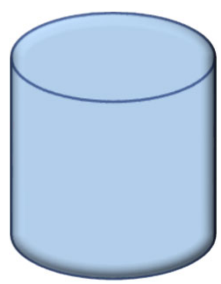
INFORMATION

observer and choice

- *Information* is defined as "a measure of the freedom from *choice* with which a message is *selected* from the set of all possible messages"
- Bit (short for *binary digit*) is the most elementary **choice** one can make
  - Between two items: "0' and "1", "heads" or "tails", "true" or "false", etc.
  - Bit is equivalent to the choice between two equally likely alternatives
    - Example, if we know that a coin is to be tossed, but are unable to see it as it falls, a message telling whether the coin came up heads or tails gives us one bit of information

1 Bit of *information*

uncertainty ***removed***, information gained



1 Bit of uncertainty

H,T?

choice between 2 symbols recognized by an observer

INDIANA UNIVERSITY
rocha@indiana.edu
informatics.indiana.edu/rocha

Fathers of uncertainty-based information

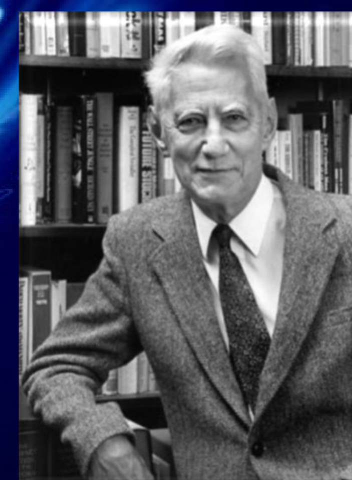
- Information is transmitted through noisy communication channels
  - Ralph Hartley and Claude Shannon (at Bell Labs), the fathers of Information Theory, worked on the problem of efficiently transmitting information; i. e. ***decreasing the uncertainty*** in the transmission of information.

Hartley, R.V.L., "Transmission of Information", *Bell System Technical Journal*, July 1928, p.535.

C. E. Shannon [1948], “A mathematical theory of communication”. *Bell System Technical Journal*, **27**:379-423 and 623-656

C. E. Shannon, “A Symbolic analysis of relay and switching circuits” .*MS Thesis*, (unpublished) MIT, 1937.

C. E. Shannon, “An algebra for theoretical genetics.” *Phd Dissertation*, MIT, 1940.

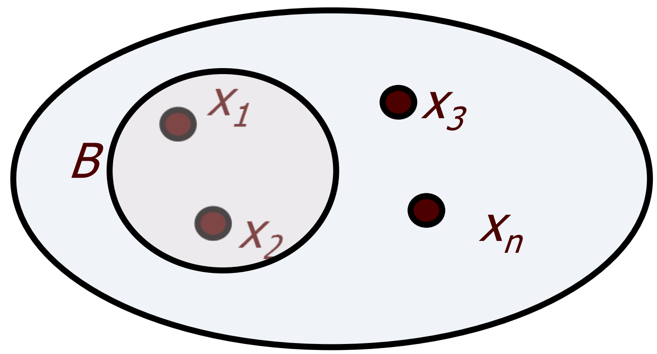


- Multiplication Principle
  - “If some choice can be made in M different ways, and some subsequent choice can be made in N different ways, then there are M x N different ways these choices can be made in succession” [Paulos]
    - 3 shirts and 4 pants = 3 x 4 = 12 outfit choices

rocha@indiana.edu
informatics.indiana.edu/rocha

- Nonspecificity
  - Hartley measure
    - The amount of uncertainty associated with a set of alternatives (e.g. messages) is measured by the **amount of information needed to <u>remove</u> the uncertainty**

A = Set of Alternatives



Quantifies how many yes-no questions need to be asked to establish what the correct alternative is

$$H(A) = \log_2 |A|$$

Measured in bits

Number of Choices

Elementary Choice is between 2 alternatives: 1 bit

$$H(B) = \log_2(2) = 1$$

$$\log_2(4) = 2 \qquad 2^2 = 4$$

$$\log_2(16) = 4 \qquad 2^4 = 16$$

$$\log_2(1) = 0$$

$$H(A) = \log_2(16) = 4$$

$$H(B) = \log_2(4) = 2$$

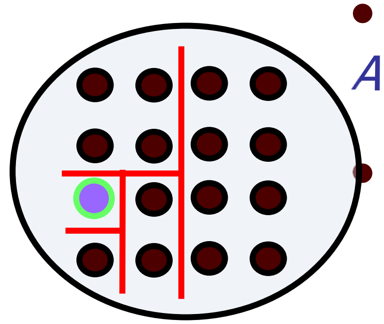
$$H(A) = \log_2 |A|$$

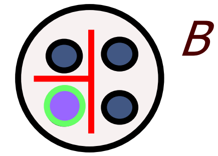
Measured in bits

Number of Choices

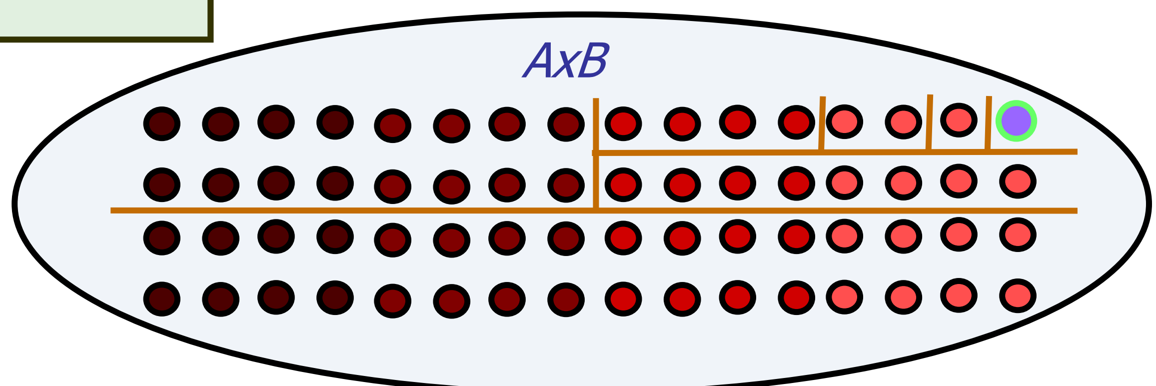
Quantifies how many yes-no questions need to be asked to establish what the correct alternative is

- Example
  - Menu Choices
    - A = 16 Entrees
    - B = 4 Desserts
  - How many dinner combinations?
    - 16 x 4 = 64



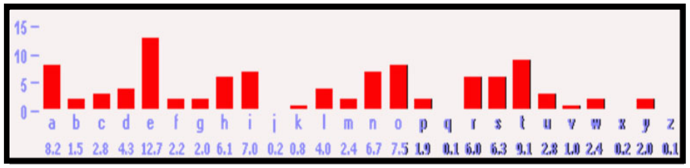


$$H(A \times B) = \log_2(16 \times 4) =$$
$$= \log_2(16) + \log_2(4) = 6$$

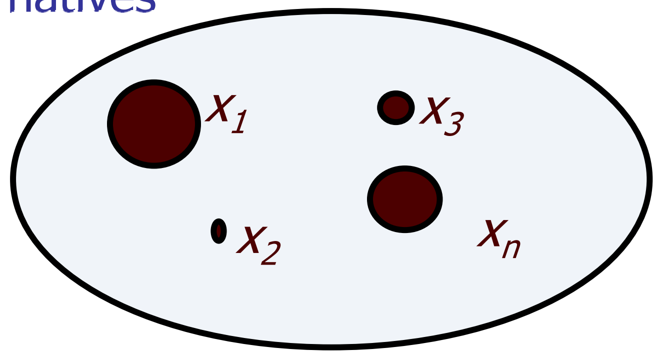


INDIANA UNIVERSITY
rocha@indiana.edu
informatics.indiana.edu/rocha

uncertainty-based information



A = Set of weighted Alternatives



- Shannon's measure
  - The ***average*** amount of uncertainty associated with a set of ***weighted*** alternatives (e.g. messages) is measured by the ***average*** amount of information needed to remove the uncertainty

$$H_S(A) = -\sum_{i=1}^{n} p(x_i)\log_2\left(p(x_i)\right)$$

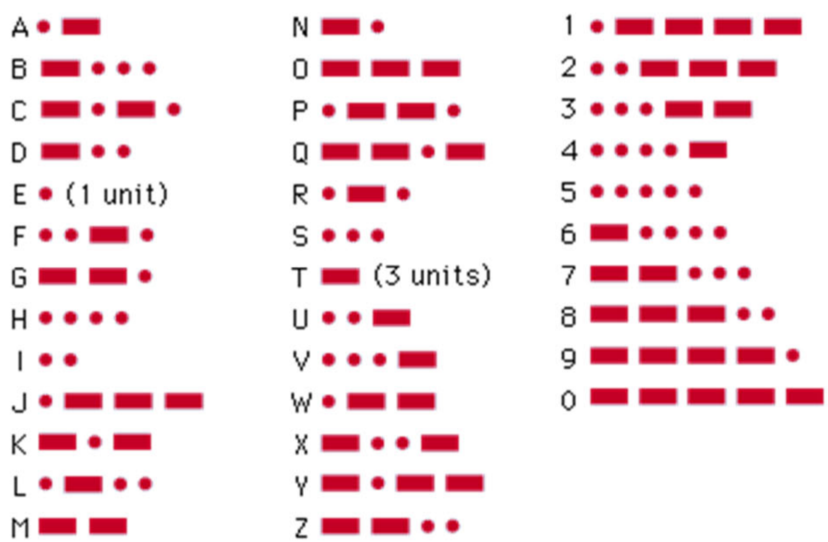
Probability of alternative

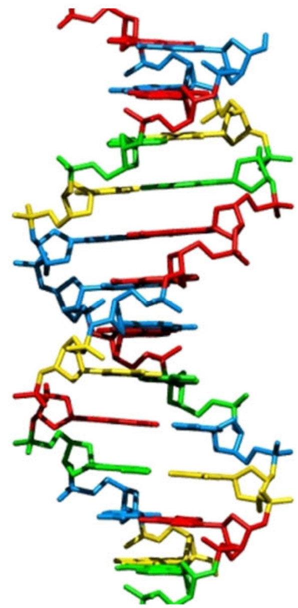
Measured in bits

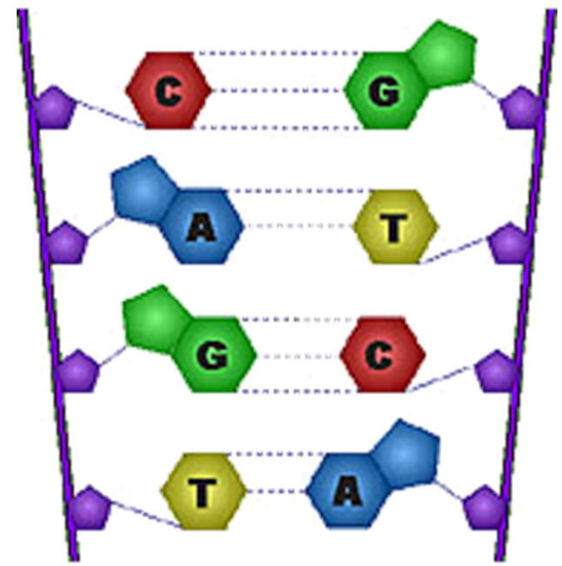
INDIANA UNIVERSITY
rocha@indiana.edu
informatics.indiana.edu/rocha

Message encoded in an alphabet of ***n*** symbols, for example:

- English (26 letters + space + punctuations)
- Morse code (dot, dash, space)
- DNA (A, T, G, C)







INDIANA UNIVERSITY
rocha@indiana.edu
informatics.indiana.edu/rocha

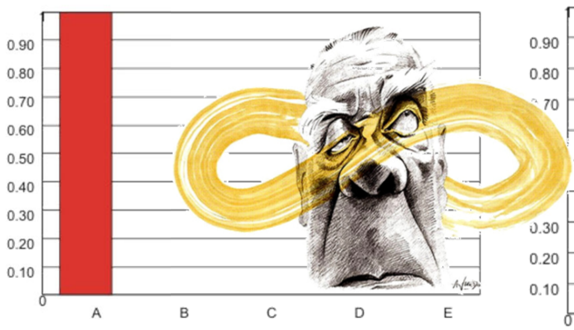
- Given a symbol set {A,B,C,D,E}
  - And occurrence probabilities P_A, P_B, P_C, P_D, P_E,
- The Shannon entropy is
  - The average minimum number of bits needed to represent a symbol

$$H_S = -\left(p_A \log_2(p_A) + p_B \log_2(p_B) + p_C \log_2(p_C) + p_D \log_2(p_D) + p_E \log_2(p_E)\right)$$

$$H_S = -(1.\log_2(1) + 0.\log_2(0) + 0.\log_2(0) + 0.\log_2(0) + 0.\log_2(0)) = -\log_2(1)$$

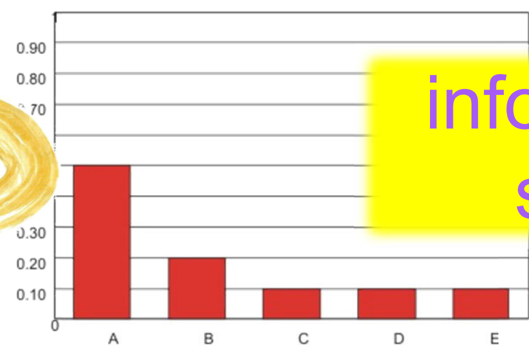
$$H_S = -5.\left(\frac{1}{5}\right).\log_2\left(\frac{1}{5}\right) = -(\log_2(1) - \log_2(5)) = \log_2(5)$$

$$H_S = -(\tfrac{1}{2}.\log_2\left(\tfrac{1}{2}\right) + \tfrac{1}{5}.\log_2\left(\tfrac{1}{5}\right) + 3.\left(\tfrac{1}{10}\right).\log_2\left(\tfrac{1}{10}\right))$$



H_S = 0 bits

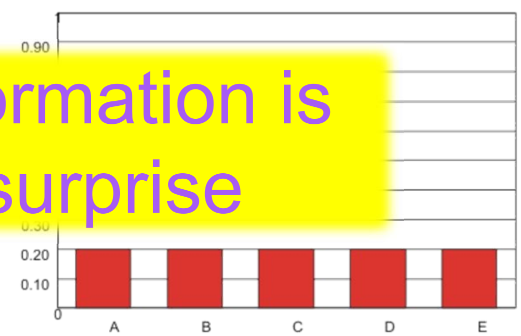
0 questions



H_S = 1.96

≈2 questions

information is surprise



H_S = 2.32 bits

INDIANA UNIVERSITY

rocha@indiana.edu

informatics.indiana.edu/rocha

what it measures

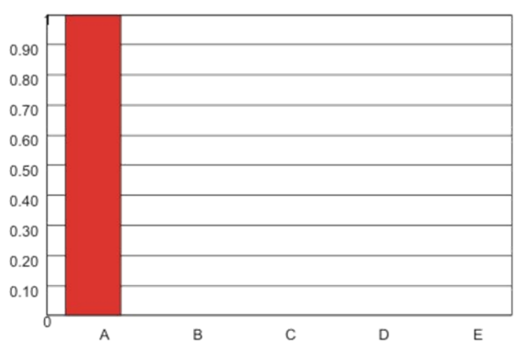
uncertainty, about outcome. How much information is gained when symbol is known

- ***on average***, how many *yes-no* questions need to be asked to establish what the symbol is
- “structure” of uncertainty in situations

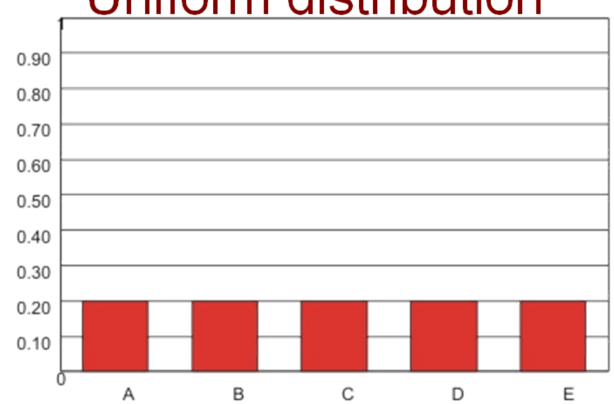
$$H_S(A) = -\sum_{i=1}^{n} p(x_i)\log_2\left(p(x_i)\right)$$

$$H_S \in \left[0, \log_2 |X|\right]$$

For one alternative

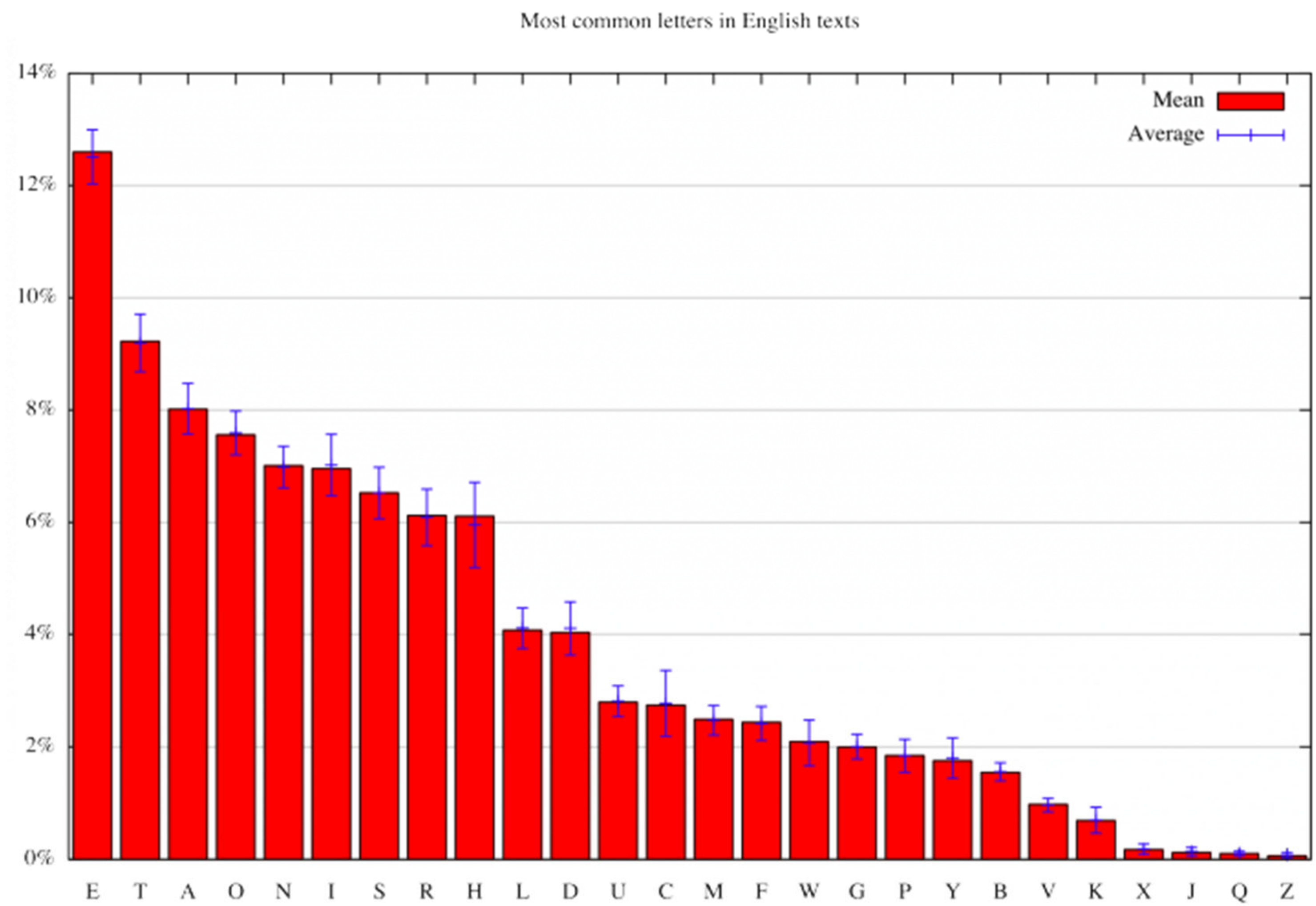


Uniform distribution



INDIANA UNIVERSITY
rocha@indiana.edu
informatics.indiana.edu/rocha

from letter frequency



http://www.macfreek.nl/memory/Letter_Distribution

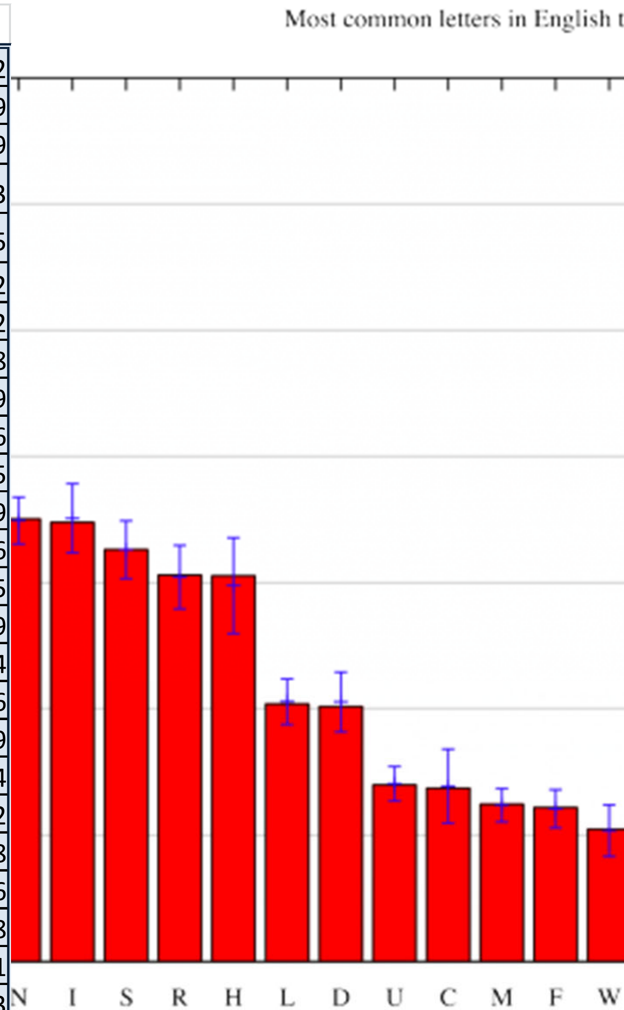
INDIANA UNIVERSITY
rocha@indiana.edu
informatics.indiana.edu/rocha

english entropy (rate)

from letter frequency

	p(x)	log2(p(x))	-p(x).log2(p(x))
e	0.124167	-3.0096463	0.373698752
t	0.096923	-3.3670246	0.326340439
a	0.082001	-3.6082129	0.295877429
i	0.076805	-3.7026522	0.284382943
n	0.076406	-3.7101797	0.283478135
o	0.07141	-3.8077402	0.271908822
s	0.070677	-3.8226195	0.270170512
r	0.066813	-3.903723	0.260820228
l	0.044831	-4.4793659	0.200813559
d	0.036371	-4.7810716	0.173891876
h	0.035039	-4.8349111	0.169408515
c	0.034439	-4.8598087	0.167367439
u	0.028777	-5.11894	0.147307736
m	0.028		094755
f	0.023		220629
p	0.020317	-5.6211617	0.114205704
y	0.018918	-5.7240814	0.108289316
g	0.018119	-5.7863688	0.104842059
w	0.013523	-6.2084943	0.083954364
v	0.012457	-6.3269343	0.078812722
b	0.010658	-6.5519059	0.069830868
k	0.00393	-7.9911852	0.031406876
x	0.002198	-8.8294354	0.019409218
j	0.001998	-8.9669389	0.017919531
q	0.000933	-10.066609	0.009387113
z	0.000599	-10.705156	0.006412389
		Entropy	**4.14225193**

Hartley Measure
H(|26|) 4.7004397



	p(x)	log2(p(x))	-p(x).log2(p(x))
Space	0.18288	-2.4509943	0.448249175
E	0.10267	-3.2839625	0.337152952
T	0.07517	-3.7336995	0.280662128
A	0.06532	-3.9362945	0.257125332
O	0.06160	-4.0210249	0.247678132
N	0.05712	-4.1298574	0.235897914
I	0.05668	-4.1409036	0.234724772
S	0.05317	-4.2332423	0.225081718
R	0.04988	-4.3254212	0.215748053
H	0.04979	-4.3281265	0.215478547
L	0.		63015644
D	0.		61811184
U	0.02276	-5.4574854	0.124201198
C	0.02234	-5.4844363	0.122504535
M	0.02027	-5.6248177	0.113990747
F	0.01983	-5.6561227	0.112164711
W	0.01704	-5.8750208	0.100104113
G	0.01625	-5.9435013	0.096576215
P	0.01504	-6.0547406	0.091082933
Y	0.01428	-6.1301971	0.087518777
B	0.01259	-6.3117146	0.079456959
V	0.00796	-6.9728048	0.055511646
K	0.00561	-7.4778794	0.041948116
X	0.00141	-9.4709063	0.013346416
J	0.00098	-10.001987	0.009754119
Q	0.00084	-10.222907	0.008554069
Z	0.00051	-10.929184	0.005604998
		Entropy	**4.0849451**

Hartley Measure
H(|27|) 4.7548875

http://www.macfreek.nl/memory/Letter_Distribution

INDIANA UNIVERSITY
rocha@indiana.edu
informatics.indiana.edu/rocha

entropy and meaning

- entropy quantifies information (surprise), but it does not consider information content
  - semantic aspects of information are irrelevant to the engineering problem in Shannon's conception

$$H_S(A) = -\sum_{i=1}^{n} p(x_i)\log_2\left(p(x_i)\right)$$

If you don't wanna see me

Did a full one eighty, crazy
Thinking 'bout the way I was
Did the heartbreak change me? Maybe
But look at where I ended up

I'm all good already
So moved on it's scary
I'm not where you left me at all

So if you don't wanna see me dancing with somebody
If you wanna believe that anything could stop me

Don't show up, don't come out
Don't start caring about me now
Walk away, you know how
Don't start caring about me now

Aren't you the guy who tried to
Hurt me with the word "goodbye"?
Though it took some time to survive you
I'm better on the other side

tIf nnom w ueayso e 'naed

,aa uz y rdi IlgitDohyfec ne
nhbn ygTwtwIsuah ki'e aoti
edmhir argh? abeDkaetcehatebenMy
odandw p Buloue elrtkhte e

ndet e t tsremoitro ebhhel'

INDIANA UNIVERSITY
rocha@indiana.edu
informatics.indiana.edu/rocha

entropy according to probabilistic model

0^{th} order model: equiprobable symbols

$$H(A) = \log_2 |A|$$

Hartley Measure
H(|27|) 4.7548875

XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGXYD QPAAMKBZAACIBZLHJQD

1^{st} order model: frequency of symbols

$$H_S(A) = -\sum_{i=1}^{n} p(x_i)\log_2(p(x_i))$$

H_S =4.08

OCRO HLI RGWR NMIELWIS EU LL NBNESBEYA TH EEI ALHENHTTPA OOBTTVA NAH BRL

2^{nd} order model: frequency of digrams

ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE

Most common *digrams*: th, he, in, en, nt, re, er, an, ti, es, on, at, se, nd, or, ar, al, te, co, de, to, ra, et, ed, it, sa, em, ro.

3^{rd} order model: frequency of trigrams

IN NO IST LAT WHEY CRATICT FROURE BERS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE

Most common *trigrams*: the, and, tha, ent, ing, ion, tio, for, nde, has, nce, edt, tis, oft, sth, men

H_S = 2.8

4^{th} order model: frequency of tetragrams

THE GENERATED JOB PROVIDUAL BETTER TRAND THE DISPLAYED CODE ABOVERY UPONDULTS WELL THE CODERST IN THESTICAL IT DO HOCK BOTHE MERG INSTATES CONS ERATION NEVER ANY OF PUBLE AND TO THEORY EVENTIAL CALLEGAND TO ELAST BENERATED IN WITH PIES AS IS WITH THE

including more structure reduces surprise

http://pages.central.edu/emp/LintonT/classes/spring01/cryptography/letterfreq.html

http://everything2.com/title/entropy+of+English

INDIANA UNIVERSITY

rocha@indiana.edu
informatics.indiana.edu/rocha

Other measures

- Mutual Information
  - Amount of information about one variable that can be gained (uncertainty reduced) by observing another variable
- Information Gain (Kullback-Leibler Divergence)
  - Difference between two probability distributions *p* and *q*,
    - average number of bits per data point needed in order to represent *q* (model approximation) as it deviates from *p* ("true" or theoretical distribution)
- Transfer Entropy
  - transfer of information between two random processes in time
    - Amount of information (in bits) gained, or uncertainty lost, in knowing future values of Y, knowing the past values of X and Y.

$$I(X;Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2 \frac{p(x_i, y_j)}{p(x_i)p(y_j)}$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

$$IG(p(X), q(X)) = \sum_{i=1}^{n} p(x_i)\log_2 \frac{p(x_i)}{q(x_i)}$$

$$T_{X\rightarrow Y} = H(Y_t|Y_{t-1:t-L}) - H(Y_t|Y_{t-1:t-L}, X_{t-1:t-L})$$

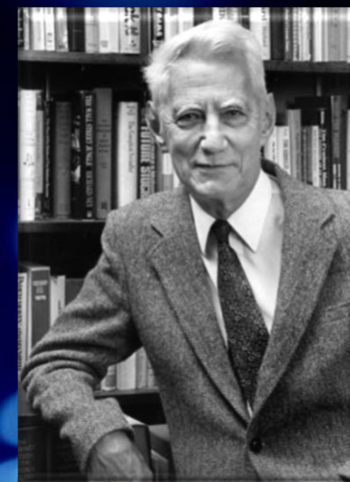
INDIANA UNIVERSITY
rocha@indiana.edu
informatics.indiana.edu/rocha

Information as decrease in uncertainty .

$$H(A) = \log_2 |A|$$

Measured in bits

Number of Choices



Hartley, R.V.L., "Transmission of Information", *Bell System Technical Journal*, July 1928, p.535.

$$H_S(A) = -\sum_{i=1}^{n} p(x_i)\log_2(p(x_i))$$

Measured in bits

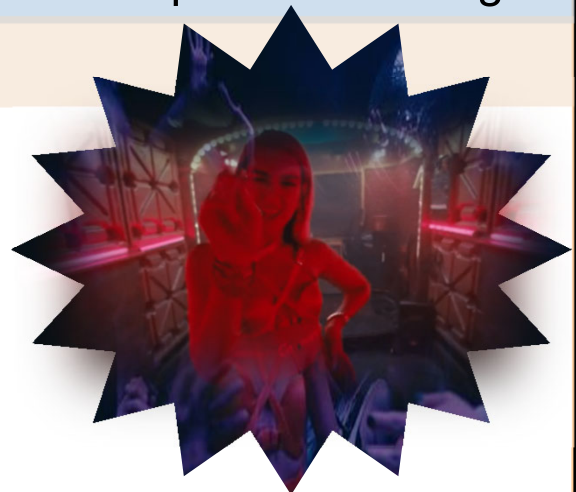
Probability of alternative

including more structure reduces surprise

information is surprise

C. E. Shannon [1948], "A mathematical theory of communication". *Bell System Technical Journal*, **27**:379-423 and 623-656

rate of removing uncertainty of each symbol



If you don't wanna see me

Did a full one eighty, crazy
Thinking 'bout the way I was
Did the heartbreak change me? Maybe
But look at where I ended up

I'm all good already
So moved on it's scary
I'm not where you left me at all

So if you don't wanna see me dancing with somebody

"syntactic" surprise But what about function and meaning (semantics)?

INDIANA UNIVERSITY
rocha@indiana.edu
informatics.indiana.edu/rocha