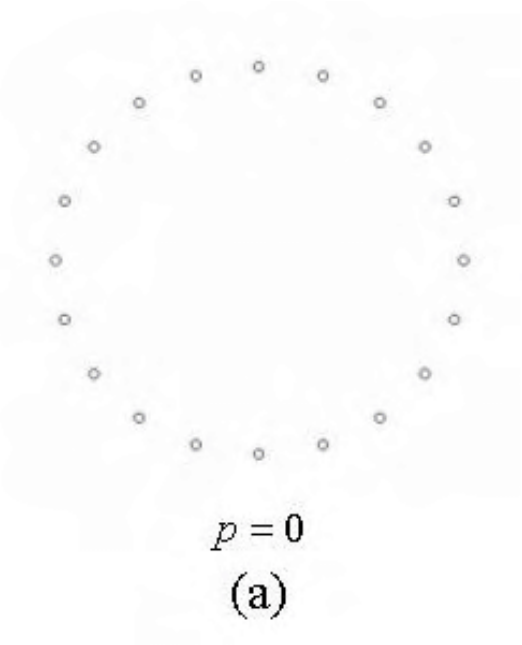

welcome
to Lecture 5
Computation and the Brain

First: What happened last Wednesday

The Brain: a *Uuuuuuge* directed graph

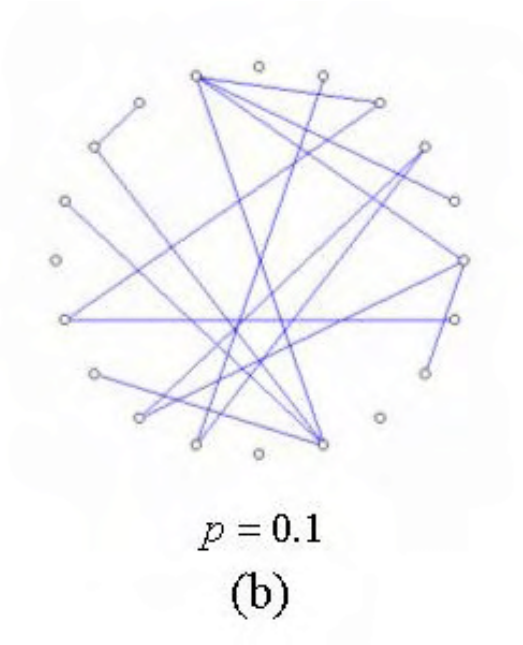
- About 1011 neurons
- Between 1014-15 synapses
- Notice: this means that every neuron has on average 1000 – 10,000 presynaptic neurons, and again as many postsynaptic ones
- *A random graph?*

Erdo”s – Renyi graphs



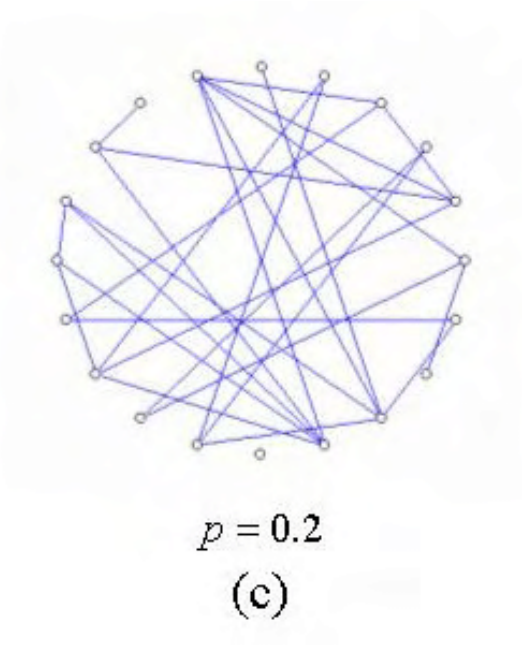
$p = 0$

(a)



$p = 0.1$

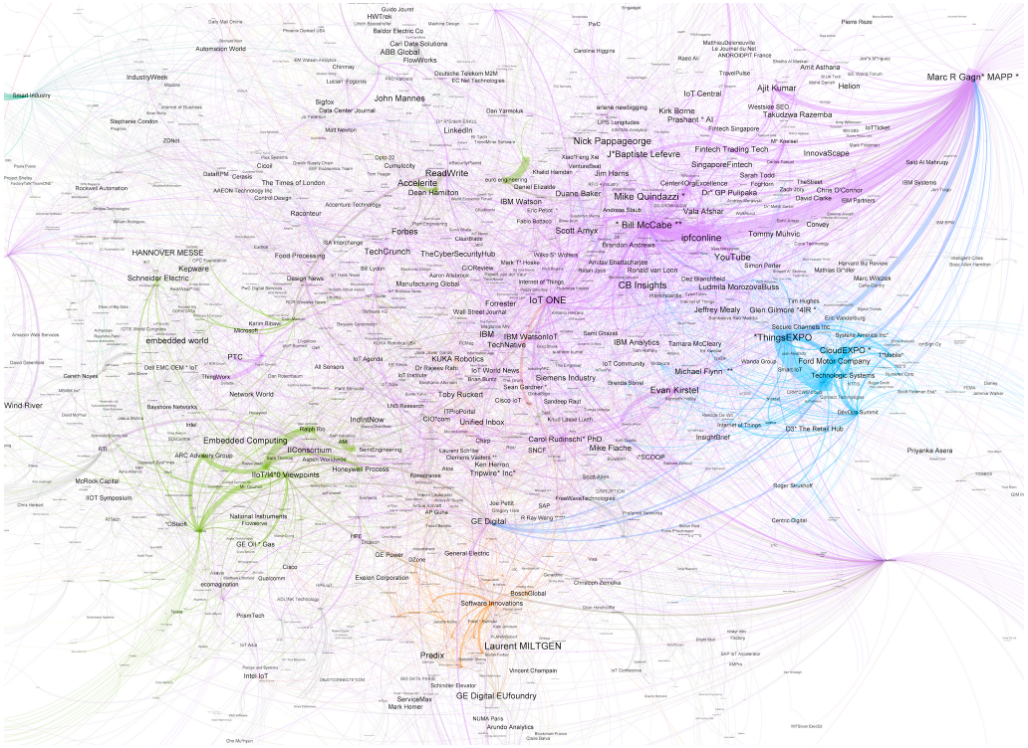
(b)



$p = 0.2$

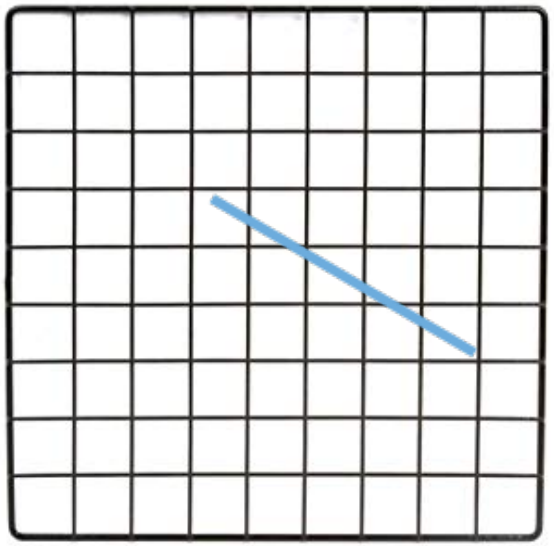
(c)

Power law / Internet like graph



Small world graphs [Kleinberg 2000]

- A grid (2D geometry!)
- Plus from each node
  very few random edges
- Going distance d away
  with probability ~ d-**2**



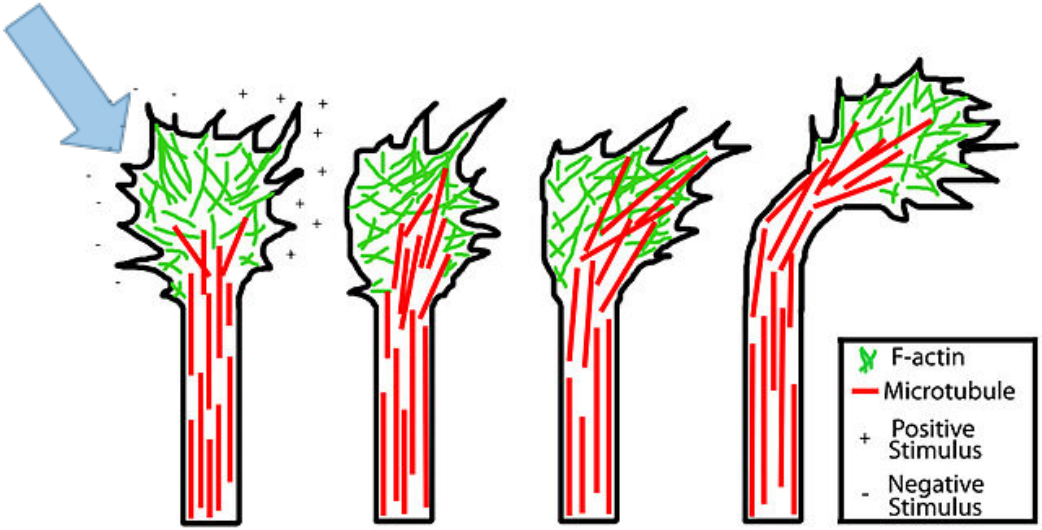
- **Theorem**: Greedy algorithm routes in very few steps

What is the generative model of the Human Connectome?

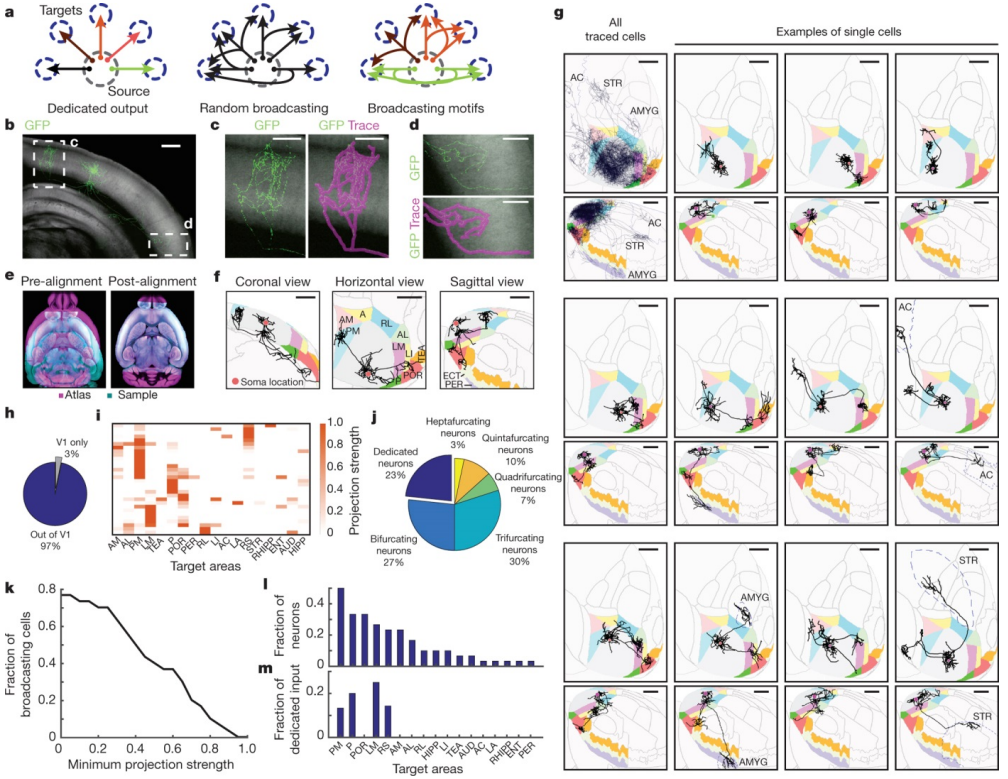
3. Then the axon's growth cone takes over

- A sensory – motor organ

- Navigates the brain following chemical cues (advance, turn, stop)

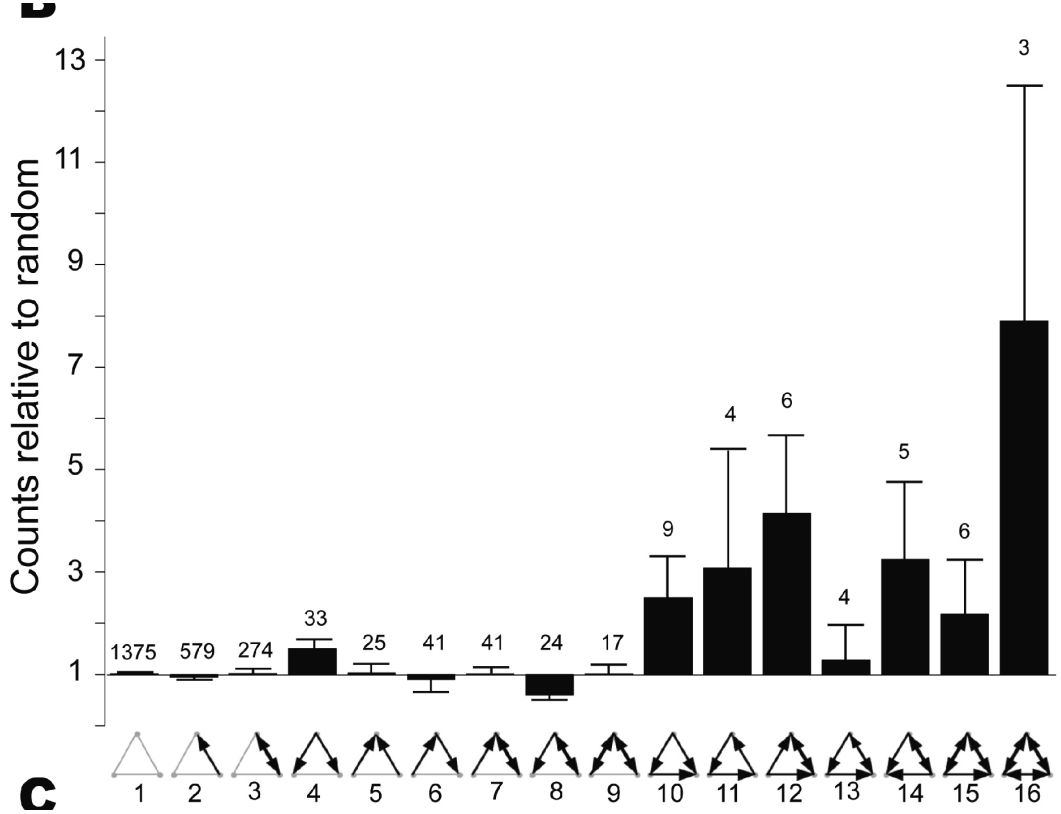


Brain-wide single-cell tracing reveals the diversity of axonal projection patterns of layer-2/3 V1 neurons,with most cells projecting to more than one target area



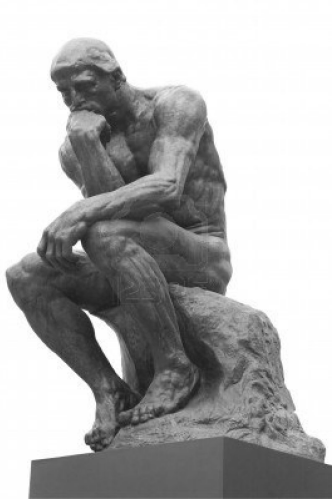
Y Han *et al. Nature* **556**, 51–56 (2018)
doi:10.1038/nature26159 nature

So, what kind of graph is the connectome? Three neuron connectivity



"Highly Nonrandom Features of Synaptic Connectivity in Local Cortical Circuits"
S. Song et al., PLOS Bio 2005

Biases from Random
Why? *



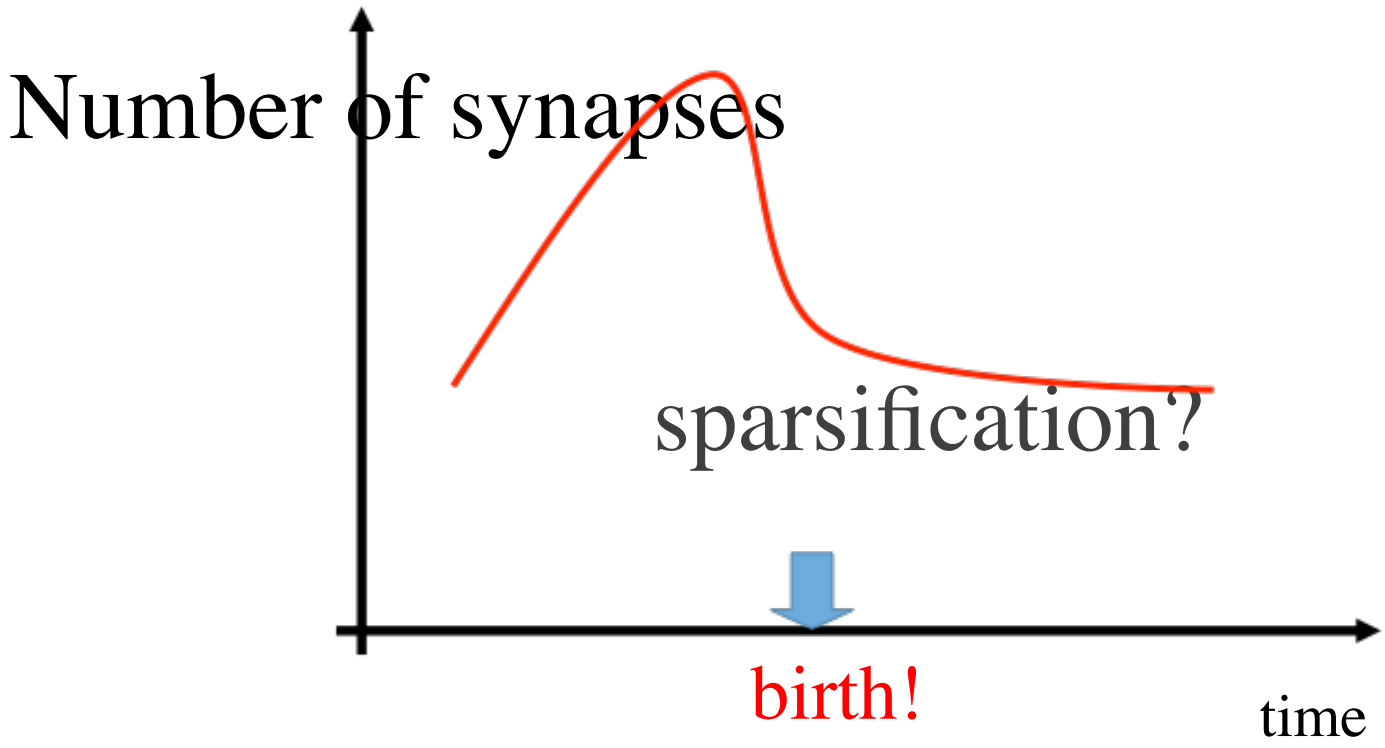
- Geometry?

"Independently Outgrowing Neurons and Geometry-Based Synapse Formation Produce Networks with Realistic Synaptic Connectivity"

van Ooyen et al. PLOS 1 2014

- Or plasticity?

cf H. Markram V2 cells responding to the same edge direction are much more likely to be connected


Number of synapses
sparsification?
birth!
time

Questions? Thoughts? Feedback?

(f) Discussion Point 1: "Is it really meaningful to watch talks that are far beyond the prior knowledge of the audience (in any research area)?"

Again this is what I want to discuss, which is about how to better arrange the topic and difficulty in the videos. I would admit that the Simon's Institute has really good lectures from great speakers in many topics, but for beginners in that area it's sometimes hard to follow and learn (especially for neuroscience topics which requires many prior knowledge or prerequisites). Previous talks were not that hard to follow, but this week's talk is really hard at least from my perspective.

Moreover, the content of the talks have no connection with the materials in the course, which makes it even harder to build a complete understanding of the big picture. I'm not wondering what's the main purpose of watching talks irrelevant to the course's main topic: if the target is self-studying something else after class, then why should we pay for the course?

Thalamus????

How to read and watch for HW: have open on your screen Wikipedia, Scholarpedia, other papers, etc.
If stuck, ask and chat in Piazza.

Start Early!

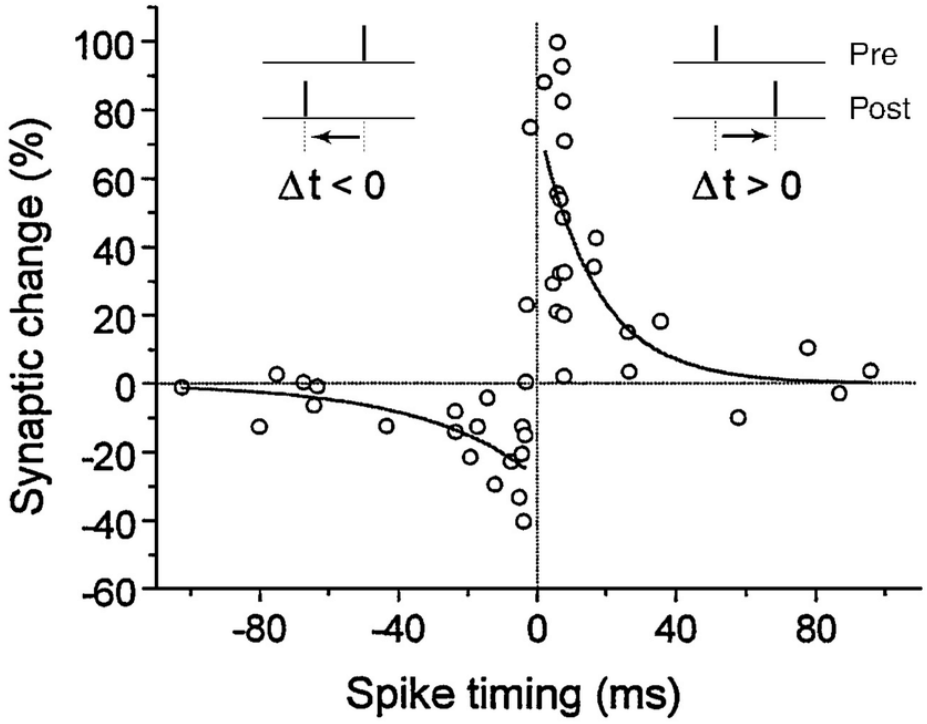
Today: Information, the Brain, and DNNs

- But first:

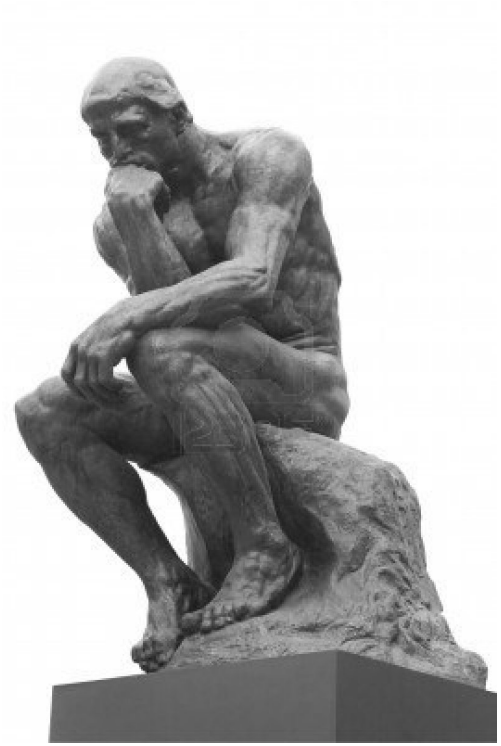
- Two talks by Jacob Portes and Kiran Vodrahalli!!

Spike timing-dependent plasticity (STDP)

If spike arrives in time, some gain. Just in time, big gain.
If it misses it, some loss.
Just misses it, big loss.



Wow! What does this mean? *



Plasticity and DNNs

- The Brain apparently learns through the plasticity of the synapses
- DNNs learn through back propagation, a kind of synaptic plasticity
- Same things then?
- ***Big*** difference: in back propagation information flows ***from*** post-synaptic ***to*** pre-synaptic neuron (biologically implausible)
- [Lillicrap et al. 1914]: constant (non-plastic) random synaptic weights in a **backwards** synapse suffice for some learning…

Today

- Information, the Brain, and DNNs

Next Wednesday: Dynamical systems and the Brain

Information Theory

- C. Shannon 1948 *"A mathematical theory of communication"*
- Want to communicate over binary, noisy channel
- Channel capacity C bits per second
- **Q:** What is the highest possible rate?
- A: It depends on C, and the entropy of the information source (the distribution of messages)

  Max rate achieved by coding

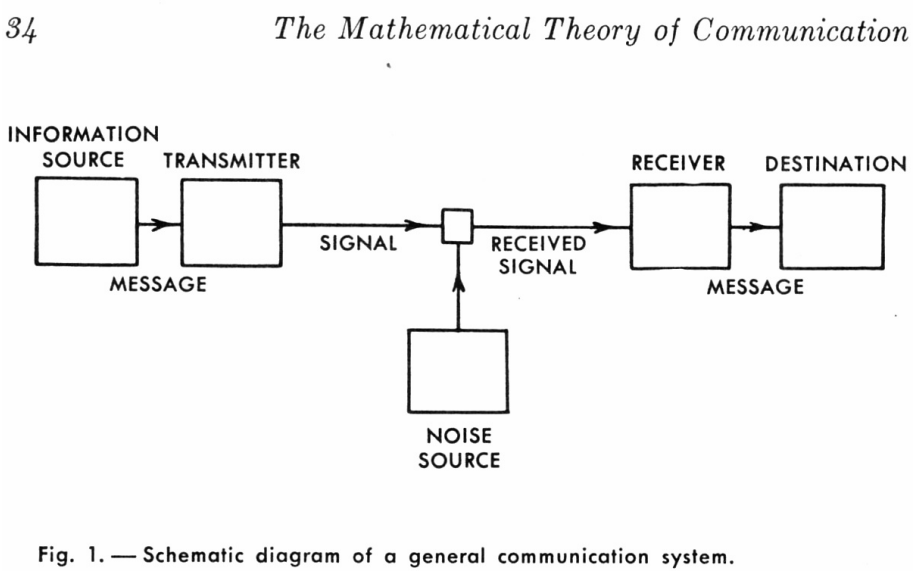


Fig. 1. — Schematic diagram of a general communication system.

Entropy of an event r

- Say, one particular response of a neuron
- Its probability is Prob[r]
- The ***entropy*** of r, H(prob[r]) ≥ 0, measures how "interesting" or "surprising" the event r is
- If r' is the response of another distant neuron, then we expect that the surprise/***entropy*** of the joint event (r, r') is the sum
- E(Prob[r, r']) = E(Prob[r] Prob[r']) = E(Prob[r]) + E(Prob[r'])
- It follows that **E(prob[r]) = - log2 (prob[r])**

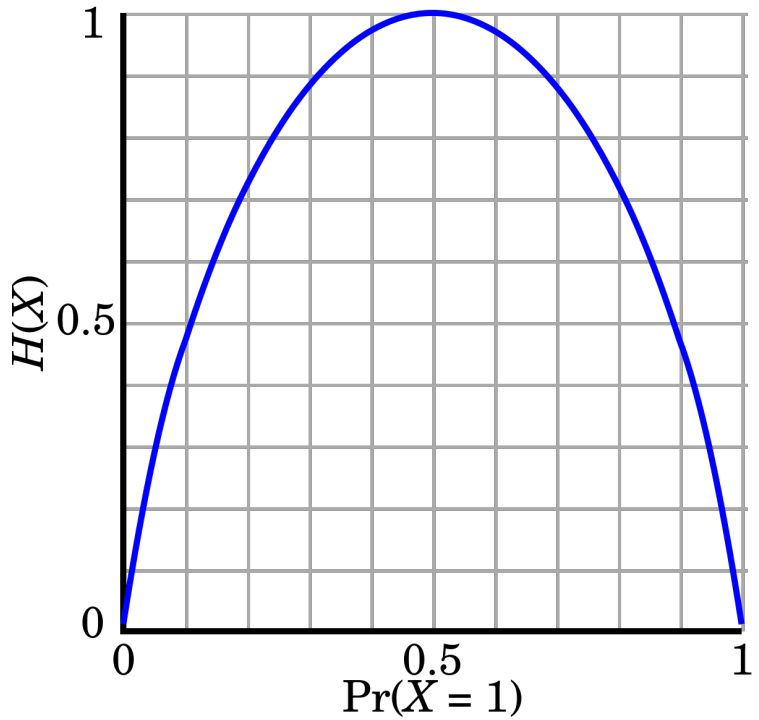
Entropy of a distribution (or of a source)

- Suppose that the response of a neuron (or a stimulus in the environment) has a discrete distribution

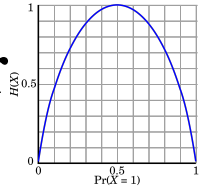
  D = {Prob[r1], …, Prob[rm]}.

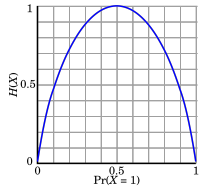
- Q: What is the expected surprise(entropy) of the distribution?
- A: **H(D) = -Σj Prob[rj] log2 (Prob[rj])**
- The ***entropy*** of a distribution
- Of the stimulus, the response, or the ***source*** more generally

Entropy of a binary variable X



Entropy of an information source

- Random bits: H = 1, remember 

- Biased bits (p, 1-p): H = h(p)
- Random symbols in {a, b, …, z, _ } H = 4.8
- Important: entropy is maximized when the symbol probabilities are equal, remember 

- Reason: log2 is a concave function
- English text: H is about 1.6 due to redundancy

From Shannon's paper:

XFOMJRXKHRJFFJUJZ LPWCFWKCYJFFJEYVKCQSGHY DQPAAMKBZAACIBZLHJQD.

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

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

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

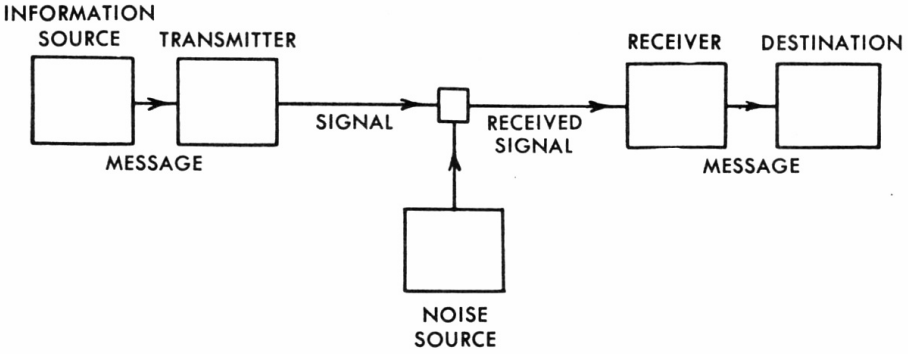
REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO.

THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS.

A parenthesis: Kolmogorov complexity

- Q: What is the complexity of 00110100110100011011101011011101 00?
- How about 011011011011011011011011011011011011011011?
- (And, is this a meaningful question?)
- A (A. Kolmogorov 1963): The length of the smallest program that can generate the string
- Random string of length n has complexity n (+ c)
- c depends on the programming language, and becomes insignificant for large strings
- The Kolmogorov complexity of a random source is its entropy

Shannon's paper (1948)



- Defines entropy in the context of communication
- Channel capacity C
- Noise in the channel: bit is changed with probability p
- Source of entropy H generates a stream of symbols
- e.g., English text
- Q: What is the highest possible rate of transmission?

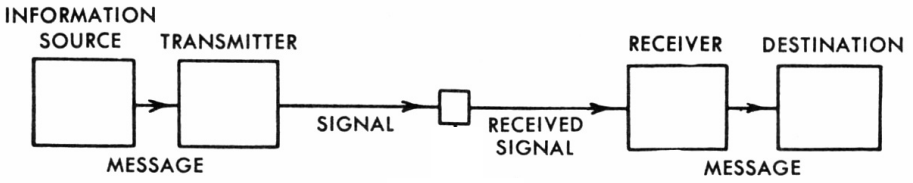
Shannon's first theorem

Theorem 1: Suppose the channel has capacity C bits/sec, the source has entropy H, and there is no noise. Then the maximum rate (symbols/sec) is C/H:

a. No rate greater than C/H can be achieved

a. For any $\varepsilon > 0$, rate $\geq$ C/H – ε can be achieved by coding

Proof of (a): Recall



Shannon's first theorem

Theorem 1: Suppose the channel has capacity C bits/sec, the source has entropy H, and there is no noise. Then the maximum rate (symbols/sec) is C/H:

a. No rate greater than C/H can be achieved

a. For any $\varepsilon > 0$, rate $\geq$ C/H – ε can be achieved by coding

Proof of (a) **Lemma**: If 

then H' $\leq$ H

So, take the box to be encoder + channel: #symbols/sec = H'/H $\leq$ C/H

Shannon's first theorem

Theorem 1: Suppose the channel has capacity C bits/sec, and there is no noise. Then the maximum rate (symbols/sec) is C/H:

a. No rate greater than C/H can be achieved

a. For any $\varepsilon > 0$, rate $\geq$ C/H – ε can be achieved by coding

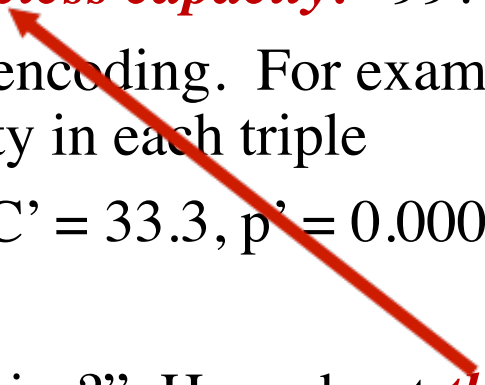
Proof of (b): Encode blocks of m input symbols (where m is large enough to achieve ε). The encoding of block B, denoted **e(B),** is this:

Order all (#symbols)m blocks in order of decreasing probability. If B belongs to the blocks of the first half (50%) of probability mass, then the first bit of e(B) is 0, otherwise it is 1. Repeat for the 25% fractions of this 50%. And so on until you hit the block itself. (Fano – Shannon code)

Lemma: The length of the encoding of B, |e(B)| ≤ log2(1/prob[B])

Channel with noise

- Suppose that the channel has capacity C bits/sec ***but*** with probability p a bit is flipped
- Say C = 100 and p = 0.01
- What is the ***equivalent noiseless capacity?*** 99? Or perhaps ***92***?
- We can fight noise through encoding. For example, repeat each bit three times and take majority in each triple
- Channel capacity becomes C' = 33.3, p' = 0.000301
- Capacity – noise tradeoff.
- Can we do better? "Zero noise?" How about ***this***?



The capacity of a channel with noise is…

$$C' = C\,(1 - H(\{p, 1-p\}))$$
$$= C\,(1 - h(p))$$

When $p = \frac{1}{2}$ then $C' = 0$

Appropriately, because in this case the chance of equivocation (uncertainty, error) becomes 1.

Equivocation? Relative and conditional entropy

- Two data streams, x and y
- (input of the encoder, output of the decoder)
- Entropy of the joint distribution **H(x,y)**
- Bayes (or chain) rule: **H(x,y) = H(x) + H(y|x) = H(y) + H(x|y)**
- **H(y|x)** is a measure of how surprised you are looking at y if you know x.
- This is the **equivocation** of the transmission

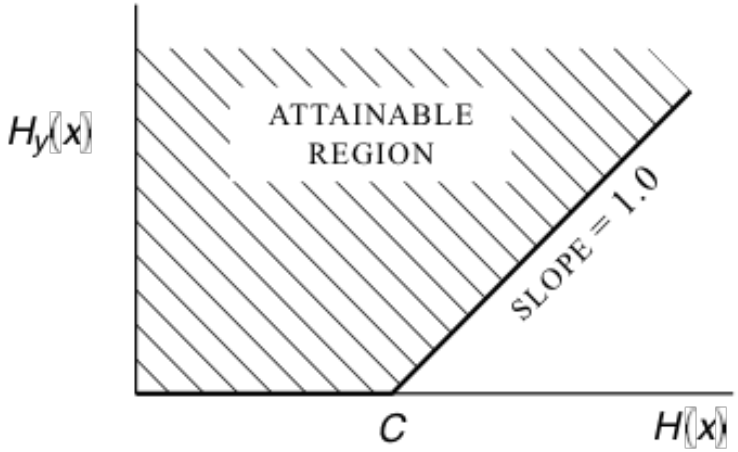
Shannon's second theorem

Theorem 2: If the channel has capacity C and noise p < ½ then

(a) Any rate R < C (1– h(p)) can be achieved by coding

(b) No rate greater than C (1 – h(p)) can be achieved

(c) If equivocation is allowed, the attainable region is as shown



Shannon's second theorem, part (a)

Theorem 2: If a channel has capacity C and noise $p < \frac{1}{2}$ then

(a) Any rate $R < C\,(1 - h(p))$ can be achieved by coding

Proof of (a): Consider a long bit string B of length m. Map each to a random bit string c(B) of length m + r, where r is the redundancy afforded by the excess of C over R.

Remarkably, after c(B) is transmitted as a corrupted bit string A, B can be recovered from A as the closest code to it.

This is because the "balls" around all c(B)s and radius p(m+r) are disjoint (with high probability).

Shannon's second theorem, parts (b) and (c)

Theorem 2: (b) No rate greater than C (1 – h(p)) can be achieved

c) If equivocation is allowed, the attainable region is as shown

Proof of (b) and (c): If the code has redundancy less than C h(p), the balls centered at any code word will intersect with large probability and equivocation results.

OK, back to information and neurons, some basic facts

- We have introduced some important concepts in entropy
- Entropy of the joint distribution **H(x,y)**
- Conditional entropy **H(x|y)**
- Chain rule: **H(x,y) = H(x) + H(y|x) = H(y) + H(x|y)**
- Mutual information:

I(x, y) = Σi,j Prob[i,j] log2(Prob[i,j]/Prob[i] Prob[j])

OK, back to information and neurons, some basic facts (cont.)

- Mutual information:

  $I(x, y) = \Sigma_{i,j}\ Prob[i,j]\ \log_2(Prob[i,j]/Prob[i]\ Prob[j])$

- Kullback-Leibler divergence of two distributions P, Q

  $KL(P,Q)$ $(\neq KL(Q,P))$ **$= \Sigma_i\ P[i]\ \log_2\ (P[i]/Q[i])$**

- Notice: **$I(x, y) = KL\ (P(x,y), P(x)\cdot P(y))$**

Finally…

- Discrete information theory can be extended to continuous random variables and distributions with minimal conceptual effort

$$\Sigma \rightarrow \int$$

Information theory and the Brain

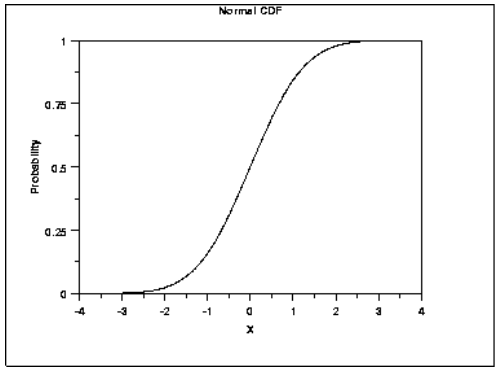
- Helps elucidate how information is communicated in the brain
- How much information
- Is the coding used good? Optimal?

One application: how flies see

- There is a neuron called large monopolar cell (LMC) in the visual system of the fly that senses the **contrast** of the scene
- It encodes the level of contrast **x** by a firing rate **r**
- Contrasts in the fly's environment come with a distribution

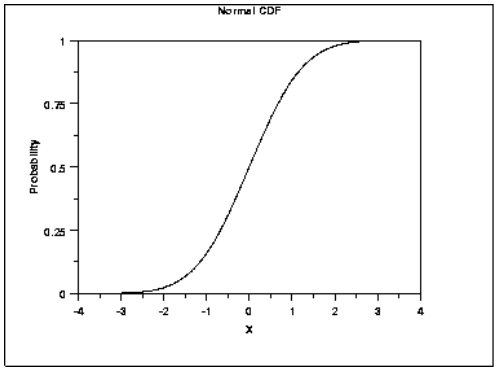
One application: how flies see (cont.)

- In 1981 neuroscientist S. Laughlin measured this distribution, and its cumulative curve looks like



- How do you think that the fly encodes these levels of contrast? For simplicity, assume that it does through a dozen different firing rates

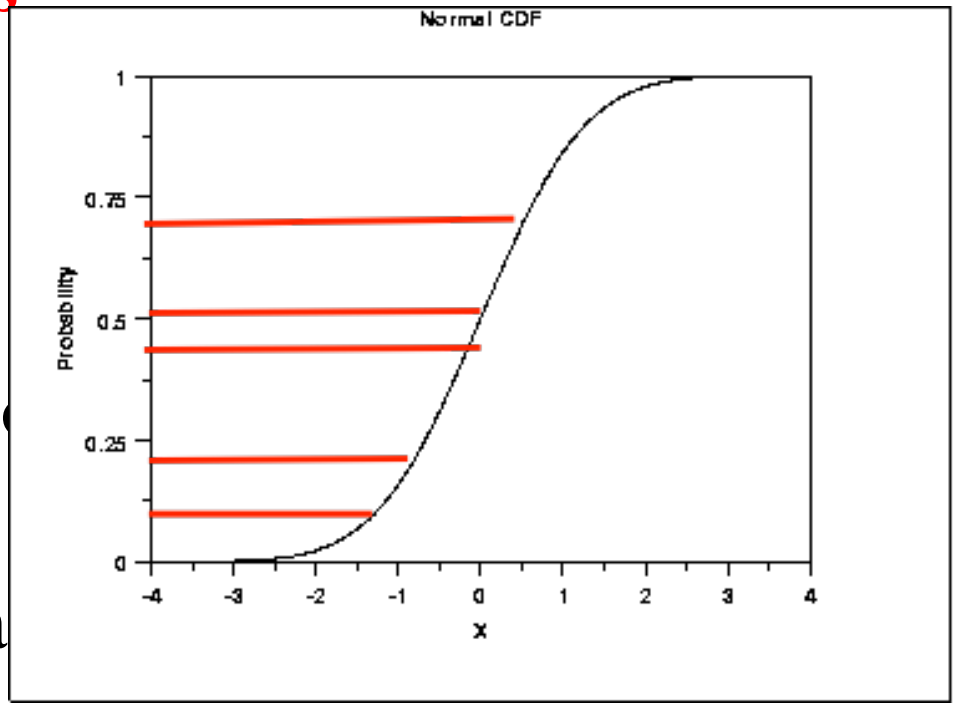
One application: how flies see (cont.)



- Let's think. The LMC is a source that informs the rest of the fly's brain about the contrast it sees
- The fly wants to maximize the entropy of this source, so it conveys, on average, the maximum possible information
- How is this achieved?

One application: how flies see (cont.)

- Let's mark the **boundaries** between different rates
- The probabilities covered should be equal
- Because we know this is h you maximize entropy
- The stripes should be equa
- **The response distribution**



One application: how flies see (cont.)

- **The response distribution should mimic the stimulus cumulative distribution!**