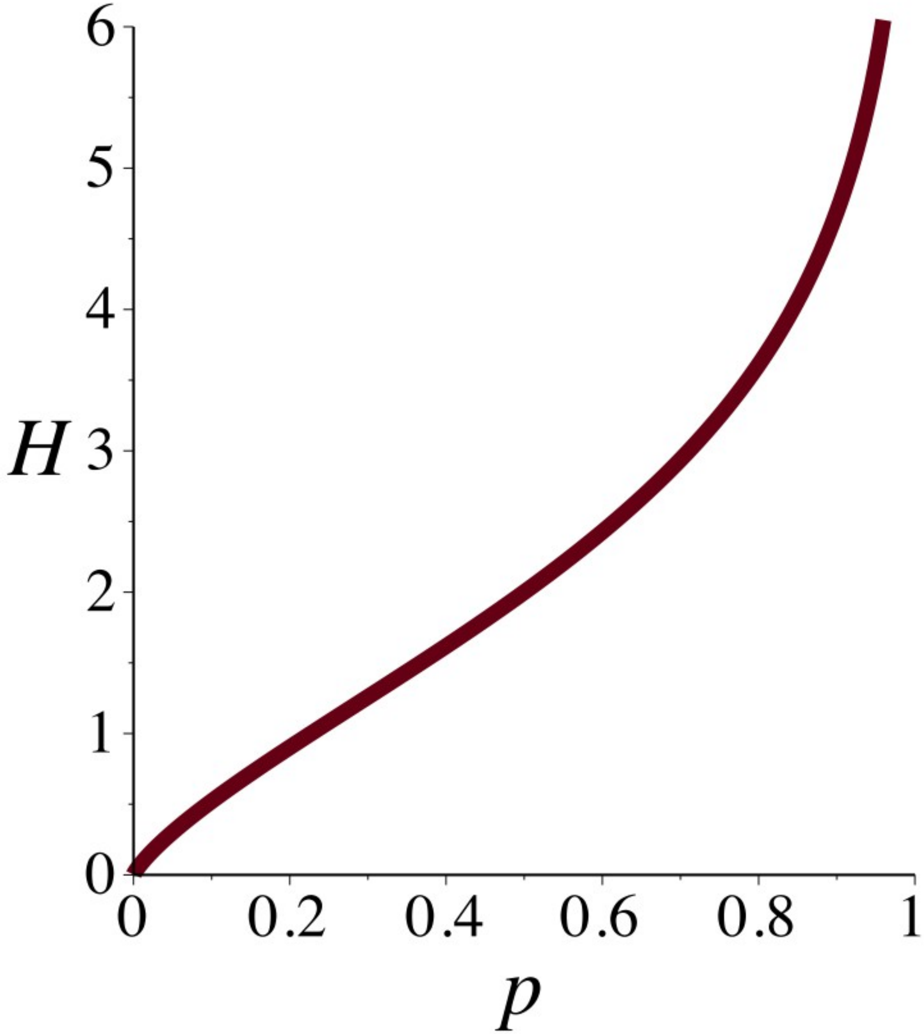


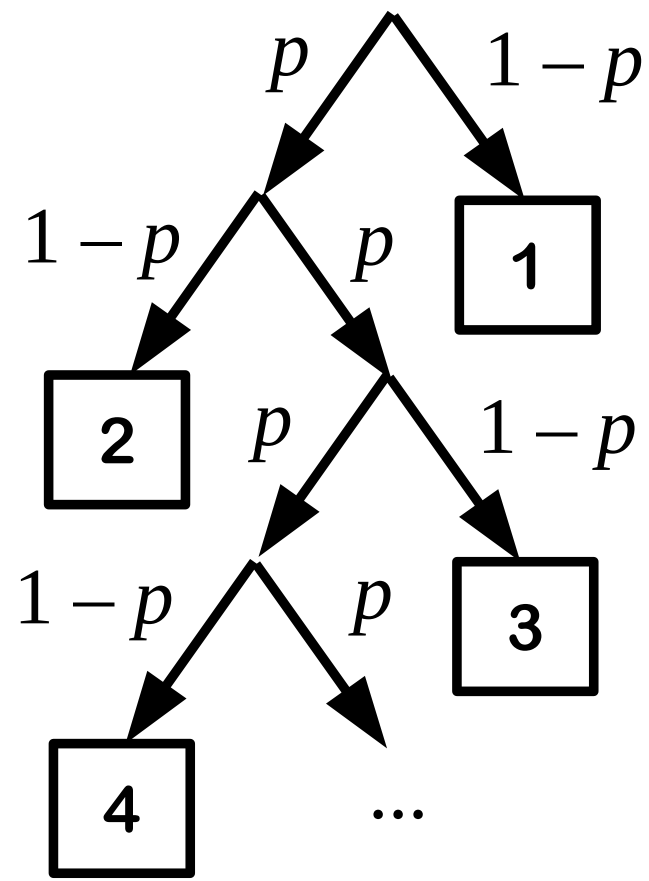
Categorical distribution with 3 outcomes



$$H(1/3, 1/3, 1/3) = \log 3$$
$$H(1/2, 1/2, 0) = 1$$
$$H(1, 0, 0) = 0$$

The geometric distribution





Nice properties of the Shannon entropy:

Entropy is positive:

$$H(X) \qquad 0.$$

Conditional entropy obeys the chain rule:

$$H(X \quad Y) == H(X) + H(X \mid Y).$$

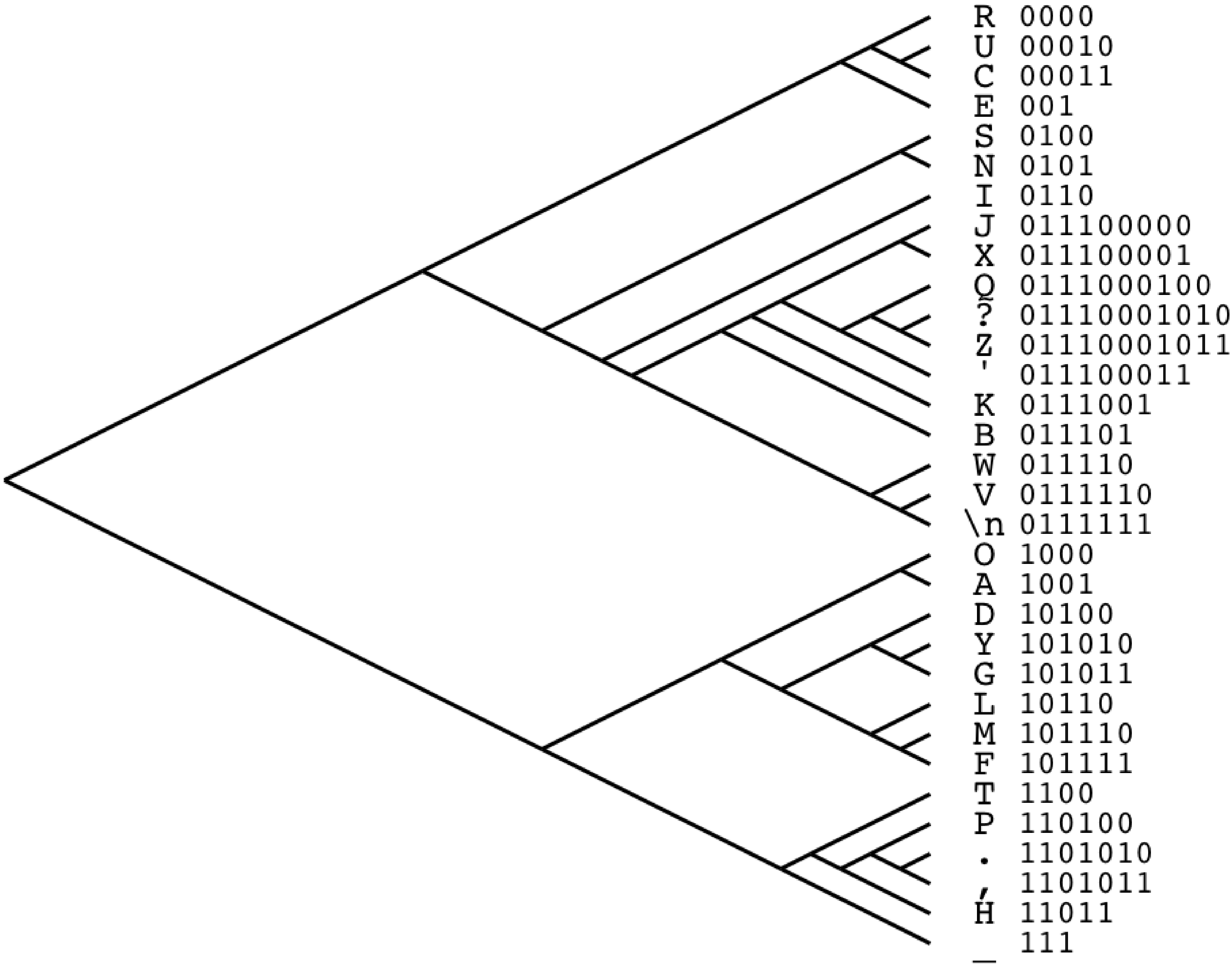
On average, conditioning decreases entropy:

$$\mathrm{H}(X) \qquad \mathrm{H}(X \mid Y).$$

($H(X \mid Y)$ is the average value of $H(X \mid Y = y)$.)

X	"a"	"b"	"c"	"d"	"e"
$\Pr(X = x)$	0.05	0.15	0.20	0.25	0.35

David A. Huffman: “A Method for the Construction of Minimum-Redundancy Codes,” Proceedings of the Institute of Radio Engineers, 1952


R 0000
U 00010
C 00011
E 001
S 0100
N 0101
I 0110
J 011100000
X 011100001
Q 0111000100
? 01110001010
Z 01110001011
' 011100011
K 0111001
B 011101
W 011110
V 0111110
\n 0111111
O 1000
A 1001
D 10100
Y 101010
G 101011
L 10110
M 101110
F 101111
T 1100
P 110100
. 1101010
, 1101011
H 11011
_ 111

Code	x	p(x)	–log p(x)	k	Code	x	p(x)	–log p(x)	k
1001	'A'	.0634	3.98	4	0111000100	'Q'	.0008	10.33	10
011101	'B'	.0135	6.21	6	0000	'R'	.0470	4.41	4
00011	'C'	.0242	5.37	5	0100	'S'	.0502	4.32	4
10100	'D'	.0321	4.96	5	1100	'T'	.0729	3.78	4
001	'E'	.0980	3.35	3	00010	'U'	.0234	5.42	5
101111	'F'	.0174	5.84	6	0111110	'V'	.0075	7.06	7
101011	'G'	.0165	5.92	6	011110	'W'	.0156	6.0	6
11011	'H'	.0438	4.51	5	011100001	'X'	.0014	9.46	9
0110	'I'	.0552	4.18	4	101010	'Y'	.0160	5.97	6
011100000	'J'	.0009	10.17	9	01110001011	'Z'	.0005	11.04	11
0111001	'K'	.0061	7.35	7	0111111	'\n'	.0084	6.89	7
10110	'L'	.0336	4.89	5	111	' '	.1741	2.52	3
101110	'M'	.0174	5.85	6	011100011	"'"	.0019	9.06	9
0101	'N'	.0551	4.18	4	1101011	','	.0117	6.42	7
1000	'O'	.0622	4.01	4	1101010	'.'	.0109	6.52	7
110100	'P'	.0180	5.8	6	01110001010	'?'	.0003	11.56	11

David A. Huffman: “A Method for the Construction of Minimum-Redundancy Codes,” Proceedings of the Institute of Radio Engineers, 1952

Entropy of a categorical variable Let X be distributed according to the following table:

x	1	2	3	4	5
$\Pr(X = x)$	1/12	1/6	1/6	1/4	1/3

1. Find $H(X)$.
2. Construct a Huffman code for X.
3. Decode the message 001011000011 according to your code.

Huffman tree for a die Let X be distributed uniformly on the set $\{1, 2, 3, 4, 5, 6\}$.

1. Huffman-encode the values of X.
2. What is the average code word length for the tree you have constructed? How does that compare with $H(X)$?
3. If you interpret a codeword length of k as an implicit probability of 2^{-k}, what is then the implicit distribution expressed by your code?

Age order (McKay, Exercise 2.35) You want to know whether A is older than B. A tells you she is older than C.

How much information does that message give you?

Knights and Knaves (McKay, Exercise 2.37) A person who lies two third of the time tells you that φ. How much information does that give you?

Shuffling cards (McKay, Exercise 6.19) Roughly how many bits of uncertainty do you create by thoroughly shuffling a deck of cards?