Information in Biology

CRI - *Centre de Recherches Interdisciplinaires*, Paris

May 2012

- Information processing is an essential part of Life.
- Thinking about it in quantitative terms may is useful.

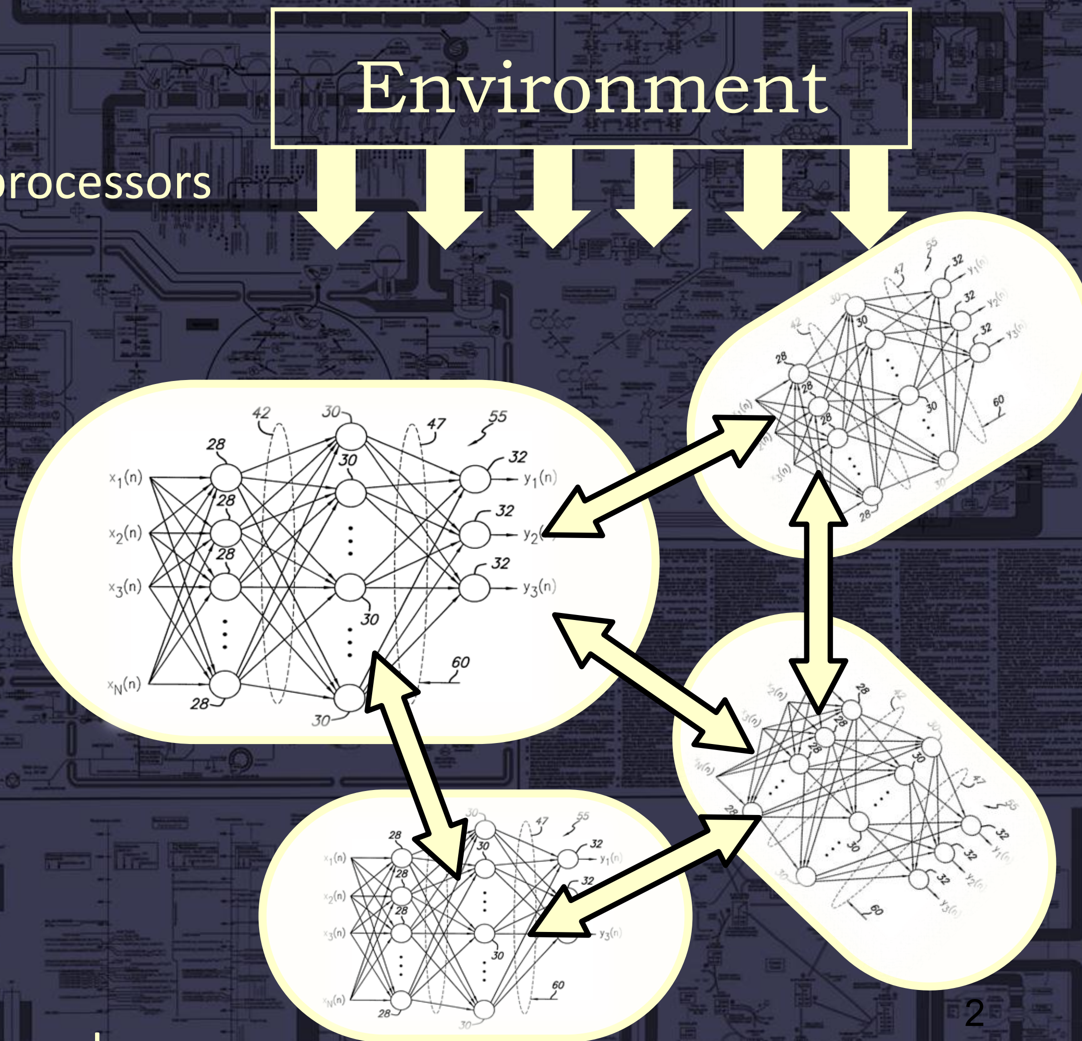
Living information is carried by molecular channels

<u>"Living systems"</u>

I. Self-replicating information processors

II. Evolve collectively.

III. Made of molecules.



- Generic properties of molecular channels subject to evolution?

- Information theory approach?

- Other biological information channels.

Outline – Information in Biology

- **Information in Biology**
  - Concept of information is found in many living systems:
    DNA, signaling, neuron, ribosomes, evolution.
  - Goals: (1) Formalize and quantify biological information.
    (2) Application to various biological systems.
    (3) Looking for common principles.

I. **Information and Statistical Mechanics:** Shannon's information theory and its relation to statistical mechanics.

II. **Overview: Living Information: molecules, neurons, population and evolution.**
Living systems as information sources, channels and processors.

III. **Molecular information and noise.**

IV. **Neural networks and coding theory.**

V. **Population dynamics, social interaction and sensing**.

…

I. Basics of Information Theory (Shannon)

Shannon's Information theory

- **Information theory**: a branch of applied math and electrical engineering.
- Developed by Claude E. Shannon.
- Main results: fundamental limits on signal processing such as,
  - How well data can be compressed?
  - What is the reliability of communicating signals?
- Numerous applications (besides communication eng.):
  - Physics (stat mech), Math (statistical inference), linguistics,

  Computer science (cryptography, complexity), Economics (portfolio theory).

- The key quantity which measures information is **entropy**:
- Quantifies the uncertainty involved in predicting the value of a random variable (e.g., a coin flip or a die).

- What are the biological implications?

Claude Elwood Shannon (1916-2001)

- 1937 master's thesis: *A Symbolic Analysis of Relay and Switching Circuits.*
- 1940 Ph.D. thesis: *An Algebra for Mandelian Genetics.*
- WWII (Bell labs) works on cryptography and fire-control systems:

  Data Smoothing and Prediction in Fire-Control Systems.

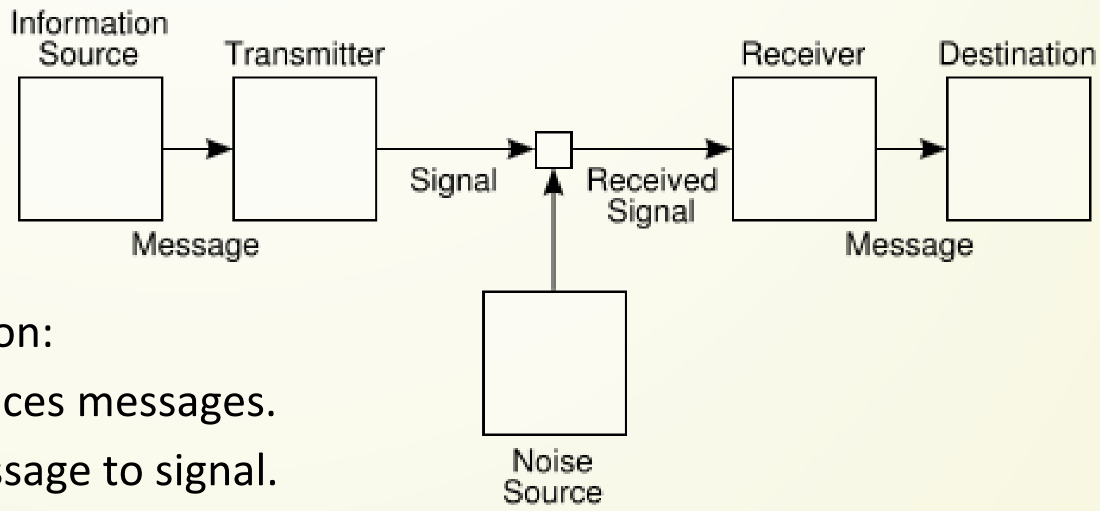
  Communication theory of secrecy systems.
- 1948: *Mathematical Theory of Communication.*
- 1949: Sampling theory: Analog to digital.
- 1951: *Prediction and Entropy of Printed English.*

- 1950: Shannon's mouse: 1st artificial learning machine.
- 1950: *Programming a Computer for Playing Chess*.
- 1960: 1st wearable computer, Las Vegas.

A Mathematical Theory of Communication

- Shannon’s seminal paper: "A Mathematical Theory of Communication".

Bell System Technical Journal **27** (3): 379–423 (1948).



- Basic scheme of communication:
  - Information **source** produces messages.
  - **Transmitter** converts message to signal.
  - **Channel** conveys the signal with **Noise.**
  - **Receiver** transforms the signal back into the message
  - **Destination:** machine, person, organism receiving the message.

- Introduces **information entropy** measured in **bits**.

What is information?

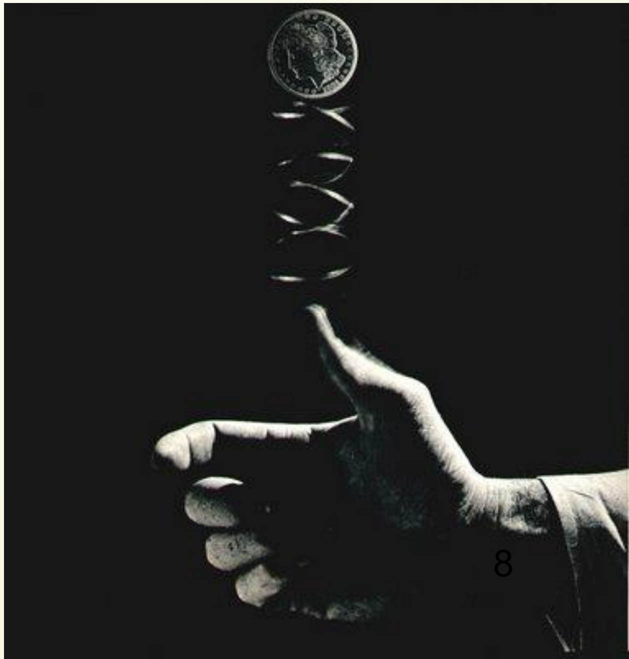
"The fundamental problem of communication is that of reproducing at one point, either exactly or approximately, a message selected at another point."

- Engineering perspective –
  - How to make good transmission channels
  - Problem with telegraph lines,

- Define information as measure for the “surprise”
  - If a binary channel transmits only 1’s

    there is no information (no surprise).
  - If the channel transmits 0’s and 1’s

    with equal probability – max. information.

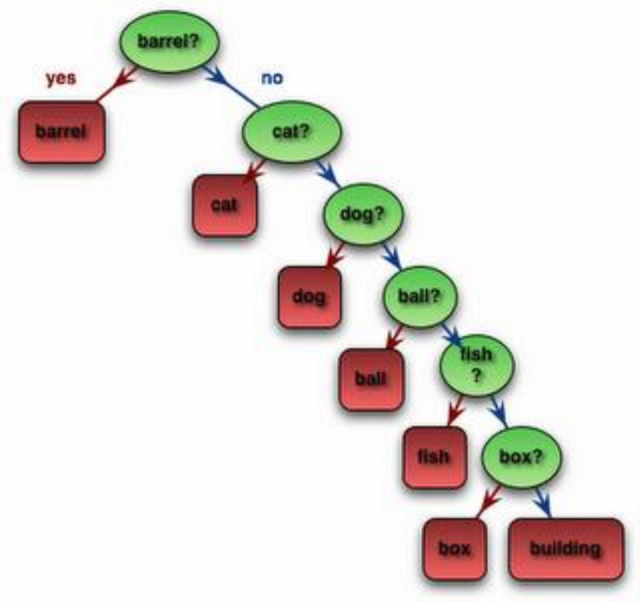


Intuition: 20 questions game

- Try to guess the object from:

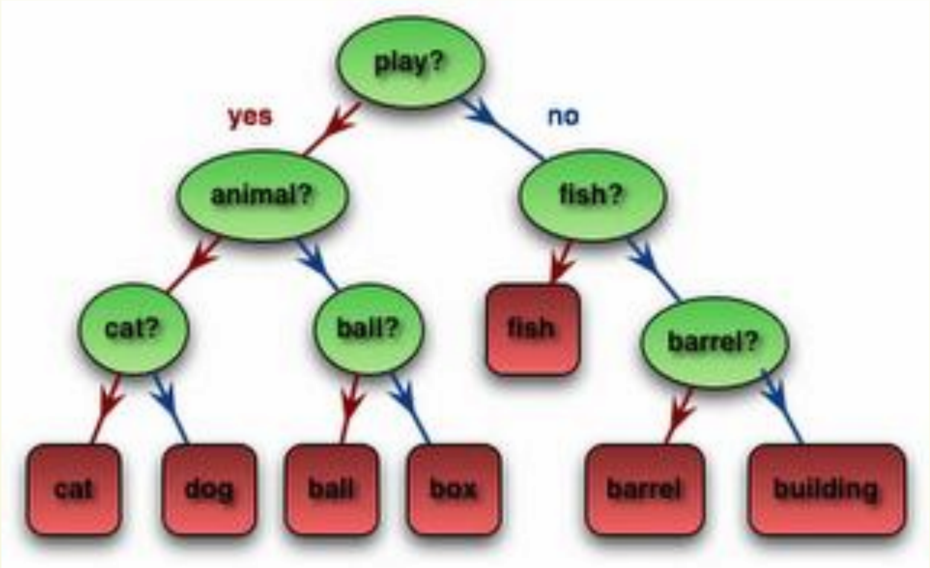
{barrel, cat, dog, ball, fish, box, building}.

- First strategy: wild guess.



Intuition: 20 questions game

- Optimal strategy: equalized tree



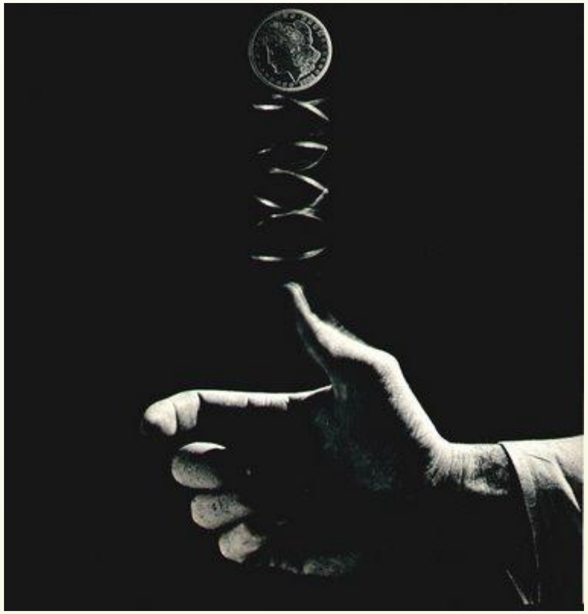
- Information = # of yes/no questions in an optimal tree.

Introducing the bit

- If I have a (equal) choice between two alternatives the information is:

$$I=1 \text{ bit} = \log_2(\#\text{Alternatives})$$

1 bit =



Harry Nyquist (1924):

Certain Factors Affecting Telegraph Speed

Example: How many bits are in a genome of length *N?*

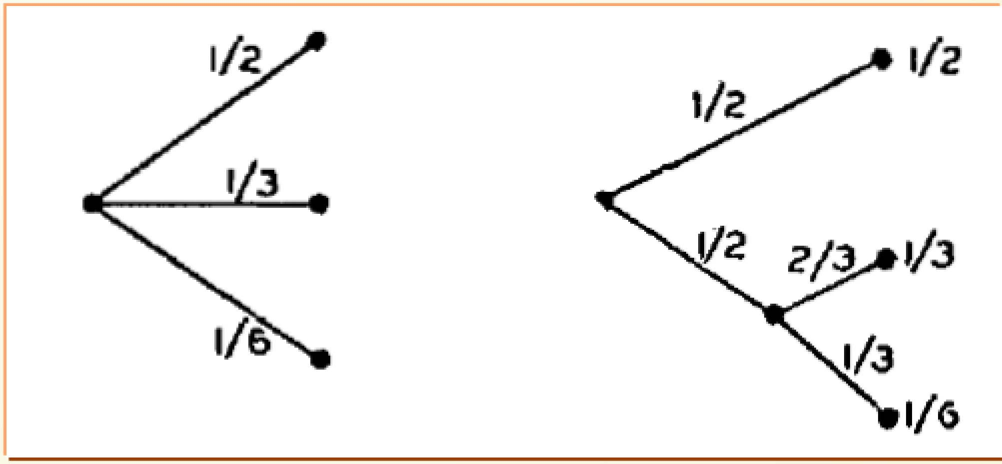
Shannon's axioms of information

- For random variable that distributes as $p_1 \ldots p_n$,search for a function $H(p_1,..,p_n)$ that measures information that obeys:

1) Continuous in p_i

2) For equal probabilities it is Nyquist expression: $H(1/n,\ldots,1/n)=\log_2(n)$.

3) "Tree property": invariant to redistribution into stages.



$$H(\tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{6}) = H(\tfrac{1}{2}, \tfrac{1}{2}) + \tfrac{1}{2}H(\tfrac{2}{3}, \tfrac{1}{3})$$

Information from Shannon's axioms

- Shannon showed that the *only* function that obeys these axioms is

$$H = -\sum_i p_i \log_2 p_i = -\langle \log_2 p_i \rangle$$

(up to proportion constant).

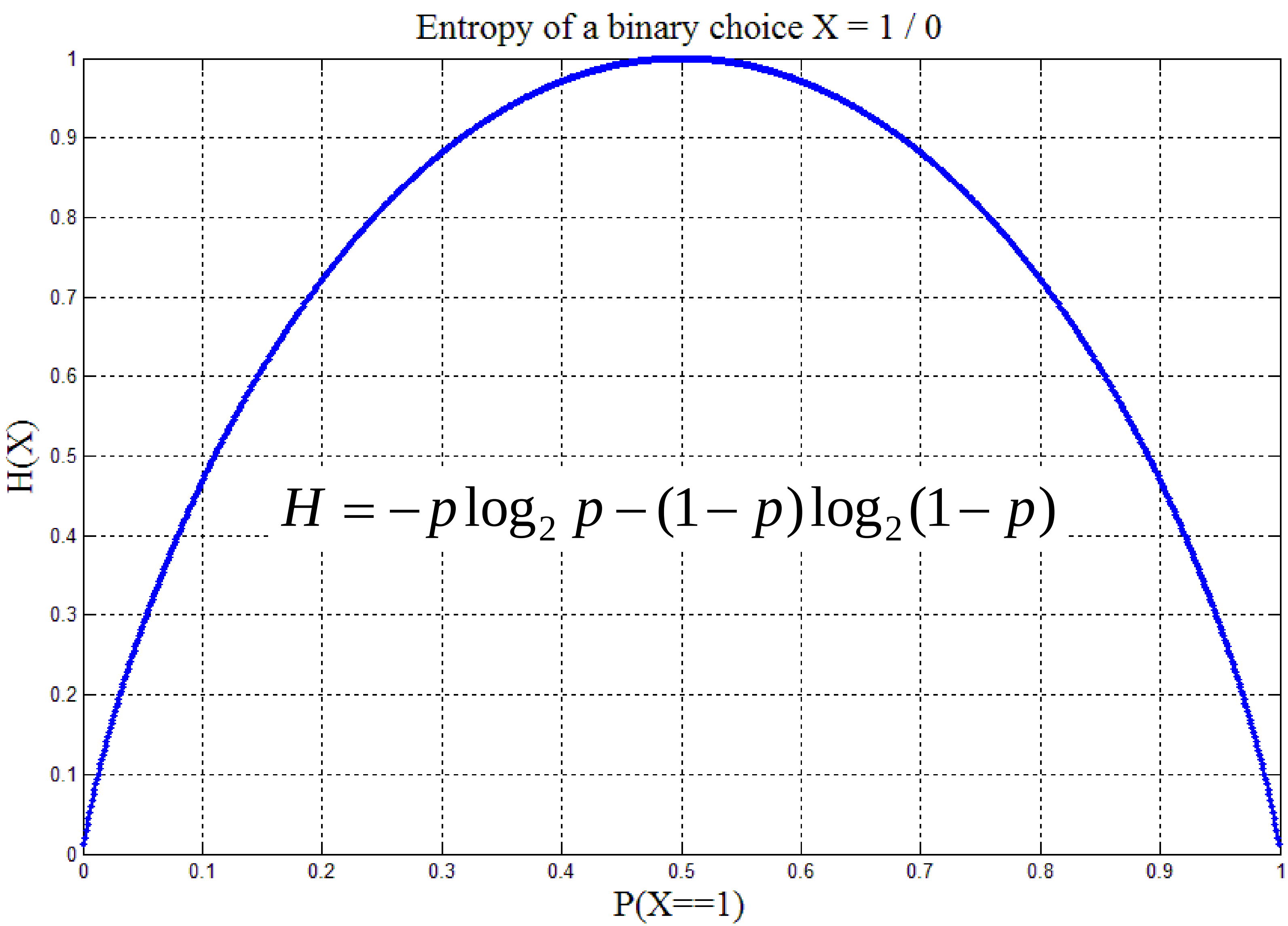
- Example : if X is uniformly distributed over 128 outcomes

$$H = -\sum_i p_i \log_2 p_i = -\sum_i \frac{1}{2^7} \log_2 \frac{1}{2^7} = \log_2 2^7 = 7 \text{ bits}$$

- Example: very uneven coin shows head only 1 in 1024 times

$$H = -\sum_i p_i \log_2 p_i = -2^{-10} \log_2 2^{-10} - (1-2^{-10}) \log_2 (1-2^{-10}) \approx (10 + \log_2 e) 2^{-10} \text{ bits}$$

Entropy of a Binary Channel



$$H = -p\log_2 p - (1-p)\log_2(1-p)$$

Why call it entropy?

- Shannon discussed this problem with John von Neumann:

“My greatest concern was what to call it. I thought of calling it ‘information’, but the word was overly used, so I decided to call it ‘uncertainty’. When I discussed it with **John von Neumann**, he had a better idea. Von Neumann told me, ‘You should call it **entropy**, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage. “

M. Tribus, E.C. McIrvine, “Energy and information”, Scientific American, 224 (1971).

Conditional probability

- Consider two random variables X,Y with a joint probability distribution P(X,Y)
  - The joint entropy is H(X,Y)
  - The mutual entropy is H(X,Y)-H(X)-H(Y)
  - The conditional entropies are H(X|Y) and H(Y|X)

Joint entropy H(X,Y) measures total entropy of the joint distribution P(X,Y)

$$H(X,Y) = -\sum_{x,y} p(x,y) \log p(x,y)$$

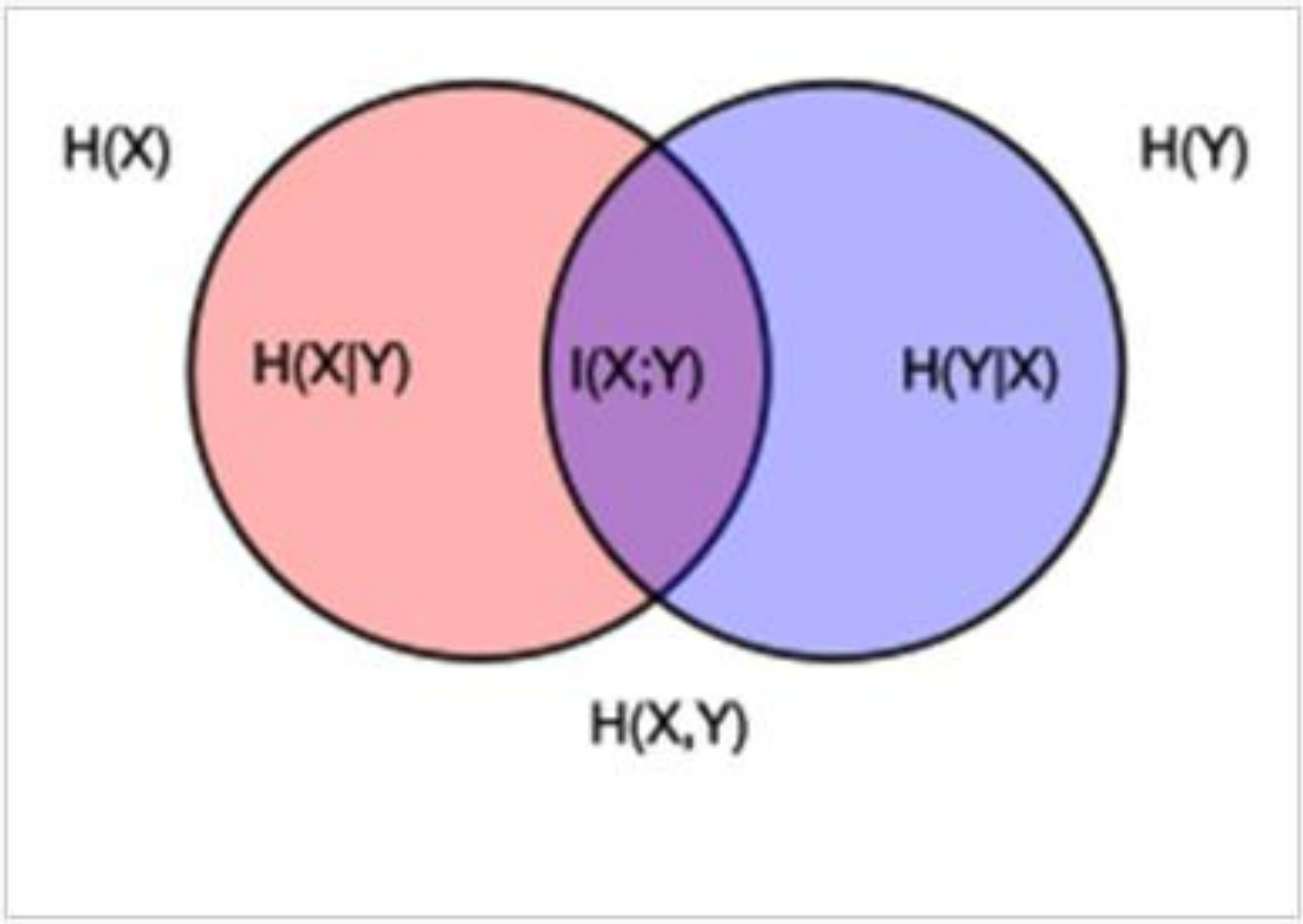
Mutual entropy I(X;Y) measures correlations (P(x,y)=P(x)P(y) => I=0)

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)\,p(y)} \right),$$

Conditional entropy H(X|Y) measures remaining uncertainty of X given Y

$$H(X|Y) = -\sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log p(x|y) = -\sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(y)}$$

More information measures



Whose entropy ?

Agonizing over this, I was driven to conclude that the different messages considered must be the set of all those that will, or might be, sent over the channel during its useful life; and therefore Shannon's H measures the degree of ignorance of the communication engineer when he designs the technical equipment in the channel. Such a viewpoint would, to say the least, seem natural to an engineer employed by the Bell Telephone Laboratories--yet it is curious that nowhere does Shannon see fit to tell the reader explicitly whose state of knowledge he is considering, although the whole content of the theory depends crucially on this.

E. T. Jaynes

Kullback Leibler entropy

Problem: Suppose a random variable X is distributed according to P(X) but we expect it to be distributed according to Q(X).

What is the level of our surprise?

Answer: The Kullback –Leibler divergence

$$D_{KL}(p \| q) = \sum_i p_i \log_2 \frac{p_i}{q_i}$$

Mutual information $$I(X,Y) = D_{KL}(p(x,y) \| p(x)p(y))$$

– Appears in many circumstances
    - Example – Markov chains

Shannon Entropy and Statistical Mechanics

I. Maxwell's Demon

- Entropy in statistical mechanics: *measure of uncertainty* of a system after specifying its macroscopic observables such as temperature and pressure.

- Given macroscopic variables, entropy measures the degree of spreading probability over different possible states.

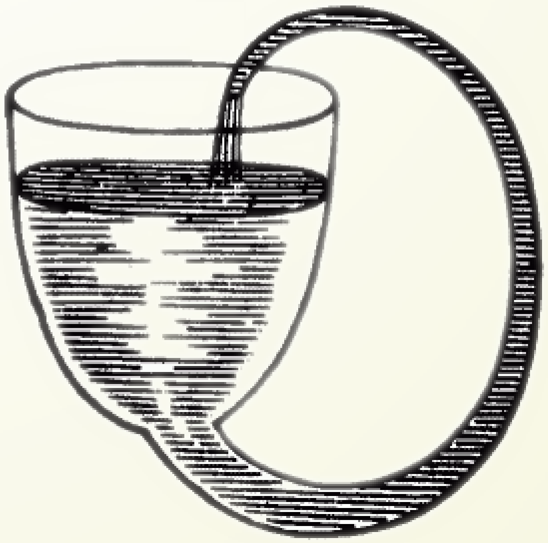
- Boltzmann' famous formula:

$$S = \ln \Omega$$

The second law of thermodynamics

- **The second law of thermodynamics:**

In general, the total entropy of a system isolated from its environment, the entropy of that system will tend not to decrease.



- Consequences:

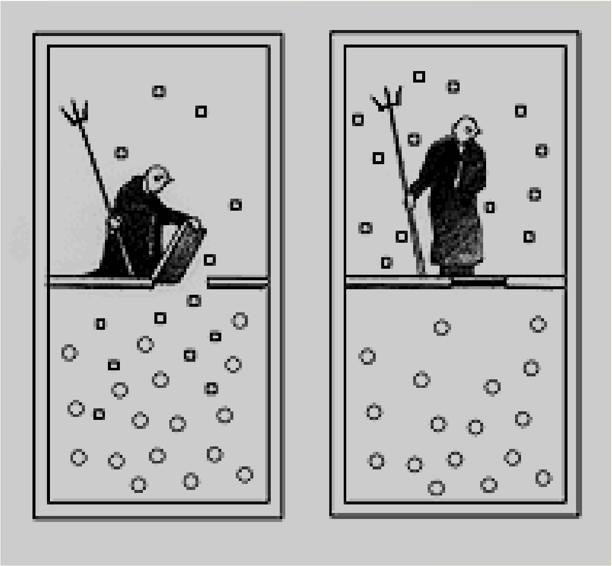
(i) heat will not flow from a colder body to a hotter body without work.

(ii) No *perpetuum mobile:* One cannot produce net work from a single temperature reservoir (production of net work requires flow of heat from a hotter reservoir to a colder reservoir).

Maxwell's thought experiment

How to violate the Second Law?

- Container divided by an insulated wall.
- Door can be opened and closed by a demon.
- Demon opens the door to allow only "hot" molecules of gas to flow to the favored side.
- One side heats up while other side cools down: decreasing entropy. Breaking 2^{nd} law!



Solution: entropy = -information

- Demon reduces the thermodynamic entropy of a system using information about individual molecules (their direction)

- Landauer (1961) showed that he demon must increase thermodynamic entropy by *at least the amount of Shannon information* he acquires and stores;

- → total thermodynamic entropy does not decrease!

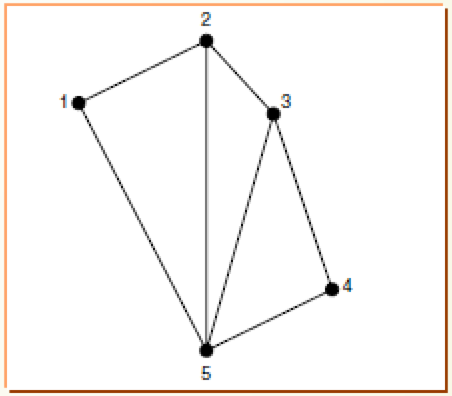
- Landauer's principle :

Any logically irreversible manipulation of information , such as the erasure of a bit, must be accompanied by a corresponding entropy increase in the information processing apparatus or its environment.

$$\min E = k_B T \ln 2$$

II. Second law in Markov chains

Random walk on a graph:.



– W is transition matrix : $p(t+1)=Wp(t)$

– p* be the steady state solution: $Wp^*=p^*$

Theorem: distribution approaches steady-state $\partial_t D(p||p^*) <= 0$

Also $\partial_t D(p||p^*)=0 <=> p=p^*$

In other words: *Markov dynamics dissipates any initial information*.

Maximum entropy inference (E. T. Jaynes)

Problem: Given partial knowledge e.g. the average value <X> of X how should we assign probabilities to outcomes P(X)?

Answer: choose the probability distribution that maximizes the entropy (surprise) and is consistent with what we already know.

Example: given energies E_i and measurement <E> what is p_i ?

Exercise: which dice X={1,2,3,4,5,6} gives <X>=3 ?

Maximum Entropy principle relates thermodynamics and information

At equilibrium, *TD entropy = Shannon information* needed to define the microscopic state of the system, given its macroscopic description.

Gain in entropy always means loss of information about this state.

Equivalently, *TD entropy = minimum number of yes/no questions* needed to be answered in order to fully specify the microstate.