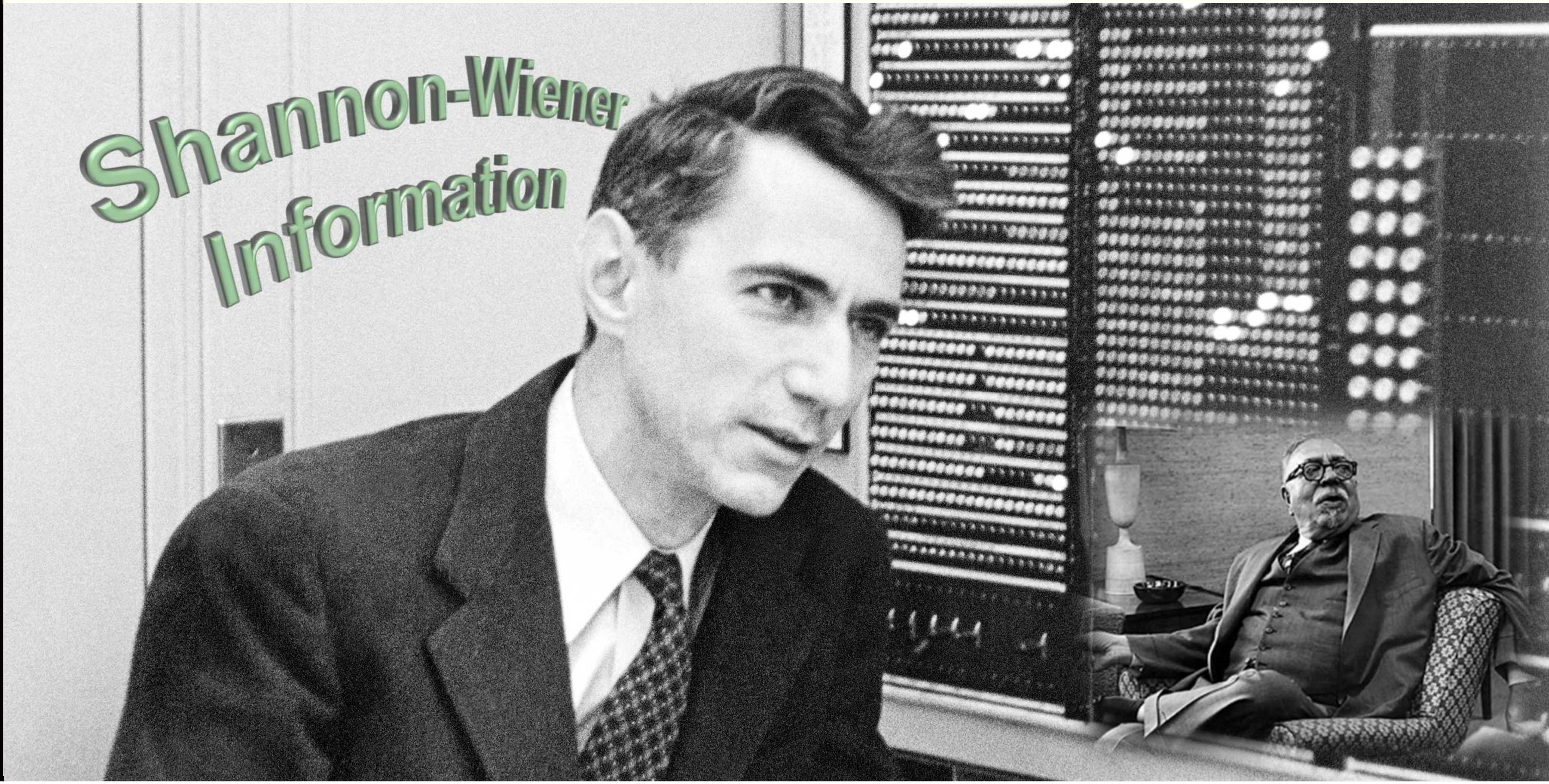
lecture 2: Shannon(-Wiener) Information



Complex adaptive systems and computational intelligence (casci lab)

Resources

- web page
  - casci.binghamton.edu/academics/ssie501
- online class
  - binghamton.zoom.us/j/93351260610
- blog: sciber
  - sciber.blogspot.com
- Brightspace
  - brightspace.binghamton.edu/d2l/home/358842

SSIE-501/ISE-440 - Fall 2024

luis m. rocha

office hours:
Tuesdays 9:00- 11:30am
binghamton.zoom.us/my/luismrocha

office hours:
Tu & Th: 10:30-13:00
K1,binghamton.zoom.us/my/

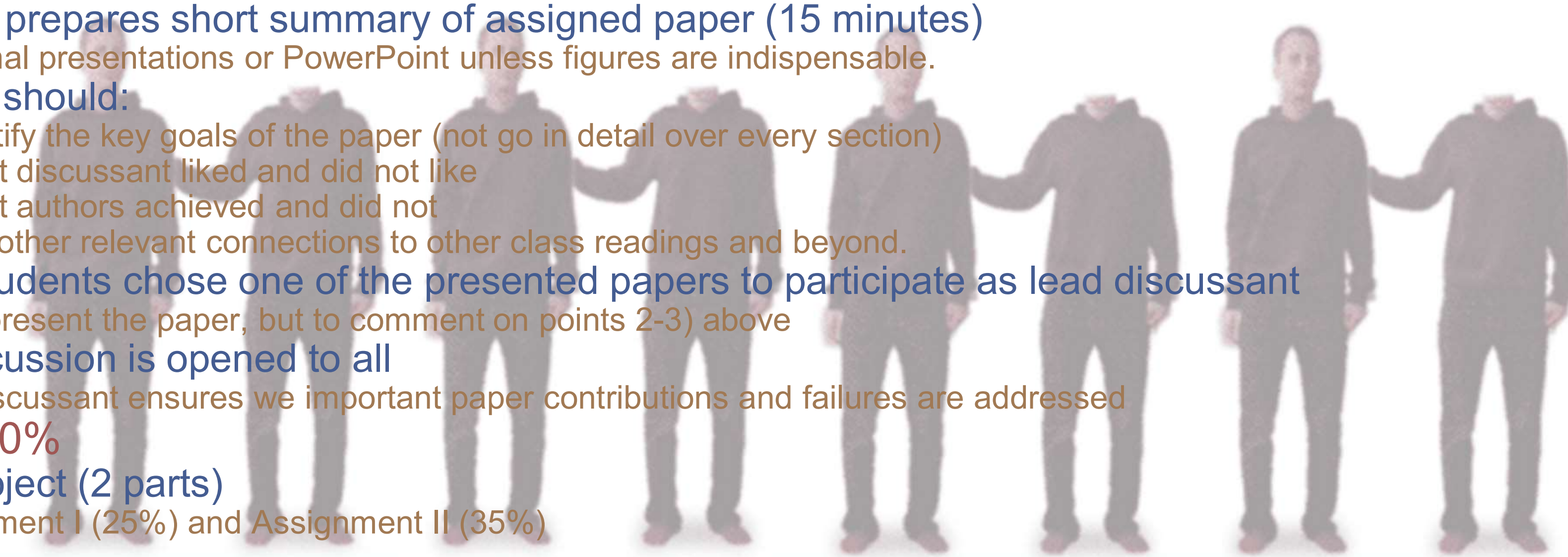
bit.ly/atBIC

rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK

evaluation

- Participation: 20%.
  - class discussion, everybody reads and discusses every paper
  - engagement in class
- Paper Presentation and Discussion: 20%
  - **SSIE501** students are assigned to papers individually or as group lead presenters and discussants
    - all students are supposed to read and participate in discussion of every paper.
  - Presenter prepares short summary of assigned paper (15 minutes)
    - no formal presentations or PowerPoint unless figures are indispensable.
  - Summary should:
    - 1) Identify the key goals of the paper (not go in detail over every section)
    - 2) What discussant liked and did not like
    - 3) What authors achieved and did not
    - 4) Any other relevant connections to other class readings and beyond.
  - **ISE440** students chose one of the presented papers to participate as lead discussant
    - not to present the paper, but to comment on points 2-3) above
  - Class discussion is opened to all
    - lead discussant ensures we important paper contributions and failures are addressed
- Black Box: 60%
  - Group Project (2 parts)
    - Assignment I (25%) and Assignment II (35%)



BINGHAMTON UNIVERSITY STATE UNIVERSITY OF NEW YORK
rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

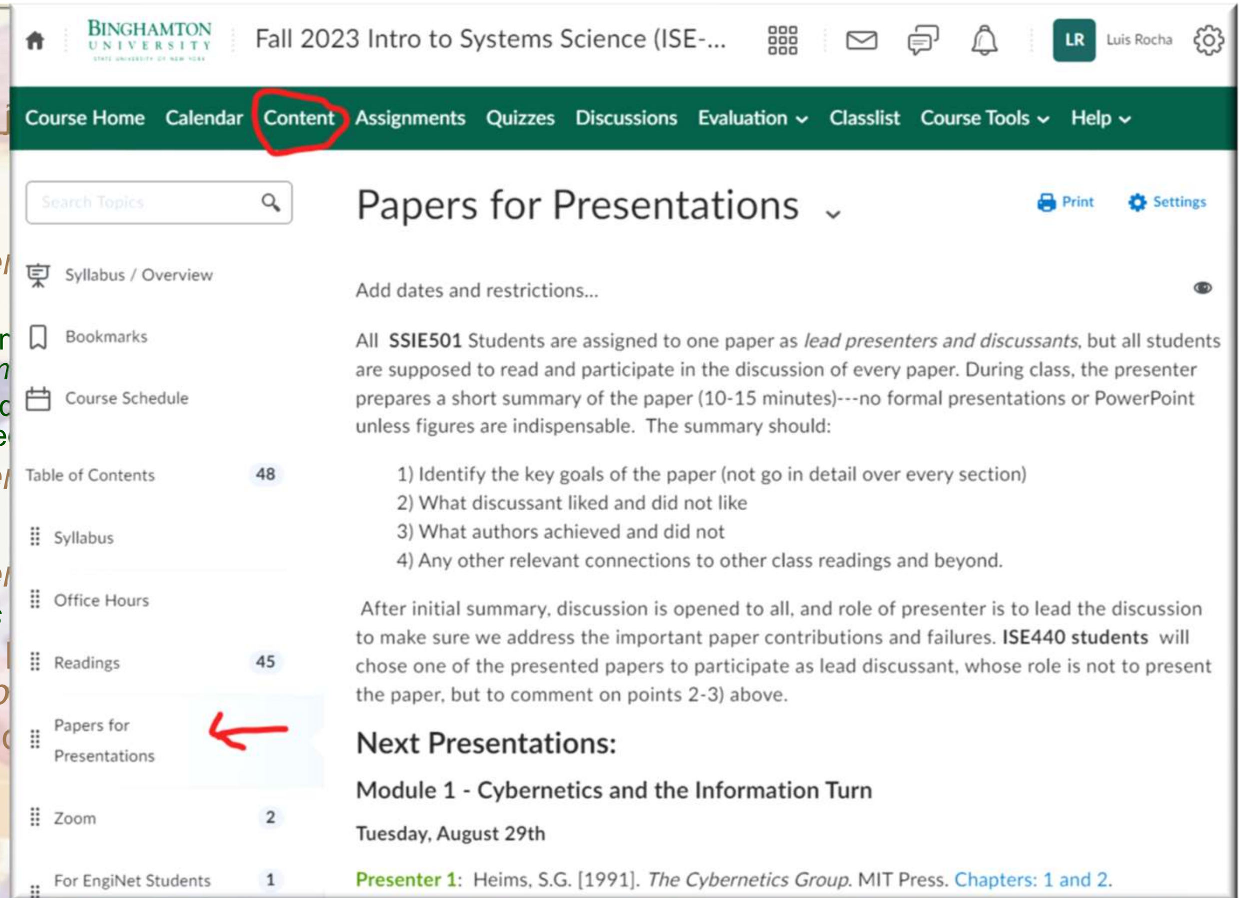
key events coming up

- Paper Presentation: 20%
  - Present (501) and lead (501&440) the discussion of an article related to the class materials
    - Enginet students post/send video or join by Zoom synchronously
- Module 1: Cybernetics and the Information Turn
- Next classes
  - Discussion Set 1 (Group 1): September 5th
    - Kline, Ronald R [2015]. *The cybernetics moment, or, why we call our age the information age*. Johns Hopkins University Press. Chapters 1-2.
      - Optional: Heims, S.G. [1991]. *The Cybernetics Group*. MIT Press. Chapters: 1,2, 11, and 12
      - Optional: McCulloch, W. and W. Pitts [1943], "A Logical Calculus of Ideas Immanent in Nervous Activity". *Bulletin of Mathematical Biophysics* **5**:115-133.
    - Gleick, J. [2011]. *The Information: A History, a Theory, a Flood*. Random House. Chapter 8.
      - Optional: Prokopenko, Mikhail, Fabio Boschetti, and Alex J. Ryan. "An information theoretic primer on complexity, self-organization, and emergence." *Complexity* **15**.1 (2009): 11-28.
  - Discussion Set 2 (Group 2)
    - Brenner, Sydney. [2012]. "History of Science. The Revolution in the Life Sciences". *Science* **338** (6113): 1427-8.
    - Brenner, Sydney. [2012]. "Turing centenary: Life's code script. *Nature* **482** (7386) (February 22): 461-461.
    - Cobb, Matthew. [2013]. "1953: When Genes Became 'Information'." *Cell* **153** (3): 503-506.
      - Optional: Searls, David B. [2010]. "The Roots of Bioinformatics". *PLoS Computational Biology* **6**(6): e1000809.
    - Weaver, W. [1948]. "Science and Complexity". *American Scientist*, **36**(4): 536-44. Also available in Klir, G.J. [2001]. *Facets of systems Science*. Springer, pp: 533-540.
  - **Discussion by all**

rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

more upcoming readings (check brightspace)

- Paper Presentation: 20%
  - Present (501) and lead (501&440)
    - Enginet students post/send video or
- Module 2: Systems Science
  - Discussion Set 3:
    - Klir, G.J. [2001]. *Facets of syste*
      - Optional:
      - Rosen, R. [1986]. "Some commen Klir, G.J. [2001]. *Facets of system*
      - Wigner, E.P. [1960], "The unreaso in mathematical sciences delivere
    - Klir, G.J. [2001]. *Facets of syste*
  - Discussion Set 4:
    - Klir, G.J. [2001]. *Facets of syste*
      - Optional: Klir, G.J. [2001]. *Facets*
    - Schuster, P. (2016). The end of of computational facilities. *Comp*
    - Von Foerster, H., P. M. Mora and Science **132**(3436):1291-5.
- Future Modules
  - See brightspace



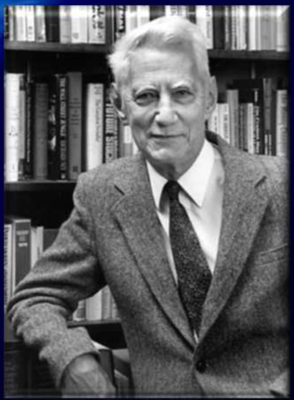
- Information is transmitted through noisy communication channels
  - Ralph Hartley and Claude Shannon (at Bell Labs), the fathers of Information Theory, worked on the problem of efficiently transmitting information; i. e. ***decreasing the uncertainty*** in the transmission of information.

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.



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

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

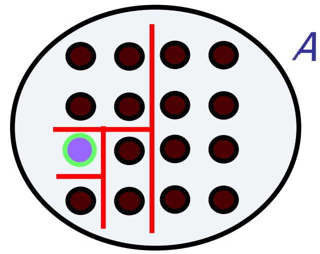
$$H(A) = \log_2 |A|$$

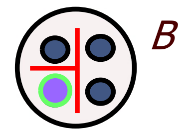
Measured in bits

Number of Choices

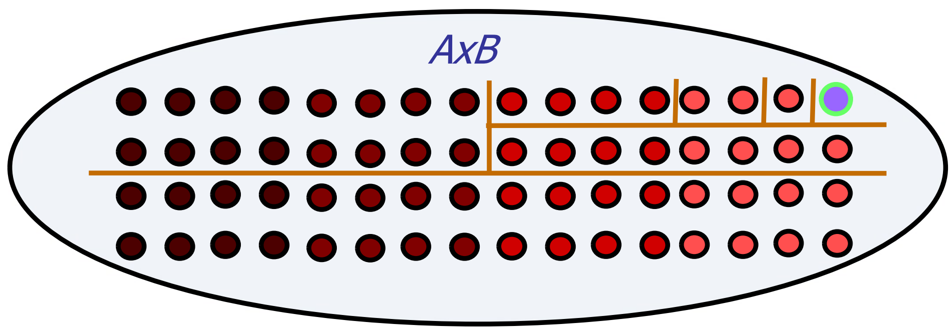
Quantifies how many yes-no questions need to be asked to establish what the correct alternative is

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



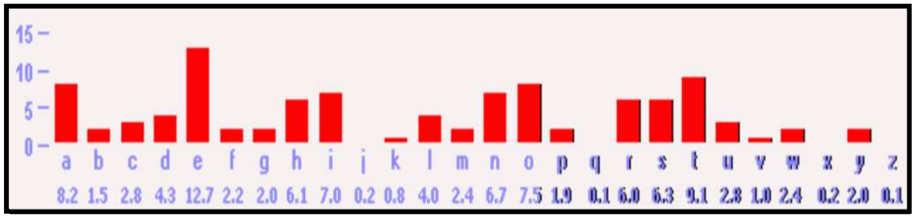


$$H(A \times B) = \log_2(16 \times 4) =$$
$$= \log_2(16) + \log_2(4) = 6$$

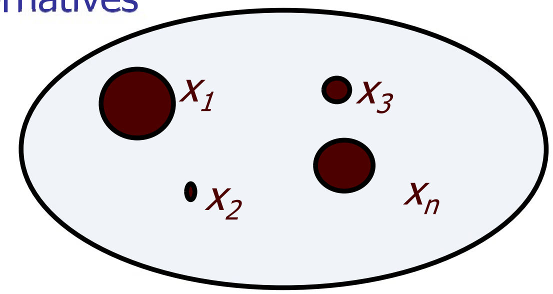


rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

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

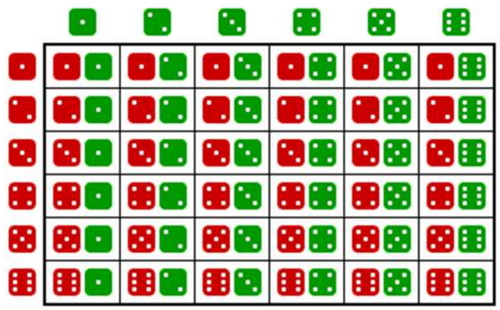
$$H_S(A) = -\sum_{i=1}^{n} p(x_i)\log_2\big(p(x_i)\big)$$

Measured in bits

Probability of alternative

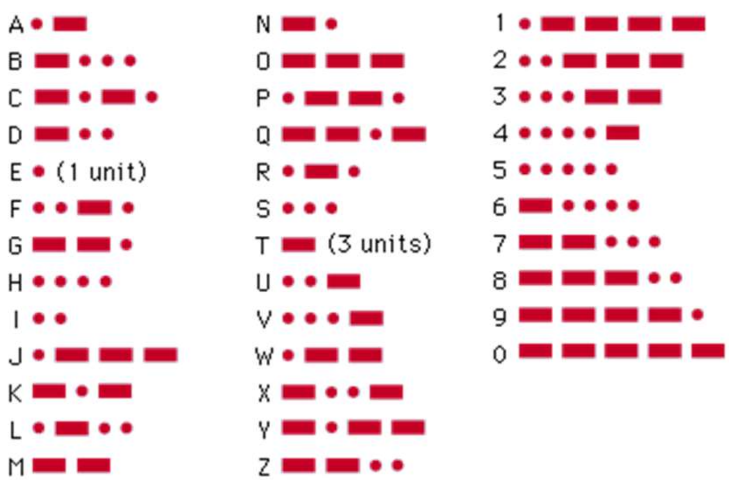
BINGHAMTON UNIVERSITY STATE UNIVERSITY OF NEW YORK
rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

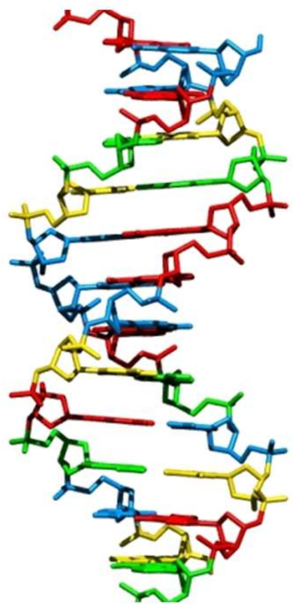
alphabet examples

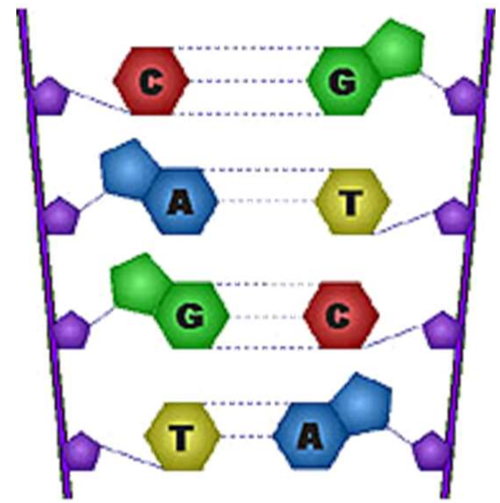


Message encoded in an alphabet of ***n*** symbols, for example:

- English (26 letters + space + punctuations)
- Morse code (dot, dash, space)
- DNA (A, T, G, C)
- Two dice (11 integers)







5-letter "english"

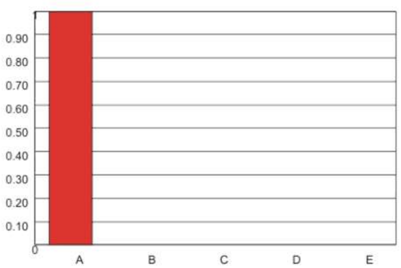
- Given a symbol set {A,B,C,D,E}
  - And occurrence probabilities P_A, P_B, P_C, P_D, P_E,
- The Shannon entropy is
  - The average minimum number of bits needed to represent a symbol

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

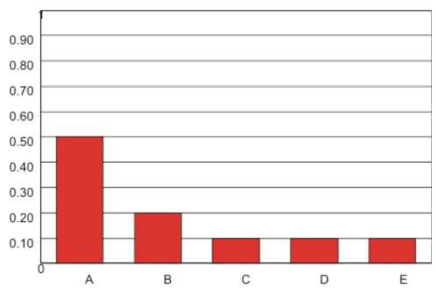
$$H_S = -(1.\log_2(1) + 0.\log_2(0) + 0.\log_2(0) + 0.\log_2(0) + 0.\log_2(0)) = -\log_2(1)$$

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

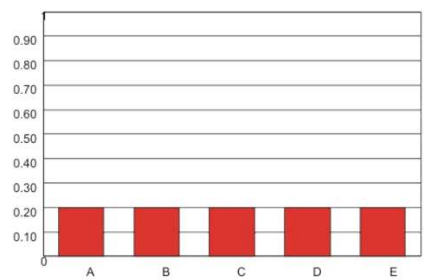
$$H_S = -(\tfrac{1}{2}.\log_2\left(\tfrac{1}{2}\right) + \tfrac{1}{5}.\log_2\left(\tfrac{1}{5}\right) + 3.\left(\tfrac{1}{10}\right).\log_2\left(\tfrac{1}{10}\right))$$



H_S = 0 bits
0 questions



H_S = 1.96
≈2 questions



H_S = 2.32 bits

information is surprise

rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

what it measures

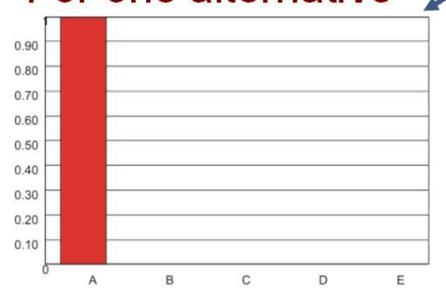
uncertainty, about outcome. How much information is gained when symbol is known

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

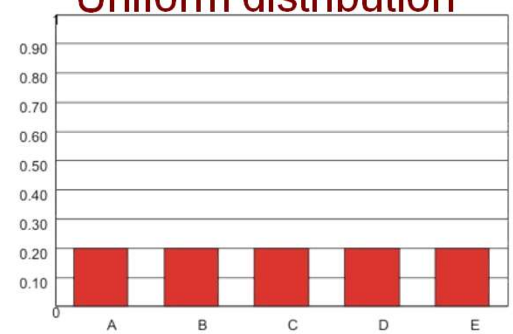
$$H_S \in = -\sum_{i=1}^{n} p(x_i)\log_2\big(p(x_i)\big)$$

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

For one alternative

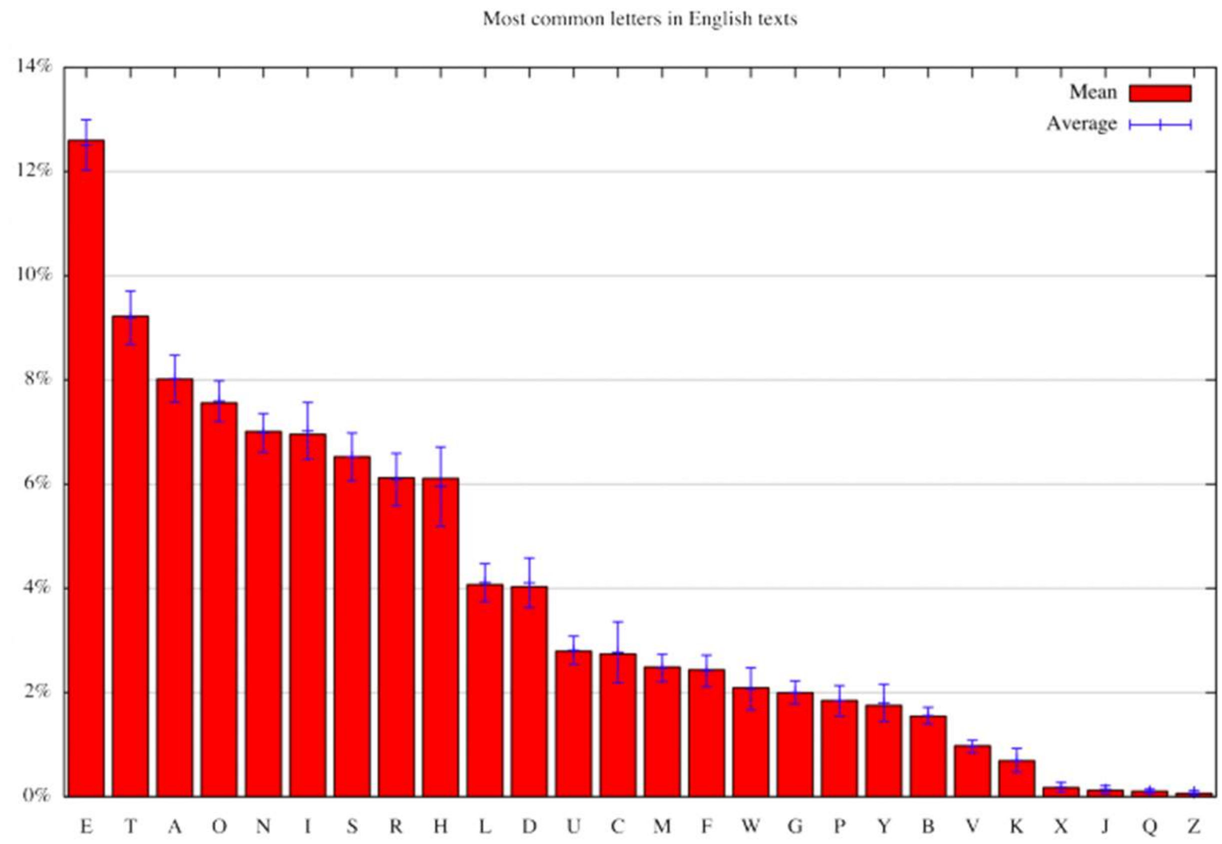


Uniform distribution



rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

from letter frequency

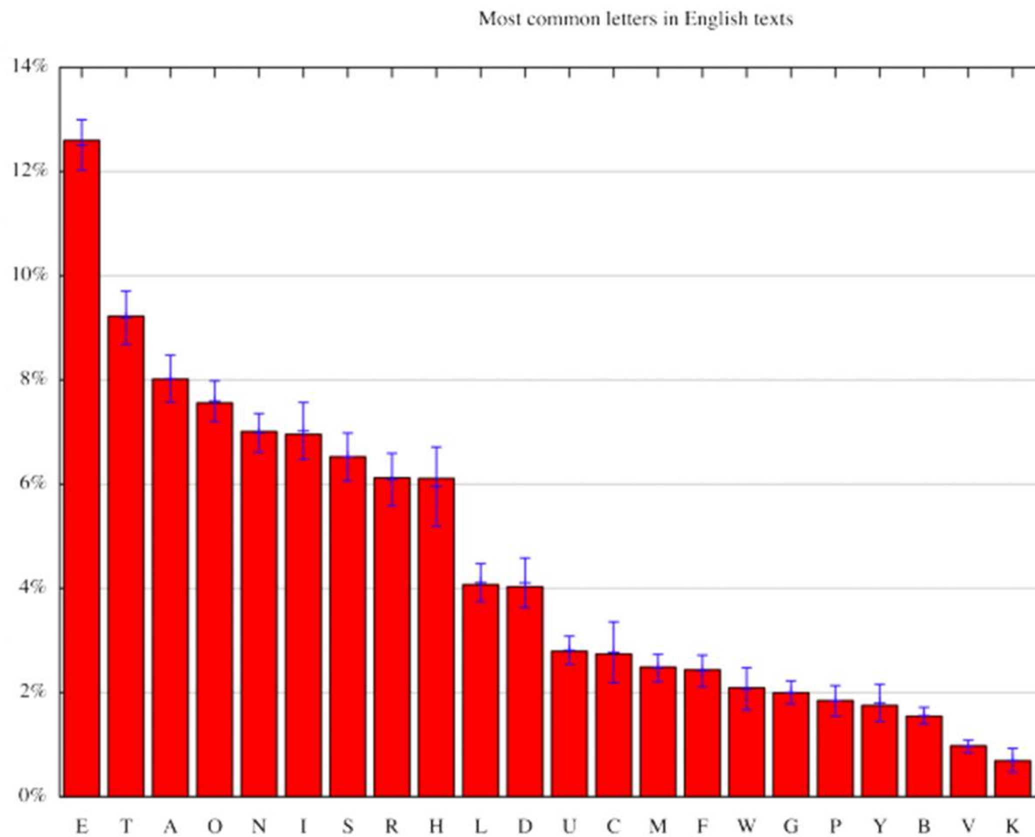


http://www.macfreek.nl/memory/Letter_Distribution

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[illegible]	[illegible]	[illegible]094755
f	0.023[illegible]	[illegible]	[illegible]220629
p	0.020517	-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.[illegible]	[illegible]	[illegible]63015644
D	0.[illegible]	[illegible]	[illegible]61811184
U	[illegible]	[illegible]	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

BINGHAMTON UNIVERSITY STATE UNIVERSITY OF NEW YORK
rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

- 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

We were good, we were gold
Kinda dream that can't be sold
We were right 'til we weren't
Built a home and watched it burn

Mm, I didn't wanna leave you
I didn't wanna lie
Started to cry, but then remembered I
I can buy myself flowers
Write my name in the sand
Talk to myself for hours
Say things you don't understand
I can take myself dancing
And I can hold my own hand
Yeah, I can love me better than you can

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

wdeo eog geWI ewr e deorw
aainhmta d rettoeKandl dsbc
eeeier ntw hWttr ewrgliwe
oriaedatmht ndc lwn thuaBeuib

eanm dtal vewdi nl o unMay
al indn nItawde i
cl rettedtebrmSrb reemntuy da oth e
uolrawe blnffmsyylc es
niWe dty ne rsehmntiama
arem Tll ssytrfu fkoooh
nyoh e gdodudtnaraustsi tnynS
atf lk emcnegyn snlicad a
hmhydcndAwannoo n dl l a
tlhl eatta nom Ybrueny h ee oaavn cce

entropy according to probabilistic model

0th order model: equiprobable symbols

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

Hartley Measure
H(|27|) 4.7548875

XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGXYD QPAAMKBZAACIBZLHJQD

1st 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

2nd order model: frequency of digrams

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.

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

3rd order model: frequency of trigrams

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

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

4th order model: frequency of tetragrams

H_S = 2.8

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

rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

other measures to infer structure and organization in nature and society

- 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.

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

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

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

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

other measures to infer structure and organization in nature and society

- Mutual Information
  - Amount of information about one variable that can be gained (uncertainty reduced) by observing another variable
- Information Gain (Kullback-Leibler Divergence)
  - ons *p* and *q*,
    - n order to represent *q* (model approximation) as it
    -)
  - processes in time
    - ainty lost, in knowing future values of Y, knowing
    - the past values of X and Y

Optional Readings: Golan, Amos, and John Harte. "Information theory: A foundation for complexity science." *Proceedings of the National Academy of Sciences* **119**.33 (2022): e2119089119.

James, R., and Crutchfield, J. (2017). "Multivariate Dependence beyond Shannon Information". *Entropy*, **19**(10), 531.

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

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

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

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

Optional Reading: Prokopenko, Mikhail, Fabio Boschetti, and Alex J. Ryan. "An information theoretic primer on complexity, self organization, and emergence." *Complexity* **15**.1 (2009): 11-28.

uncertainty-based information

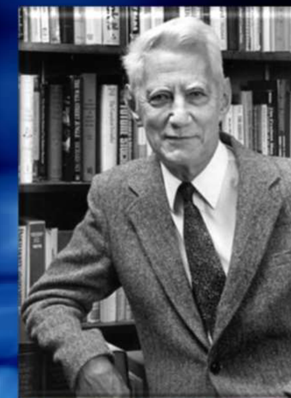
information as decrease in uncertainty .

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

Measured in bits

Number of Choices

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



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

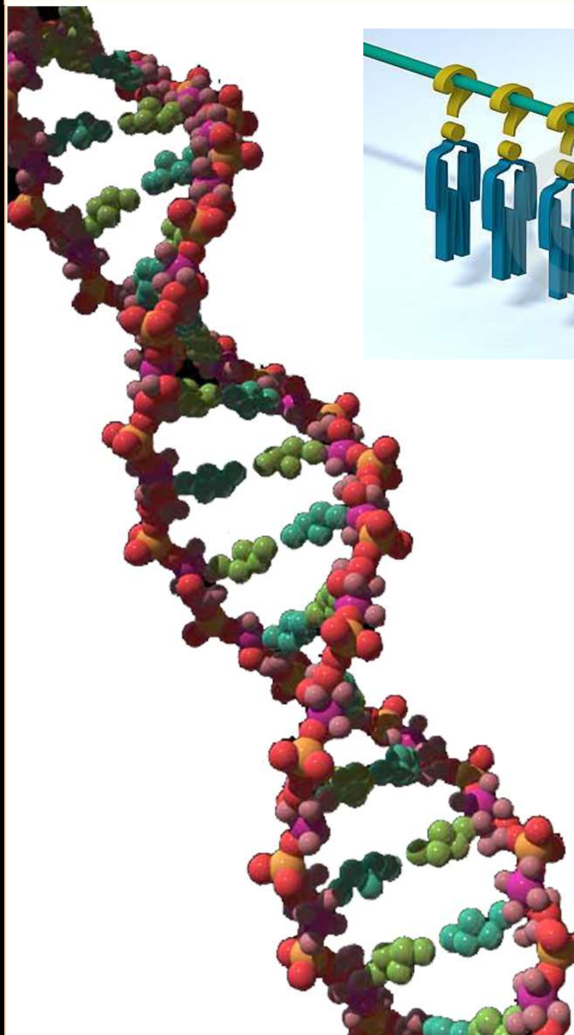
Measured in bits

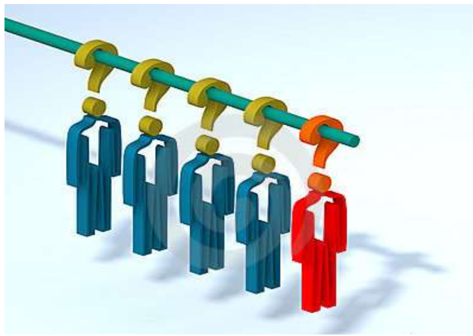
Probability of alternative

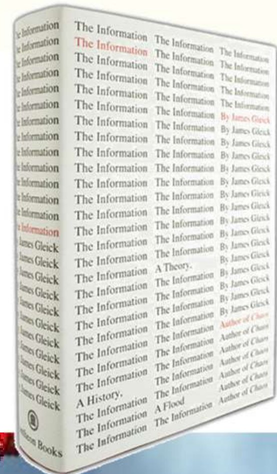
including more structure reduces surprise

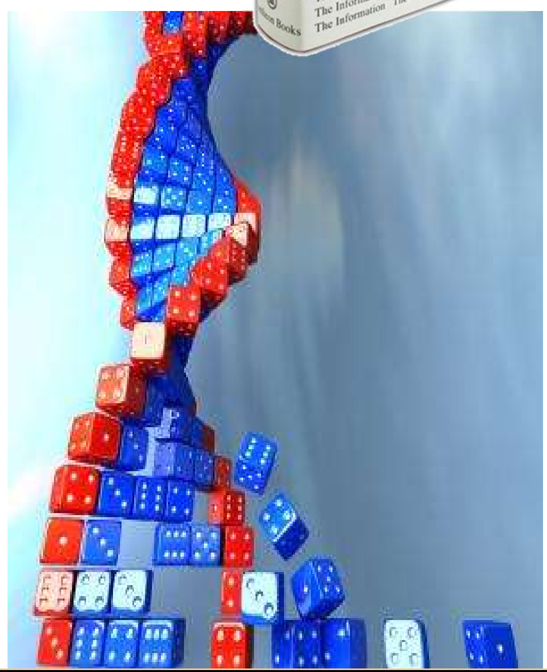
information is surprise

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







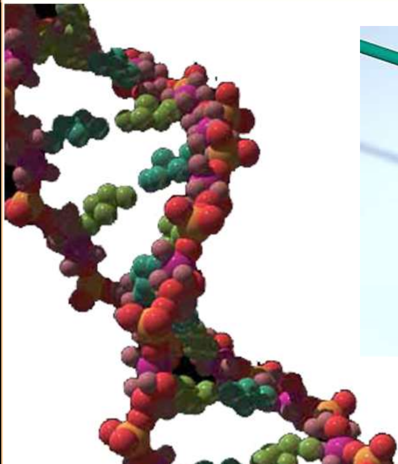


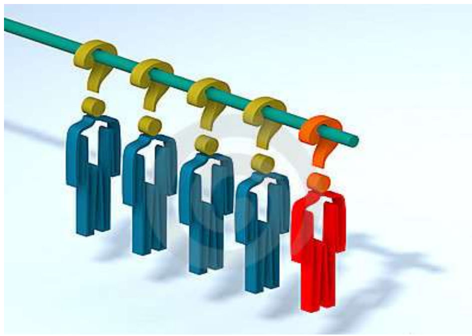
Holdin' me back
Gravity's holdin' me back
I want you to hold out the palm of your hand
Why don't we leave it at that?
Nothin' to say
When everything gets in the way
Seems you cannot be replaced
And I'm the one who will stay, oh

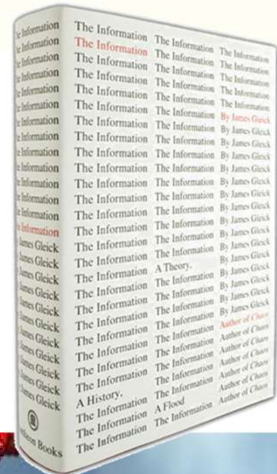
was
was
to

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

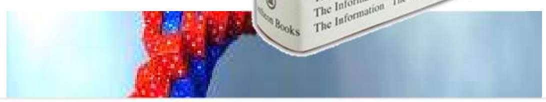
rate of removing uncertainty of each symbol







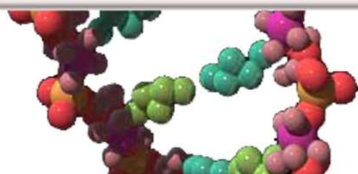
Holdin' me back
Gravity's holdin' me back
I want you to hold out the palm of your hand
Why don't we leave it at that?
Nothin' to say
When everything gets in the way
Seems you cannot be replaced
And I'm the one who will stay, oh



Optional Readings:

Prokopenko, Mikhail, Fabio Boschetti, and Alex J. Ryan. "An information theoretic primer on complexity, self organization, and emergence." *Complexity* **15**.1 (2009): 11-28.

James, R., and Crutchfield, J. (2017). "Multivariate Dependence beyond Shannon Information". *Entropy*, **19**(10), 531.



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

BINGHAMTON UNIVERSITY STATE UNIVERSITY OF NEW YORK
rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

readings

- Class Book
  - Klir, G.J. [2001]. *Facets of systems science*. Springer.
- Papers and other materials
  - Discussion Set 1 (Group 1) – September 3rd or 5th
    - Kline, Ronald R [2015]. *The cybernetics moment, or, why we call our age the information age*. Johns Hopkins University Press. Chapters 1-2.
      - Heims, S.G. [1991]. *The Cybernetics Group*. MIT Press. Chapters: 1,2, 11, 12.
      - Optional: McCulloch, W. and W. Pitts [1943]. "A Logical Calculus of Ideas Immanent in Nervous Activity". *Bulletin of Mathematical Biophysics* **5**:115-133.
    - Gleick, J. [2011]. *The Information: A History, a Theory, a Flood*. Random House. Chapter 8.
      - Optional: Prokopenko, Mikhail, Fabio Boschetti, and Alex J. Ryan. "An information theoretic primer on complexity, self-organization, and emergence." *Complexity* **15**.1 (2009): 11-28.
  - Discussion Set 2 (Group 2)
    - Brenner, Sydney. [2012]. "History of Science. The Revolution in the Life Sciences". *Science* **338** (6113): 1427-8.
    - Brenner, Sydney. [2012]. "Turing centenary: Life's code script. *Nature* **482** (7386) (February 22): 461-461.
    - Cobb, Matthew. [2013]. "1953: When Genes Became 'Information'." *Cell* **153** (3): 503-506.
      - Optional: Searls, David B. [2010]. "The Roots of Bioinformatics". *PLoS Computational Biology* **6**(6): e1000809.
    - Weaver, W. [1948]. "Science and Complexity". *American Scientist*, **36**(4): 536-44. Also available in Klir, G.J. [2001]. *Facets of systems Science*. Springer, pp: 533-540.