Natural Distributed Algorithms

- Lecture 3 -
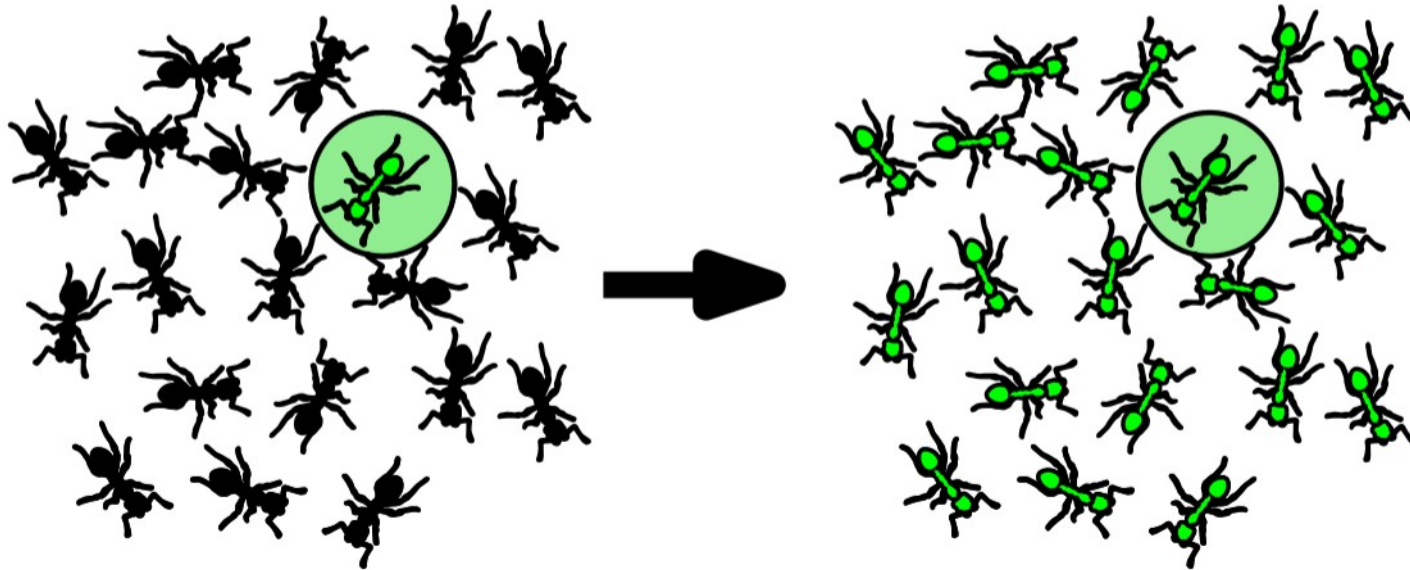## Rumor Spreading in the
Noisy PULL Model

Emanuele Natale
CNRS - UCA

CdL in Informatica
Università degli Studi di Roma "Tor Vergata"

Rumor Spreading Problem

- One **source** node in a **special state**
- **Goal configuration**: all agents in the **special state**

Stochastic Interactions: **PULL** Model
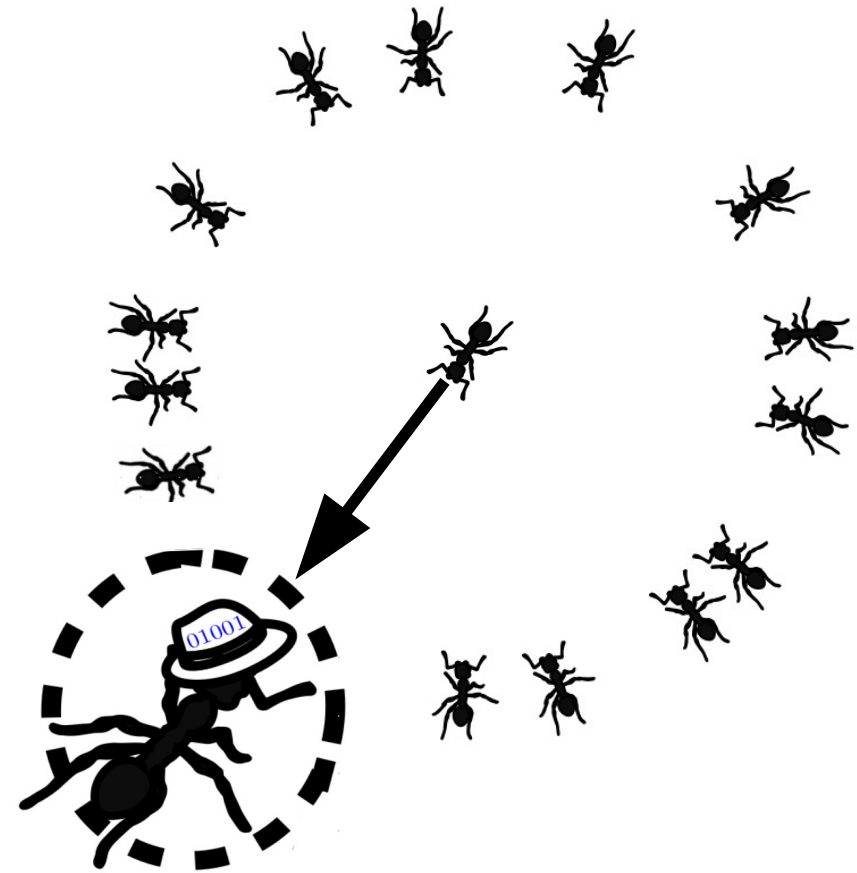
RECALL

Desired features
- Stochastic
- Parsimonious (Anonymous)
- Passive (uni-directional)

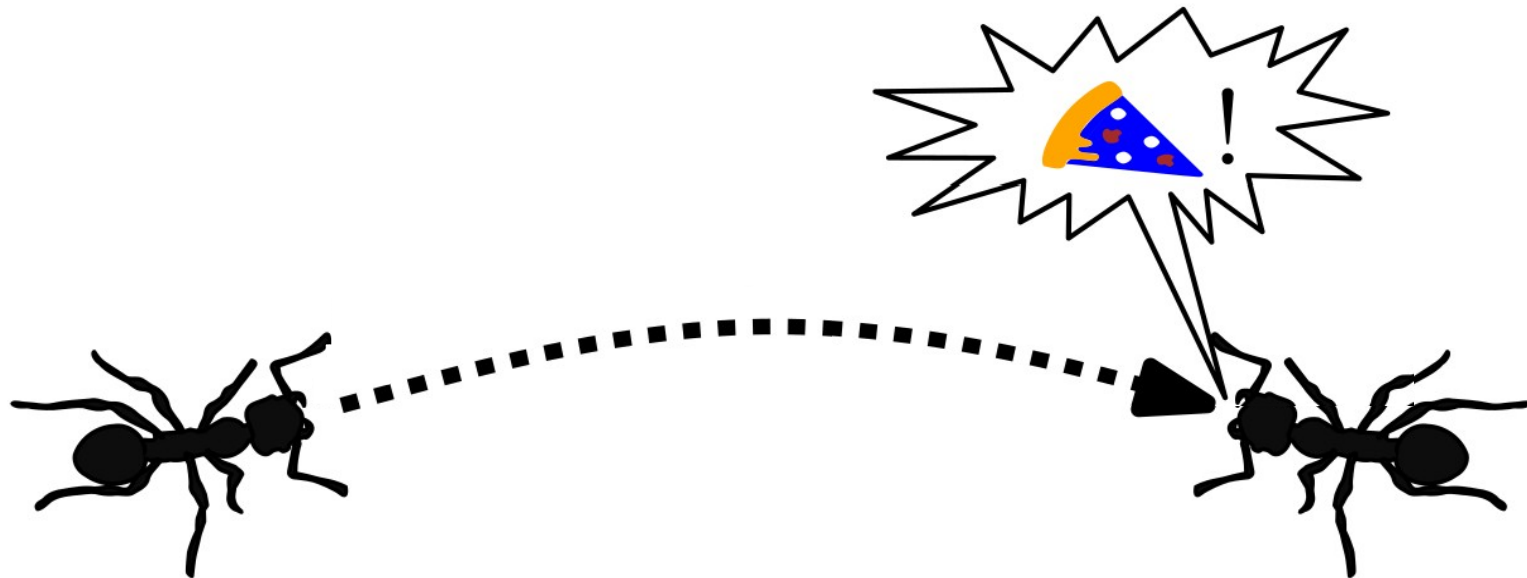**(Uniform) PULL model**
[Demers '88]
(single binary message)
- Discrete parallel time
- At each round each agent **shows** **one bit** message
- The agent can observe the message shown by **one agent** chosen independently and uniformly at random

Noisy Communication

Before being received,
each bit is flipped with probability $\frac{1}{2} - \epsilon$

1 bit per message,
noise flips it with prob $\frac{1}{2} + \epsilon$

Rumor Spreading in the **Noisy** PULL Model:
The Source is Lost in Too Much Noise

In the **PUSH** model we could estimate **the distance that a message travelled from the source**, thanks to the fact that in the B.B.S. algorithm agents can **avoid communicating**
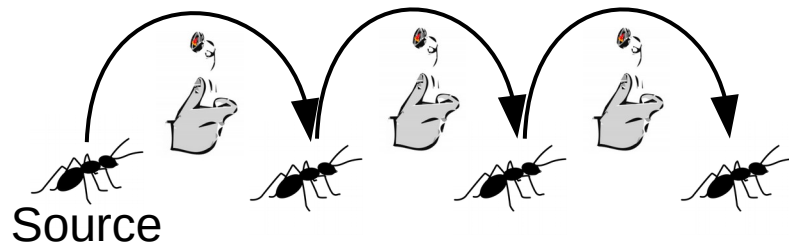
Source

In the **PULL** model we can only ask **one bit to a random node**:

? ? ?
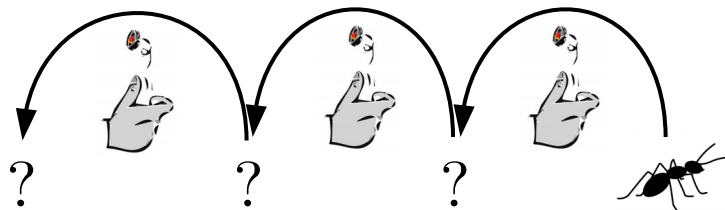
What is the probability to receive the right opinion message from the source?

How many messages do we need then?

Broadcast-PULL: A Simpler Model

In the PULL Model each agent gets a message from another randomly-chosen agent

⬇

Different agents get different information…
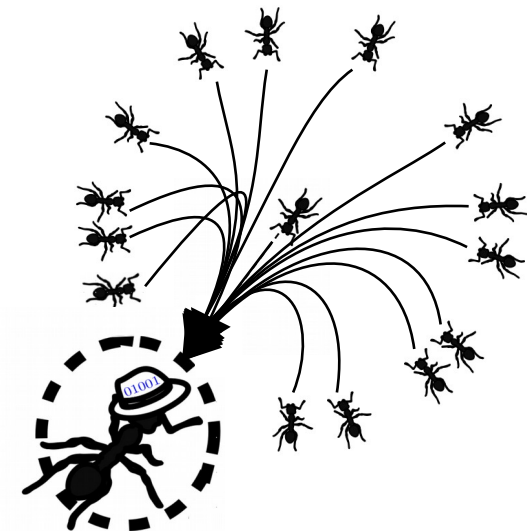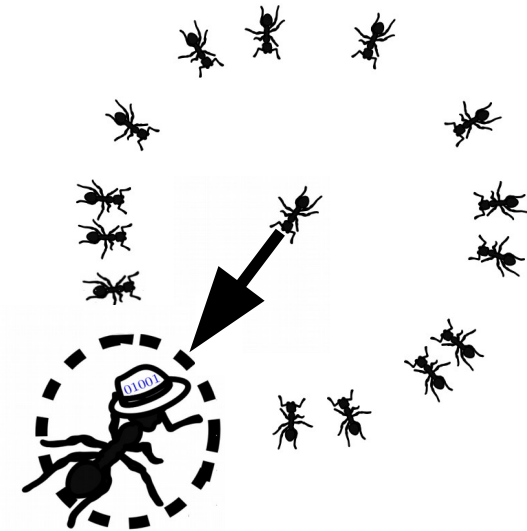Difficult to analyze!

⬇

Pòlya's principle:
"If you can't solve a problem, then there is an easier problem you can solve: find it."

⬇

Broadcast-PULL Model:
Each agent sees all messages…
and everything all other (non-source) agents see!

Distinguishing the Opinion of the Source

Assume w.l.o.g. that at each round the agents pull a message one-by-one (**sequential** version).
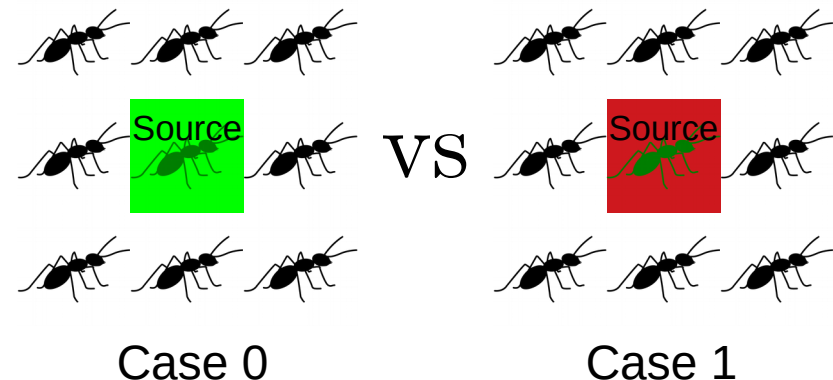
Recall: **every agent sees everything** that each non-source agents sees, even "random" choices (shared randomness)

The only difference between the two cases is the opinion of the source

Given **identical message histories**, the two systems behave identically

Case 0   VS   Case 1

Lower Bound on Number of Messages via Relative Entropy: Preliminaries

The relative entropy between random variables X and Y is

$$D\left(X,Y\right) = \sum_{x} P\left(X = x\right) \log \frac{P\left(X = x\right)}{P\left(Y = x\right)}.$$

Important properties are the Chain Rule

$$D((X, X'), (Y, Y')) = D(X', Y')$$
$$+ \mathbb{E}_{x' \sim X'}[D((X|X' = x'), (Y|Y' = x'))]$$

and the Information Processing Inequality: for any function f

$$D(X, Y) \geq D(f(X), f(Y))$$

Lower Bound on the Number of Messages

Theorem

$X = (X_1, ..., X_n)$ and $Y = (Y_1, ..., Y_n)$ random variables, $f : \mathbb{R}^n \to \{0, 1\}$ a function that outputs 1 if it guesses that the input is distributed like X, 0 if it is distributed like Y, and such that $P(f(X) = 0), P(f(Y) = 1) \leq \delta$.

Denote $X_{<i} = (X_1, ..., X_{i-1})$ and the same for $Y_{<i}$.
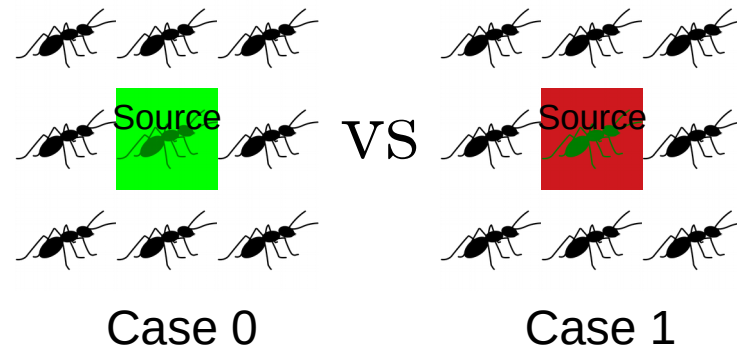
Then, it is necessary that

$$n \geq \frac{\log \frac{1}{2\delta}}{2 \max_i D\left((X_i | X_{<i}), (Y_i | Y_{<i})\right)}.$$

Proof ➡

Application of the Lower Bound 1/2

$$\mathrm{D}((X_i|X_{<i}),(Y_i|Y_{<i}))$$
$$= \mathbb{E}_{Z_{<i} \sim X_{<i}} \left[D\left((X_i|X_{<i}=Z_{<i}),(Y_i|Y_{<i}=Z_{<i})\right)\right]$$

Given the same history $Z_{<i}$, $p_{x,i}(z) := P(X_i = z | X_{<i} = z_{<i})$ and $p_{y,i}(z) := P(Y_i = z | Y_{<i} = z_{<i})$ are identical, except for the different sources.

Case 0　　　　　　Case 1

Let $\epsilon_i(z) = p_{x,i}(z) - p_{y,i}(z)$.
$|\epsilon_i(z)| \leq \frac{1}{n}$ since $p_{x,i}(z)$ and $p_{y,i}(z)$ differ iff the message come from the source, which happens with probability $\frac{1}{n}$.
Note also that $\epsilon_i(1) = -\epsilon_i(0)$.

Application of the Lower Bound 2/2

$$D\left(p_{x,i}\left(z\right), p_{y,i}\left(z\right)\right) = \sum_{z \in \{0,1\}} p_{x,i}\left(z\right) \log \frac{p_{x,i}(z)}{p_{y,i}(z)}$$

$$= \sum_{z \in \{0,1\}} p_{x,i}\left(z\right) \log \frac{p_{x,i}(z)}{p_{x,i}(z) + \epsilon_i(z)}$$

(Taylor expansion of $\log \frac{1}{1-x}$)
(constant noise: $p_{x,i}\left(z\right) \geq \frac{1}{3}$)

$$\leq \sum_{z \in \{0,1\}} p_{x,i}\left(z\right)\left(-\frac{\epsilon_i(z)}{p_{x,i}(z)} + \left(\frac{\epsilon_i(z)}{p_{x,i}(z)}\right)^2\right) \leq O\left(\frac{1}{n^2}\right)$$

that is, the required number of messages to attain the given probability of error is at least

$$\frac{\log \frac{1}{2\delta}}{2 \max_i D\left(\left(X_i | X_{<i}\right), \left(Y_i | Y_{<i}\right)\right)} \geq \frac{n^2}{2} \log \frac{1}{2\delta} \quad (\geq n \text{ rounds!})$$

Each round of PULL model corresponds to n observation of broadcast-PULL

Project Idea:

Propose algorithms to solve Rumor Spreading in the Noisy PULL Model on Erdős-Rényi graphs with parameter *p* and simulate them varying p and the noise parameter ε.

- Clementi, Andrea, Luciano Gualà, Emanuele Natale, Francesco Pasquale, Giacomo Scornavacca, and Luca Trevisan. 2018. "Consensus Needs Broadcast in Noiseless Models but Can Be Exponentially Easier in the Presence of Noise." Report. CNRS. https://hal.inria.fr/hal-01958994/document

- Boczkowski, Lucas, Ofer Feinerman, Amos Korman, and Emanuele Natale. 2018. "Limits for Rumor Spreading in Stochastic Populations." In 9th Innovations in Theoretical Computer Science Conference (ITCS 2018), edited by Anna R. Karlin, 94:49:1–49:21. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. https://doi.org/10.4230/LIPIcs.ITCS.2018.49

Consider only the binary-message case.

Simulations should be performed using open-source software with some effort to make them efficient (e.g. coded in Python using Numpy), and the source code should be made publicly available (e.g. on Gitlab) and GPL licensed.

Proof of the Lower Bound on the Number of Messages

$$D(f(X), f(Y)) = \Pr(f(X) = 1) \log \frac{\Pr(f(X)=1)}{\Pr(f(Y)=1)}$$

$$+ \Pr(f(X) = 0) \log \frac{\Pr(f(X) = 0)}{\Pr(f(Y) = 0)}$$

$$(\text{since } P(f(X) = 0), P(f(Y) = 1) \leq \delta \text{ and } \lim_{x \to 0} x \log x = 0)$$

$$\geq \frac{1}{2} \log \frac{\Pr(f(X) = 1)}{\Pr(f(Y) = 1)} \geq \frac{1}{2} \log \frac{1 - \delta}{\delta} \geq \frac{1}{2} \log \frac{1}{2\delta}.$$

By the chain rule of the relative entropy, denoting $X_{<i} = (X_1, ..., X_{i-1})$ and the same for $Y_{<i}$, we have

$$D(X, Y) = \sum_i D\left((X_i | X_{<i}), (Y_i | Y_{<i})\right)$$

$$\leq n \max_i D\left((X_i | X_{<i}), (Y_i | Y_{<i})\right).$$

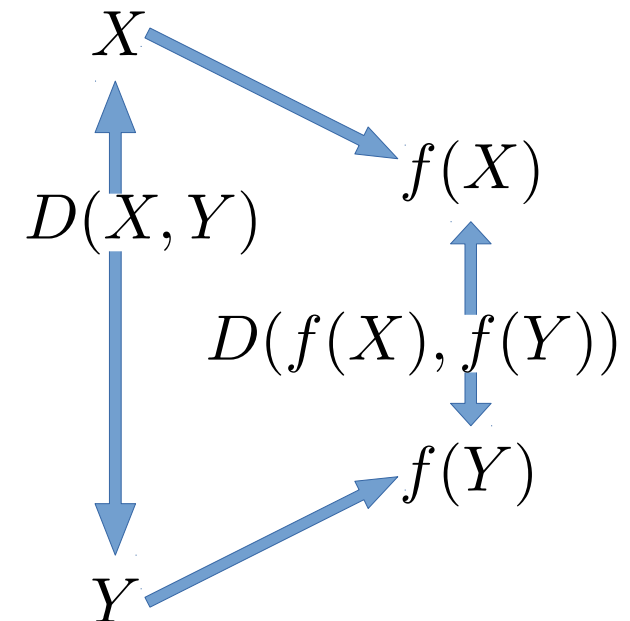Combine the bounds with the information processing inequality.

Information Processing Inequality for Relative Entropy

Let

- X and Y be two random variable on the same space Ω,

- $f : \Omega \to \{0,1\}$ be any binary function.

It holds

$$D\left(X,Y\right) \geq D\left(f(X), f(Y)\right).$$

Proof

Sketch of Proof of the Information Processing Inequality

Notice that

$$P\left(X = z\right) = P\left(X = z, f(X) = f(z)\right)$$
$$= P\left(X = z | f(X) = f(z)\right) P(f(X) = f(z))$$

and the same for Y.

$$D\left(X, Y\right) = \sum_{i \in \{0,1\}} \sum_{z:f(z)=i} P\left(f(X) = i\right) P\left(X = z \mid f(X) = i\right)$$

$$\cdot \log \frac{P\left(f(Y) = i\right) P\left(Y = z \mid f(Y) = i\right)}{P\left(f(X) = i\right) P\left(X = z \mid f(X) = i\right)}$$

$$= \sum_{i \in \{0,1\}} P\left(f(X) = i\right) \log \frac{P\left(f(Y) = i\right)}{P\left(f(X) = i\right)}$$

$$+ \sum_{i \in \{0,1\}} P\left(f(X) = i\right) D\left(\left(X \mid f(X) = i\right), \left(Y \mid f(Y) = i\right)\right)$$
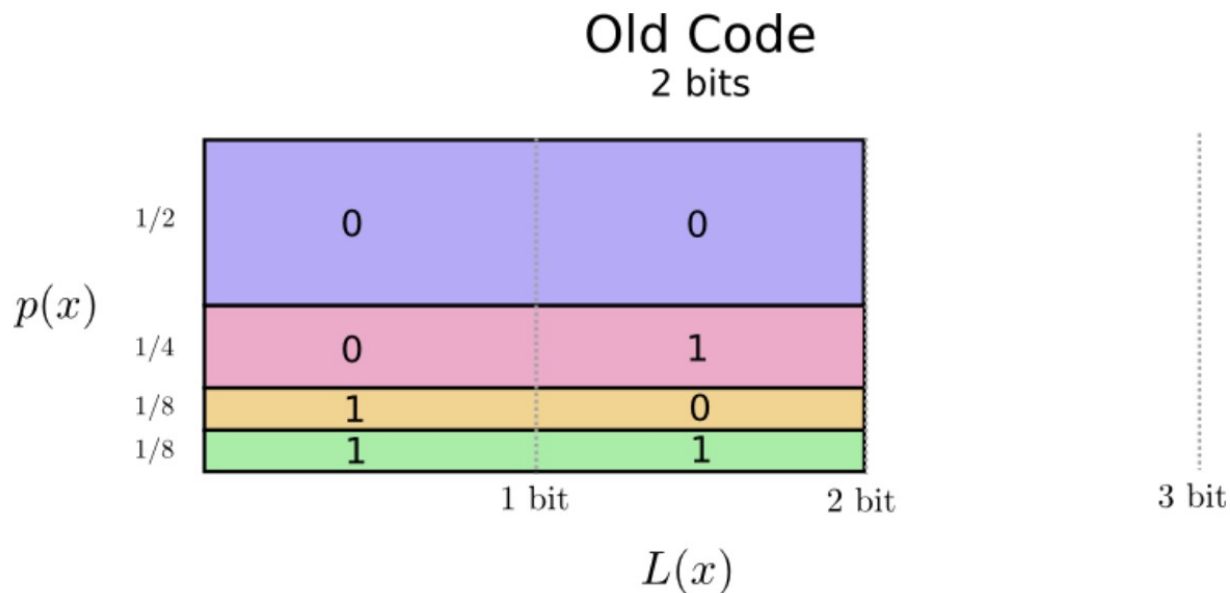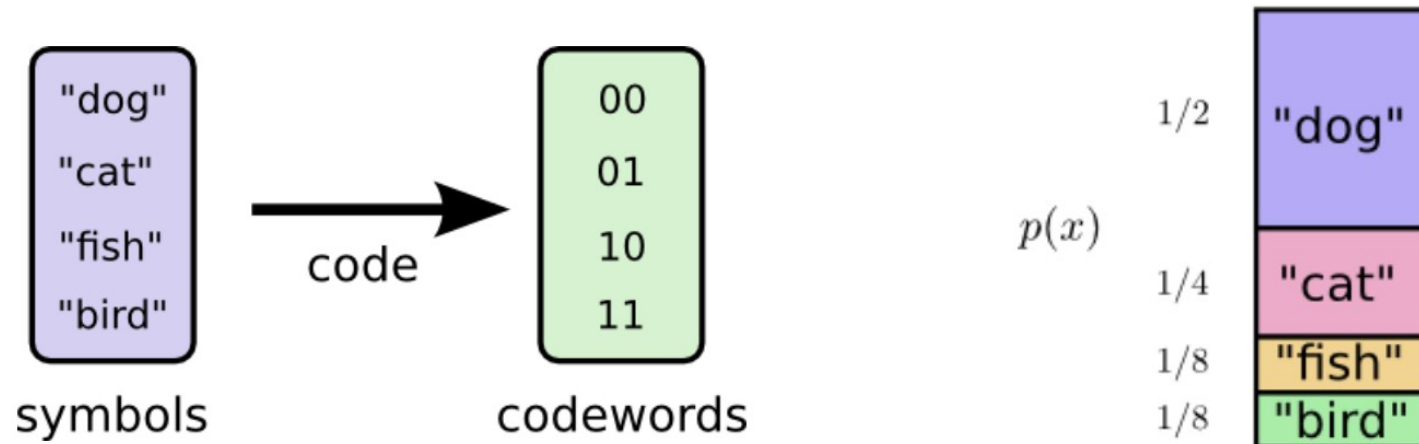
$$\geq D\left(f\left(X\right), f\left(Y\right)\right)$$

Back to Relative Entropy

Chain Rule for Relative Entropy

$$D\left((X, X'), (Y, Y')\right) = \sum_{x, x'} P\left(X = x, X' = x'\right) \log \frac{P\left(X = x, X' = x'\right)}{P\left(Y = y, Y' = y'\right)}$$

$$= \sum_{x, x'} P\left(X = x\right) P\left(X' = x' | X = x\right) \log \frac{P\left(X' = x' | X = x\right) P\left(X = x\right)}{P\left(Y' = x' | Y = x\right) P\left(Y = x\right)}$$

$$= \sum_{x, x'} P\left(X = x\right) P\left(X' = x' | X = x\right) \log \frac{P\left(X = x\right)}{P\left(Y = x\right)}$$

$$\qquad + \sum_{x, x'} P\left(X = x\right) P\left(X' = x' | X = x\right) \log \frac{P\left(X' = x' | X = x\right)}{P\left(Y' = x' | Y = x\right)}$$

$$= \sum_{x} P\left(X = x\right) \log \frac{P\left(X = x\right)}{P\left(Y = x\right)}$$

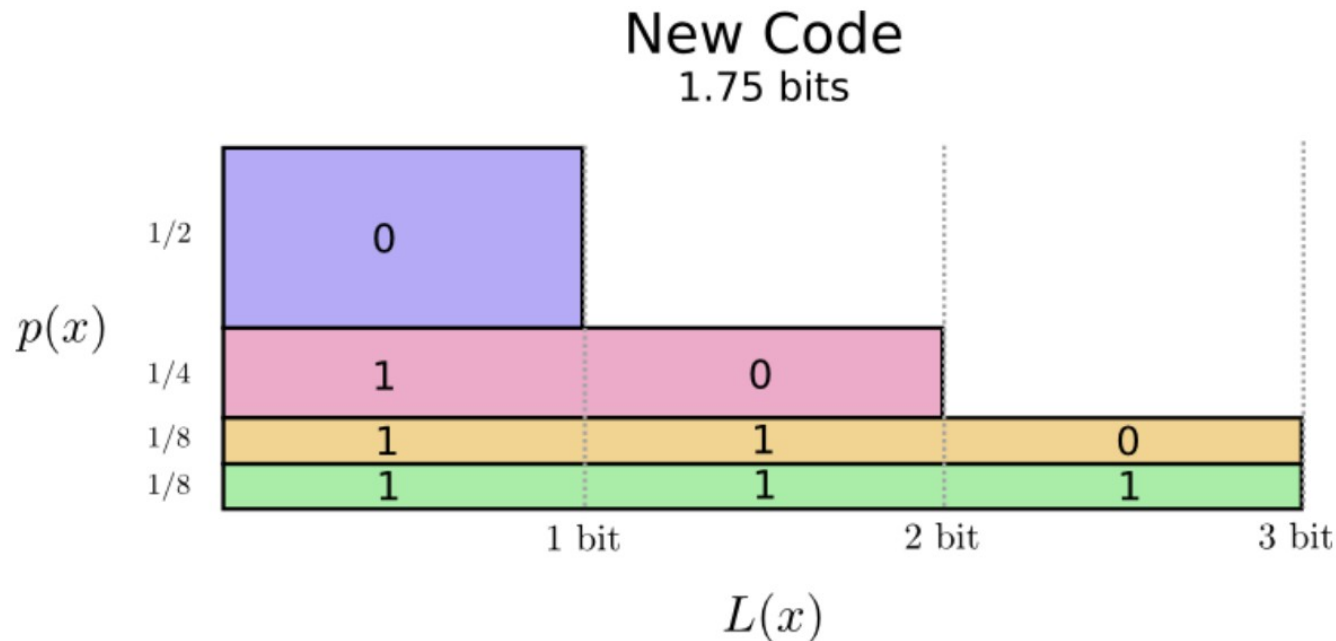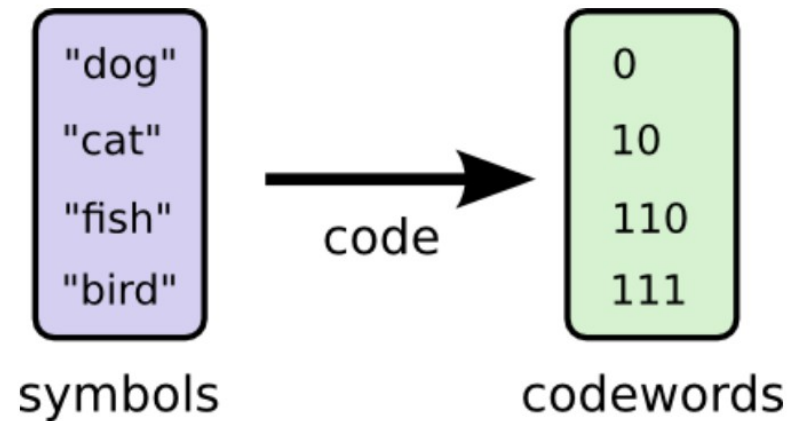$$\qquad + \sum_{x} P\left(X = x\right) D\left((X' | X = x), (Y' | Y = x)\right)$$

Visual Information Theory - Crash Course

Based on the Colah's essay Visual Information Theory
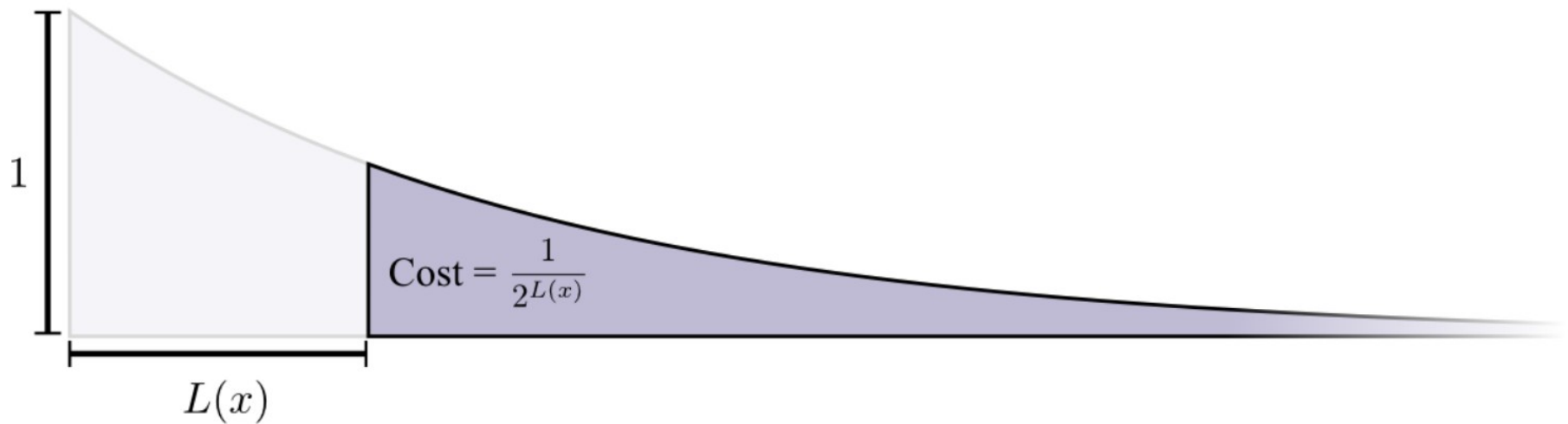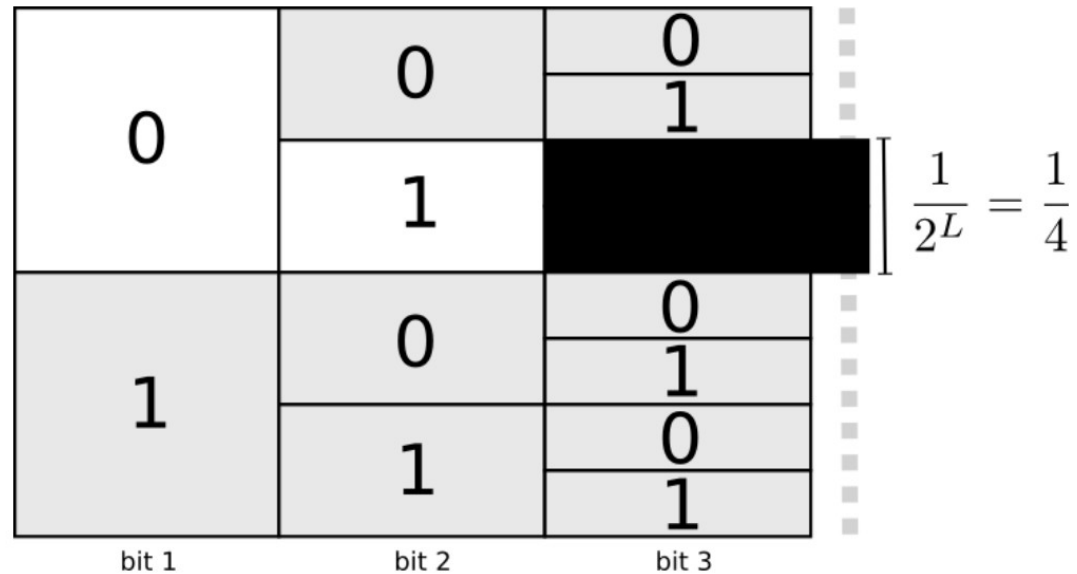(all following images are from there)

Minimizing Average Length
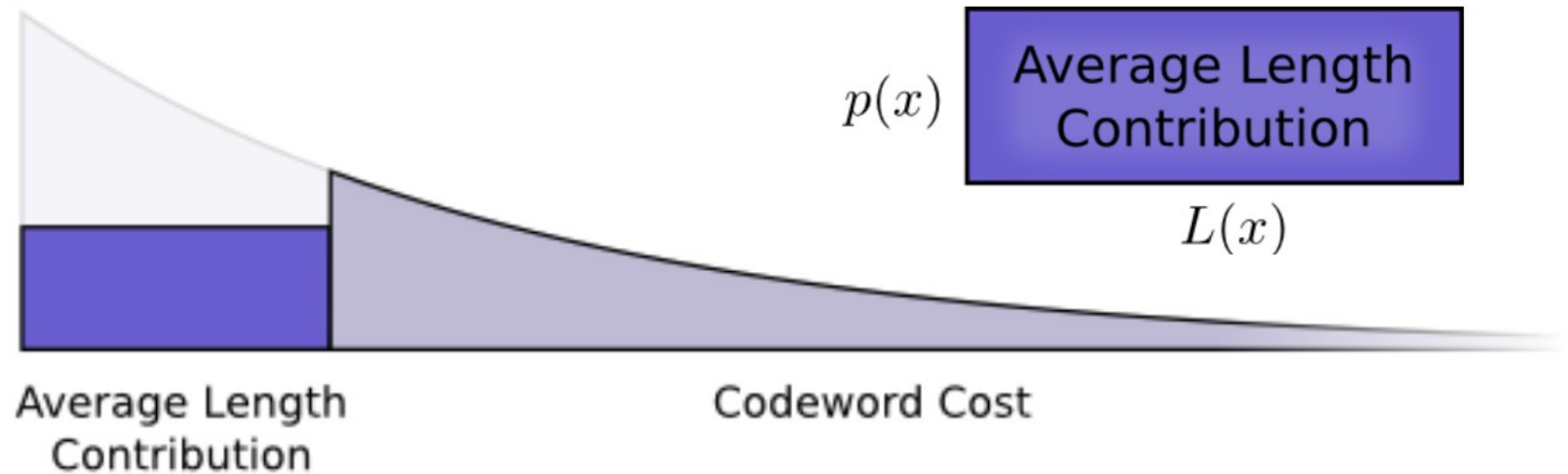
NDA - E. Natale (CNRS, UCA) - U. of Rome "Tor Vergata"

Prefix-free Codes

$$\frac{1}{2^L} = \frac{1}{4}$$

bit 1   bit 2   bit 3

$$\text{Cost} = \frac{1}{2^{L(x)}}$$

1

$L(x)$
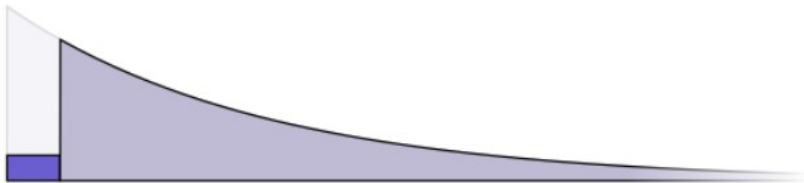
Average Length vs Cost
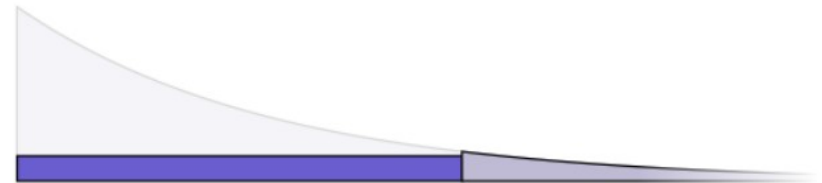
Information Content gives Optimal Length

$$p(x) = 2^{-L(x)} \iff L(x) = \log \frac{1}{p(x)}$$

Proving Optimality

$$\frac{p(a)}{p(a)+\epsilon} < 1$$

$$p(a)+\epsilon$$

$$p(b)-\epsilon$$

$$\frac{p(b)}{p(b)-\epsilon} > 1$$

$$H(p) = \sum_{x} p(x) \log_2 \left(\frac{1}{p(x)} \right)$$
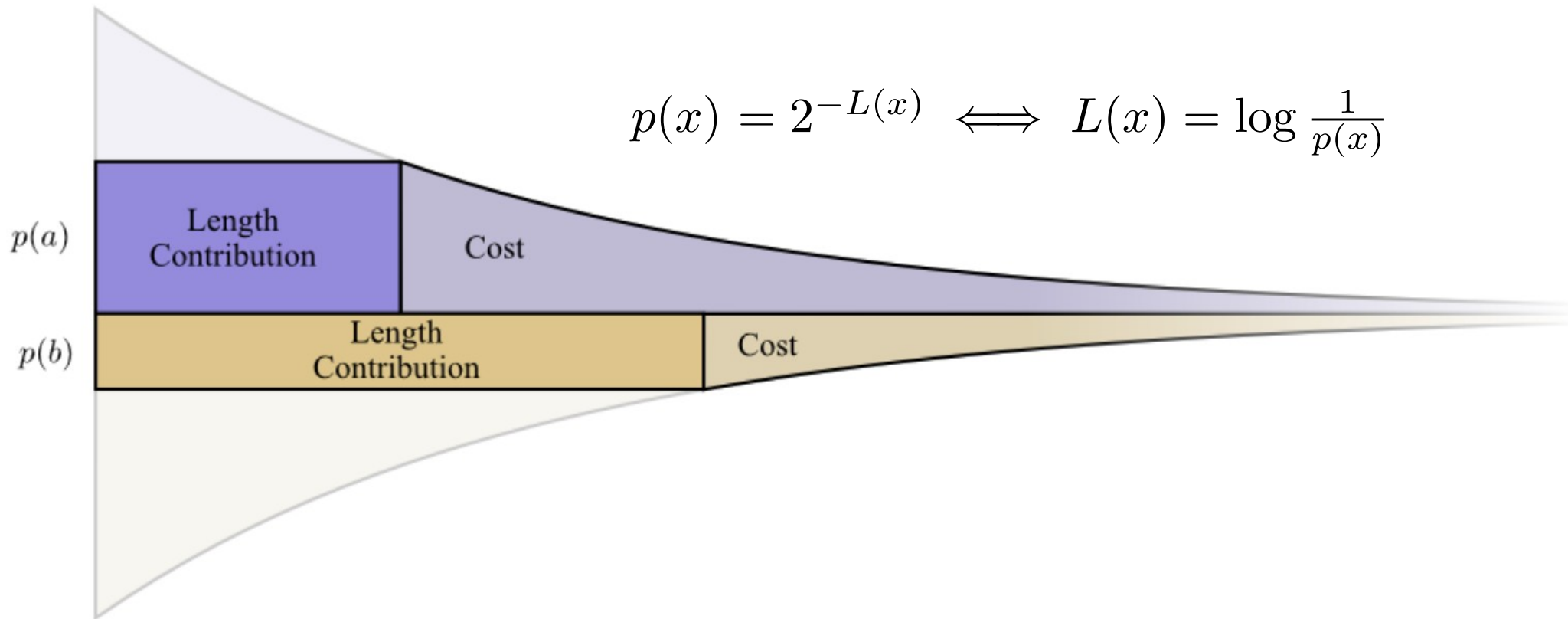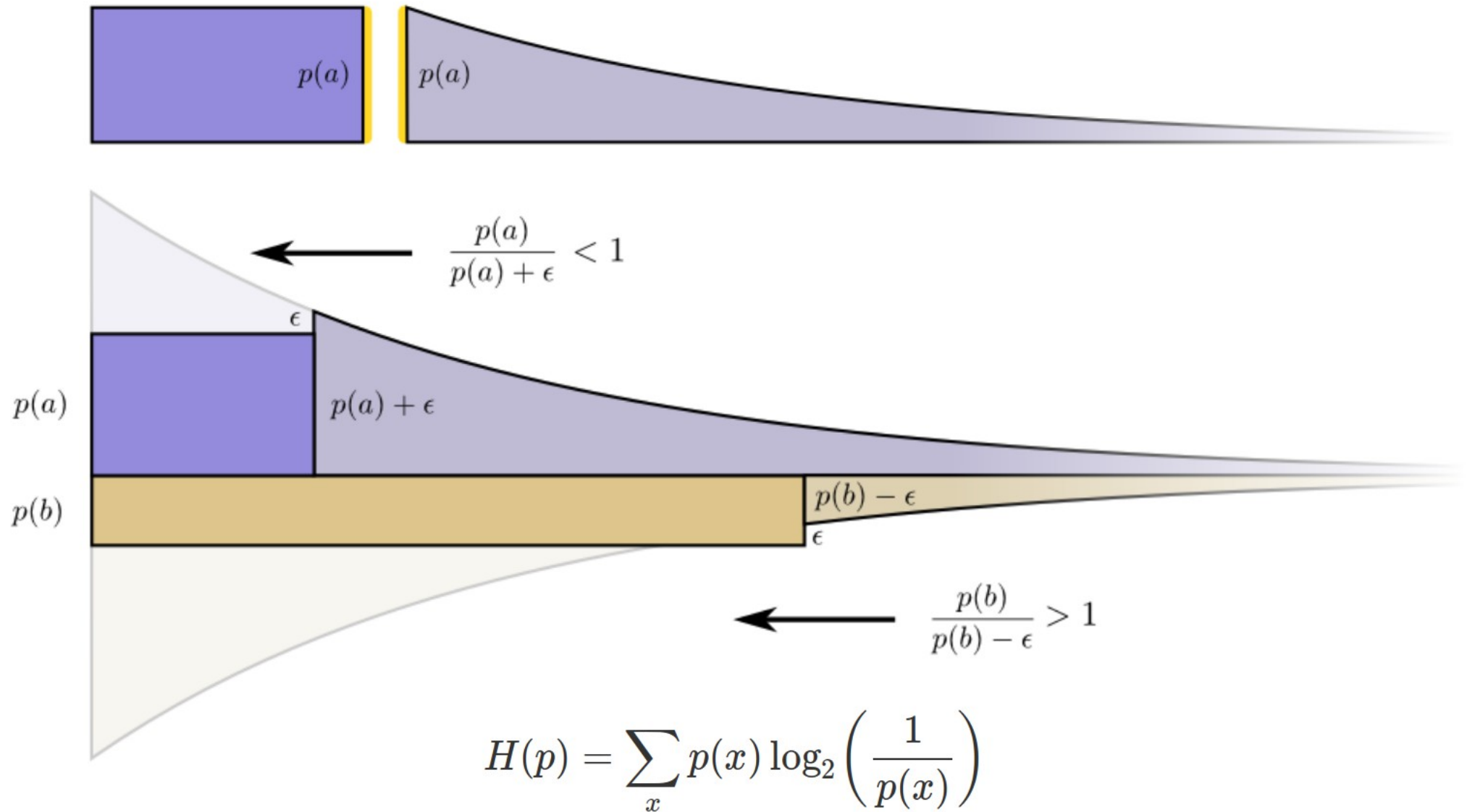
Cross and Relative Entropy

$p(x)$

1/2	"dog"
1/4	"cat"
1/8	"fish"
1/8	"bird"

Dog Lover's
Word Frequency

$q(x)$

1/8	"dog"
1/2	"cat"
1/4	"fish"
1/8	"bird"

Cat Lover's
Word Frequency

$H(p)$

$H_q(p)$

$D_q(p)$
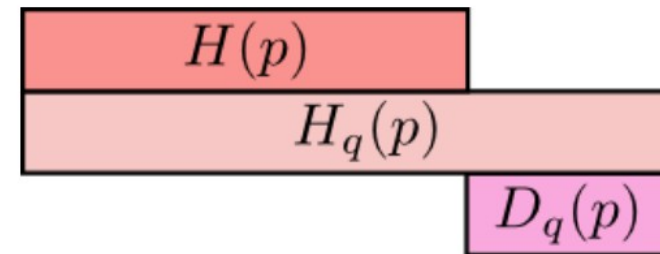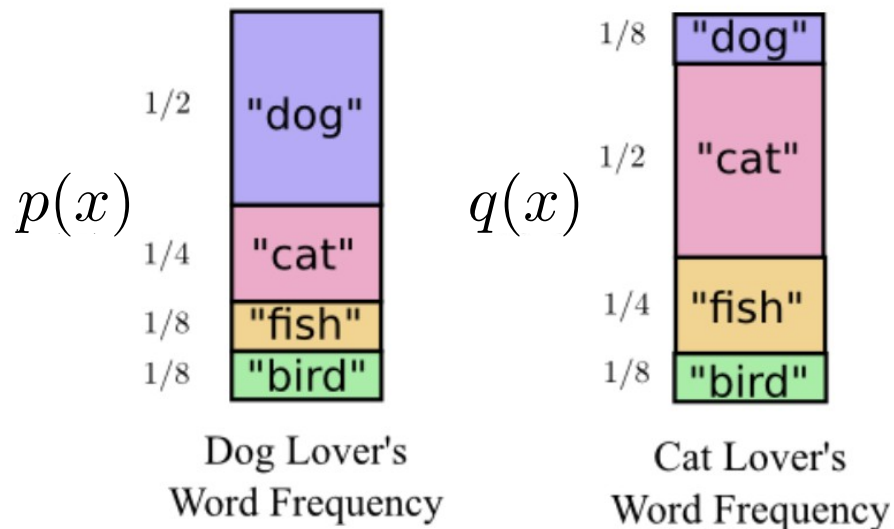
Cross Entropy: Average Length of message from $q(x)$ using code for $p(x)$

$$H_p(q) = \sum_x q(x) \log_2 \left(\frac{1}{p(x)} \right)$$

Relative Entropy: Average **Extra** Length of message from $q(x)$ using code for $p(x)$

$$D(q, p) = \sum q(x) \log \frac{q(x)}{p(x)}$$

$$= H_p(q) - H(q)$$