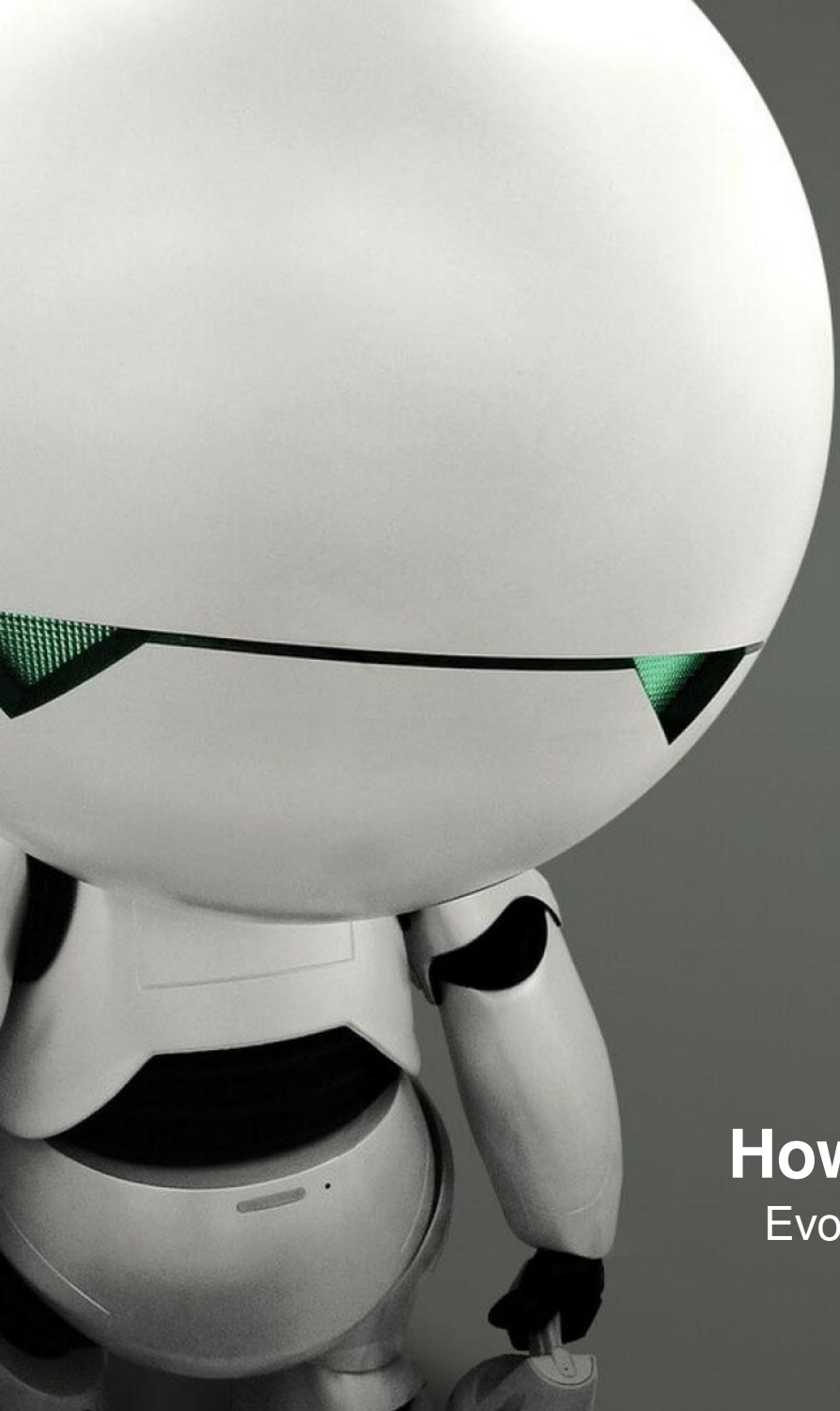
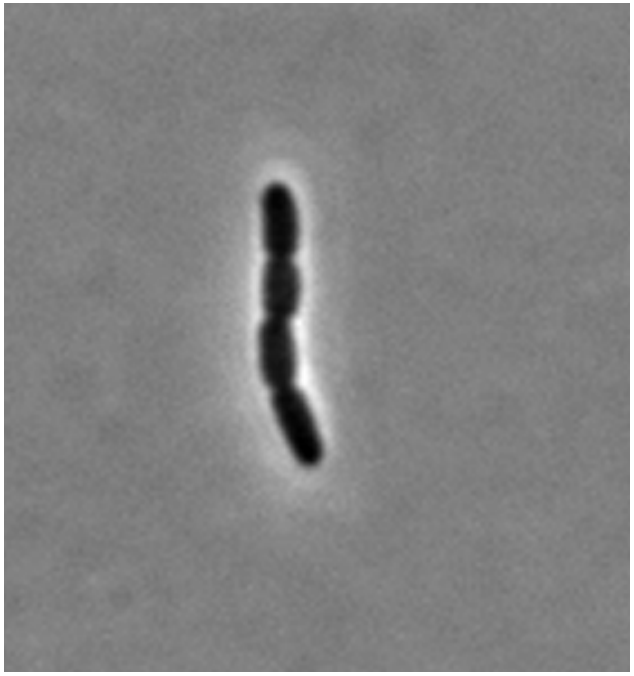


How to create a sentient machine

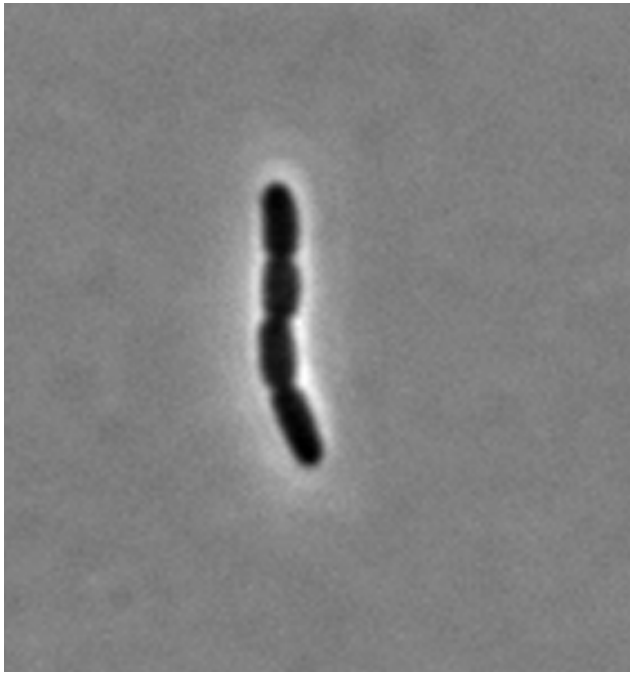
Evolving complex information processing systems

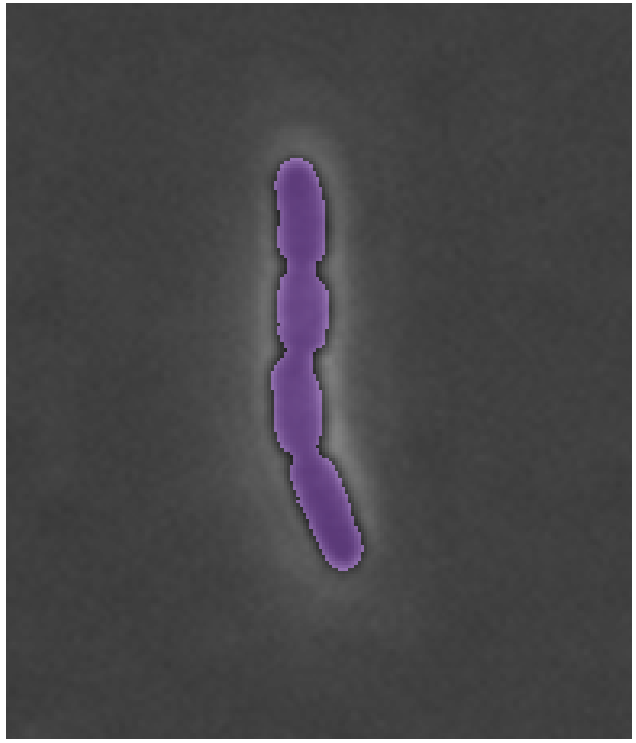
Tiago Ramalho

Where's E. Coli?



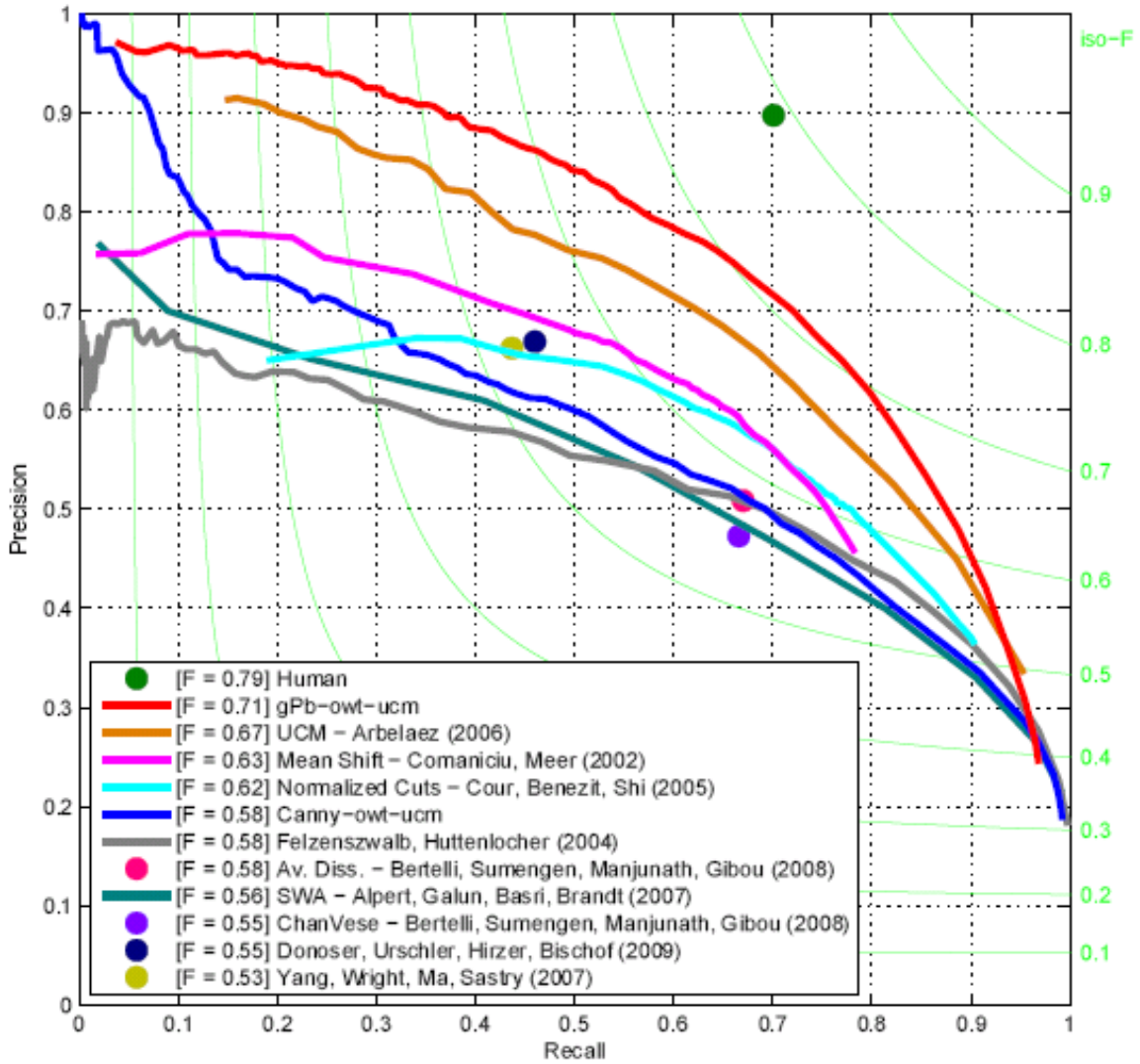
Where's E. Coli?



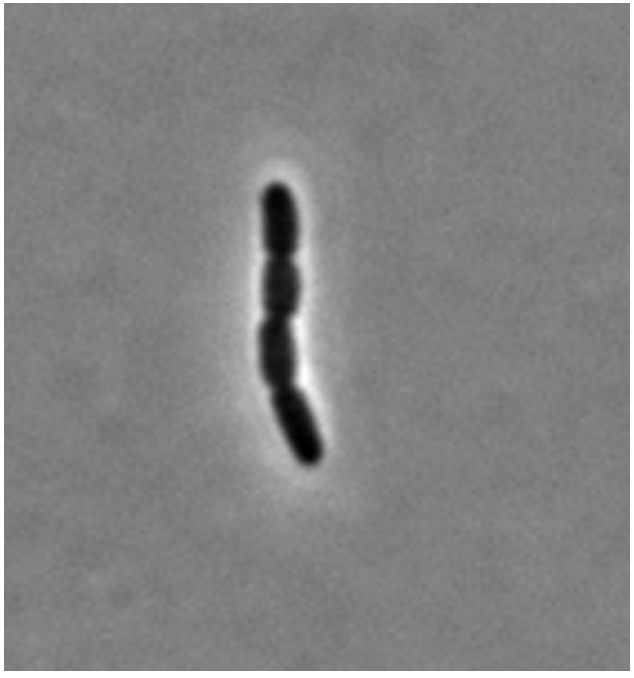


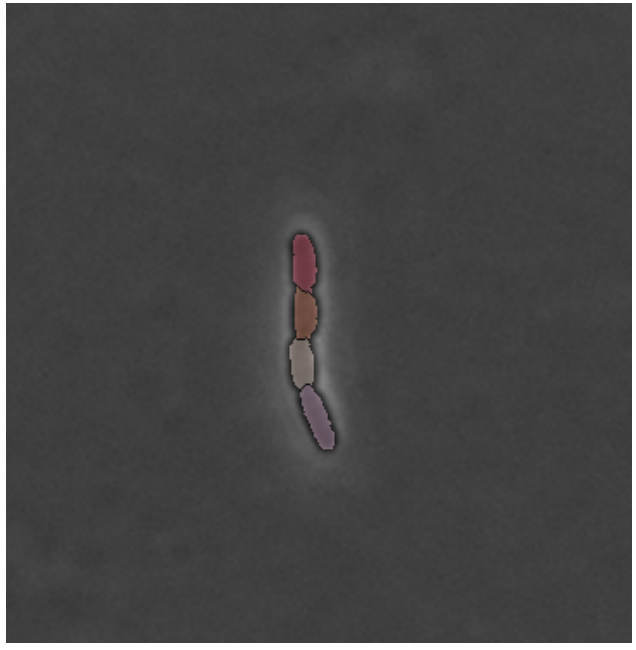
State of the art

[Arbelaez, 2010]



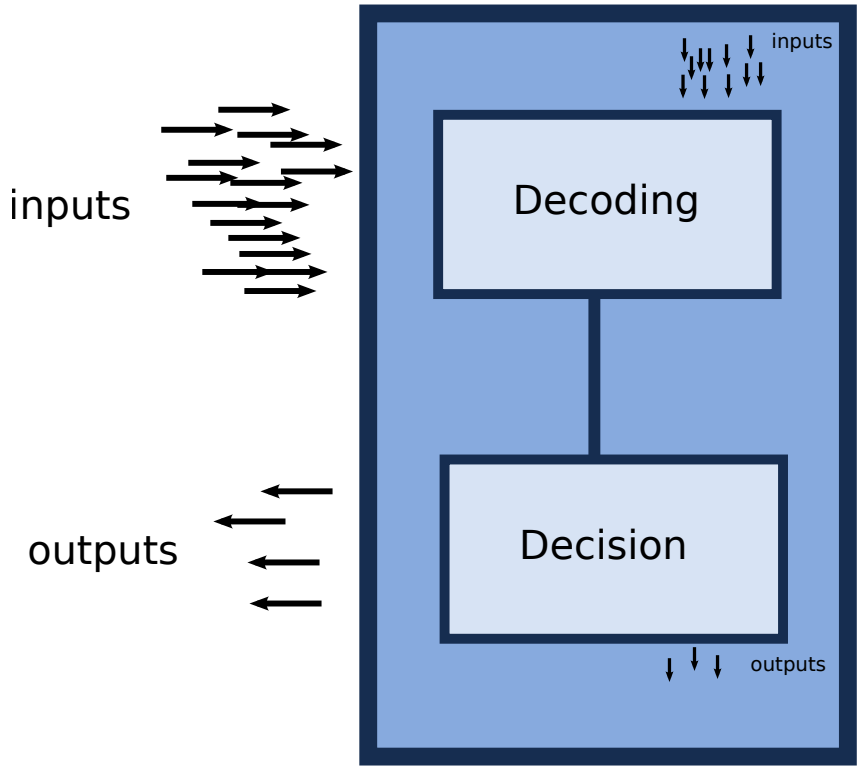
Artificial stupidity





Define intelligent

- Acquisition (signals)
- Decoding (data)
- Decision (minimize loss)
- Action (output)

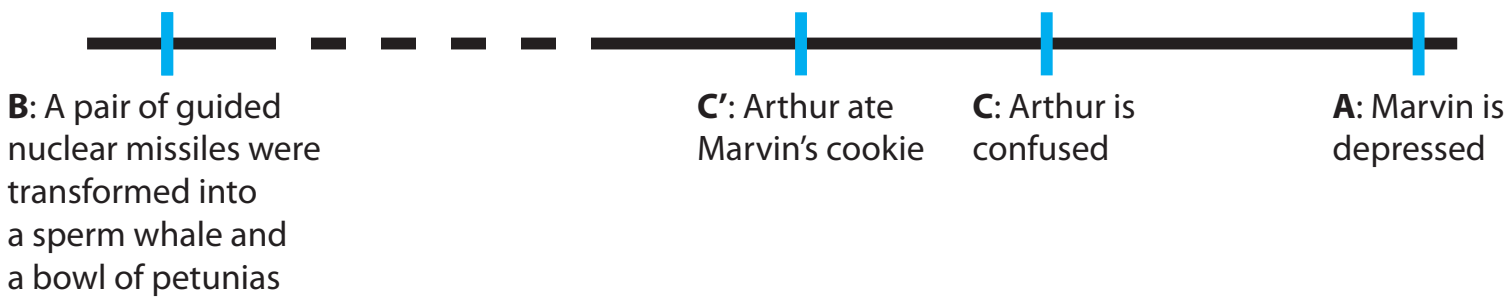


Deductive reasoning

[Jaynes, 2003]

System of propositions $\{A,B,C...\}$

- plausibility of A is a real number



Deductive reasoning

[Jaynes, 2003]

System of propositions $\{$A,B,C...$\}$

- plausibility of A is a real number
- common sense

$C \to C'$ such that
$p(A|C') > p(A|C)$
Then:

- $p(B|AC') = p(B|AC)$
- $p(AB|C') \geq p(AB|C)$
- $p(\bar{A}|C') < p(\bar{A}|C)$

B: A pair of guided nuclear missiles were transformed into a sperm whale and a bowl of petunias

C': Arthur ate Marvin's cookie

C: Arthur is confused

A: Marvin is depressed

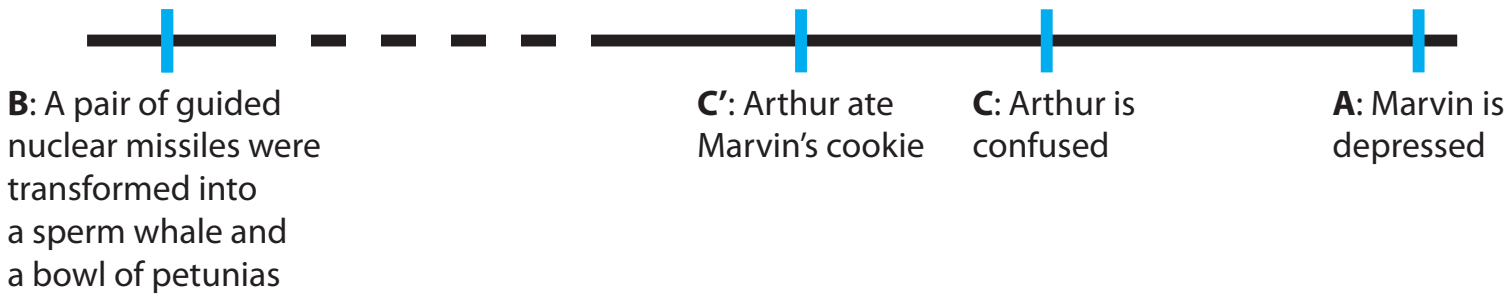
Deductive reasoning

[Jaynes, 2003]

System of propositions $\{A,B,C...\}$

- plausibility of A is a real number
- common sense
- consistent

- Same conclusions, different paths $\rightarrow$ same result
- Always use all information
- Eq. knowledge = Eq. plausibilities



Probability

Choose $P(A) \in [0,1]$, probability.
Then:

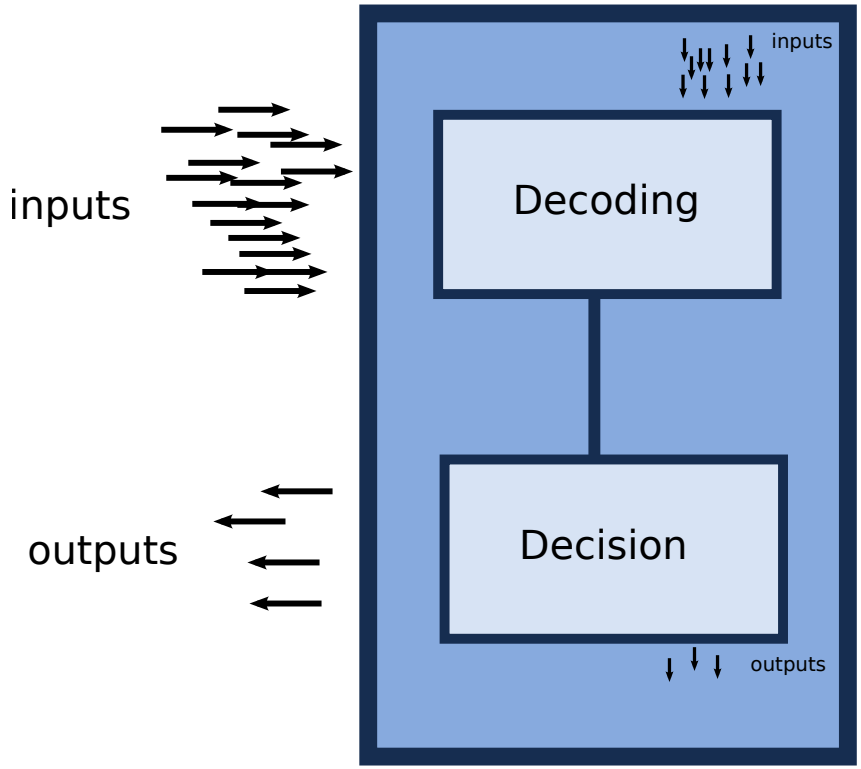
- $P(A|B) + P(\bar{A}|B) = 1$
- $P(AB|C) = P(A|C)P(B|AC)$
- $P(A+B|C) = P(A|C) + P(B|C) - P(AB|C)$
- $P(A|B) = P(B|A)P(A)/P(B)$

The actual science of logic is conversant at present only with things either certain, impossible, or entirely doubtful, none of which we (fortunately) have to reason on. Therefore the true logic for this world is the calculus of probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind.

James Clerk Maxwell, 1850

Define intelligent

- Acquisition (signals)
- Decoding (data)
- Decision (minimize loss)
- Action (output)



Decisions, decisions...

Data contains probability $\{p_i\}$ for events $\{E_i\}$. Do some output with probability $\{q_i\}$. Payoff is $A \log nq_i$ if event E_i occurs.
Your expected return is

$$A[\log(n) - k(q,p) - H(p)]$$

with

$$k(q,p) = \sum_i q_i \log \frac{q_i}{p_i} \textit{ (Kullback Leibler divergence)}$$

$$H(p) = \sum_i p_i \log p_i \textit{ (Shannon entropy)}$$

Take this home

- Your actions have to match reality!
- You have to decode reality as well as possible

Take this home

- Your actions have to match reality!
- You have to decode reality as well as possible

Important question:

Take this home

- Your actions have to match reality!
- You have to decode reality as well as possible

Important question:

You just made the log thing up didn't you?

The log thing makes sense

[Leibler, 2005]

Let's give our machines the ability to reproduce.
Then their number is $N_t = N_0(1+r)^t$.
Long term behavior:

$$\lim_{t\to\infty} \frac{\log N_t}{t} = N_0 \log(1+r)$$

Growth rate r directly influenced by the output of the machine.

Ok, so how do we do it?

It's been done before!

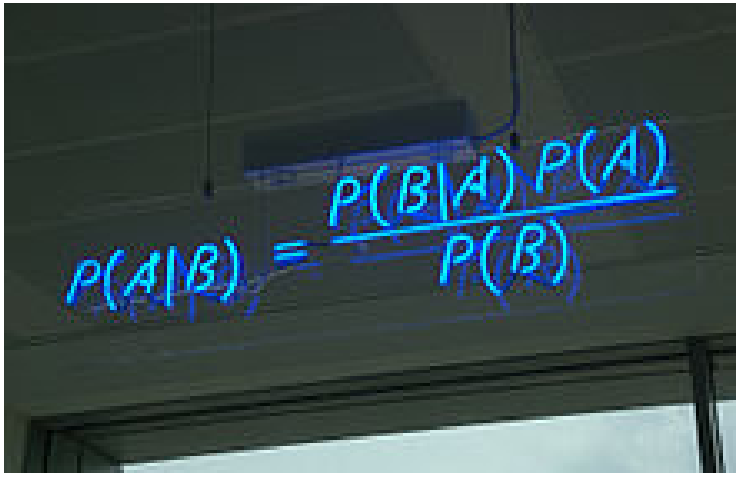
Metal things

Squishy things

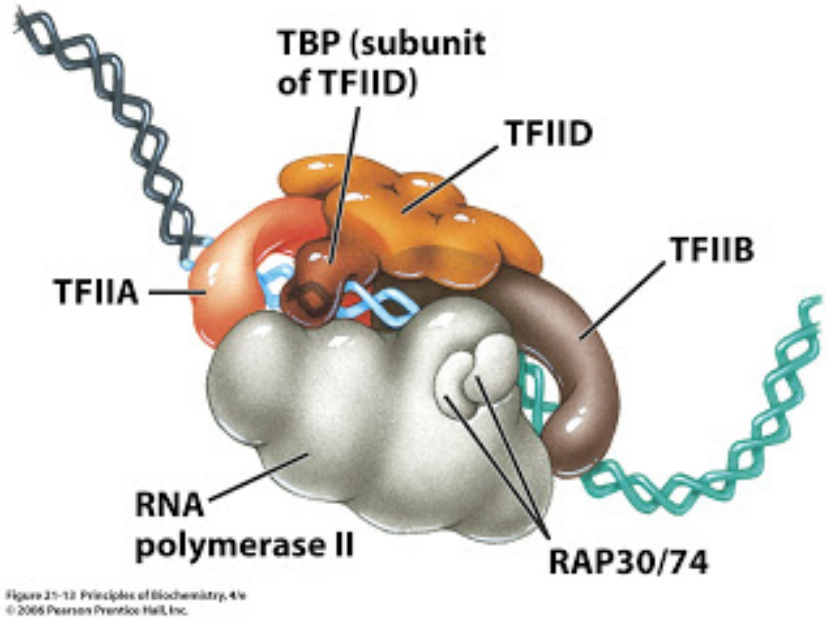
Decoding signals

[Koller; Friedman, 2009]

- Causal relationships
- Independence structure
- Simple representation for complex models
- Obeys probabilistic reasoning



Bayesian network



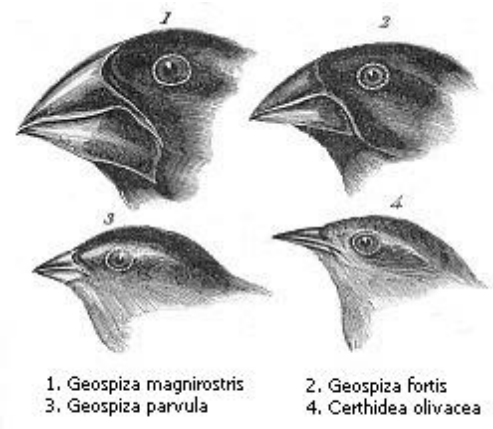
Gene regulatory network

Evolution

[Storn; Price, 1997]

- Randomly pick parameter vectors from the population and cross over their values
- Randomly mutate individual parameters
- Keep individuals with the highest scores
- Iterate

Differential evolution



Finches from Galapagos Archipelago

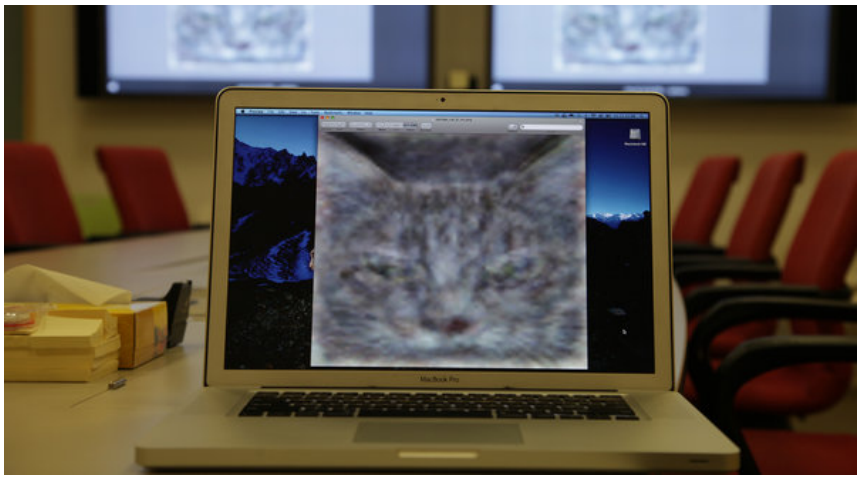
Natural selection

Endgame

[Ng; Dean, 2012]

- Too many inputs $\rightarrow$ feature extraction.
- Learning $\rightarrow$ adapt the prior to the statistics of the environment.
- Implementation: markov random fields.

$$P(F) = \frac{1}{Z} \exp\left(\sum_i w_i f_i(\{x\})\right)$$



Deep neural nets

Your brain

Questions

- Can we learn something from statistical models?
- Is it true intelligence?
- Can we grasp our own complexity?