ECE 515
Information Theory

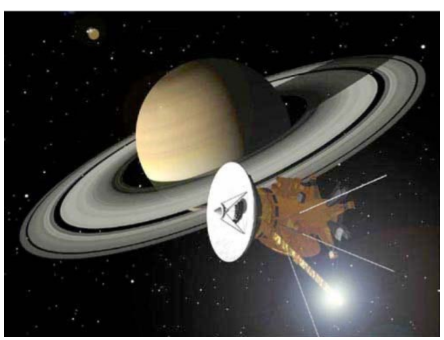
Aaron Gulliver
Dept. of Electrical and Computer Engineering

Outline

- Introduction and Motivation
- Measuring Information
  - Entropy
  - Mutual Information
- Distortionless Source Coding
- Channel Coding for Noisy Channels
- Error Correcting Codes

Information Theory

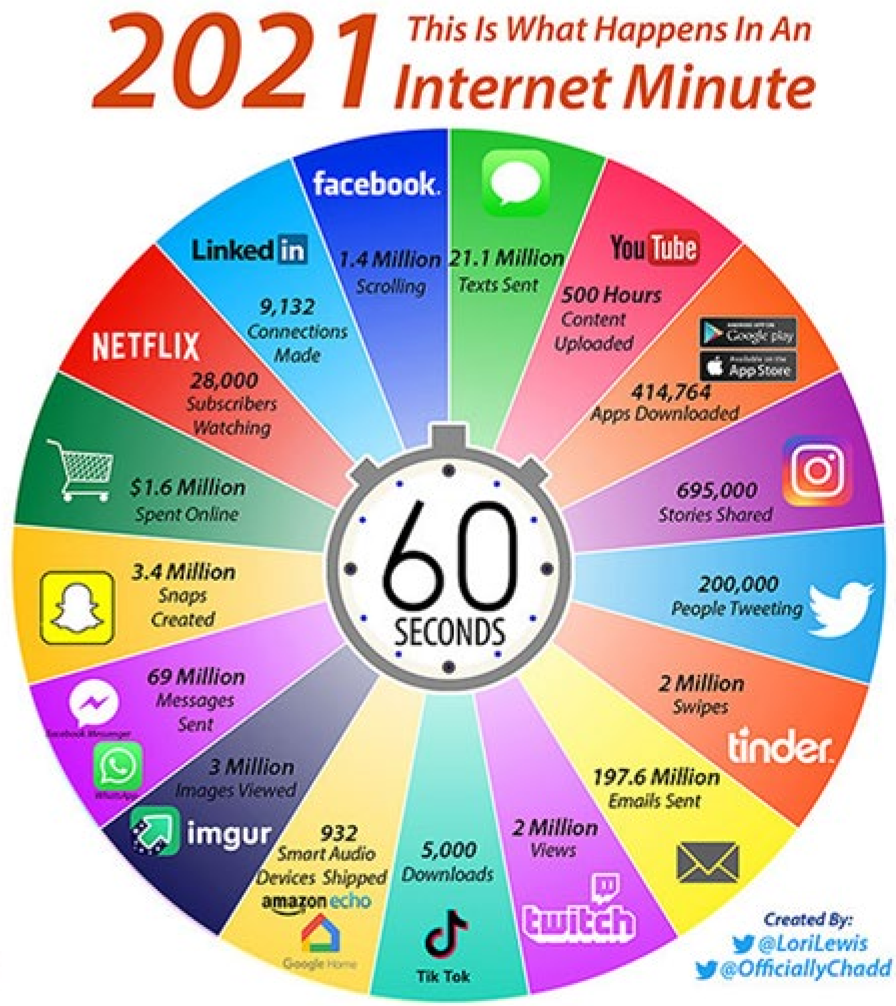
Information theory has influenced the design of virtually every system that stores, processes, or transmits information in digital form, from computers to dvds, from smart phones to deep space probes.








2020 This Is What Happens In An Internet Minute
facebook.
1.3 Million Logging In
19 Million Texts Sent
You Tube
4.7 Million Videos Viewed
Google play
App Store
400,000 Apps Downloaded
694,444 Scrolling Instagram
194,444 People Tweeting
1.6 Million Swipes
tinder.
190 Million Emails Sent
twitch
1.2 Million Views
1,400 Downloads
Tik Tok
305 Smart Speakers Shipped
amazon echo
imgur
2.5 Million Images Viewed
59 Million Messages Sent
2.5 Million Snaps Created
$1.1 Million Spent Online
NETFLIX
764,000 Hours Watched
Google
4.1 Million Search Queries
60 SECONDS
Created By:
@LoriLewis
@OfficiallyChadd


2021 This Is What Happens In An Internet Minute
facebook.
1.4 Million Scrolling
21.1 Million Texts Sent
You Tube
500 Hours Content Uploaded
Google play
App Store
414,764 Apps Downloaded
695,000 Stories Shared
200,000 People Tweeting
2 Million Swipes
tinder.
197.6 Million Emails Sent
twitch
2 Million Views
5,000 Downloads
Tik Tok
932 Smart Audio Devices Shipped
amazon echo
imgur
3 Million Images Viewed
69 Million Messages Sent
3.4 Million Snaps Created
$1.6 Million Spent Online
NETFLIX
28,000 Subscribers Watching
Linked in
9,132 Connections Made
60 SECONDS
Created By:
@LoriLewis
@OfficiallyChadd

- Butter’s Law
  - The amount of data coming out of an optical fiber will double every nine months
- Nielsen’s Law
  - Network connection speeds for high-end home users will double every 21 months
- Cooper's Law
  - The maximum number of voice conversations or equivalent data transactions that can be conducted in all of the useful radio spectrum over a given area will double every 30 months

- In the Information Society, demand will increase to consume the supply of
  - Transmission bandwidth
  - Storage facilities
  - Computing power

- The problem is how to transmit and store information
  - Efficiently
  - Reliably
  - Securely

Information Theory Applications

- Information theory has been employed in the fields of
  - Communications
  - Statistical physics and Astrophysics
  - Complexity analysis
  - Linguistics
  - Psychology
  - Biology (Genomics)
  - Economics and Investing
  - Gambling
  - …

Information

The word information has been given many different meanings in the general field of information theory.

It is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of this general field.

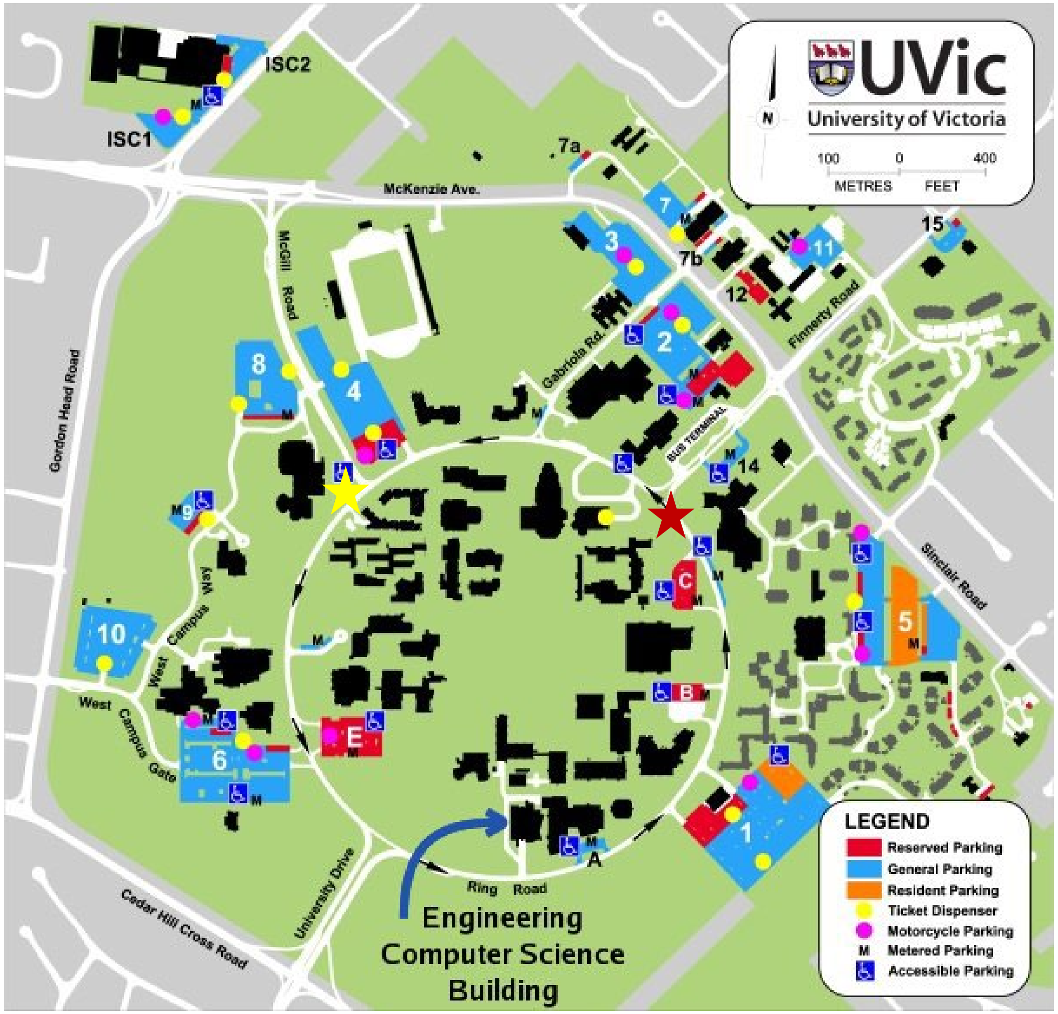
[Claude Shannon, 1953]

What is Information?

Consider the following two messages:

1. There was a traffic accident on Ring Road.
2. There was a traffic accident on Ring Road near Finnerty Road. ★

Message 2 is more precise, so perhaps it provides more information.



What is Information?

Consider the following two messages:

1. There was a traffic accident on Ring Road near Finnerty Road. ★
2. There was a traffic accident on Ring Road near McGill Road. ★

It is not clear whether message 1 or 2 provides more information.

Types of Information

- Semantic Information
  - Related to the meaning of the message
- Pragmatic Information
  - Related to the usage and effect of the message
- Syntactic Information
  - Related to the symbols and structure of the message

What is Information?

Consider the following definition of information.

- The number of letters in the message.

1. There was a traffic accident on Ring Road.

   (34 letters)

2. There was a traffic accident on Ring Road near Finnerty Road.

   (50 letters)

Definitely something we can measure and compare.
But is this useful to us?

Measuring Information

- How much information does a message provide?

1. Consider a person in a room looking out the window. She can clearly see that the sun is shining. At this moment, she receives a call from a neighbor who says: *It is now daytime*

2. A group bought a lottery ticket. One member calls the others and says: *We have won first prize*

Measuring Information

1. The message provides no information. Why? Because she is already certain that it is daytime.
2. The message provides a lot of information because the probability of winning first prize is very small.

Conclusion

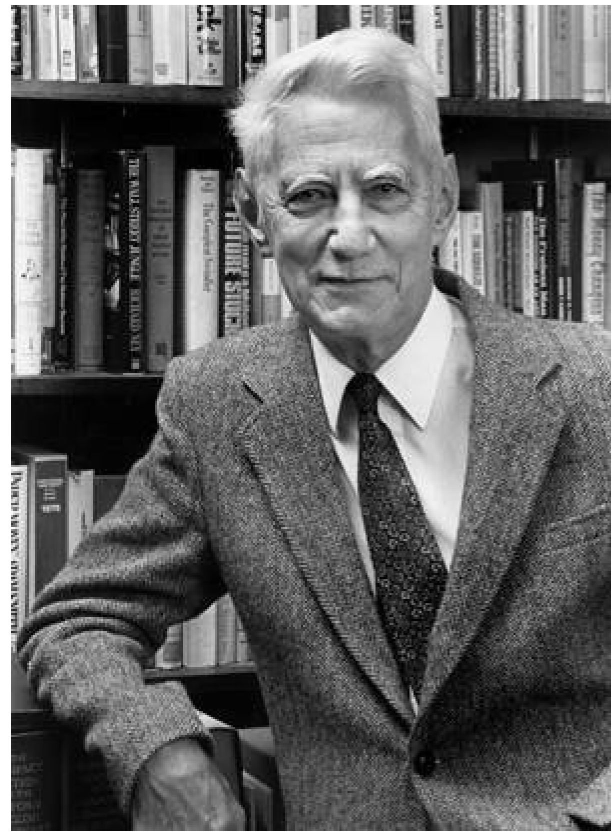
- The information provided by a message is inversely related to the probability of occurrence of that message.
- If a message is very probable, it does not provide much information. If it is very unlikely, it provides a lot of information.

Measuring Information

An information measure is needed to

1. Determine limits on the amount of information that can be transmitted through a channel
2. Determine limits on by how much data can be compressed
3. Construct systems that approach these limits
   - channel coding
   - source coding

Claude Shannon (1916-2001)



Shannon did not consider meaning:

These semantic aspects of communications are irrelevant to the engineering aspects

A Mathematical Theory of Communication, BSTJ July, 1948

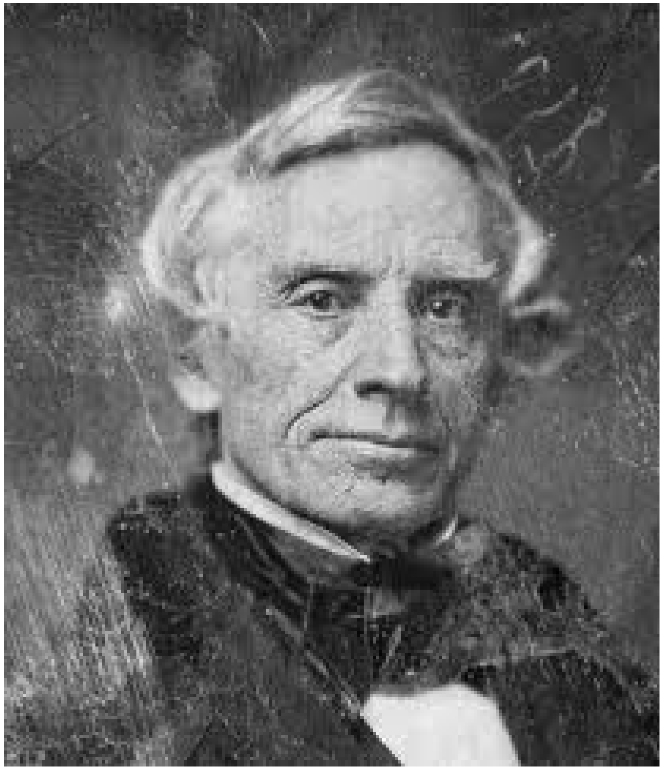
The fundamental problem of communication is that of reproducing at one point exactly or approximately a message selected at another point. …

If the channel is noisy it is not in general possible to reconstruct the original message or the transmitted signal with certainty by any operation on the received signal.

History of Information Theory

1838	S.F.B. Morse
1924	H.T. Nyquist
1928	R.V.L. Hartley
1940	A.M. Turing
1946	R.E. Hamming
1948	C.E. Shannon
1951	S. Kullback and R. Leibler; D.A. Huffman

Samuel F. B. Morse (1791-1872)



International Morse Code

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

(a) Letters

1	· — — — —
2	· · — — —
3	· · · — —
4	· · · · —
5	· · · · ·
6	— · · · ·
7	— — · · ·
8	— — — · ·
9	— — — — ·
0	— — — — —

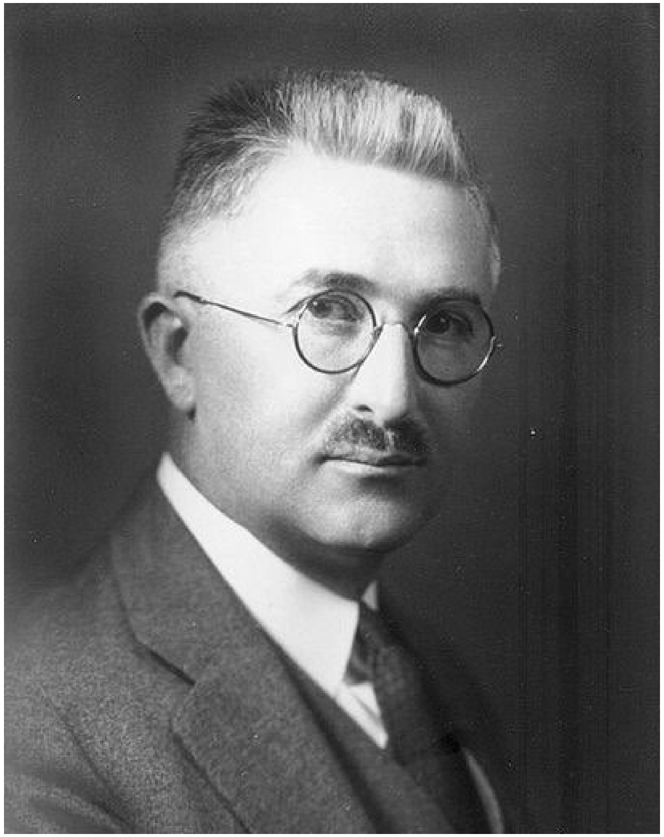
(b) Numbers

Period (·)	· — · — · —	Wait sign (AS)	· — · · ·
Comma (,)	— — · · — —	Double dash (break)	— · · · —
Interrogation (?)	· · — — · ·	Error sign	· · · · · · · ·
Quotation Mark (”)	· — · · — ·	Fraction bar (/)	— · · — ·
Colon (:)	— — — · · ·	End of message (AR)	· — · — ·
Semicolon (;)	— · — · — ·	End of transmission (SK)	· · · — · —
Parenthesis ( )	— · — — · —		

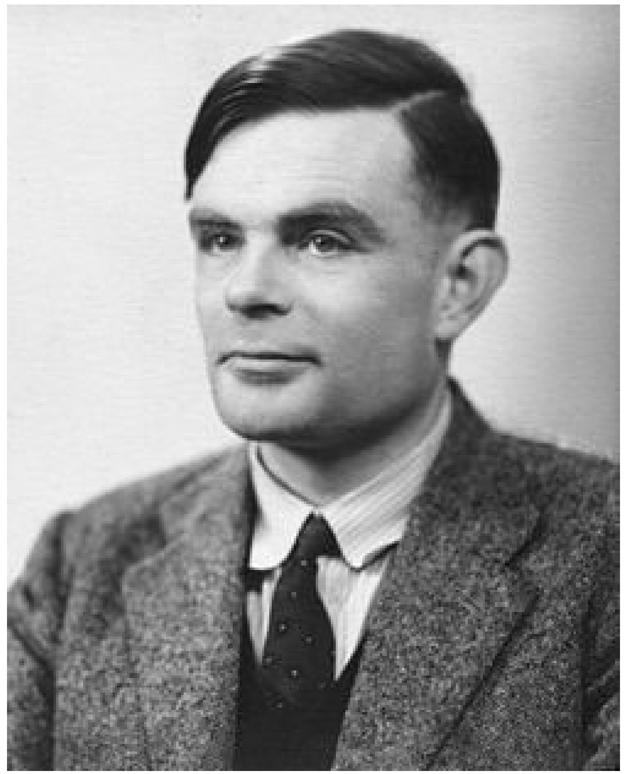
(c) Punctuation and Special Characters

Harry T. Nyquist (1889-1976)

Ralph V. L. Hartley (1888-1970)

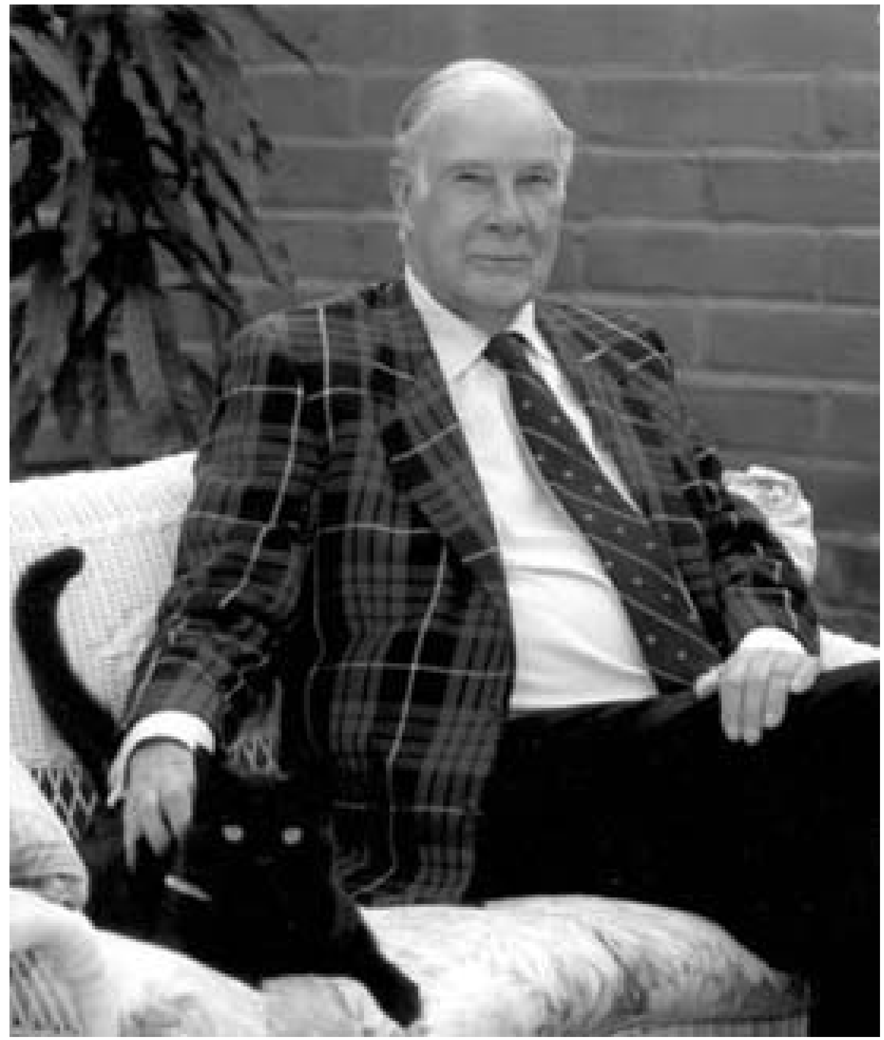


Alan M. Turing (1912-1954)

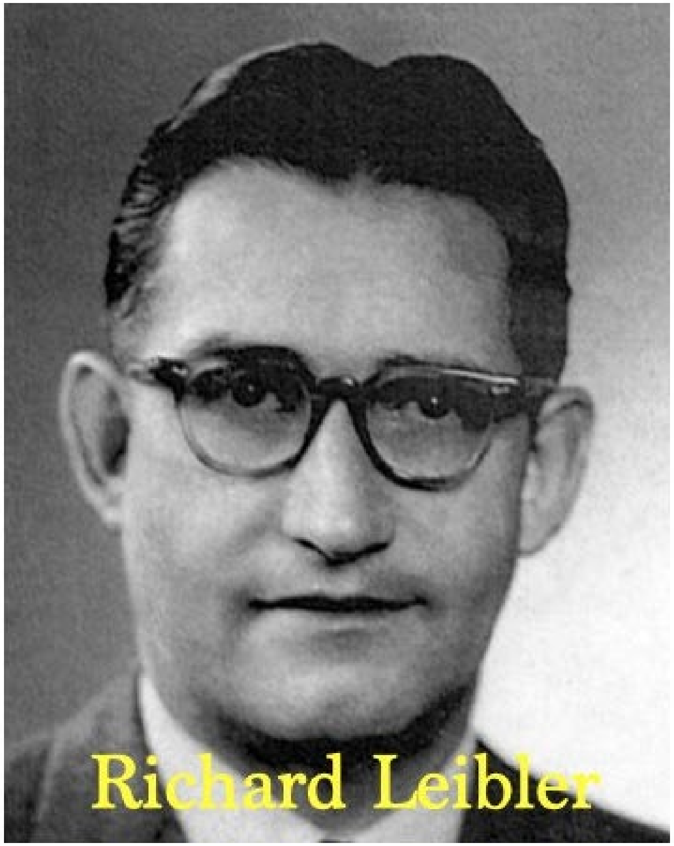


Enigma Machine

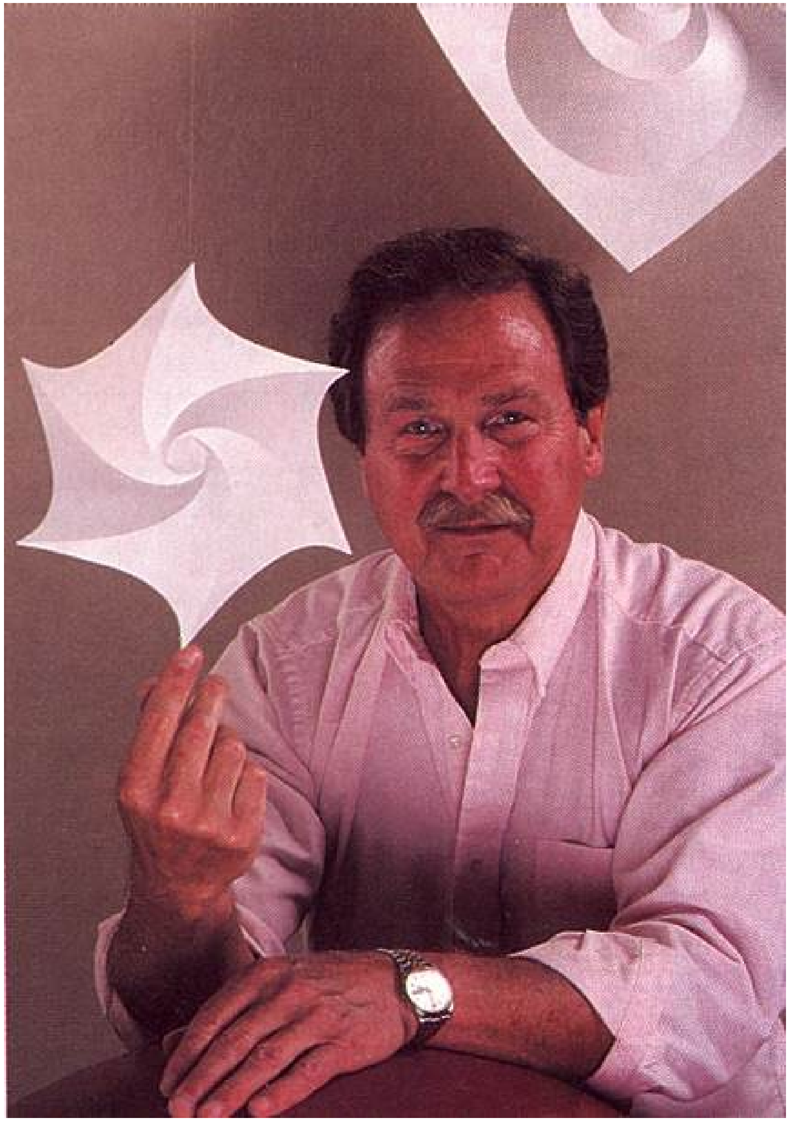
Richard W. Hamming (1915-1998)

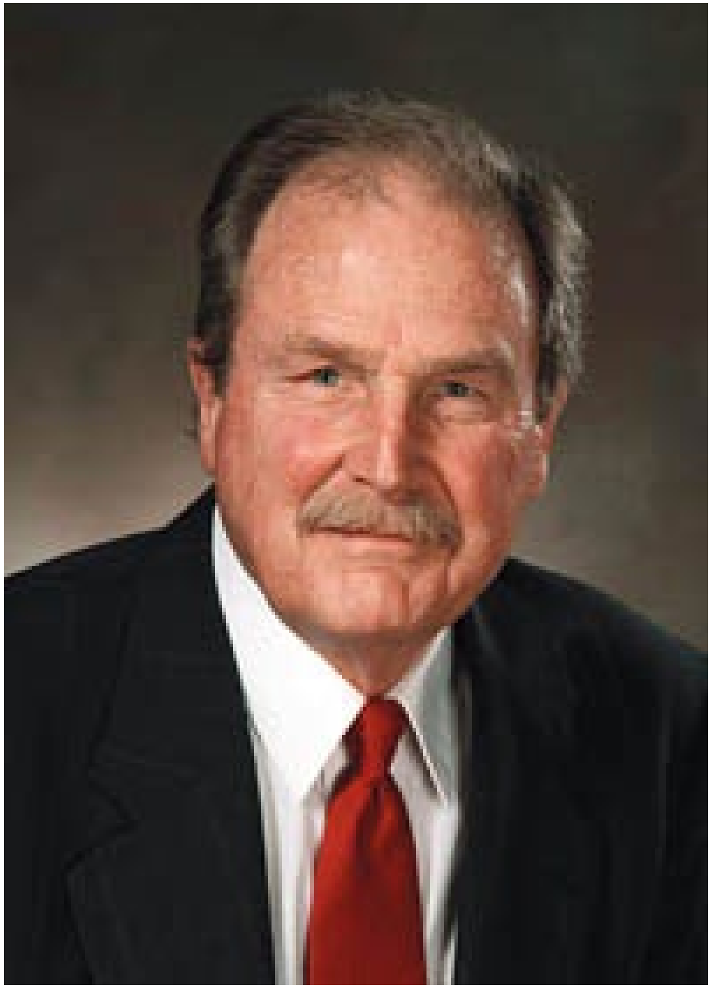


Solomon Kullback (1907-1994)
Richard A. Leibler (1914-2003)



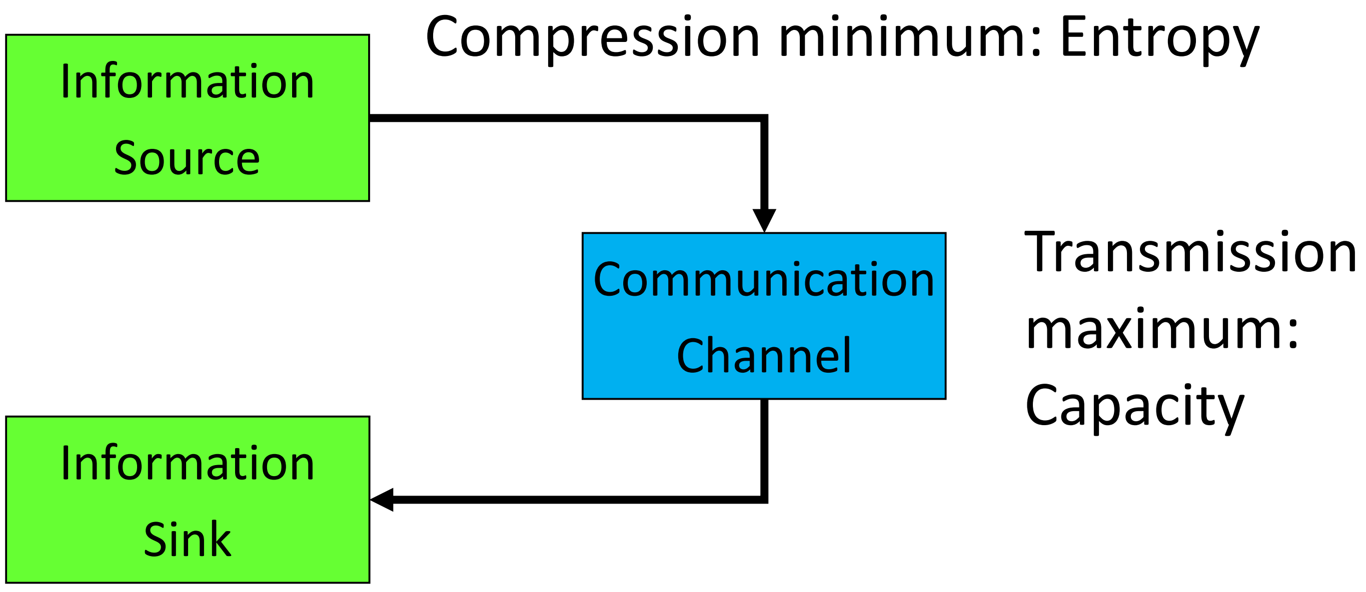
David A. Huffman (1925-1999)





Communication Systems

Shannon considered the efficient and reliable transmission of information through a channel.



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.

WHILE the frequency relations involved in electrical communication are interesting in themselves, I should hardly be justified in discussing them on this occasion unless we could deduce from them something of fairly general practical application to the engineering of communication systems. What I hope to accomplish in this direction is to set up a quantitative measure whereby the capacities of various systems to transmit information may be compared. In doing this I shall discuss its application to systems of telegraphy, telephony, picture transmission and television over both wire and radio paths. It will, of course, be found that in very many cases it is not economically practical to make use of the full physical possibilities of a system. Such a criterion is, however, often useful for estimating the possible increase in performance which may be expected to result from improvements in apparatus or circuits, and also for detecting fallacies in the theory of operation of a proposed system.

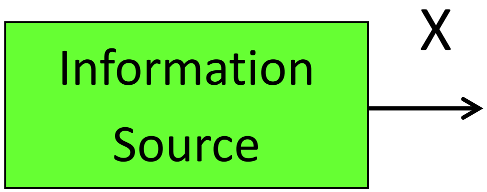
Inasmuch as the results to be obtained are to represent the limits of what may be expected under rather idealized conditions, it will be permissible to simplify the discussion by neglecting certain factors which, while often important in practice, have the effect only of causing the performance to fall somewhat further short of the ideal. For example, external interference, which can never be entirely eliminated in practice, always reduces the effectiveness of the system. We may, however, arbitrarily assume it to be absent, and consider the limitations which still remain due to the transmission system itself.

In order to lay the groundwork for the more practical applications of these frequency relationships, it will first be necessary to discuss a few somewhat abstract considerations.

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

``It is desirable therefore to eliminate the psychological factors involved and to establish a measure of information in terms of purely physical quantities.''

Hartley Measure of Information



- X is a discrete random variable with N possible outcomes.

Hartley Measure of Information

- Consider a six-sided die (N=6)
- One roll has 6 possible outcomes
- Three rolls have 6^3 = 216 possible outcomes
- The amount of information provided by one roll is $\log_b(6)$
- The amount of information provided by three rolls is $\log_b(216) = 3\times\log_b(6)$
  - Three times the amount provided by one roll

Hartley Measure of Information

- The Hartley measure of information provided by the observation of a discrete random variable X is

$$\log_b N$$

  - N is the number of possible outcomes for X
  - b is the base of the logarithm
- For m observations (trials) the amount of information is

$$\log_b(N^m) = m(\log_b N)$$

Information Measures

- b = 2 bits or Sh (Shannons)
- b = e logons or nats (natural units)
- b = 10 hartleys (in honour of R.V.L. Hartley) or bans (town of Banbury)

$$
\begin{aligned}
1\ logon &= \frac{1}{\ln 2} &= \log_2 e &= 1.443\ Sh \\
1\ hartley &= \frac{1}{\log_{10} 2} &= \log_2 10 &= 3.322\ Sh \\
1\ hartley &= \frac{1}{\log_{10} e} &= \ln 10 &= 2.303\ logons
\end{aligned}
$$

Telephone Numbers

- North American telephone numbers have m=10 decimal digits, i.e. 250-721-6028
- N=10 and each digit provides

$$\log_2(10) = 3.322 \text{ bits of information}$$

- There are $N^m=10^{10}$ possible telephone numbers
- The amount of information provided by a telephone number is

$$\log_2(10^{10}) = 10\times\log_2(10) = 33.22 \text{ bits}$$

Humans

- 7.97 billion people on earth

$$\log_2(7.97\times10^9) = 32.89 \text{ bits}$$

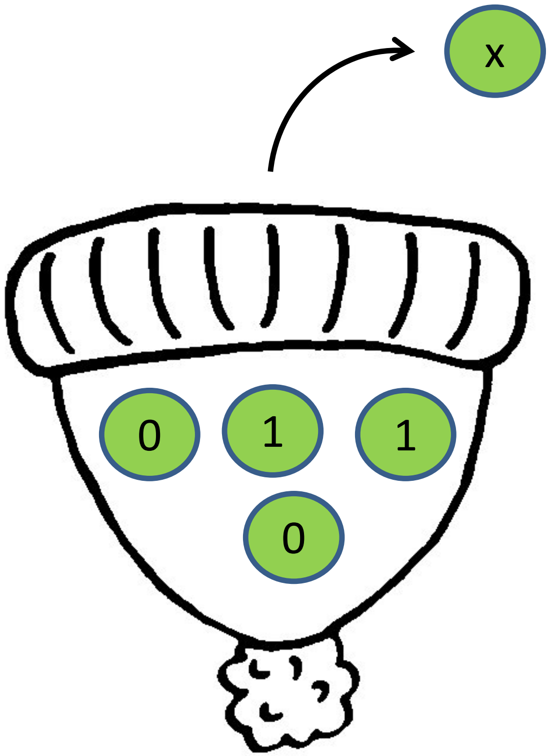
- Each person can be uniquely represented by only 33 bits
- This is just 33 times the amount of information in 1 bit

V Numbers

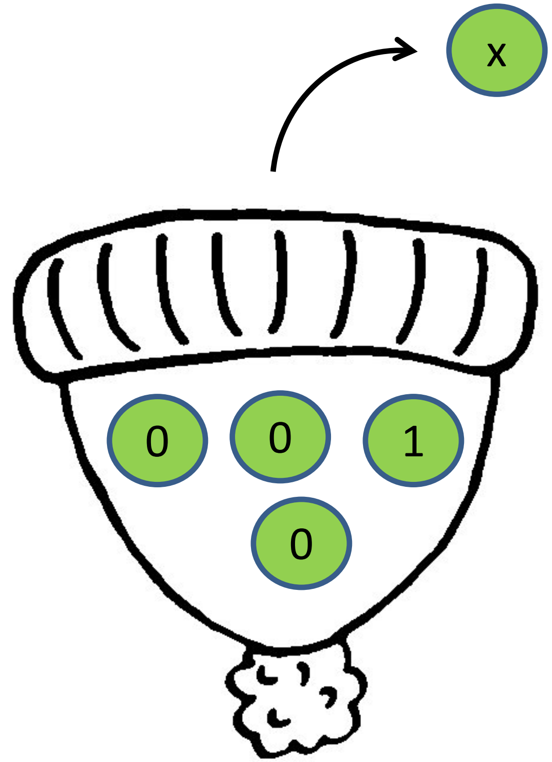
- V00767124
- $\log_2(26\times10^8) = 31.28$ bits

- V0 is already known and the next digit is 0 or 1, so the amount of information is only

$$\log_2(2\times10^6) = 6\times\log_2(10) + 1 = 20.93 \text{ bits}$$

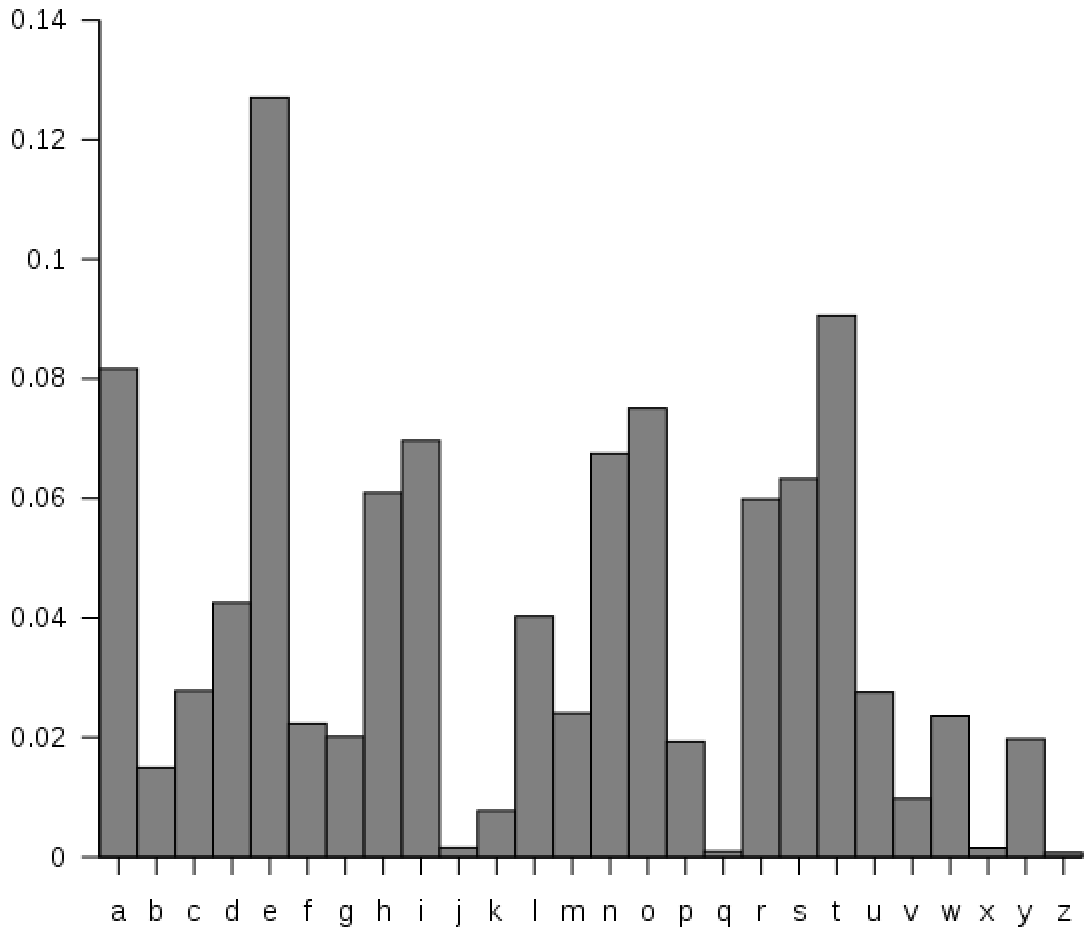


(a)



(b)

Frequency of English Letters



Self Information

$$I_X(x_i) = -\ \log_b p(x_i)$$

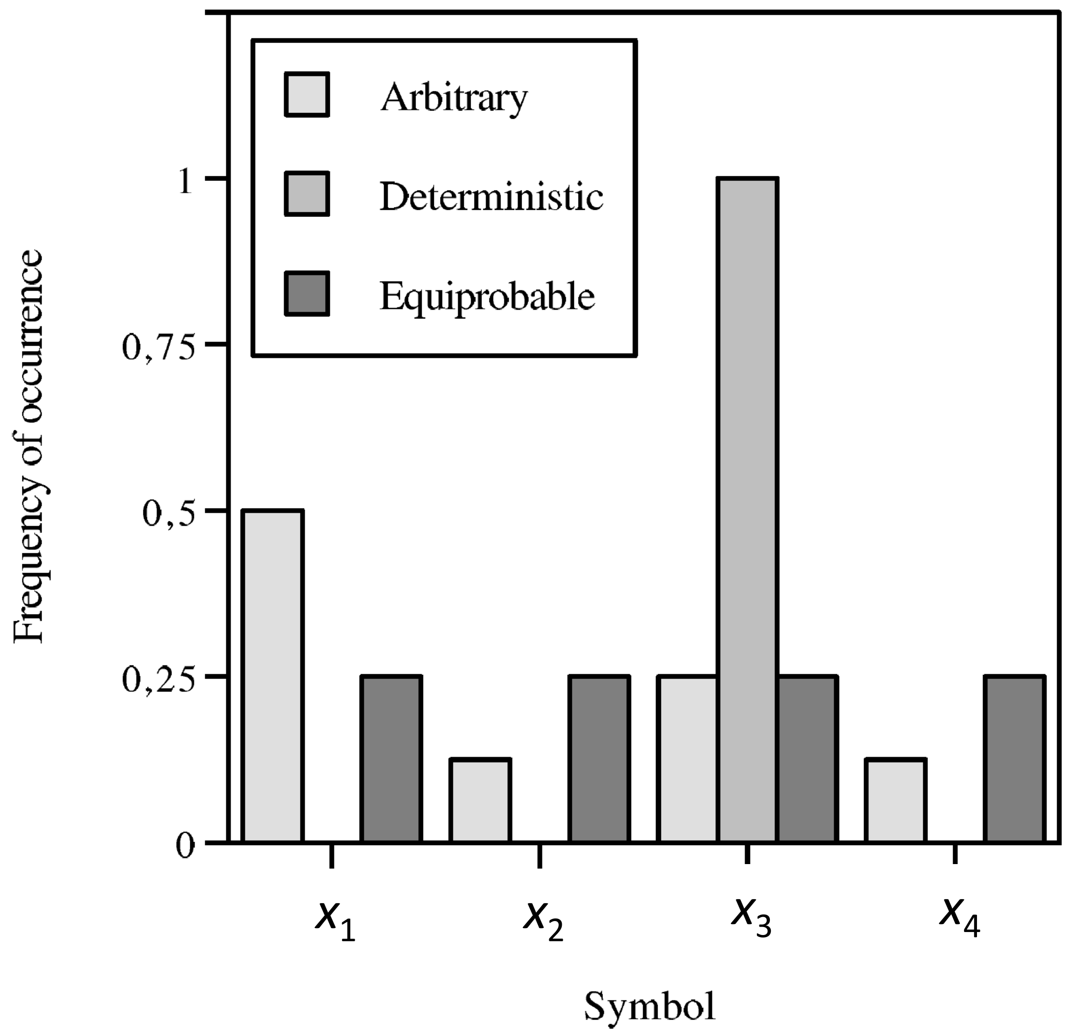
Entropy

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

Entropy

- The measure H(X) satisfies the three criteria
  1. Considers the symbol probabilities $p(x_i)$
  2. Considers the number of possible outcomes N
  3. Is additive

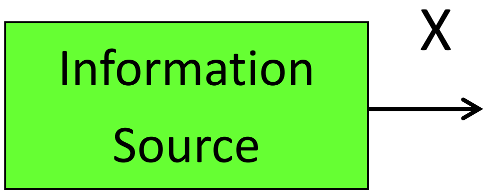
Quaternary Distributions



Entropy of Quaternary Distributions

- $H_{Arbitrary}$ = 1.75 bits
- $H_{Deterministic}$ = 0.00 bits
- $H_{Equiprobable} = -\log_2(1/4) = \log_2(4) = 2.00$ bits

Information



- Before a discrete source outputs a symbol there is uncertainly as to which one it will be, e.g. a letter from an alphabet of size N.
- After it has been received, the uncertainty is resolved.
- Information is the uncertainty which is resolved when a symbol is received.

Winning the Lottery

- Probability of winning first prize

$$p(\text{first prize}) = \frac{1}{\binom{50}{7} \times 3} = \frac{1}{33,294,800}$$

- Probability of winning a prize

$$p(\text{a prize}) = \frac{1}{7.0}$$

- Probability of winning first prize knowing that a prize has been won

$$p(\text{first prize} \mid \text{a prize}) = \frac{7.0}{33,294,800} = \frac{1}{4,756,400}$$

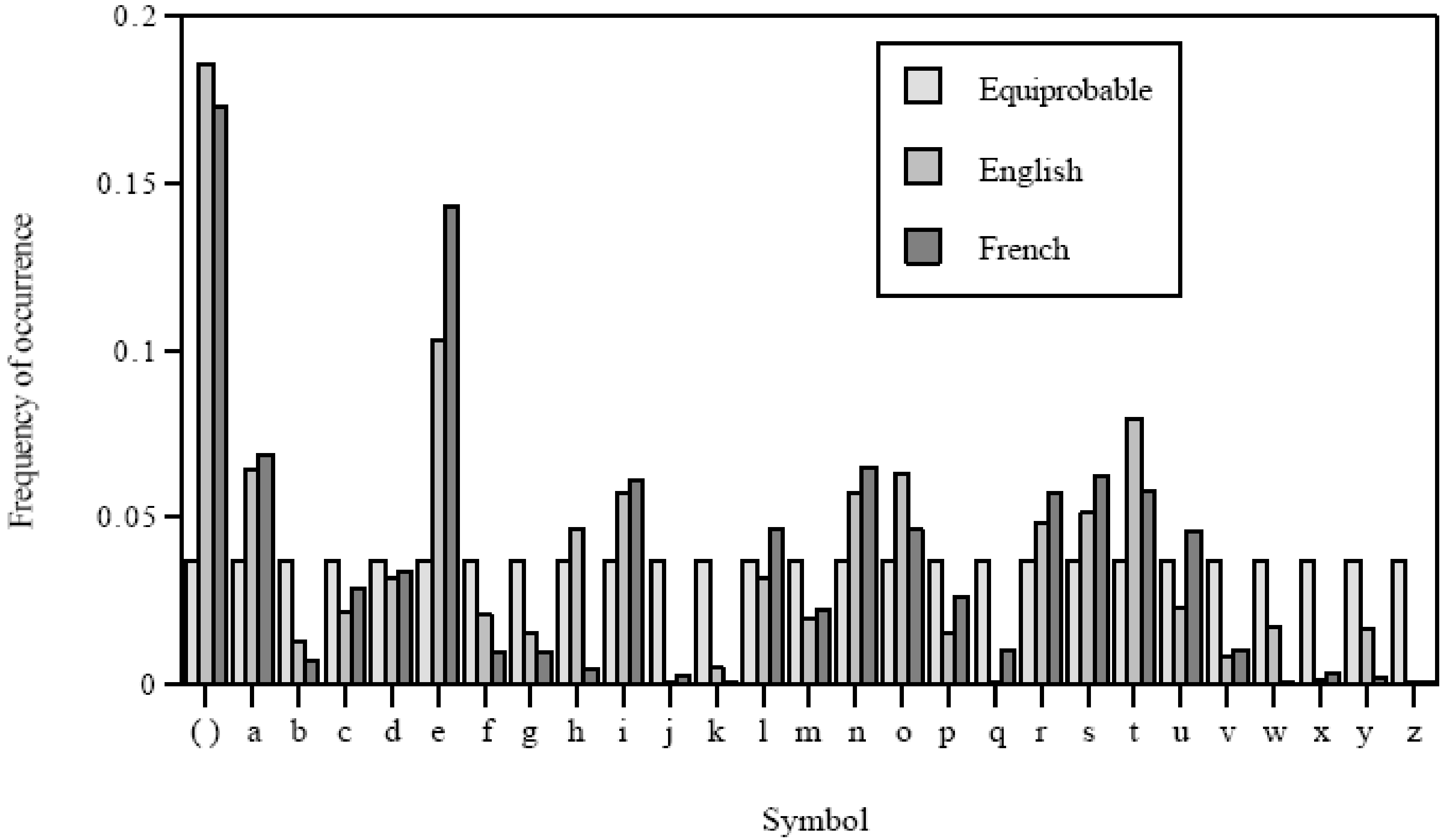
Winning the Lottery

- $H_{before} = \log_2(33{,}294{,}800) = 24.99$ bits

- After knowing that a prize has been won
  $H_{after} = \log_2(4{,}756{,}400) = 22.18$ bits

- Information is the reduction in uncertainty
  $H_{before} - H_{after} = 24.99 – 22.18 = 2.81$ bits

- Saying that you have won a prize provides 2.81 bits of information

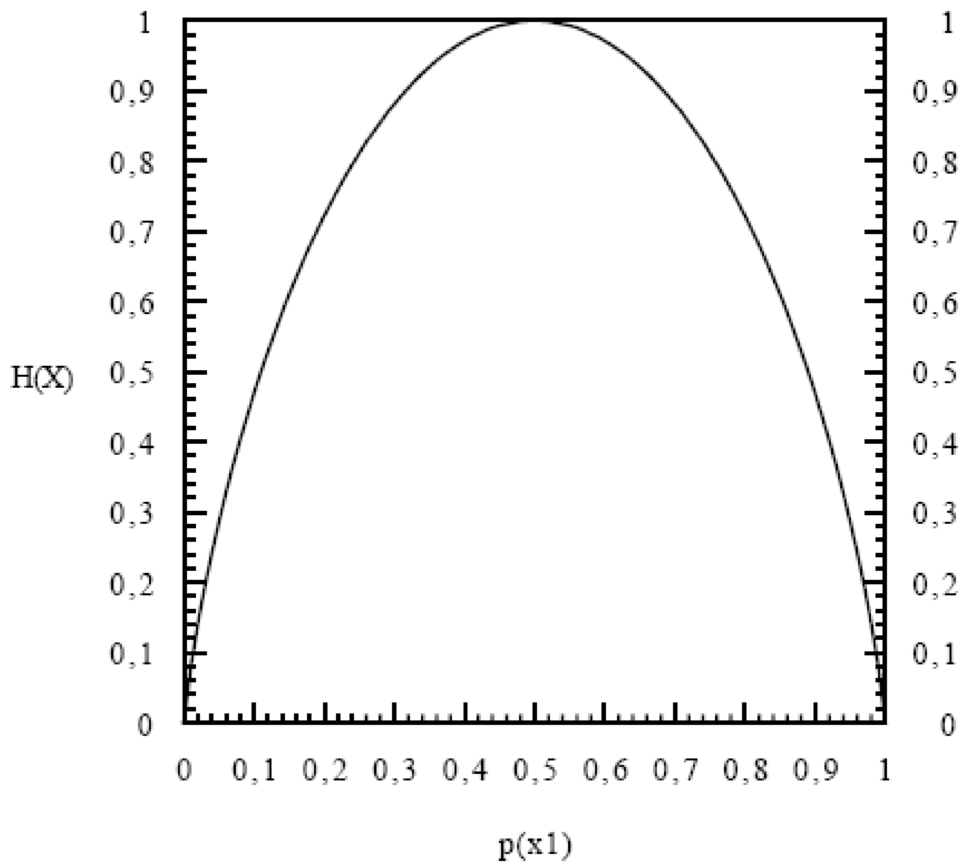

0.2
0.15
0.1
0.05
0
Frequency of occurrence
Equiprobable
English
French
( ) a b c d e f g h i j k l m n o p q r s t u v w x y z
Symbol

Letter	Equiprobable		English language		French language	
x_i	$p(x_i) = \frac{1}{27}$	$-\log_2 \frac{1}{27}$	$p(x_i)$	$-\log_2 p(x_i)$	$p(x_i)$	$-\log_2 p(x_i)$
□	0.0370	4.7549	0.1859	2.4274	0.1732	2.5295
a	0.0370	4.7549	0.0642	3.9613	0.0690	3.8573
b	0.0370	4.7549	0.0127	6.2990	0.0068	7.2002
c	0.0370	4.7549	0.0218	5.5195	0.0285	5.1329
d	0.0370	4.7549	0.0317	4.9794	0.0339	4.8826
e	0.0370	4.7549	0.1031	3.2779	0.1428	2.8079
f	0.0370	4.7549	0.0208	5.5873	0.0095	6.7179
g	0.0370	4.7549	0.0152	6.0398	0.0098	6.6730
h	0.0370	4.7549	0.0467	4.4204	0.0048	7.7027
i	0.0370	4.7549	0.0575	4.1203	0.0614	4.0256
j	0.0370	4.7549	0.0008	10.2877	0.0024	8.7027
k	0.0370	4.7549	0.0049	7.6730	0.0006	10.7027
l	0.0370	4.7549	0.0321	4.9613	0.0467	4.4204
m	0.0370	4.7549	0.0198	5.6584	0.0222	5.4933
n	0.0370	4.7549	0.0574	4.1228	0.0650	3.9434
o	0.0370	4.7549	0.0632	3.9839	0.0464	4.4297
p	0.0370	4.7549	0.0152	6.0398	0.0261	5.2598
q	0.0370	4.7549	0.0008	10.2877	0.0104	6.5873
r	0.0370	4.7549	0.0484	4.3688	0.0572	4.1278
s	0.0370	4.7549	0.0514	4.2821	0.0624	4.0023
t	0.0370	4.7549	0.0796	3.6511	0.0580	4.1078
u	0.0370	4.7549	0.0228	5.4548	0.0461	4.4391
v	0.0370	4.7549	0.0083	6.9127	0.0104	6.5873
w	0.0370	4.7549	0.0175	5.8365	0.0005	10.9658
x	0.0370	4.7549	0.0013	9.5873	0.0035	8.1584
y	0.0370	4.7549	0.0164	5.9302	0.0018	9.1178
z	0.0370	4.7549	0.0005	10.9658	0.0006	10.7027

Languages

- $H_{English}$ = 4.08 bits
- H_{French} = 3.96 bits
- $H_{Equiprobable} = \log_2(27)$ = 4.76 bits

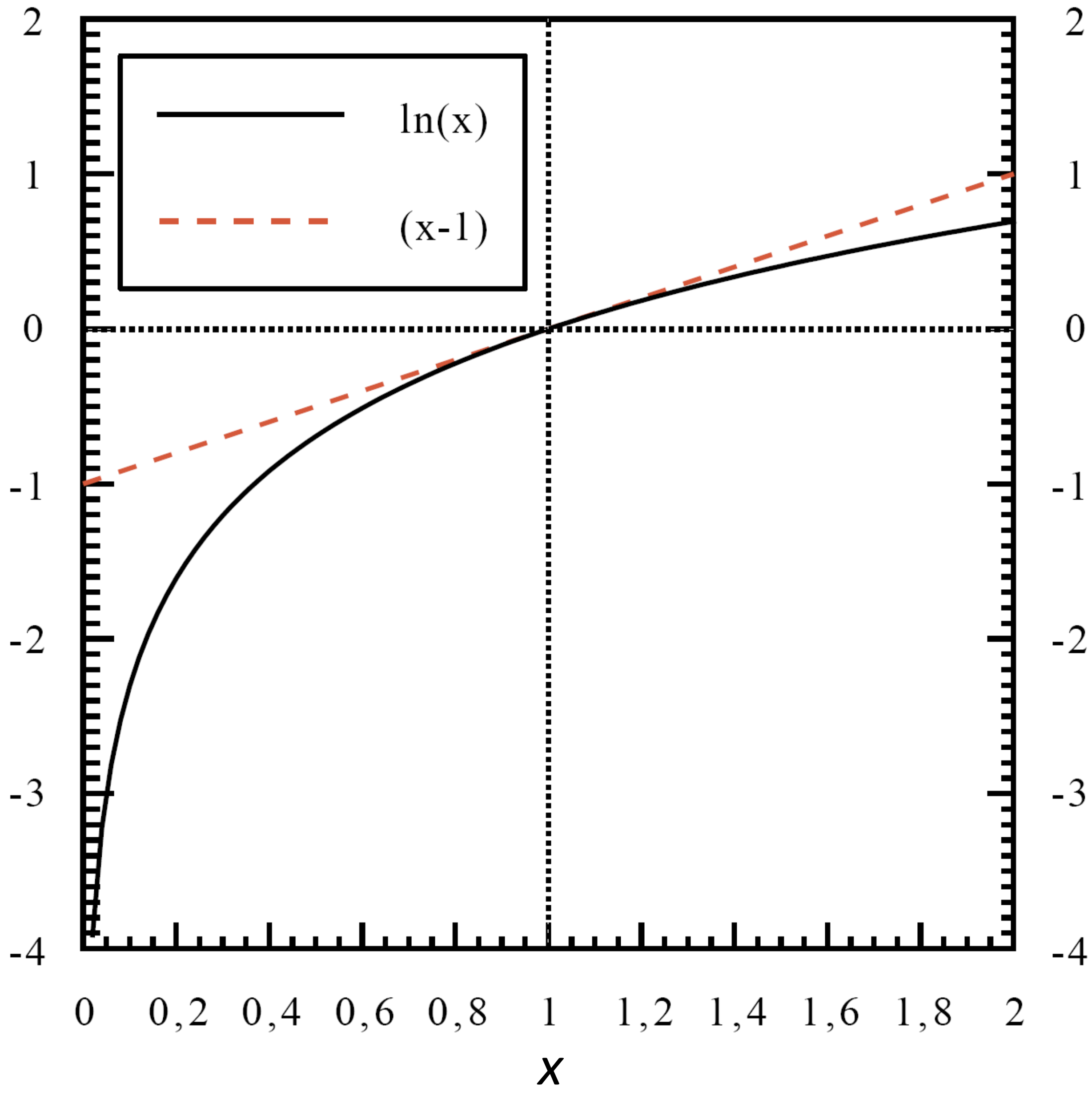
$$H(X) = -\left[p(x_1)\log_b p(x_1) + (1 - p(x_1))\log_b(1 - p(x_1))\right]$$



Information Theory Inequality

$$\log(x) \leq (x-1)\log e, \qquad x > 0$$

with equality iff $x = 1$



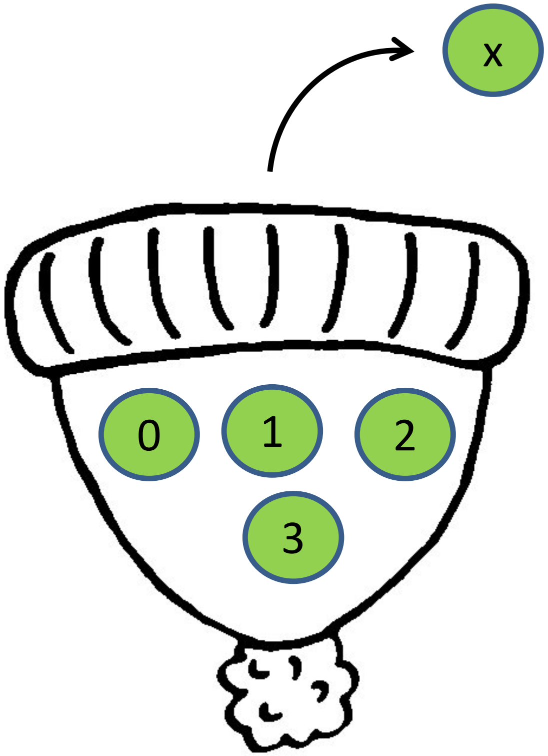
ln(x)
(x-1)
2
1
0
-1
-2
-3
-4
0
0,2
0,4
0,6
0,8
1
1,2
1,4
1,6
1,8
2
x

Entropy H(X)

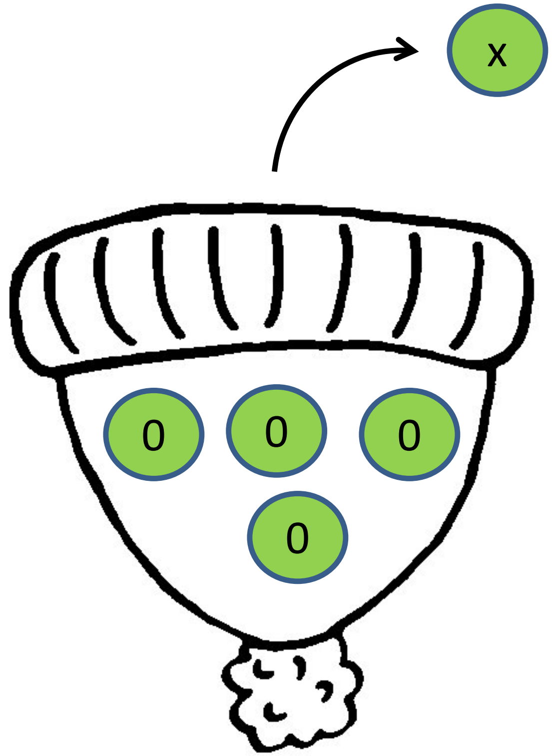
- $0 \leq H(X) \leq \log N$

- The maximum entropy ($H(X) = \log N$) is reached when the symbols are equiprobable

$$p(x_i) = 1/N$$



(c)



(d)

What is H(X) for these sources?

(c) H(X) = 2 bits (d) H(X) = 0 bits