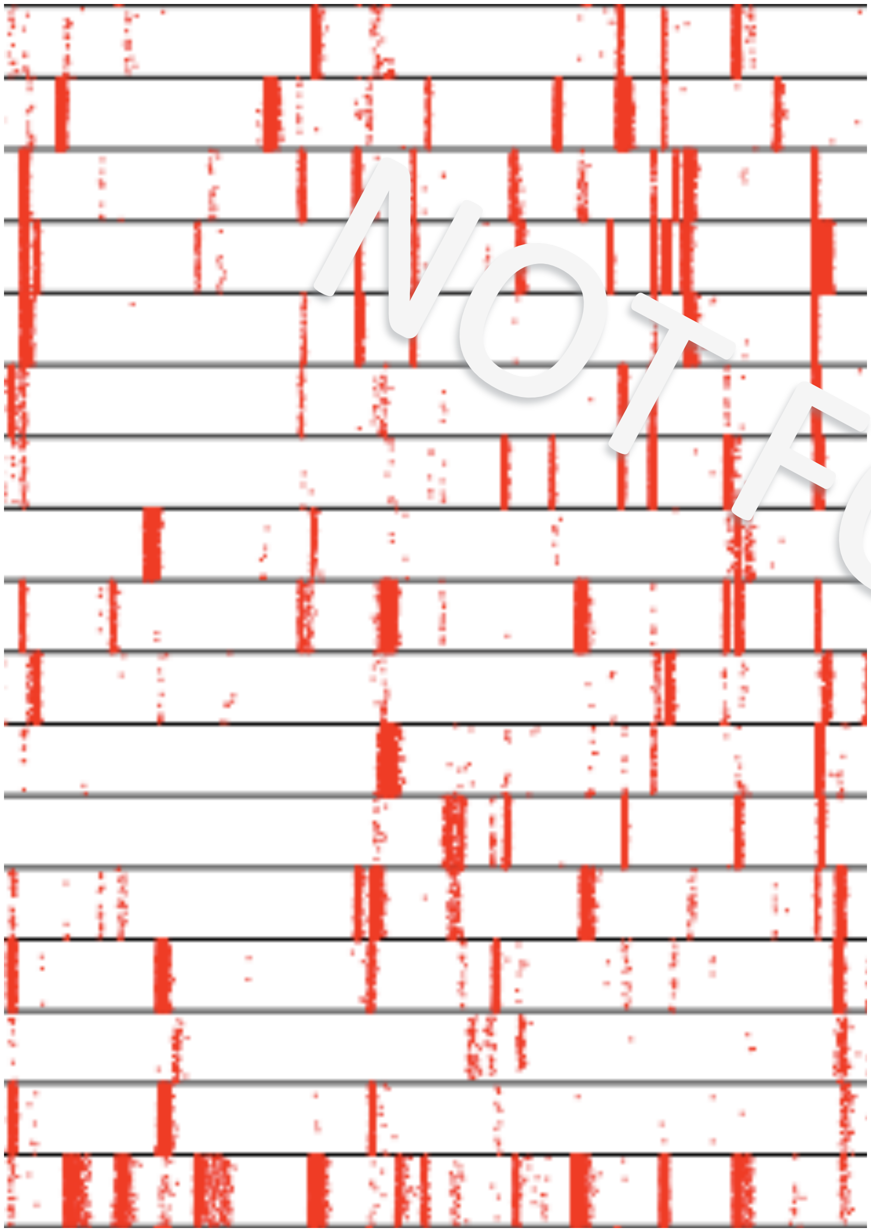
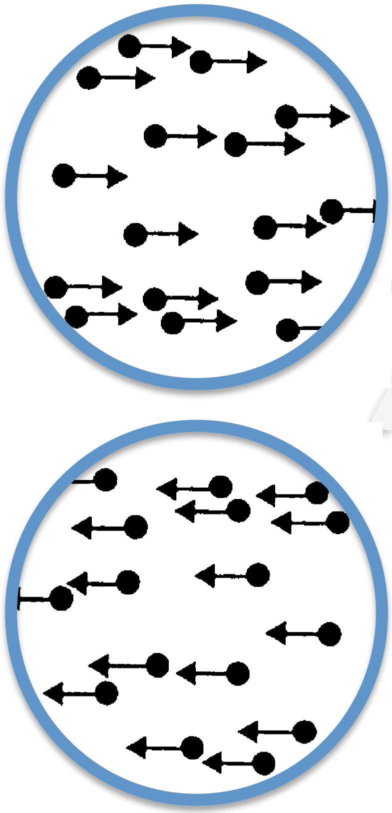
Information and coding principles

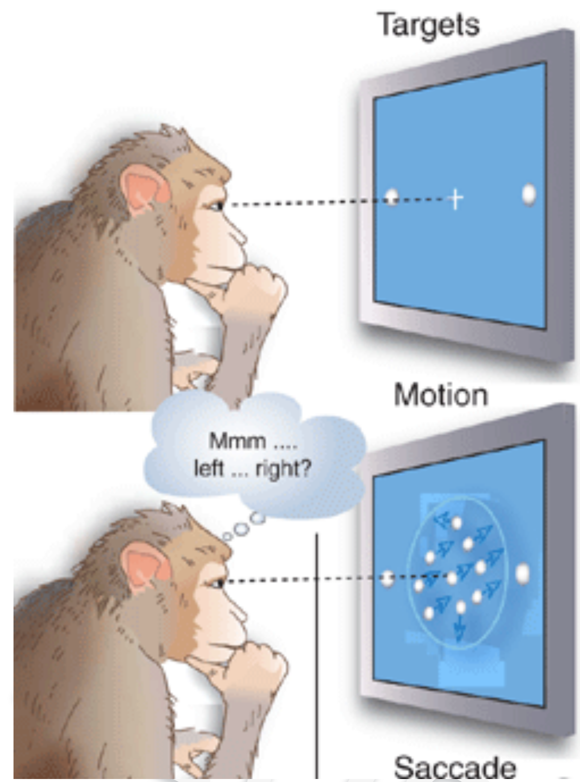


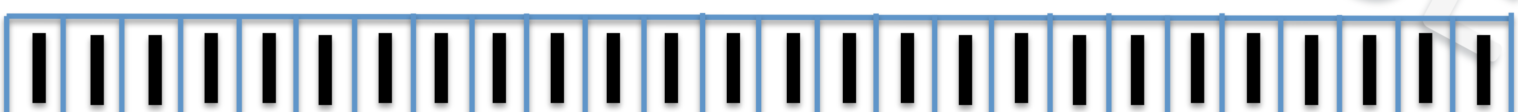
Information and coding principles

- Defining entropy and information
- Computing information for neural spike trains
- What can information tell us about coding?

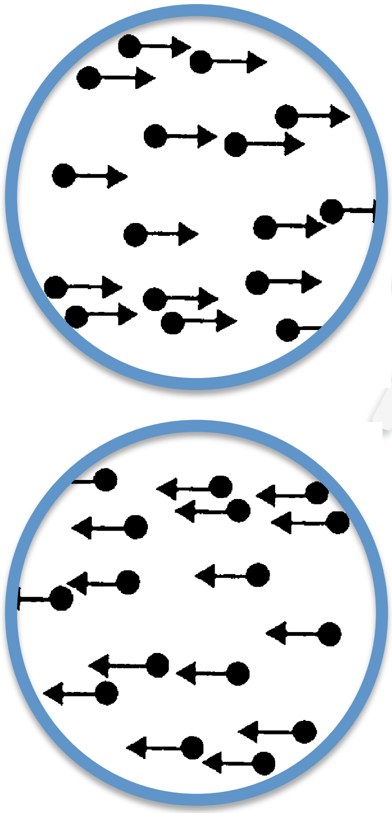
How good is my code?

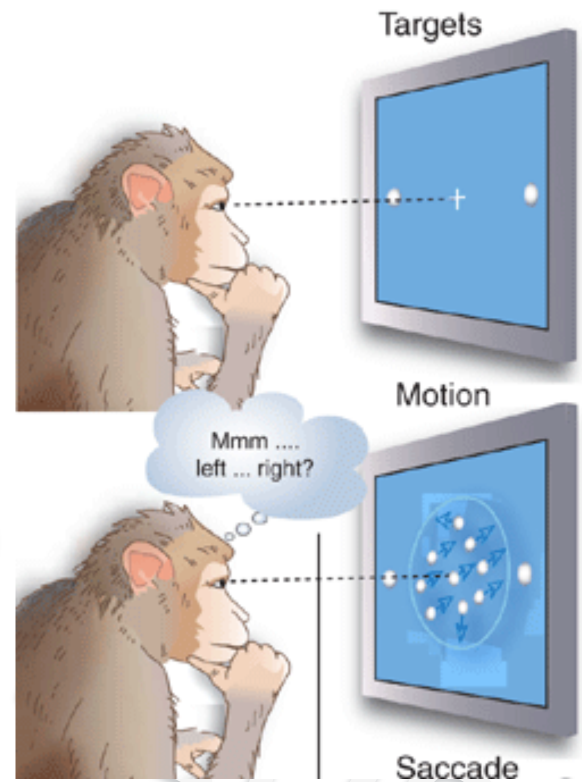


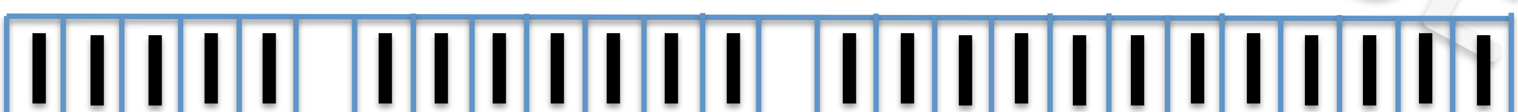




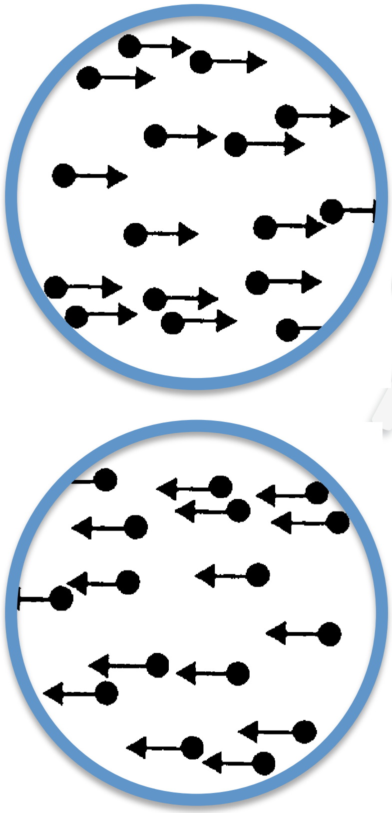
How good is my code?

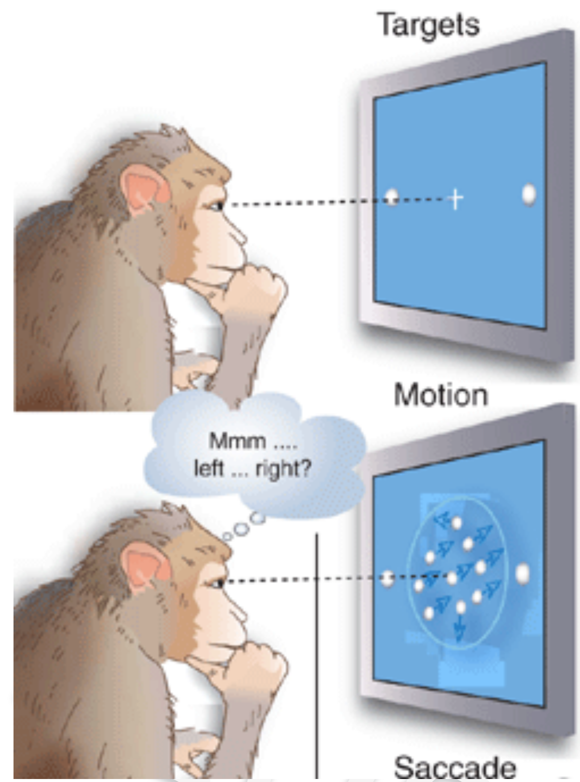


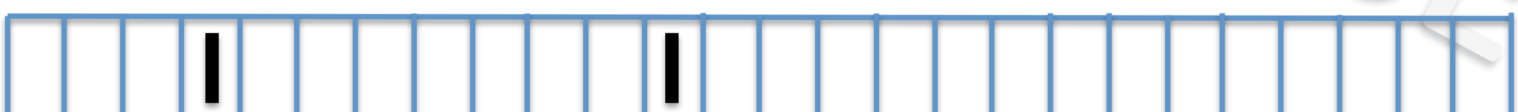




How good is my code?

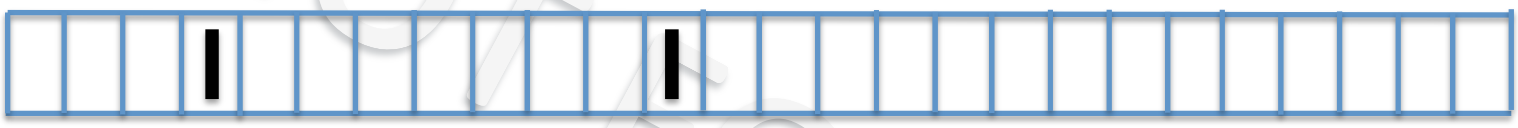






Surprise!

Information



$$P(1) = p$$

$$P(0) = 1 - p$$

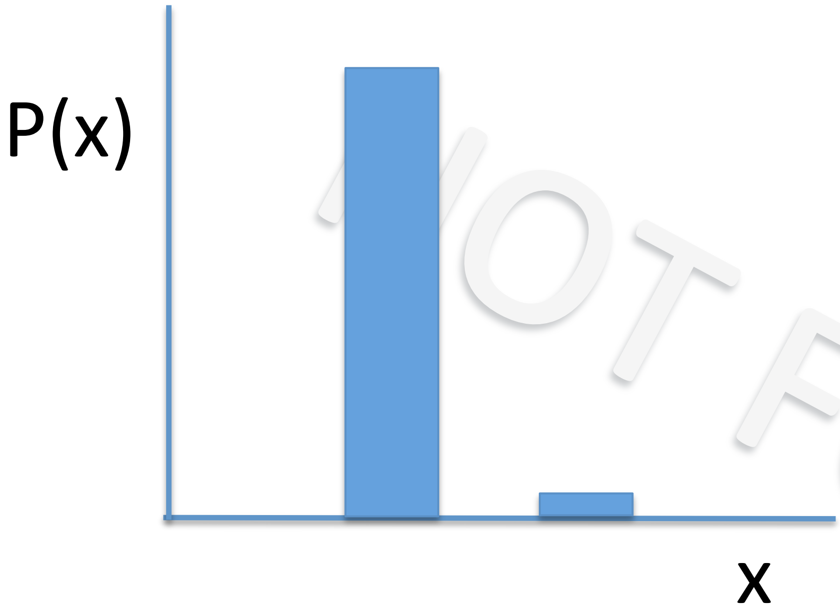
$$\text{information}(1) = -\log_2 p$$

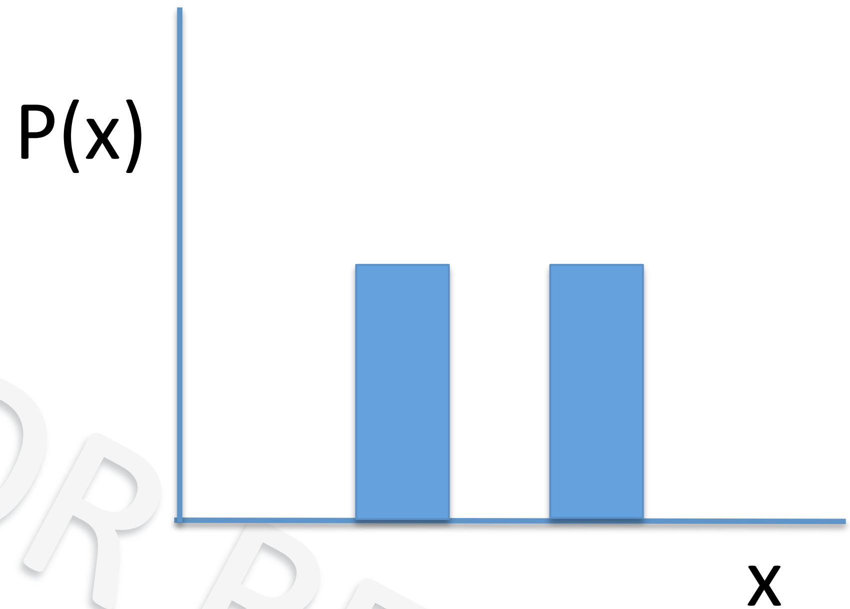
$$\text{information}(0) = -\log_2 (1-p)$$

Information

Each *bit* of information specifies location by an additional factor of 2.

What is entropy?





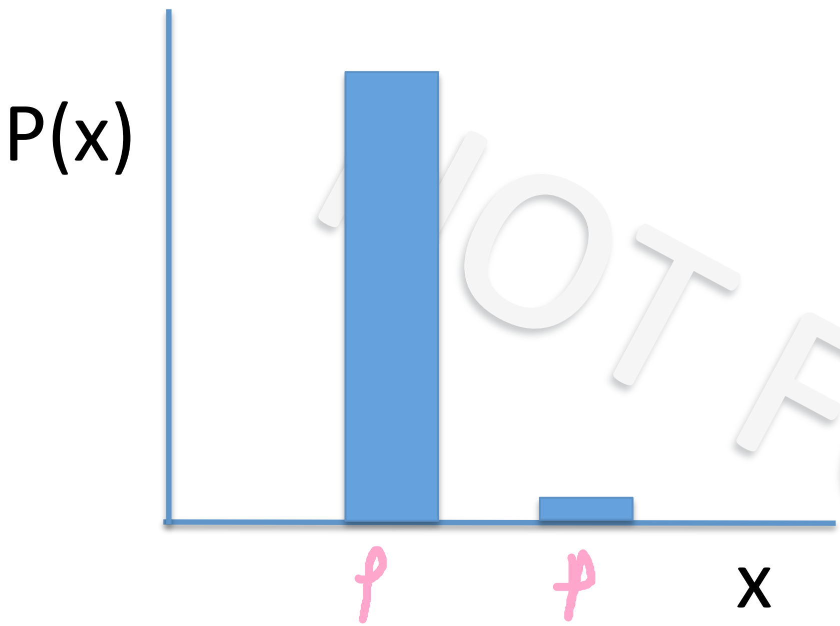
Entropy $= \text{average information}$

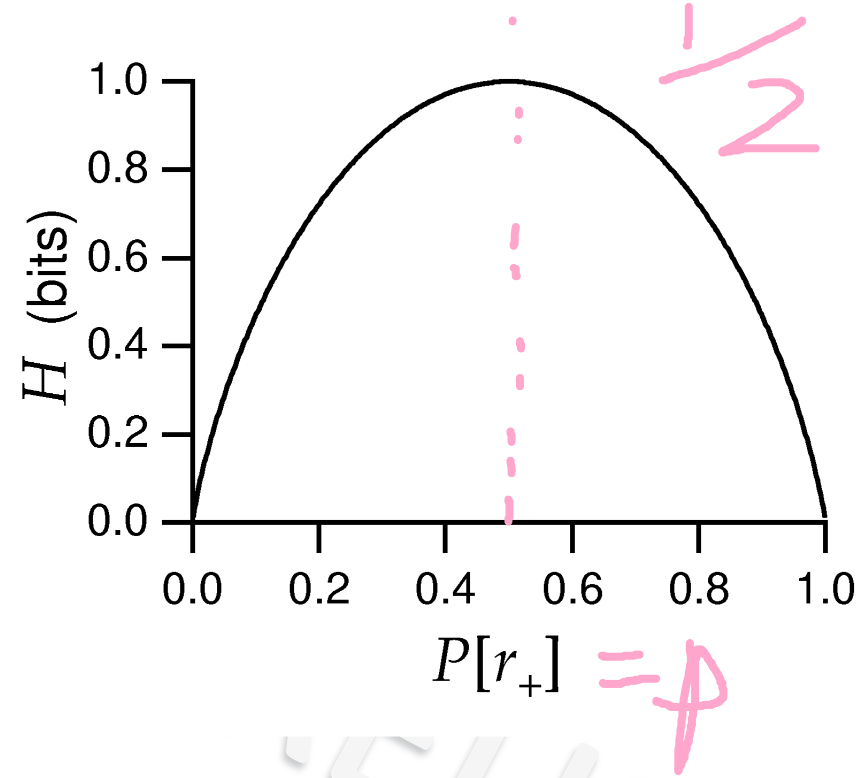
$$= - \sum_i p_i \log_2 p_i$$

$$= - \int dx\, p(x) \log_2 p(x)$$

Units are *bits*

Maximizing the entropy

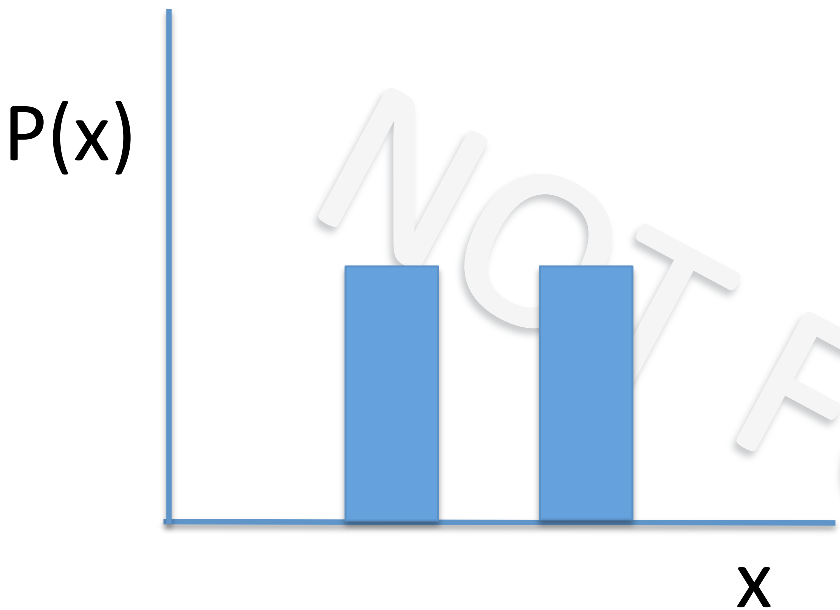


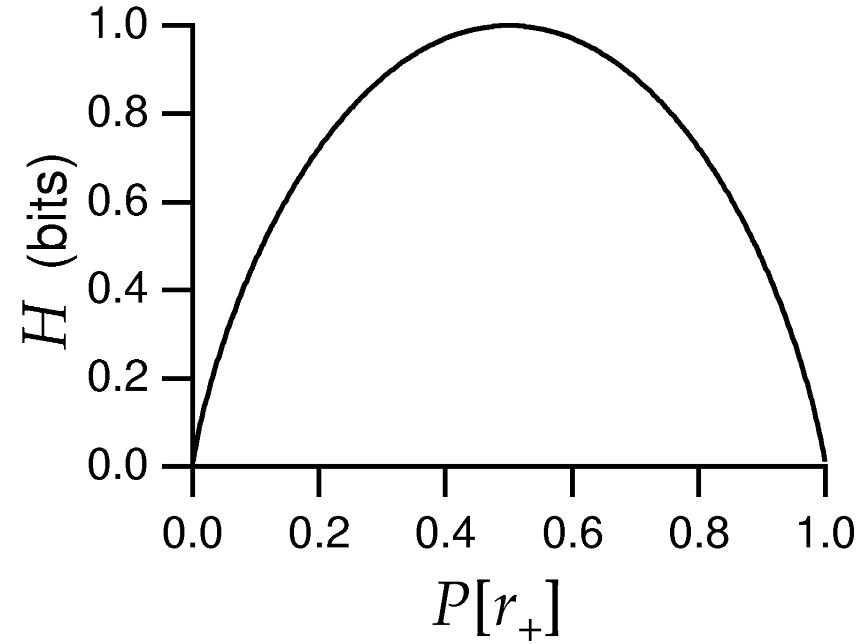


Entropy = $- \Sigma\, p_i \log_2 p_i$

$= -p \log p - (1-p) \log (1-p)$

Maximizing the entropy



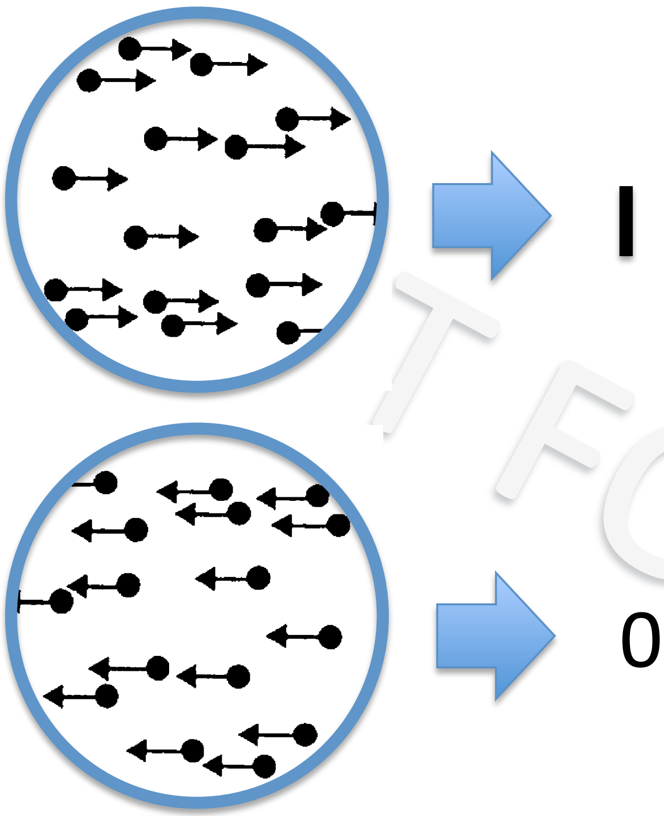


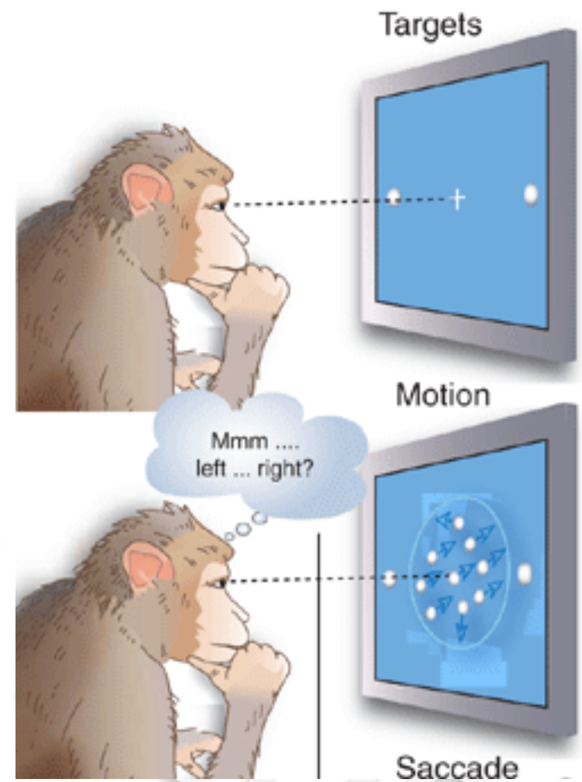
Entropy = - $\Sigma\ p_i \log_2 p_i$

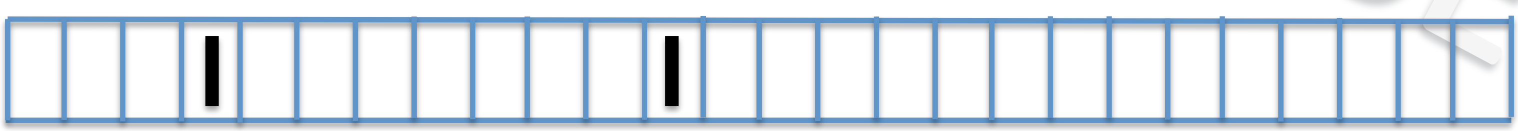
Maximum entropy squash

Generally P(x) is not uniform… but it would be best for you if it were!

Back to our spike code: how about the stimulus?



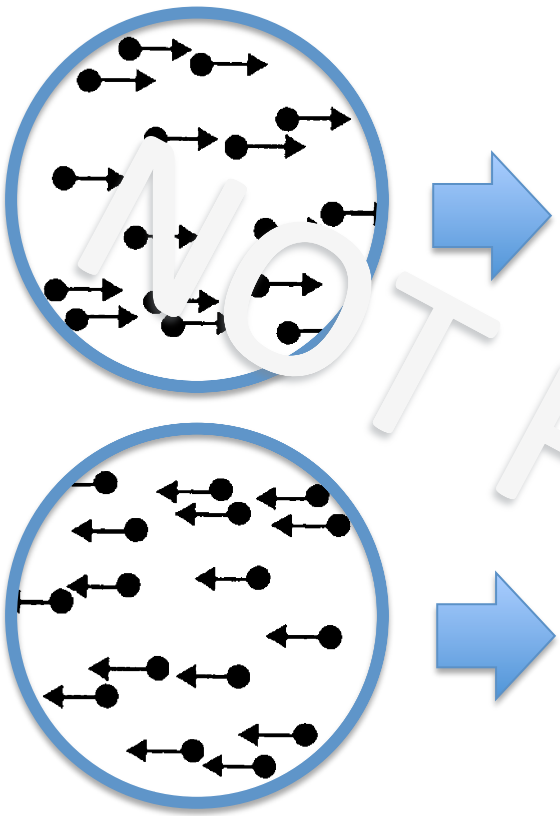




How about the stimulus?

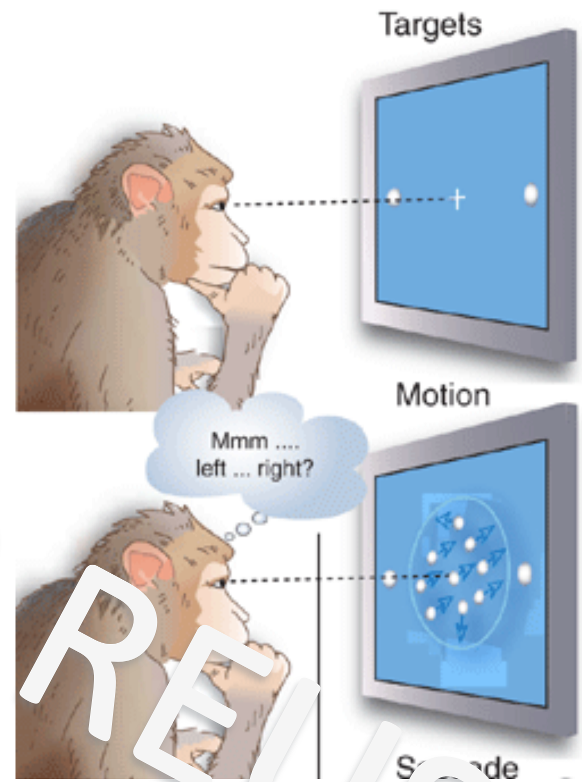
S = +

S = -

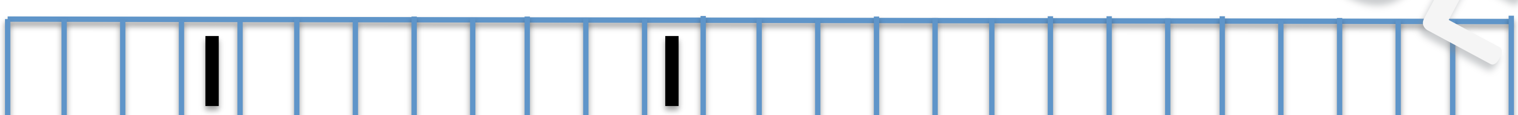


I

0



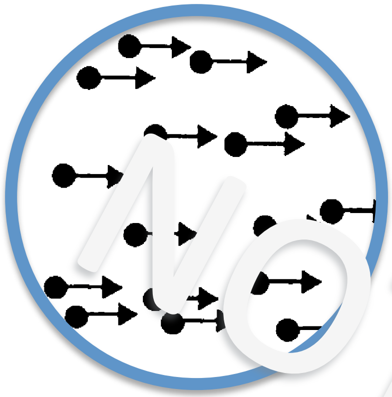
r



s - - - + - - - - - - - + - - - - - - - - - - - - - -

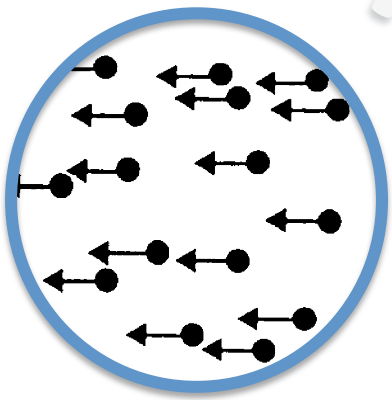
How about the stimulus?

S = +

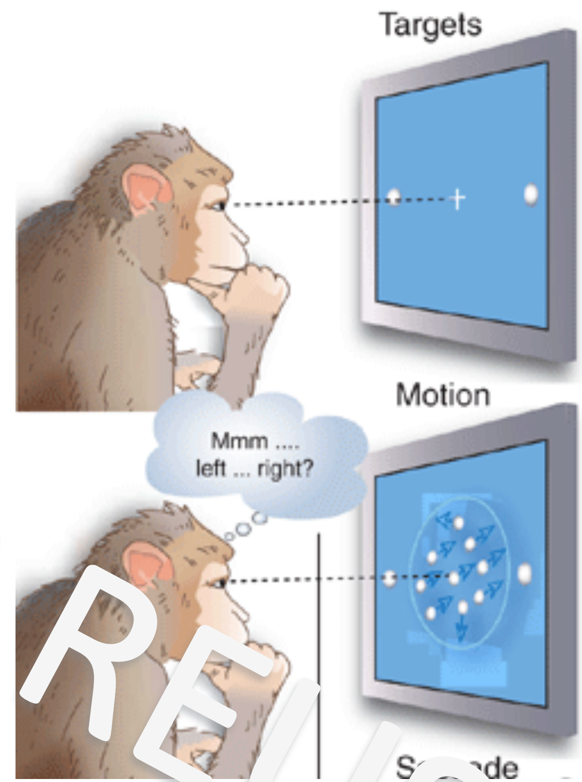


|

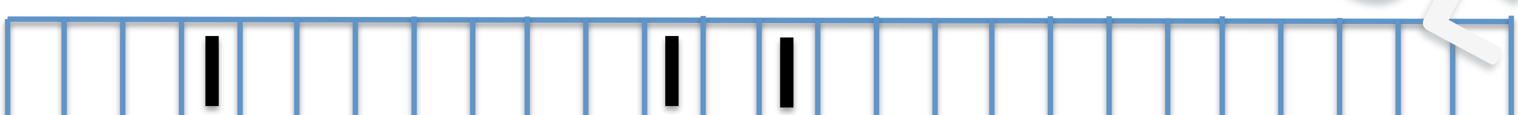
S = -



0



r

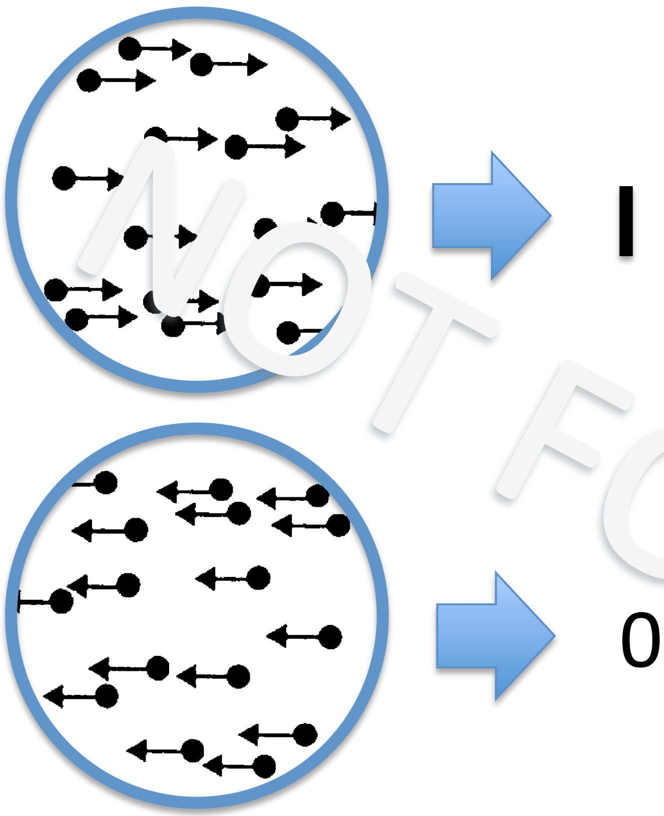


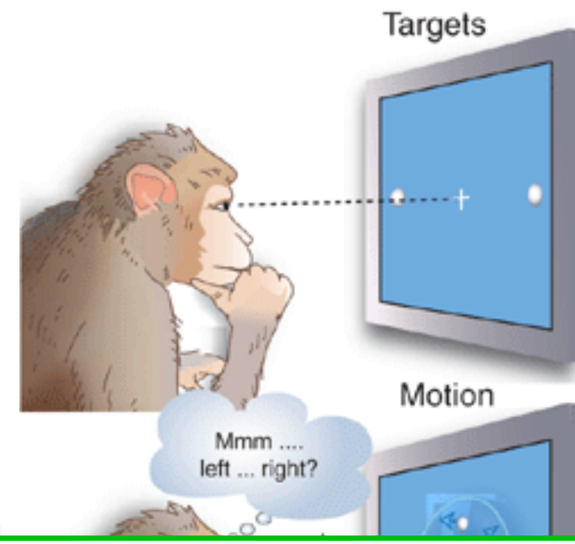
s - - - + - - - - - - - + - - - - + - - - - - - - - -

How about the stimulus?

S = +

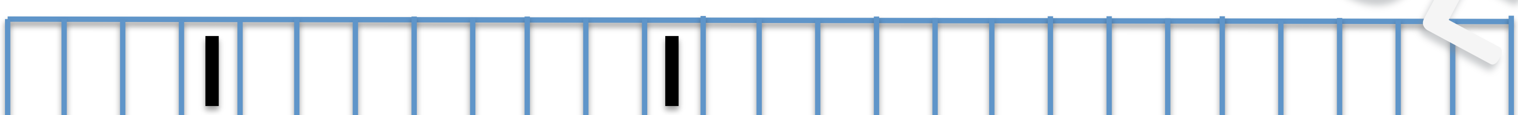
S = -





What if the problem were not so much that our code were noisy, but that we haven't exactly understood what the code is doing.
如果不是因为我们探测的刺激有太多噪音，而是我们不理解他们怎么编码的呢？

That is, maybe there's some temporal sequencing S that should be more appropriately thought about as the true stimulus.

r 

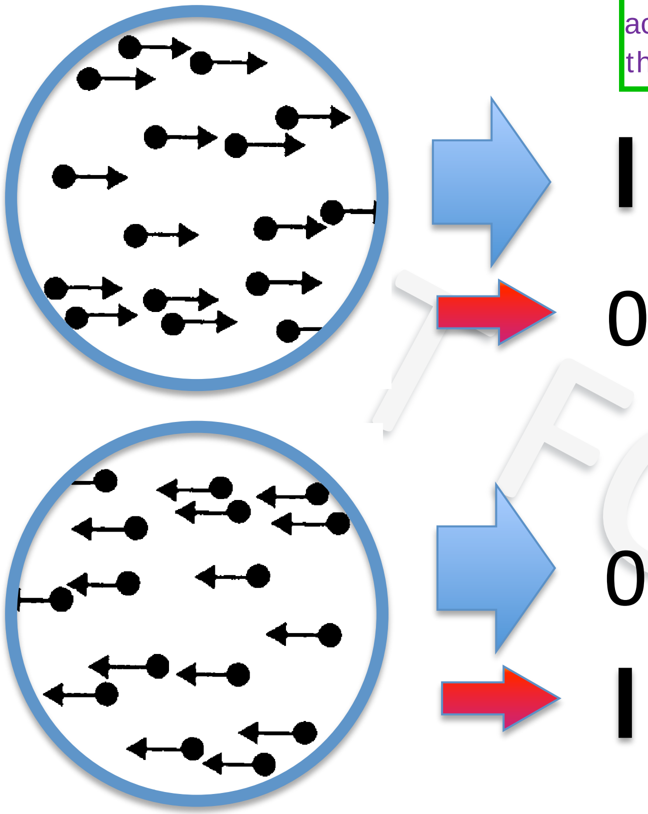
s - + - - - + - - - + - + + - + - + - - - + - - - + -

How much of the variability in *r* is encoding *s*?

How much of the variability that we see here in R is actually used for encoding S? We need to incorporate the possibility for error.

S = +

S = -

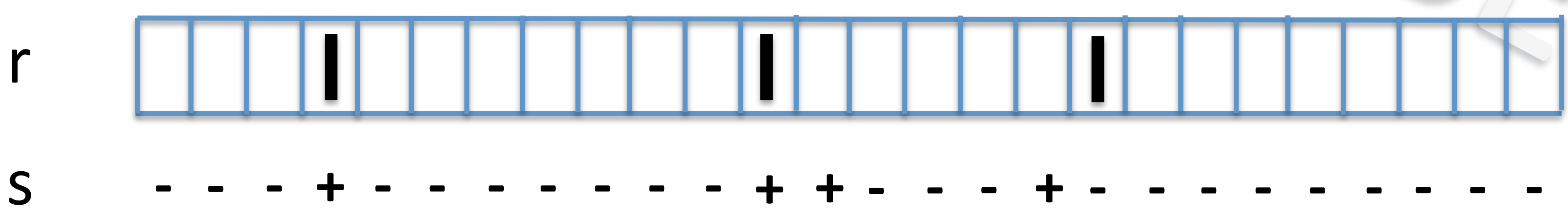


$P(r_+|+) = 1\text{-}q$

$P(r_-|+) = q$

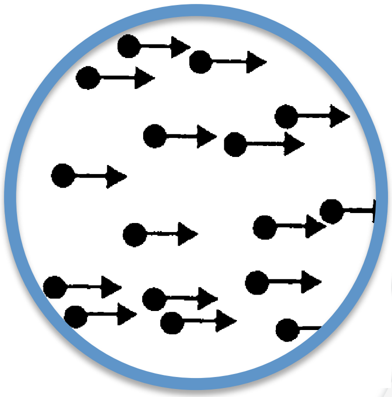
$P(r_-|-) = q$

$P(r_+|-) = 1\text{-}q$



How much of the variability in *r* is encoding *s*?

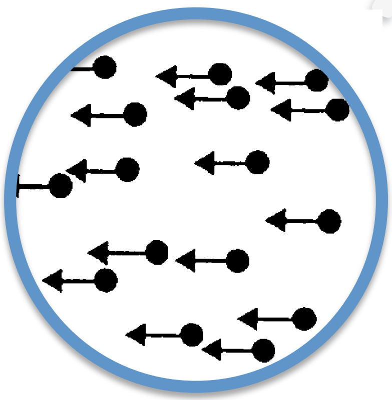
S = +



1 $P(r_+|+) = 1\text{-}q$

0 $P(r_-|+) = q$

S = -



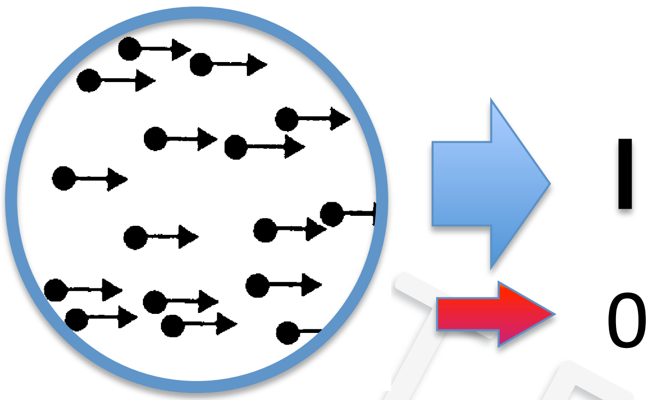
0 $P(r_-|-) = q$

1 $P(r_+|-) = 1\text{-}q$

$$H = -\sum_{i=1}^{2} p_i \log_2 p_i$$

How much of the variability in *r* is encoding *s*?

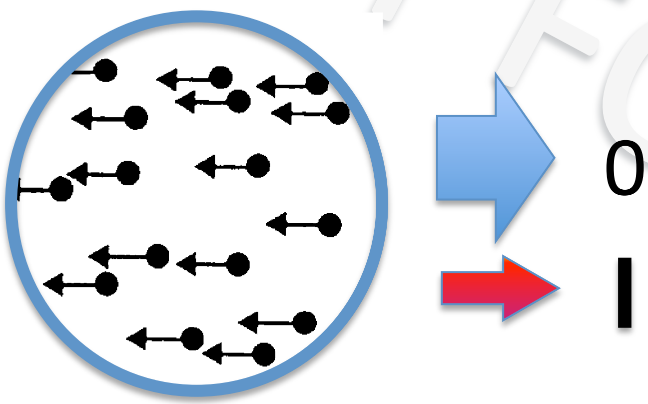
S = +



$P(r_+|+) = 1\text{-}q$

$P(r_-|+) = q$

S = -



$P(r_-|-) = q$

$P(r_+|-) = 1\text{-}q$

本来spike分布的entropy

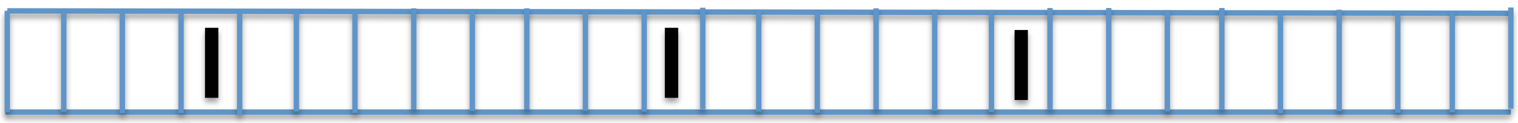
Total entropy: $H[R] = -\, P(r_+) \log P(r_+) - P(r_-) \log P(r_-)$

Noise entropy: $H[R|+] = -\, q \log q - (1\text{-}q) \log (1\text{-}q)$

当你知道stimulus是什么的时候，spike分布的entropy

Mutual information

r



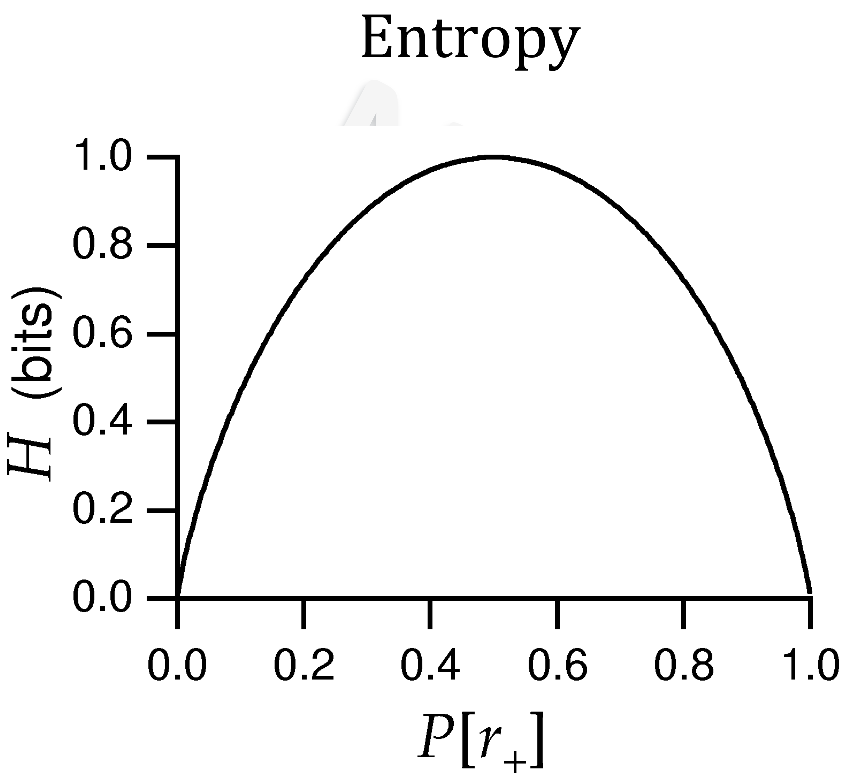
s - - - + - - - - - - - + + - - - + - - - - - - - - -

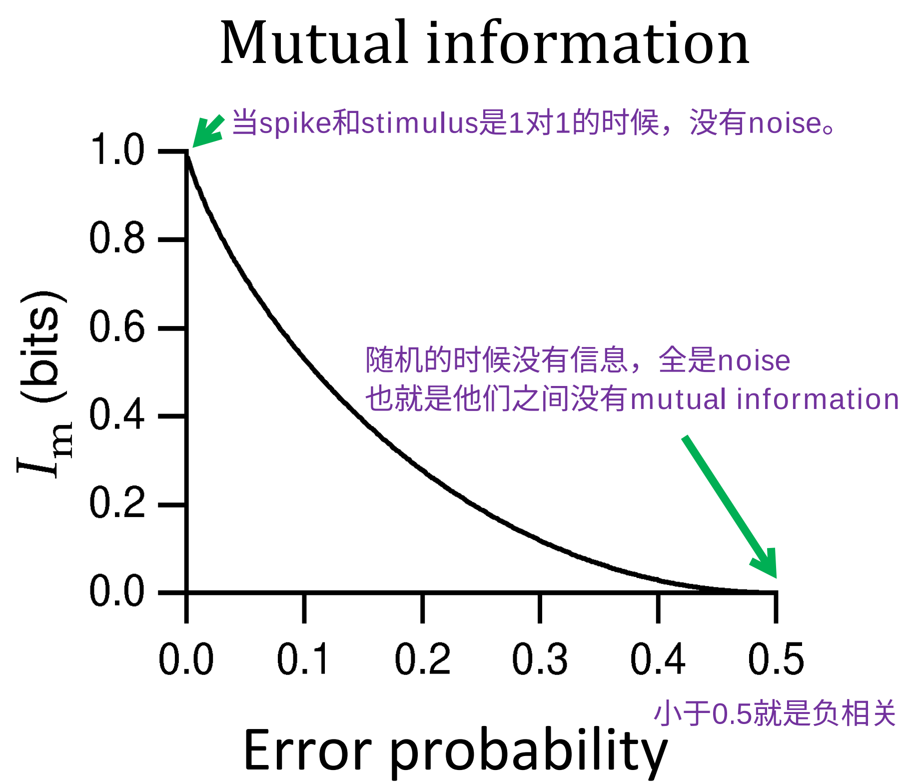
The amount of entropy that is used in coding the stimulus

$$\boxed{\text{MI(S,R) = Total entropy – average noise entropy}}$$

$$MI = - \sum_r p(r) \log_2 p(r) - \sum_s p(s) \left[- \sum_r p(r|s) \log_2 p(r|s)\right]$$

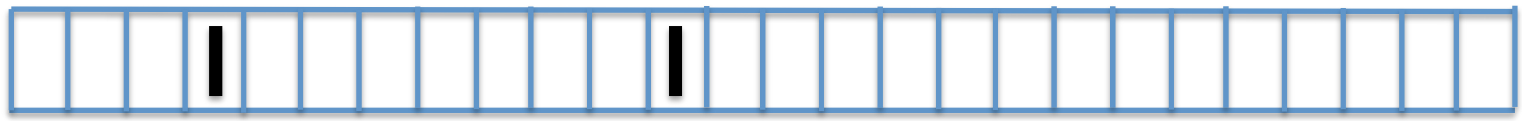
Entropy and information





Mutual information you can calculate in your head

1. r

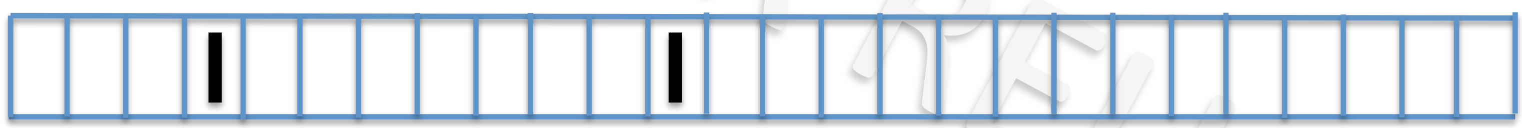


s - + - - - + - - - + - + + - + - + - - - + - - - + -

Response is unrelated to stimulus

- What is p(r|s)? P(r)
- What is the MI ? 0

2. r

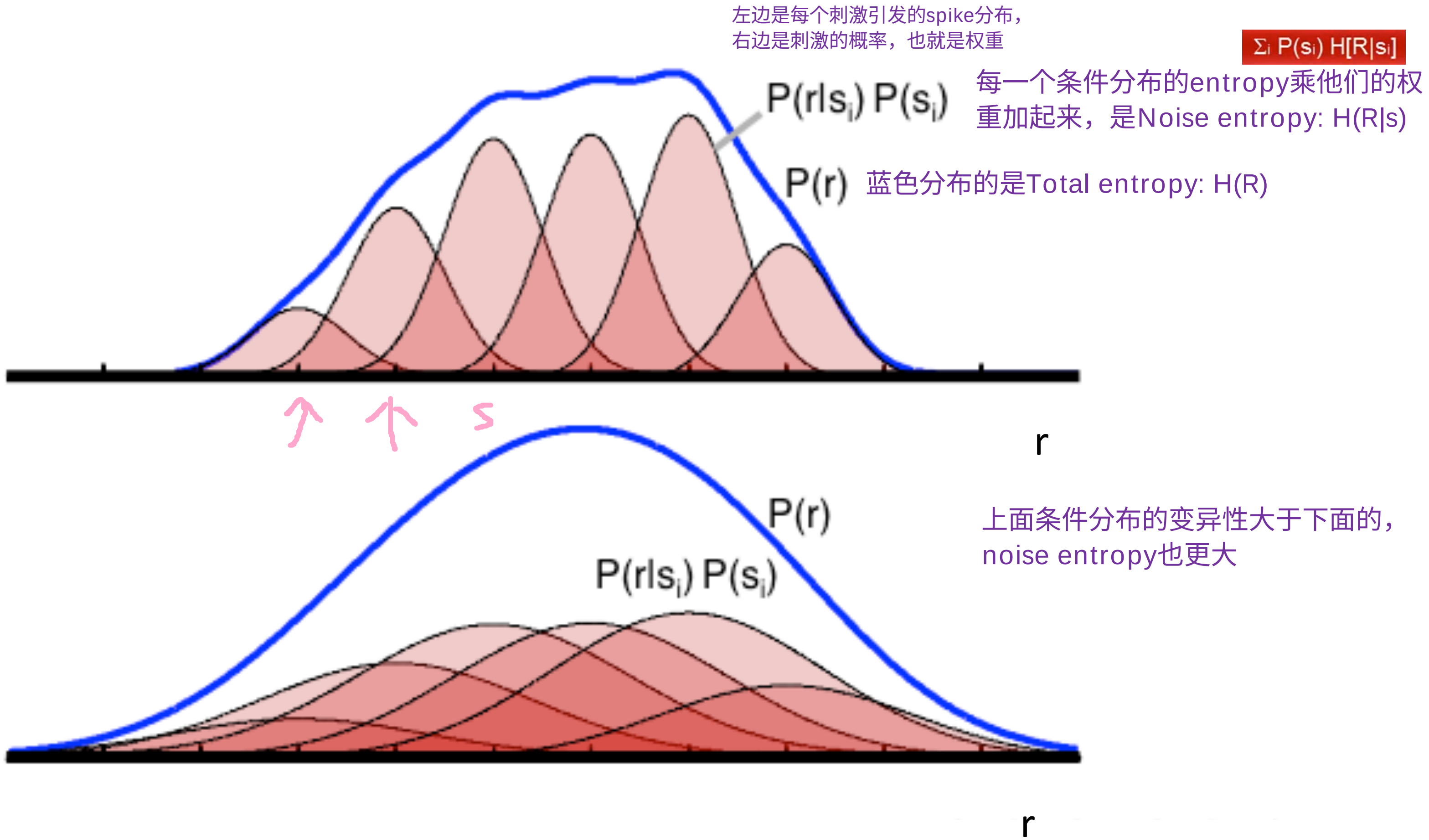


s - - - + - - - - - - - + - - - - - - - - - - - - - -

Response is perfectly predicted by stimulus

MI = all Entropy

Entropy and information in continuous variables



Fairhall, Shea-Brown, Barreiro (2012)

Mutual information measures relationships

$$I(R,S) = -\sum_r p(r)\log_2 p(r) - \sum_s p(s)\left[\sum_r p(r|s)\log_2 p(r|s)\right]$$

$$= H[R] - \sum_s p(s)\, H[R|s]$$

Information quantifies how *independent* R and S are:

$$I(S;R) = D_{KL}\left[P(R,S), P(R)P(S)\right]$$

测量R和S的联合分布的概率，和假设他们独立时各自的概率相乘的值相似度

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x\in\mathcal{X}} P(x)\log\left(\frac{P(x)}{Q(x)}\right).$$

Mutual information

Information quantifies how *independent* R and S are:

$$I(S,R) = D_{KL}\,[P(R,S),\, P(R)P(S)]$$

$$\int ds\,dr\; P(s,r) \log \frac{P(r,s)}{P(r)P(s)} = \int ds\,dr\; P(s,r) \log \frac{P(r|s)\cancel{P(s)}}{P(r)\cancel{P(s)}}$$

$$= \int ds\,dr\; P(s,r)\left[\log P(r|s) - \log P(r)\right]$$

$$= -\int ds\,dr\; P(s,r) \log P(r) \Leftarrow H[P(r)]$$

$$+ \underbrace{\int ds\,dr\; P(s)P(r|s) \log P(r|s)}_{\int ds\, P(s)\, H[P(r|s)]}$$

$$I(S,R) = H[R] - \Sigma_s\, P(s)\, H[R|s]\,.$$

$$I(S,R) = H[S] - \Sigma_r\, P(r)\, H[S|r]\,.$$

Calculating mutual information

Mutual information is the difference between the total response entropy and the mean noise entropy:

$$I(S;R) = H[R] - \Sigma_s P(s)\, H[R|s)] .$$

Grandma's famous mutual information recipe

Take one stimulus *s* and repeat many times to obtain $P(R|s)$.

Compute variability due to noise: *noise entropy* $H[R|s]$

Repeat for all s and average: $\Sigma_s P(s)\, H[R|s)]$.

Compute $P(R) = \Sigma_s P(s)\, P(R|s)$ **and the total entropy** $H[R]$