Introduction to Information Theory Part 1

Deepak Kumar
Bryn Mawr College

1

1948, Bell Labs

- Two very significant events
- Two neologisms

2

1948, Bell Labs

- Two very significant events
  1. Invention of an amazingly simple device that could do anything a vacuum tube could do and more efficiently.

     Semiconductor triode? *Iotatron*? …
- Two neologisms
  1. *Transistor*

Spawned the electronic revolution!

3

1948, Bell Labs

- Two very significant events
  1. Invention of an amazingly simple device that could do anything a vacuum tube could do and more efficiently.

     Semiconductor triode? *Iotatron*? …
- Two neologisms
  1. *Transistor*

Spawned the electronic revolution!

4

Shockley Semiconductor Lab

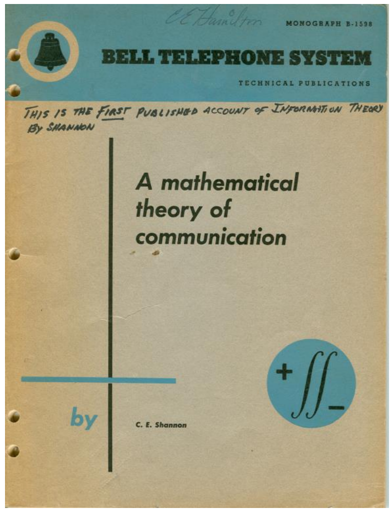
5

Shockley Semiconductor Lab

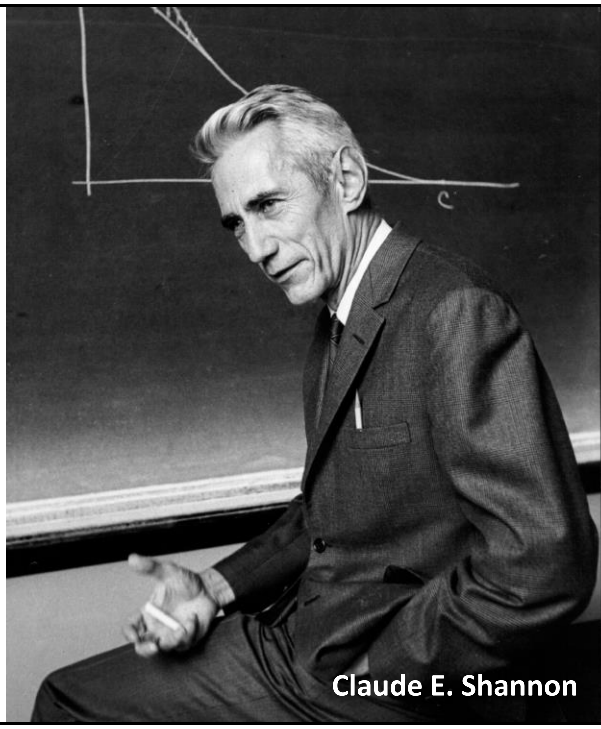
1948, Bell Labs

- Two very significant events
  1. Invention of transistor
  2. Publication of a monograph

- Two neologisms
  1. *Transistor*
  2. *bit* – unit of measuring information
     Not the same as binary digit.

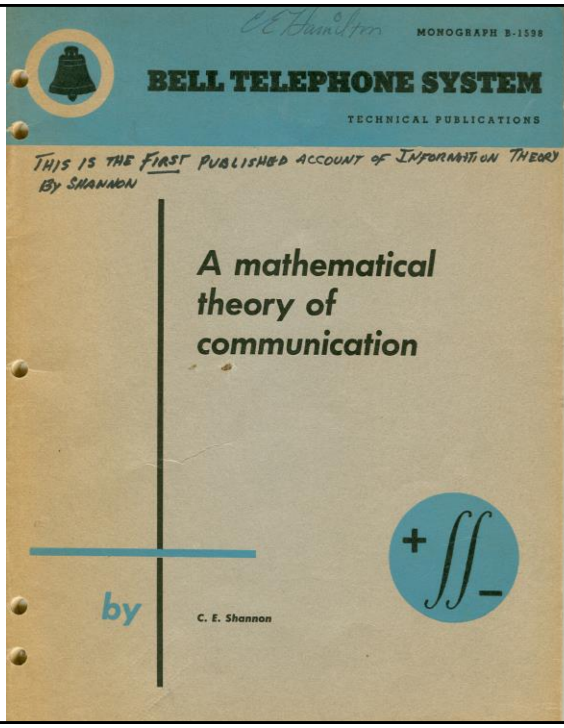


7

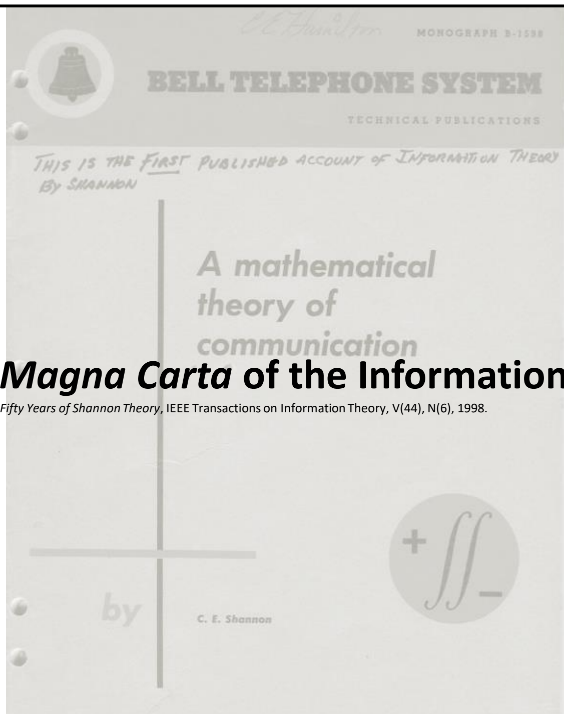


Claude E. Shannon

8



9



The *Magna Carta* of the Information Age

Sergio Verdu, *Fifty Years of Shannon Theory*, IEEE Transactions on Information Theory, V(44), N(6), 1998.

10

A General Communication System

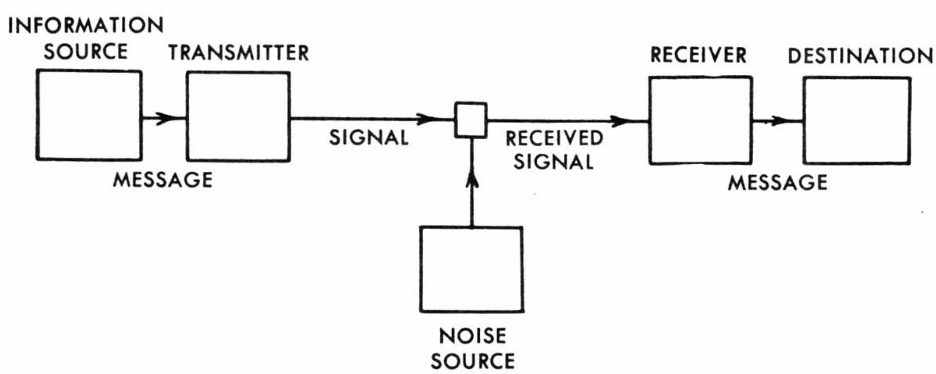
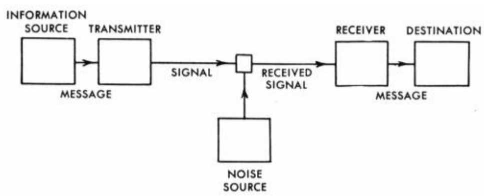


Figure 1.

A Mathematical Theory of Communication, by Claude E. Shannon, *The Bell System Technical Journal,* Vol. 27, pp. 379–423, 623–656, July, October, 1948.

11

Shannon's Information Theory



- Conceptualization of information & modeling of information sources
- Sending of information across the channel:
  - What are the limits on the amount of information that can be sent?
  - What is the effect of noise on this communication.

12

A General Communication System

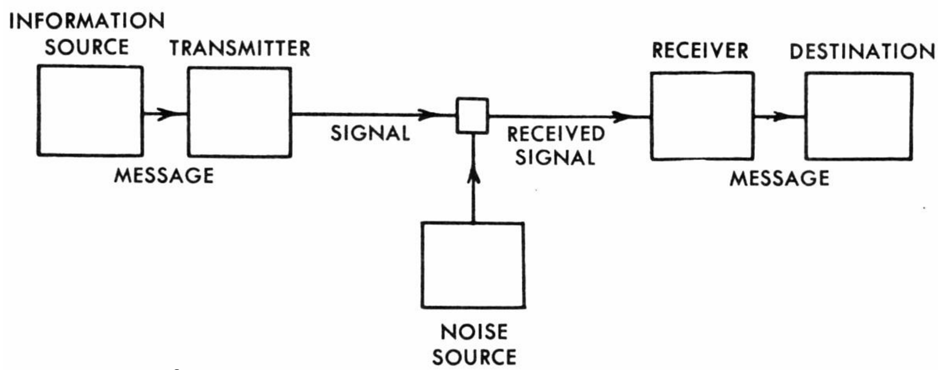
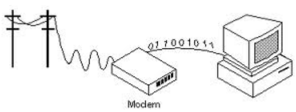


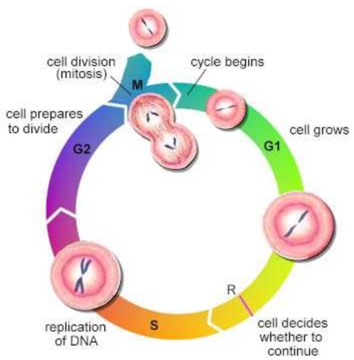
Figure 1.

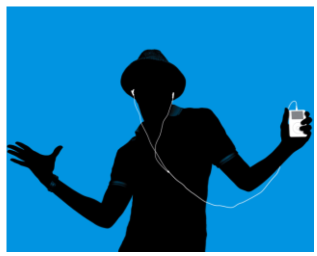
A Mathematical Theory of Communication, by Claude E. Shannon, *The Bell System Technical Journal*, Vol. 27, pp. 379–423, 623–656, July, October, 1948.

13

Examples of Communication Systems







14

How much information?

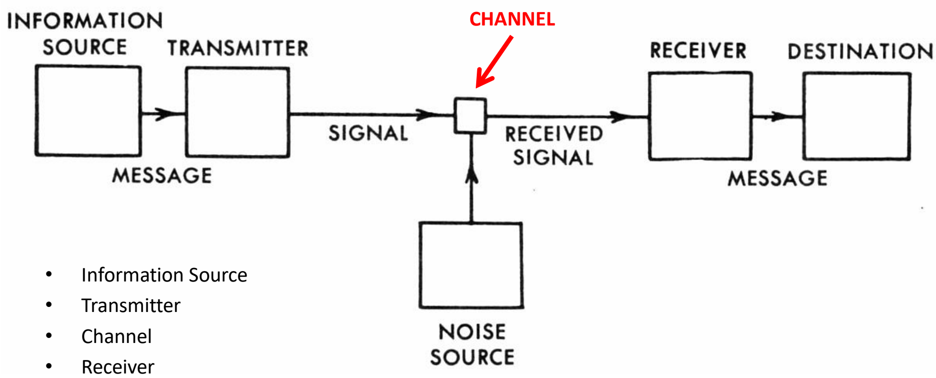
Some more Trivia: Origins of Bluetooth

H (ᚼ) + B (ᛒ) =

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

A Mathematical Theory of Communication, by
Claude E. Shannon, *The Bell System Technical Journal,*
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

A General Communication System

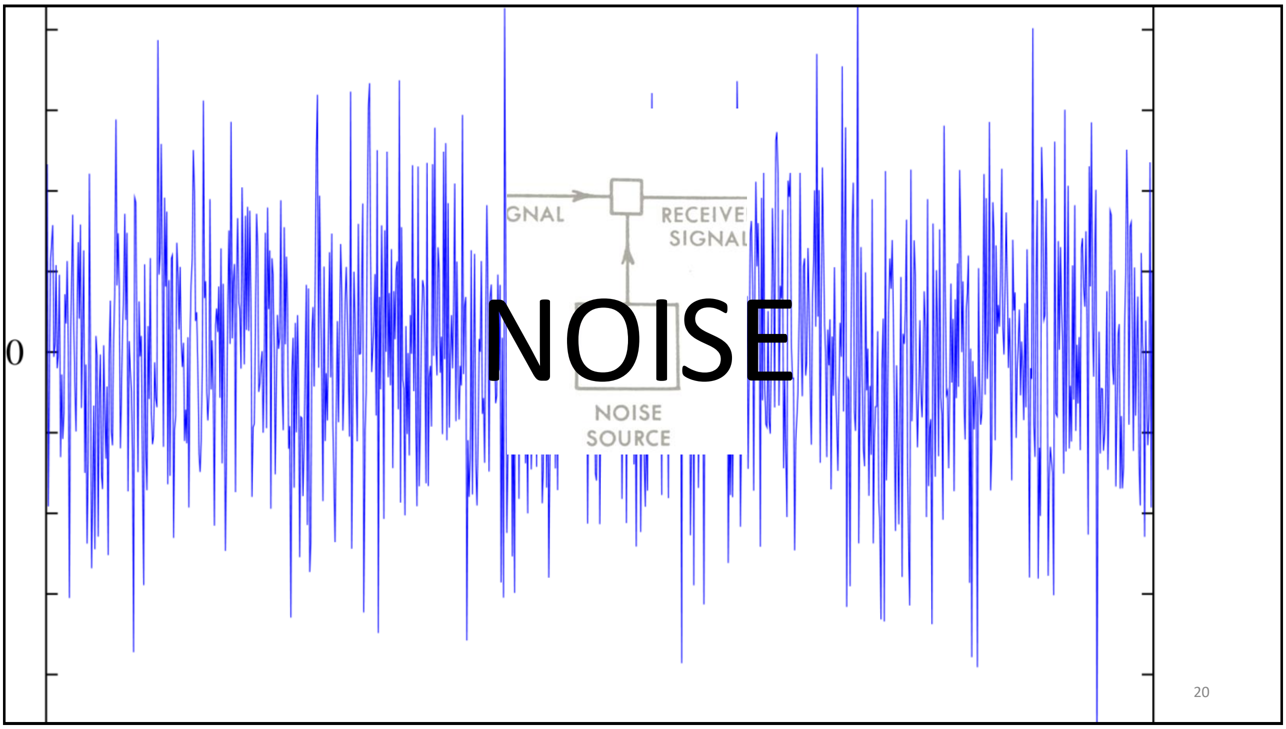


- Information Source
- Transmitter
- Channel
- Receiver
- Destination

A Mathematical Theory of Communication, by
Claude E. Shannon, *The Bell System Technical Journal,*
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

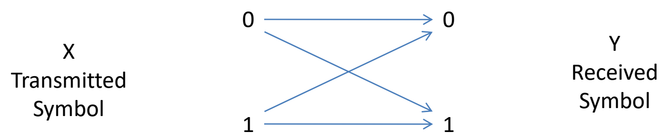
Perfect Communication
(*Noiseless Channel*)

19



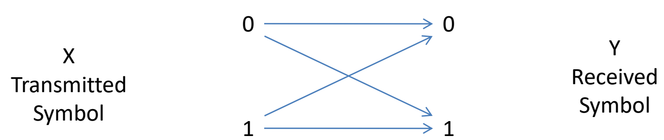
20

Motivating Noise…



21

Motivating Noise…

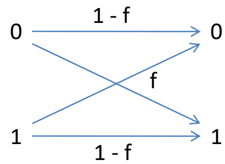


P(Y=1|X=0) = f P(Y=0|X=0) = 1 - f
P(Y=0|X=1) = f P(Y=1|X=1) = 1 - f

22

Motivating Noise…

f = 0.1, n = ~10,000



Question

How can we achieve perfect communication over an imperfect, noisy communication channel?

- Use more reliable components;
- Stabilize the environment;
- Use larger areas;
- Use power/cooling to reduce thermal noise.

These are all costly solutions.

Alternately…

How can we achieve perfect communication over an imperfect, noisy communication channel?

- Accept that there will be noise
- Add error detection and correction
- Introduce the concepts of ENCODER/DECODER

Information Theory
Theoretical limitations of such systems

Coding Theory
Creation of practical encoding/decoding systems

25

Alternately…

How can we achieve perfect communication over an imperfect, noisy communication channel?

- Accept that there will be noise
- Add error detection and correction
- Introduce the concepts of ENCODER/DECODER

REDUNDANCY IS KEY!

Information Theory
Theoretical limitations of such systems

Coding Theory
Creation of practical encoding/decoding systems

26

Shannon's Insight

High Reliability → Low Transmission Rate

I.e. Perfect reliability → Zero Transmission Rate

For a given level of noise there is an associated rate of transmission that can be achieved with arbitrarily good reliability.

e.g. Sending a lone T versus THIS…

Meaning? What meaning?

"Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities.

These semantic aspects of communication are irrelevant to the engineering problem."

A Mathematical Theory of Communication, by Claude E. Shannon, *The Bell System Technical Journal,* Vol. 27, pp. 379–423, 623–656, July, October, 1948.

What is information?

- Just the physical aspects…Shannon, 1948
- The General Definition of Information…Floridi, 2010.

GDI) σ is an instance of information, understood as semantic content, if and only if:

GDI.1) σ consists of *n data,* for $n \geq 1$;

GDI.2) the data are *well formed;*

GDI.3) the well-formed data are *meaningful.*

What is information?

- Just the physical aspects…Shannon, 1948
- The General Definition of Information…Floridi, 2010.

GDI) σ is an instance of information, understood as semantic content, if and only if:

GDI.1) σ consists of *n data,* for $n \geq 1$;

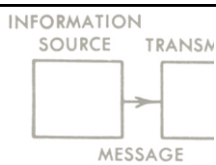
GDI.2) the data are *well formed;*

GDI.3) the well-formed data are *meaningful.*

Shannon Information is…

- Uncertainty
  Can be measured by counting the number of possible messages.
- Surprise
  Some messages are more likely than others.
- Difficult
  What is significant is the difficulty of transmitting the message from one point to the other.
- Entropy
  A fundamental measure of information.

31

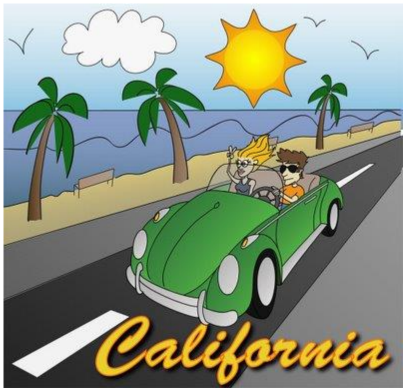


Information Source

- An information source generates a finite number of messages (or symbols).
- Information is quantified using probabilities.
- Given a finite set of possible messages, associate a probability with each message.
- A message with low probability represents more information than one with high probability.

32

Understanding Information



It is sunny in California today!

Information: Definition

- Information is quantified using probabilities.
- Given a finite set of possible messages, associate a probability with each message.
- A message with low probability represents more information than one with high probability.

Definition of Information:

$$\boldsymbol{I(p) = log\left(\frac{1}{p}\right) = -log(p)}$$

Where $\boldsymbol{p}$ is the probability of the message
Base 2 is used for the logarithm so $\boldsymbol{I}$ is measured in **bits**
Trits for base 3, **nats** for base e, **Hartleys** for base 10…

Information in a coin flip

$$P(HEADS) = ½$$

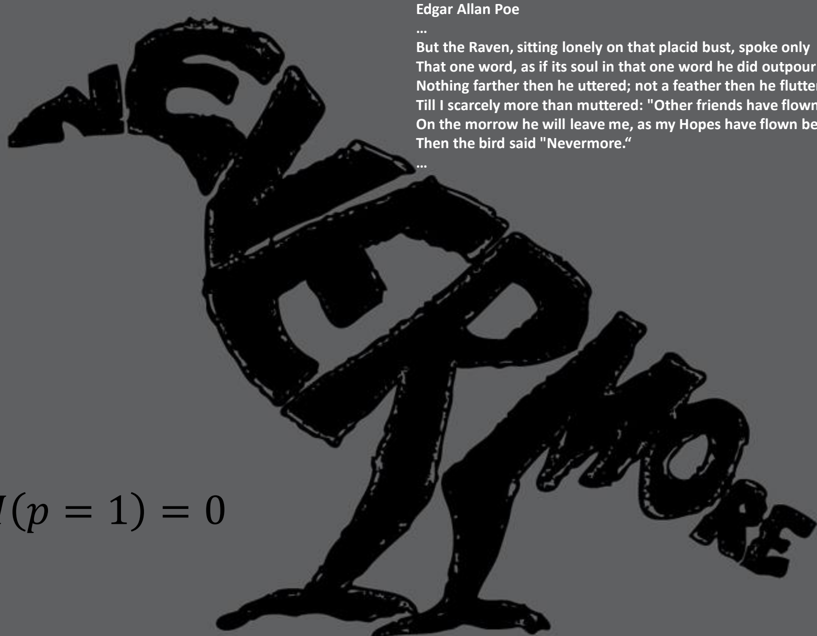
$$I = -\log(½) = 1 \text{ bit of information}$$

Given a sequence of 14 coin flips: hthhtththhttht

We will need 14 bits: 10110010110010

35

The Raven
Edgar Allan Poe
…
But the Raven, sitting lonely on that placid bust, spoke only
That one word, as if its soul in that one word he did outpour
Nothing farther then he uttered; not a feather then he fluttered--
Till I scarcely more than muttered: "Other friends have flown before--
On the morrow he will leave me, as my Hopes have flown before."
Then the bird said "Nevermore."
…



$$I(p = 1) = 0$$

Picture From: http://adrianamelnic.ro/2012/06/theraven/

36

$$\boldsymbol{I(p) = \log\left({}^{1}\!/_{p}\right) = -\log(p)}$$

Some properties of I

1. $I(p) \geq 0$
   Information is non-negative.

2. $I(p_1 * p_2) = I(p_1) + I(p_1)$
   Information we get from observing two independent events occurring is the sum of two information(s).

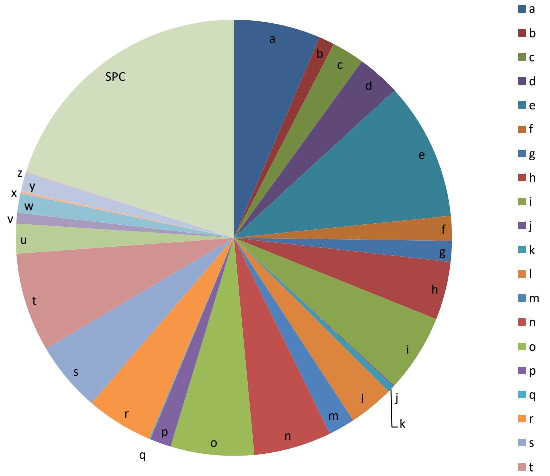
3. $I(p)$ is monotonic and continuous in p
   Slight changes in probability incur slight changes in information.

4. $I(p = 1) = 0$
   We get zero information from an event whose probability is 1.

37

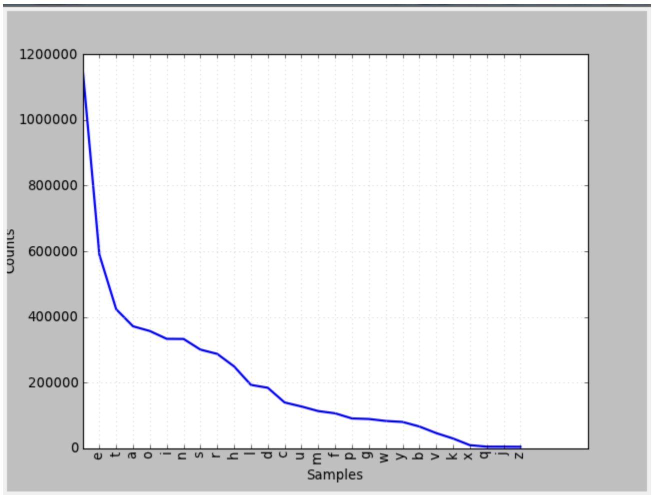
Example: Text Analysis

a	0.06428
b	0.01147
c	0.02413
d	0.03188
e	0.10210
f	0.01842
g	0.01543
h	0.04313
i	0.05767
j	0.00082
k	0.00514
l	0.03338
m	0.01959
n	0.05761
o	0.06179
p	0.01571
q	0.00084
r	0.04973
s	0.05199
t	0.07327
u	0.02201
v	0.00800
w	0.01439
x	0.00162
y	0.01387
z	0.00077
SPC	0.20096



38

Example: Text Analysis



39

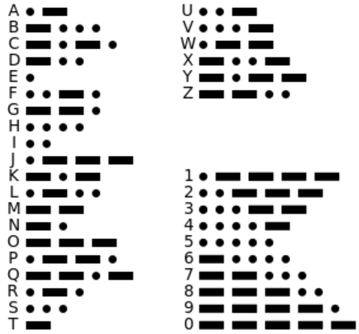
al Kindi
(9th Centry C.E.)

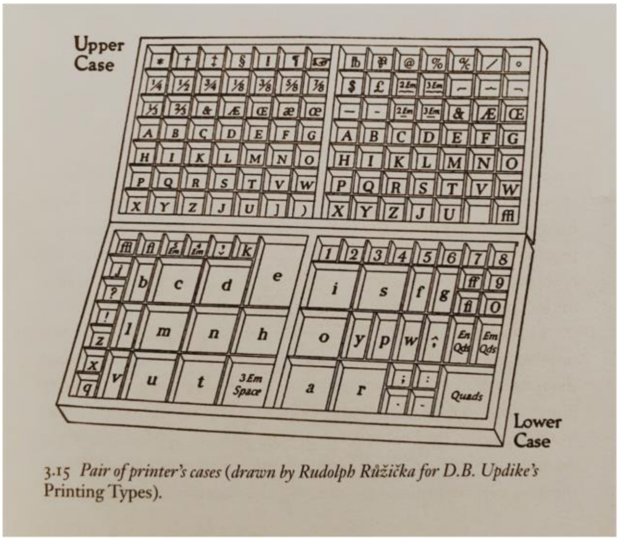
From: Mathematicians on Stamps: http://jeff560.tripod.com/stamps.html

40

Morse Code

- Short codes for common letters, longer codes for less common letters
- Sounds: [Click here]

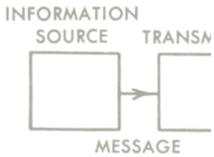




3.15 *Pair of printer's cases (drawn by Rudolph Růžička for D.B. Updike's* Printing Types).

Example: Text Analysis

Letter	Freq.	I
a	0.06428	3.95951
b	0.01147	6.44597
c	0.02413	5.37297
d	0.03188	4.97116
e	0.10210	3.29188
f	0.01842	5.76293
g	0.01543	6.01840
h	0.04313	4.53514
i	0.05767	4.11611
j	0.00082	10.24909
k	0.00514	7.60474
l	0.03338	4.90474
m	0.01959	5.67385
n	0.05761	4.11743
o	0.06179	4.01654
p	0.01571	5.99226
q	0.00084	10.21486
r	0.04973	4.32981
s	0.05199	4.26552
t	0.07327	3.77056
u	0.02201	5.50592
v	0.00800	6.96640
w	0.01439	6.11899
x	0.00162	9.26697
y	0.01387	6.17152
z	0.00077	10.34877
SPC	0.20096	2.31502



Information Source

- An information source generates a finite number of messages (or symbols)

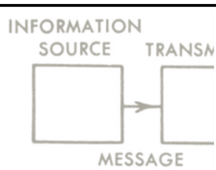
$$\{a_1, a_2, \dots, a_n\}$$

- Source emits the symbols with probabilities

$$\mathrm{P} = \{p_1, p_2, \dots, p_n\}$$

- Assume independence: successive symbols do not depend on past symbols.
- What is the average amount of information?

43



Information Source

- An information source generates a finite number of messages (or symbols)

$$\{a_1, a_2, \dots, a_n\}$$

- Source emits the symbols with probabilities

$$\mathrm{P} = \{p_1, p_2, \dots, p_n\}$$

- Assume independence: successive symbols do not depend on past symbols.
- What is the average amount of information?
  ANSWER: Entropy!

44

Definition of Entropy

- Information (I) is associated with known events/messages
- Entropy (H) is the average information w.r.to all possible outcomes

$$H(P) = \sum_i p_i \log \frac{1}{p_i}$$

Example: Text Analysis

Letter	Freq.	I
a	0.06428	3.95951
b	0.01147	6.44597
c	0.02413	5.37297
d	0.03188	4.97116
e	0.10210	3.29188
f	0.01842	5.76293
g	0.01543	6.01840
h	0.04313	4.53514
i	0.05767	4.11611
j	0.00082	10.24909
k	0.00514	7.60474
l	0.03338	4.90474
m	0.01959	5.67385
n	0.05761	4.11743
o	0.06179	4.01654
p	0.01571	5.99226
q	0.00084	10.21486
r	0.04973	4.32981
s	0.05199	4.26552
t	0.07327	3.77056
u	0.02201	5.50592
v	0.00800	6.96640
w	0.01439	6.11899
x	0.00162	9.26697
y	0.01387	6.17152
z	0.00077	10.34877
SPC	0.20096	2.31502

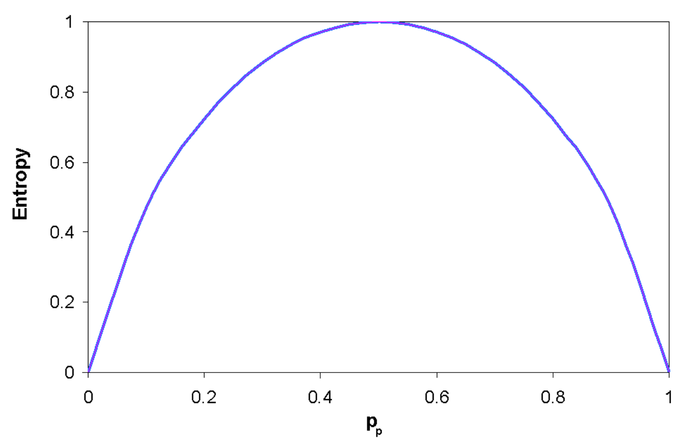
$$H(P) = \sum_i p_i \log \frac{1}{p_i} = 4.047$$

Entropy: What about it?

- Does $H(P)$ have a maximum? Where?
- Is entropy a good name for this stuff? How is it related to entropy in thermodynamics?
- How does entropy help in communication? What else can we do with it?
- Why use the letter H? ☺

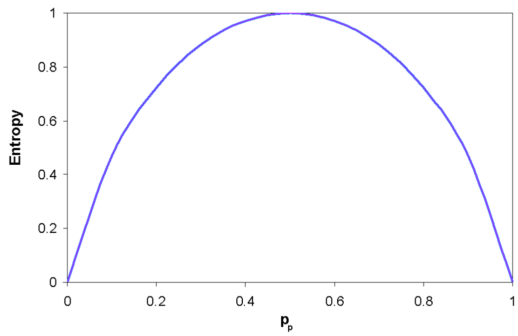
47

Entropy (2 outcomes)



48

Entropy: Properties



$$0 \le H(P_n) \le \log(n)$$

Entropy is maximized if P is uniform.

H for Entropy?

“The enthalpy is [often] written U. V is the volume, and Z is the partition function. P and Q are the position and momentum of a particle. R is the gas constant, and of course T is temperature. W is the number of ways of configuring our system (the number of states), and we have to keep X and Y in case we need more variables. Going back to the first half of the alphabet, A, F, and G are all different kinds of free energies (the last named for Gibbs). B is a virial coefficient or a magnetic field. I will be used as a symbol for information; J and L are angular momenta. K is Kelvin, which is the proper unit of T. M is magnetization, and N is a number, possibly Avogadro's, and O is too easily confused with 0. This leaves S . . ."

From Spikes: Exploring the Neural Code, by Reike et al, Bradford 1999.

…and H.

In Spikes they also eliminate H (e.g., as the Hamiltonian). But, others, I, along with Shannon, prefer to honor Hartley. Thus, H for entropy . . .

From: Tom Carter’s *Lecture Notes on Information Theory and Entropy,* Prepared for Complex Systems Summer School, Santa Fe, June 2012. sustan.csustan.edu/~tom/

References

- Eugene Chiu, Jocelyn Lin, Brok Mcferron, Noshirwan Petigara, Satwiksai Seshasai: *Mathematical Theory of Claude Shannon: A study of the style and context of his work up to the genesis of information theory. MIT* 6.933J / STS.420J The Structure of Engineering Revolutions
- Luciano Floridi, 2010: *Information: A Very Short Introduction*, Oxford University Press, 2011.
- Luciano Floridi, 2011: *The Philosophy of Information*, Oxford University Press, 2011.
- James Gleick, 2011: *The Information: A History, A Theory, A Flood*, Pantheon Books, 2011.
- David Luenberger, 2006: *Information Science*, Princeton University Press, 2006.
- David J.C. MacKay, 2003: *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.
- Claude Shannon & Warren Weaver, 1949: *The Mathematical Theory of Communication*, University of Illinois Press, 1949.
- W. N. Francis and H. Kucera: *Brown University Standard Corpus of Present-Day American English*, Brown University, 1967.
- Simon Singh: *The Code Book, The Science of Secrecy From Ancient Egypt to Quantum Cryptography*. Anchor Books, 1999.
- Tom Carter, *Lecture Notes on Information Theory and Entropy*. Prepared for Complex Systems Summer School, Santa Fe, June 2012.