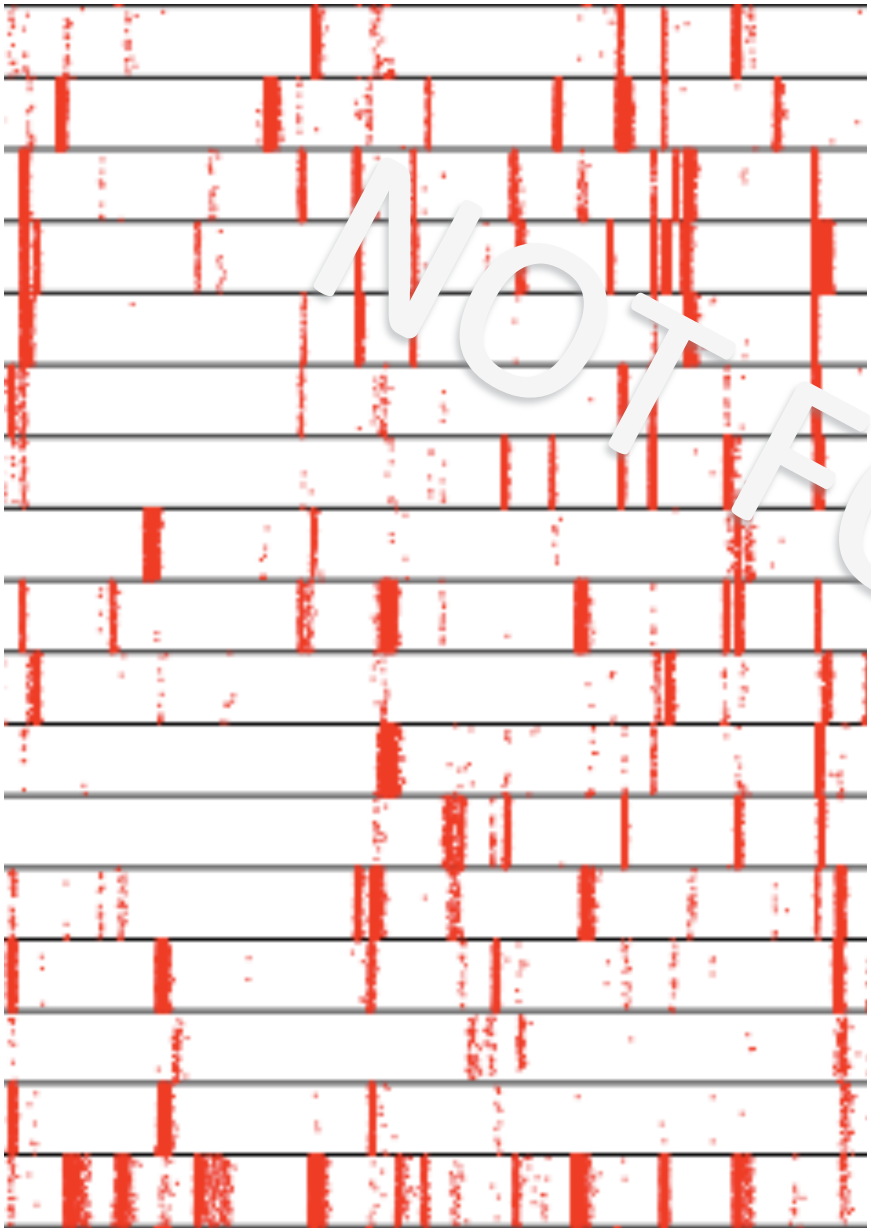
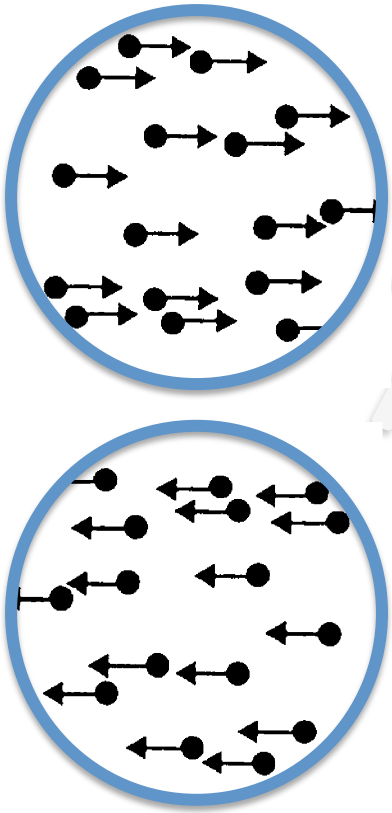
Information and coding principles

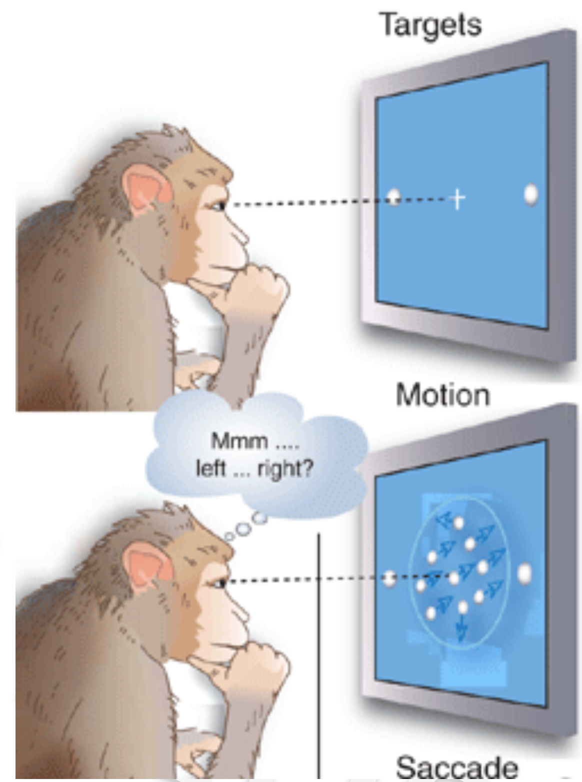


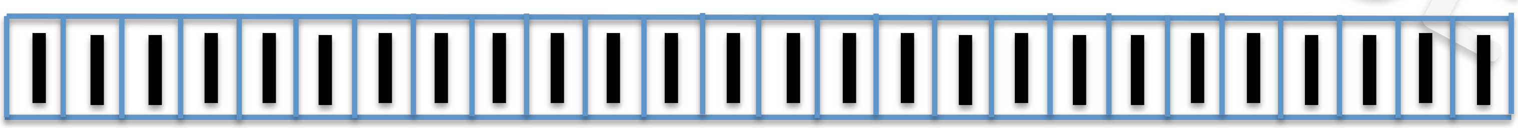
Information and coding principles

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

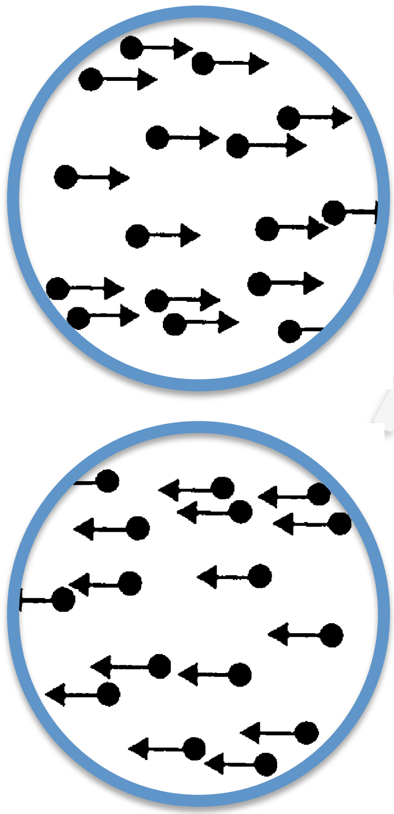
How good is my code?

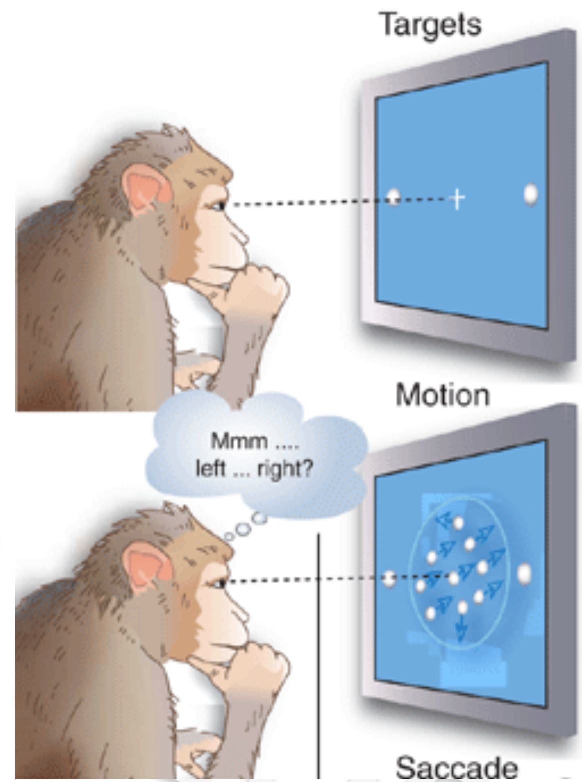


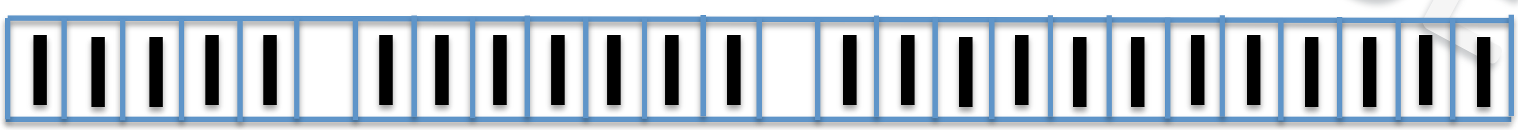




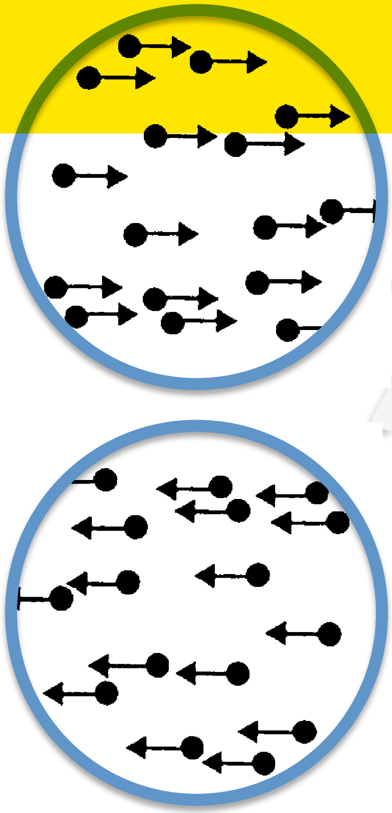
How good is my code?

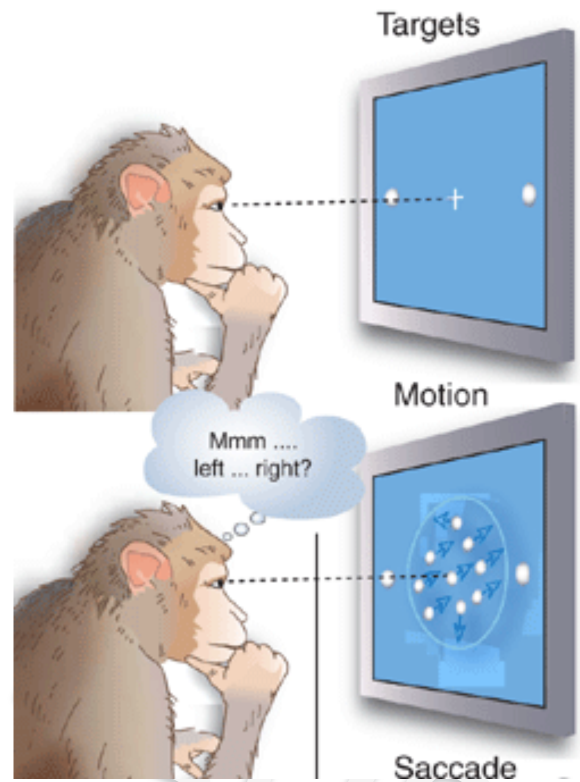


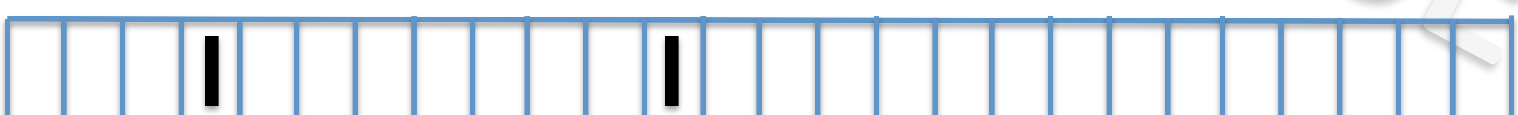




How good is my code?

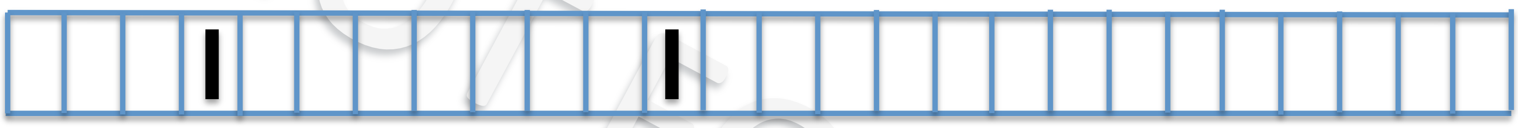






Surprise!

Information



$$P(1) = p$$

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

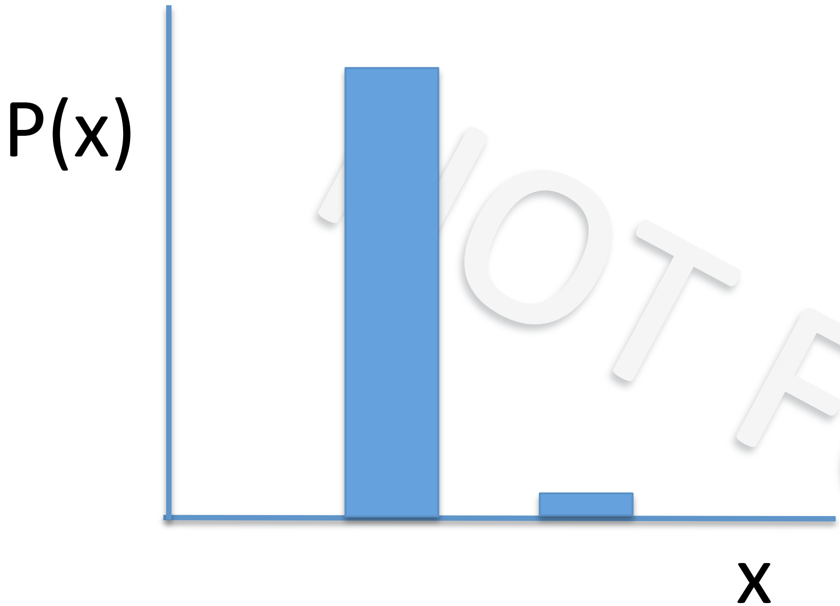
$$\text{information}(1) = -\log_2 p$$

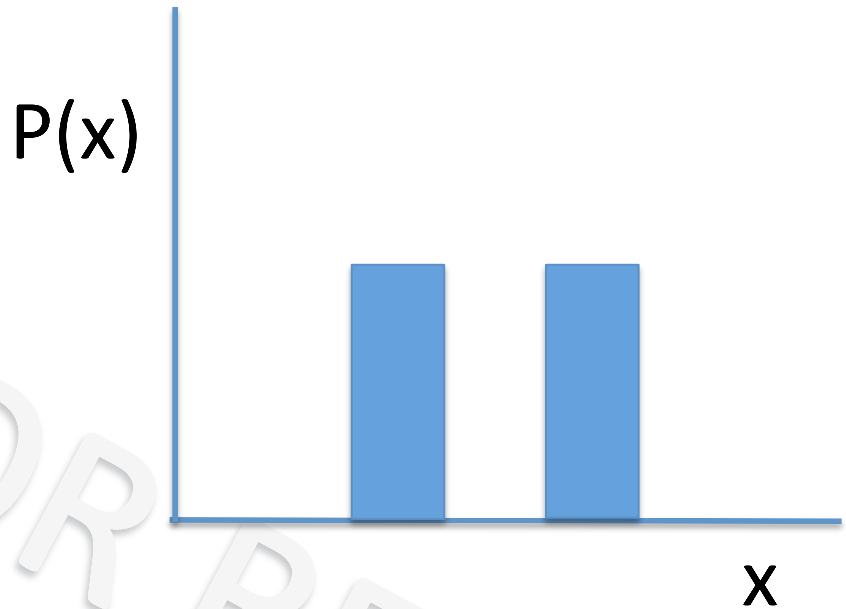
$$\text{information}(0) = -\log_2 (1-p)$$

Information

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

What is entropy?





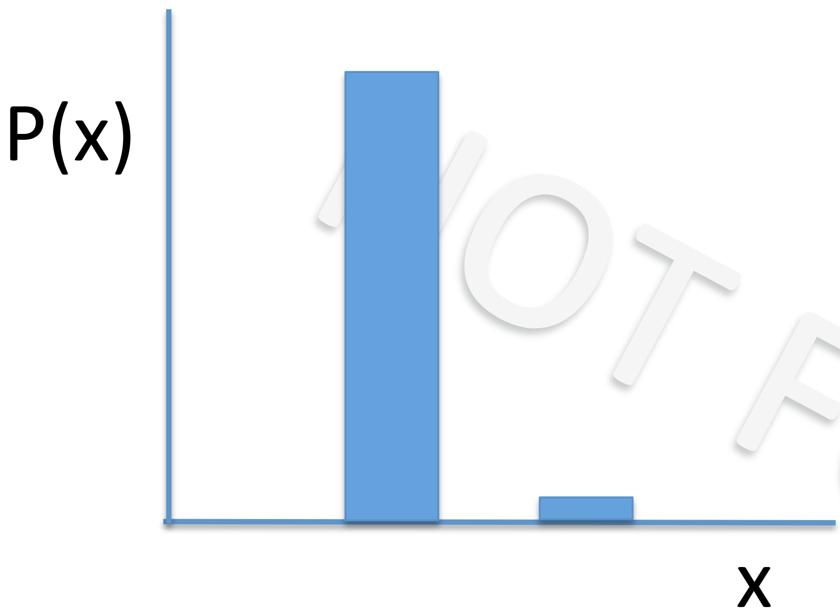
Entropy $= $ average information

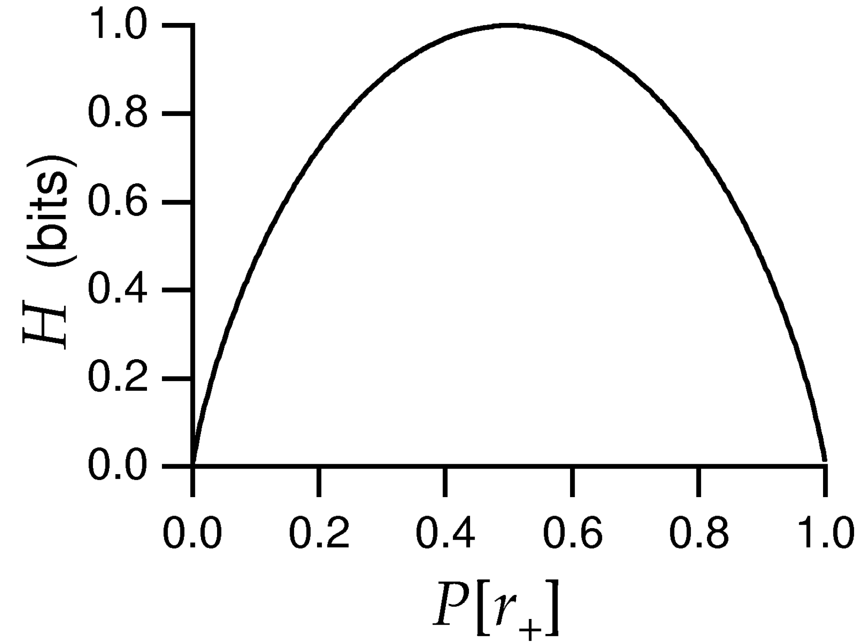
$= - \sum_i p_i \log_2 p_i$

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

Units are *bits*

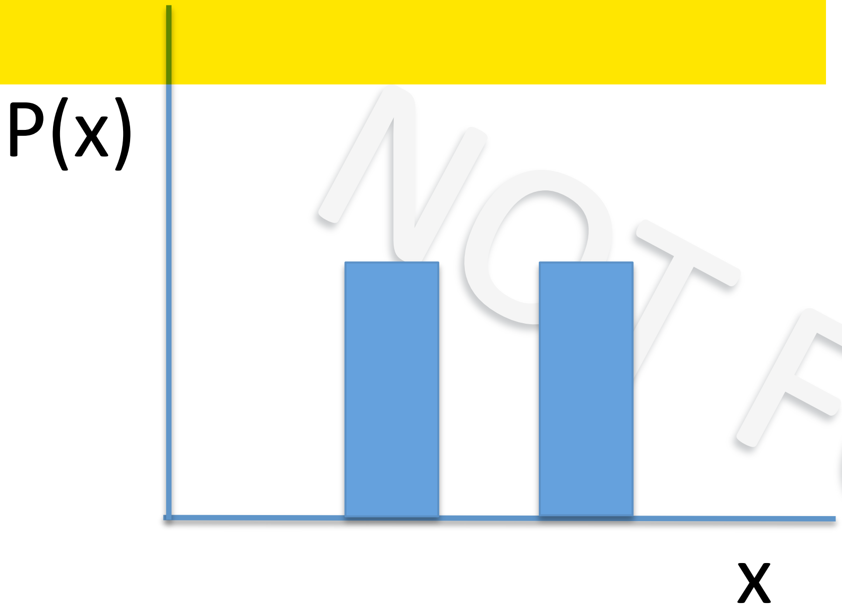
Maximizing the entropy

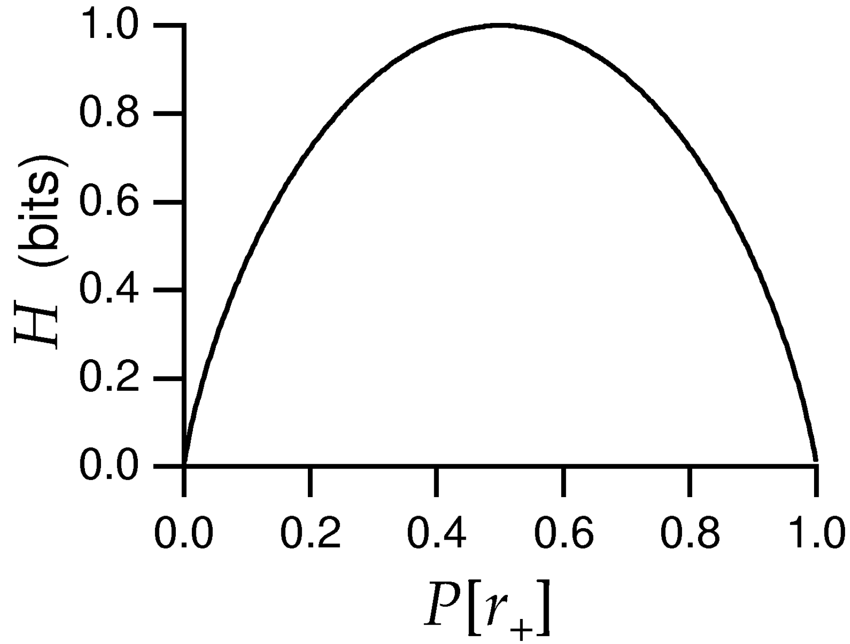




Entropy = $- \Sigma\, p_i \log_2 p_i$

Maximizing the entropy



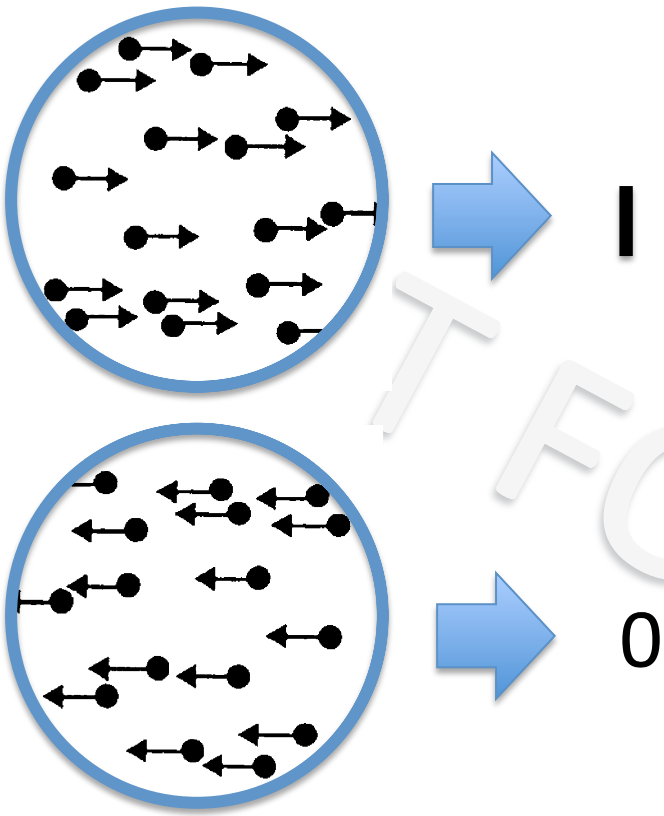


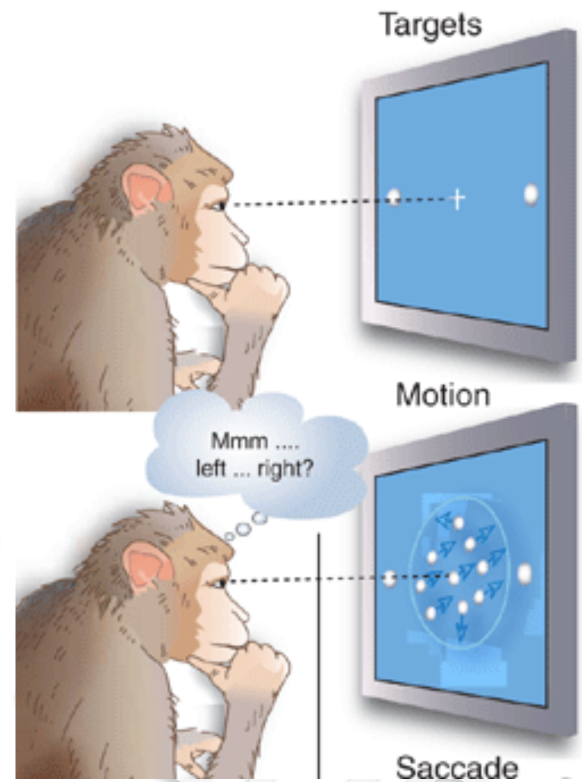
Entropy = $- \Sigma p_i \log_2 p_i$

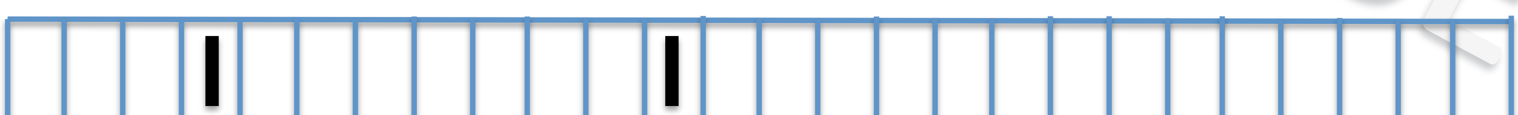
Maximum entropy squash

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?

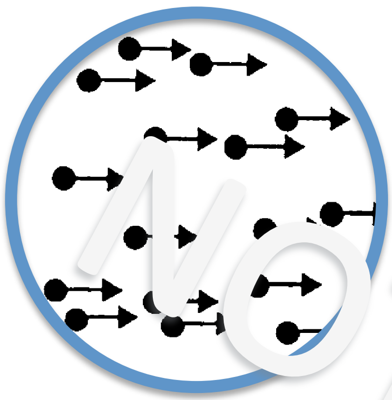






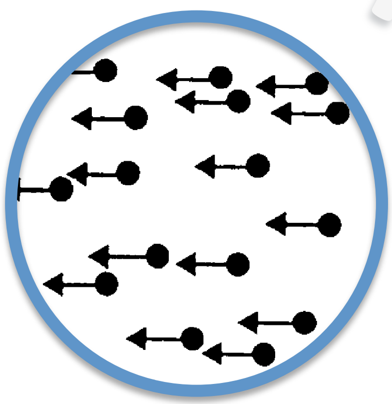
How about the stimulus?

S = +

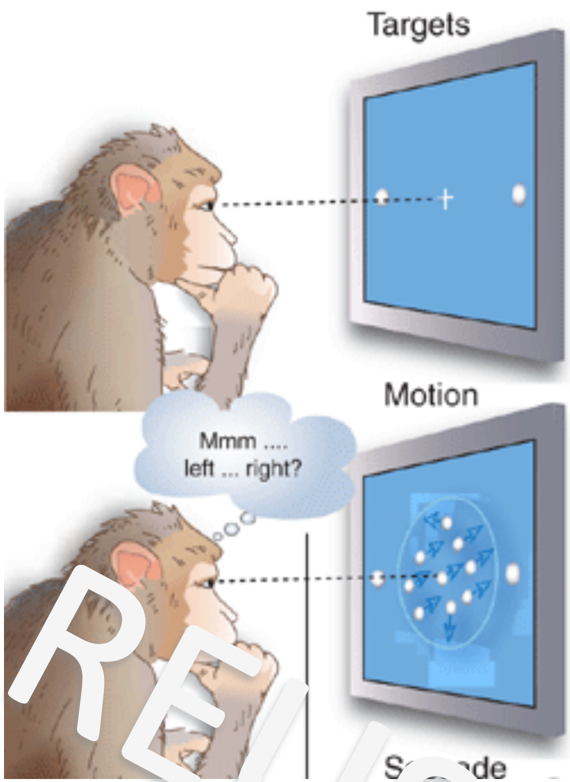


1

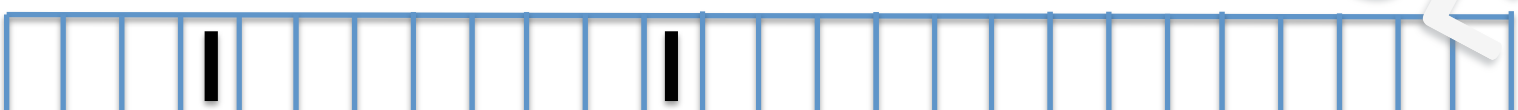
S = -



0



r

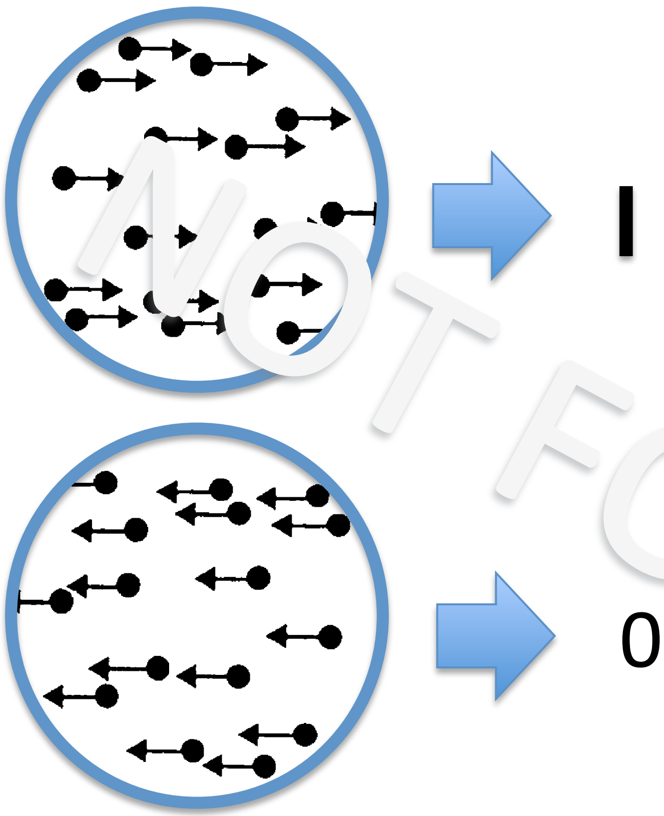


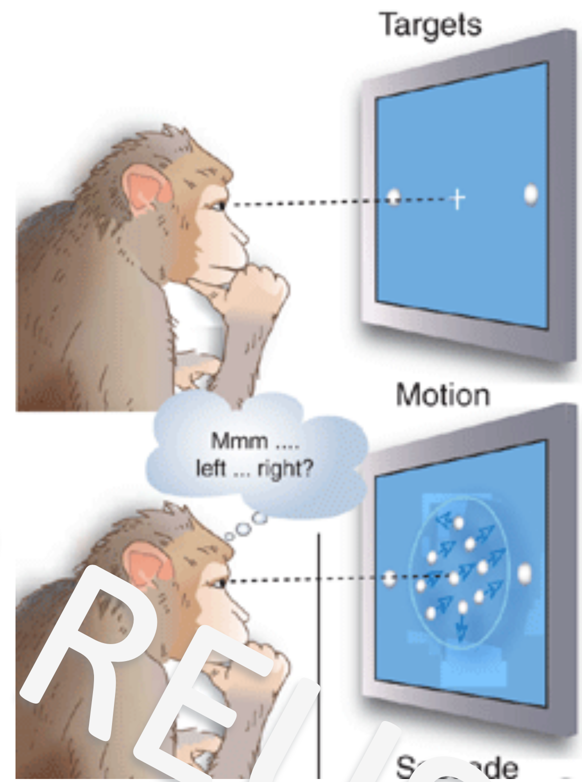
s - - - + - - - - - - - + - - - - - - - - - - - - - -

How about the stimulus?

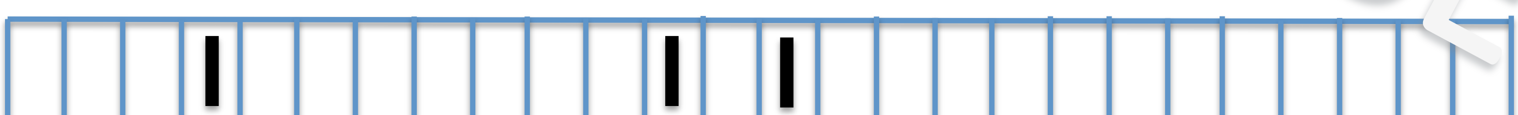
S = +

S = -



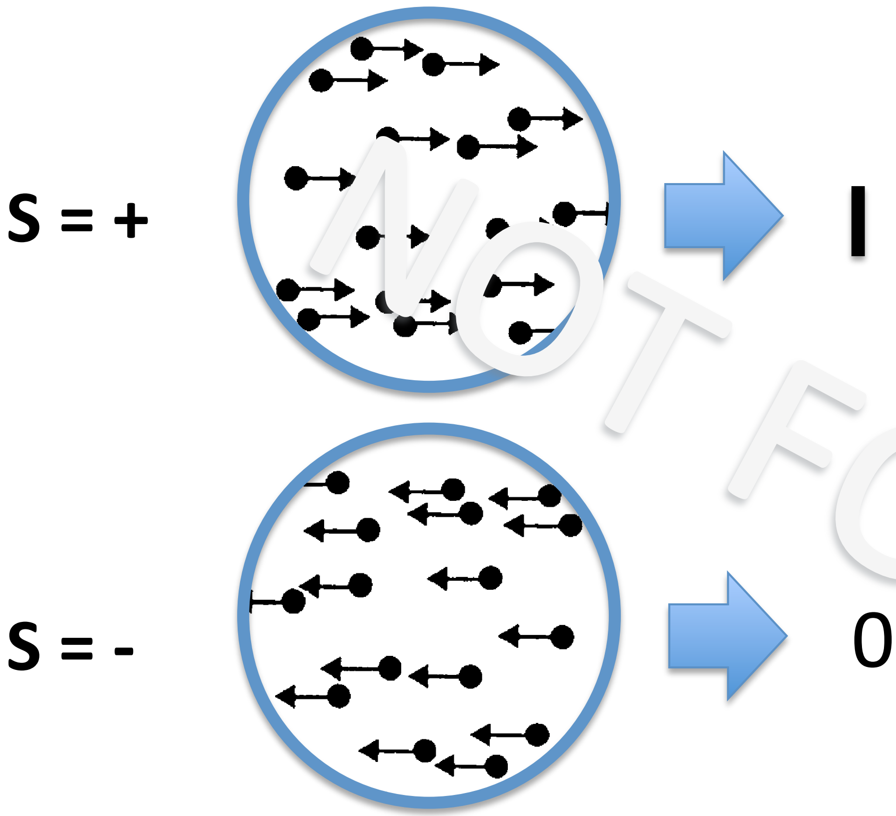


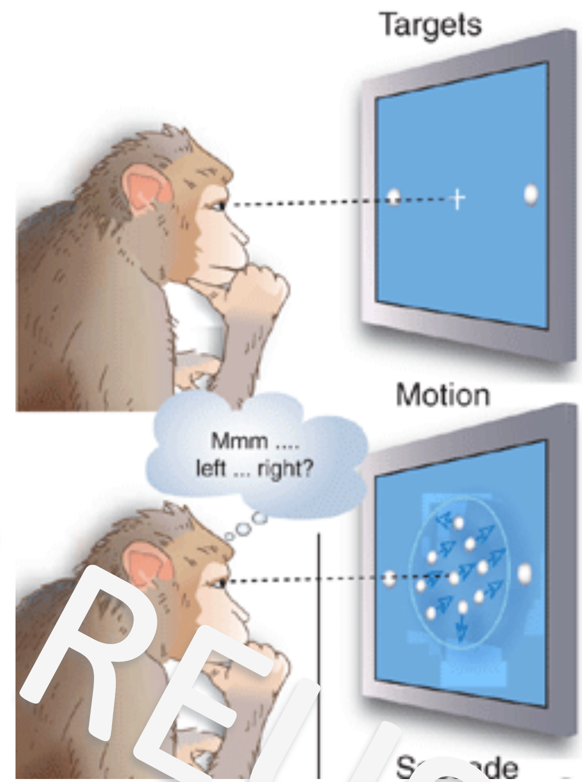
r

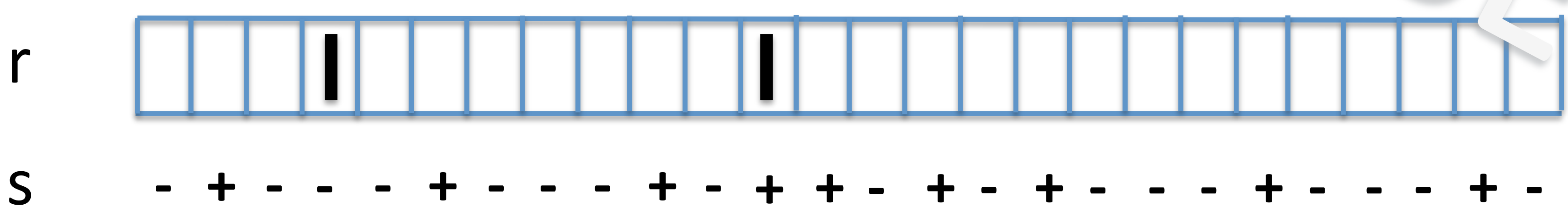


s - - - + - - - - - - - + - - - - + - - - - - - - - -

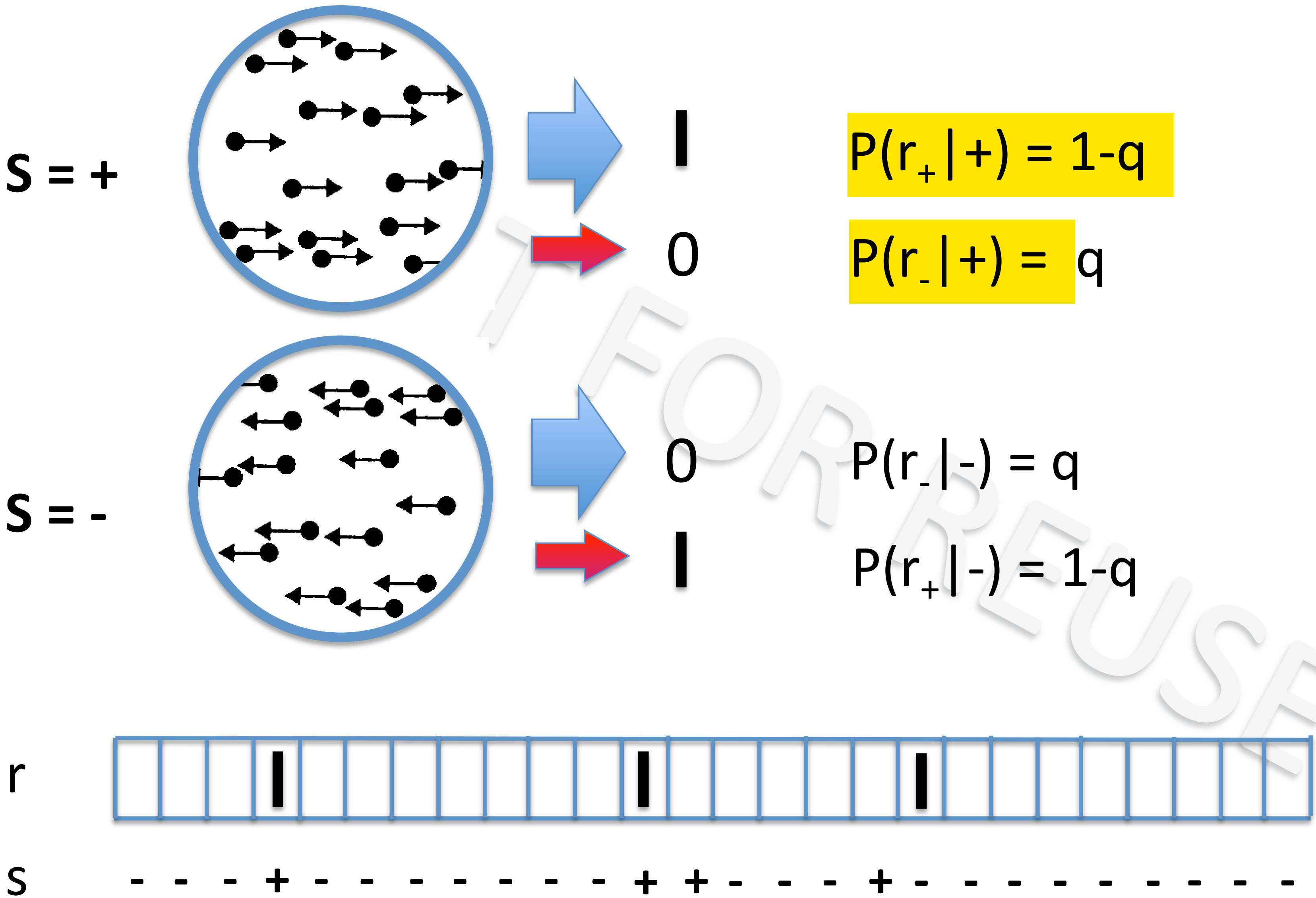
How about the stimulus?



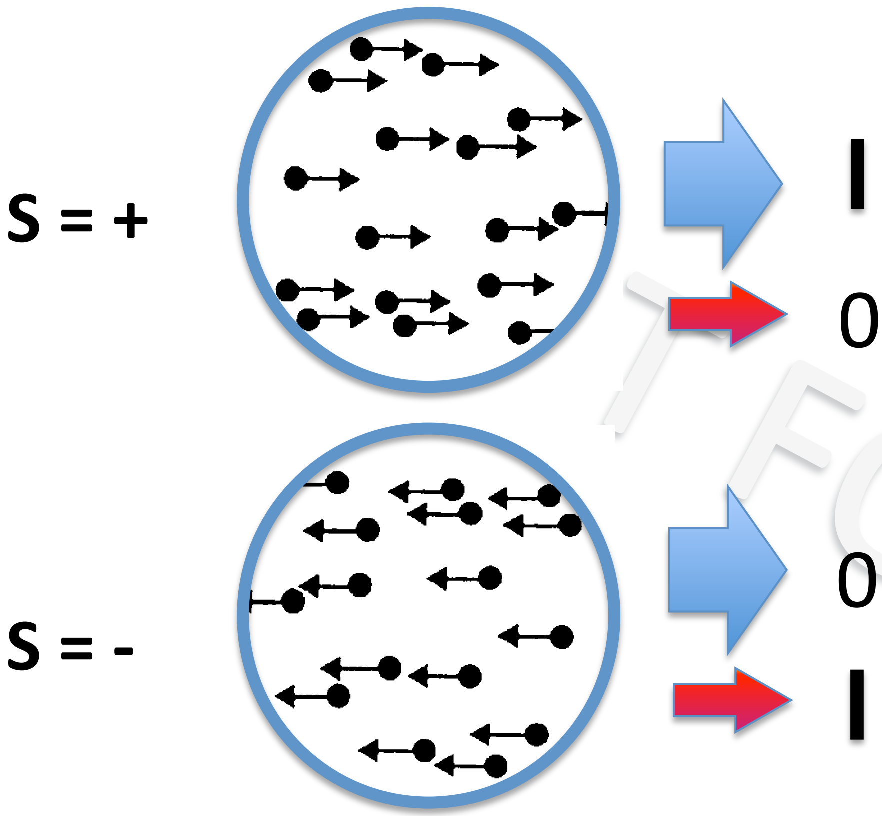




How much of the variability in *r* is encoding *s*?



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



$P(r_+|+) = 1-q$

$P(r_-|+) = q$

$P(r_-|-) = q$

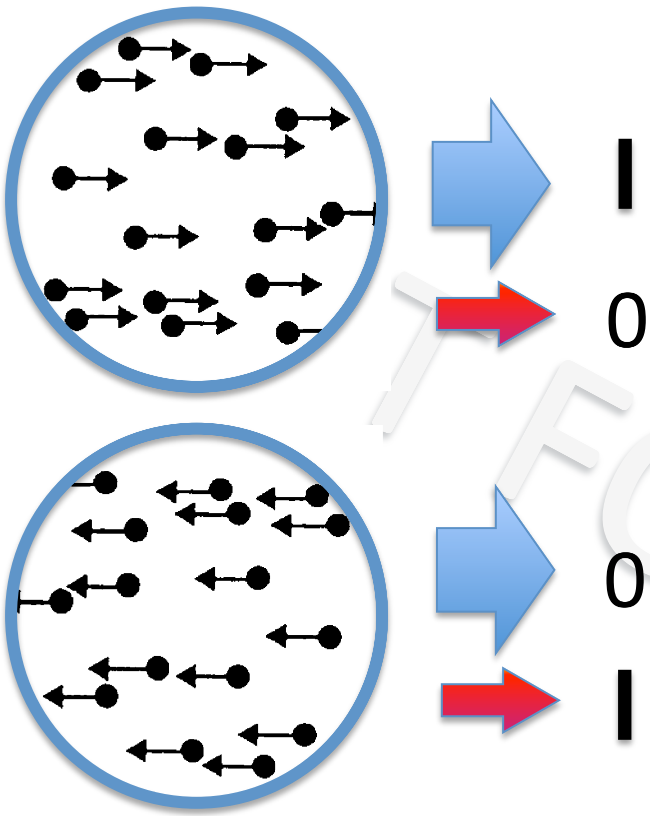
$P(r_+|-) = 1-q$

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

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

S = +

S = -



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

$P(r_-|+) = q$

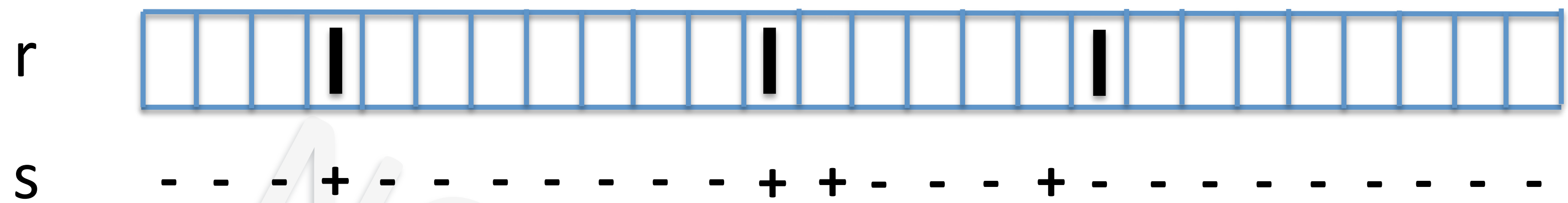
$P(r_-|-) = q$

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

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

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

Mutual information

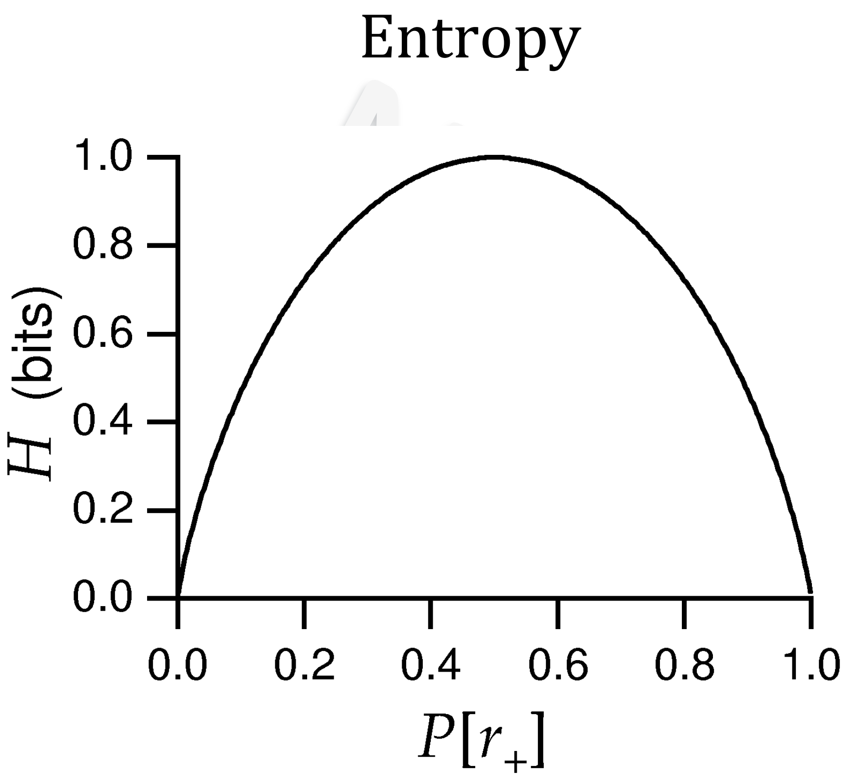


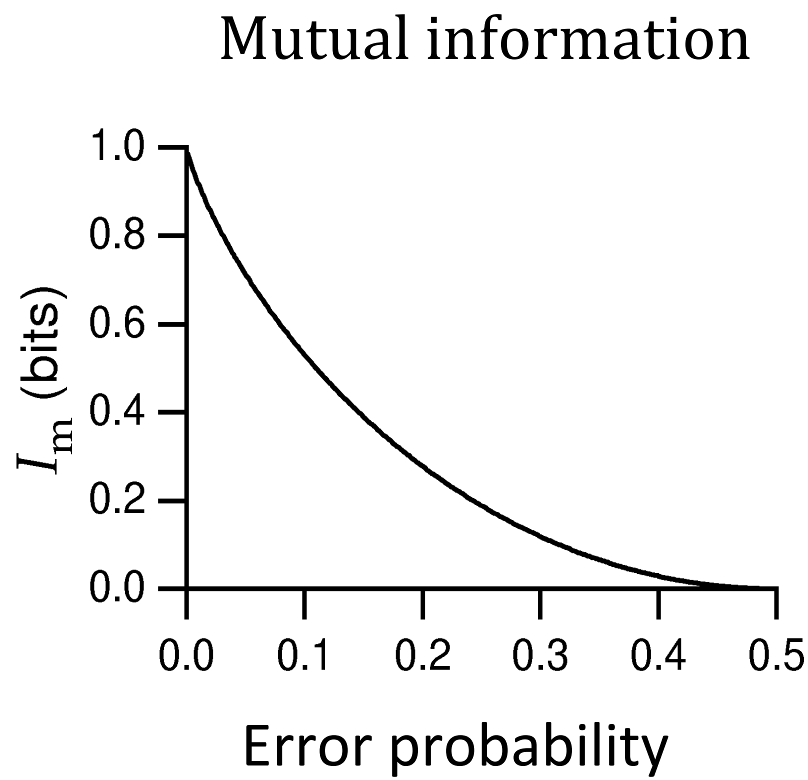
The amount of entropy that is used in coding the stimulus

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

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

Entropy and information

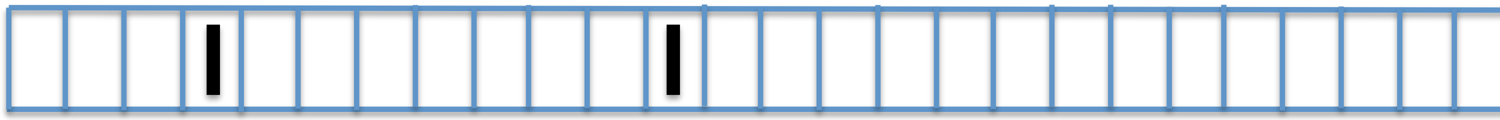




Mutual information you can calculate in your head

1.

r



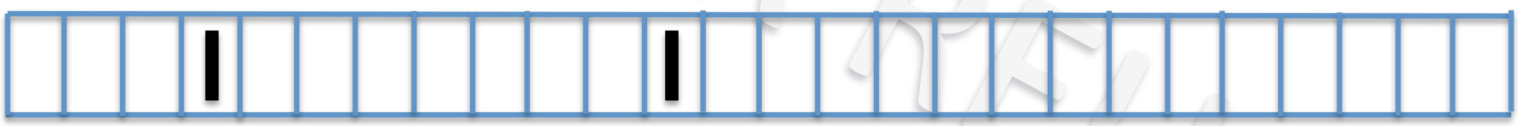
s - + - - - + - - - + - + + - + - + - - - + - - - + -

Response is unrelated to stimulus

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

2.

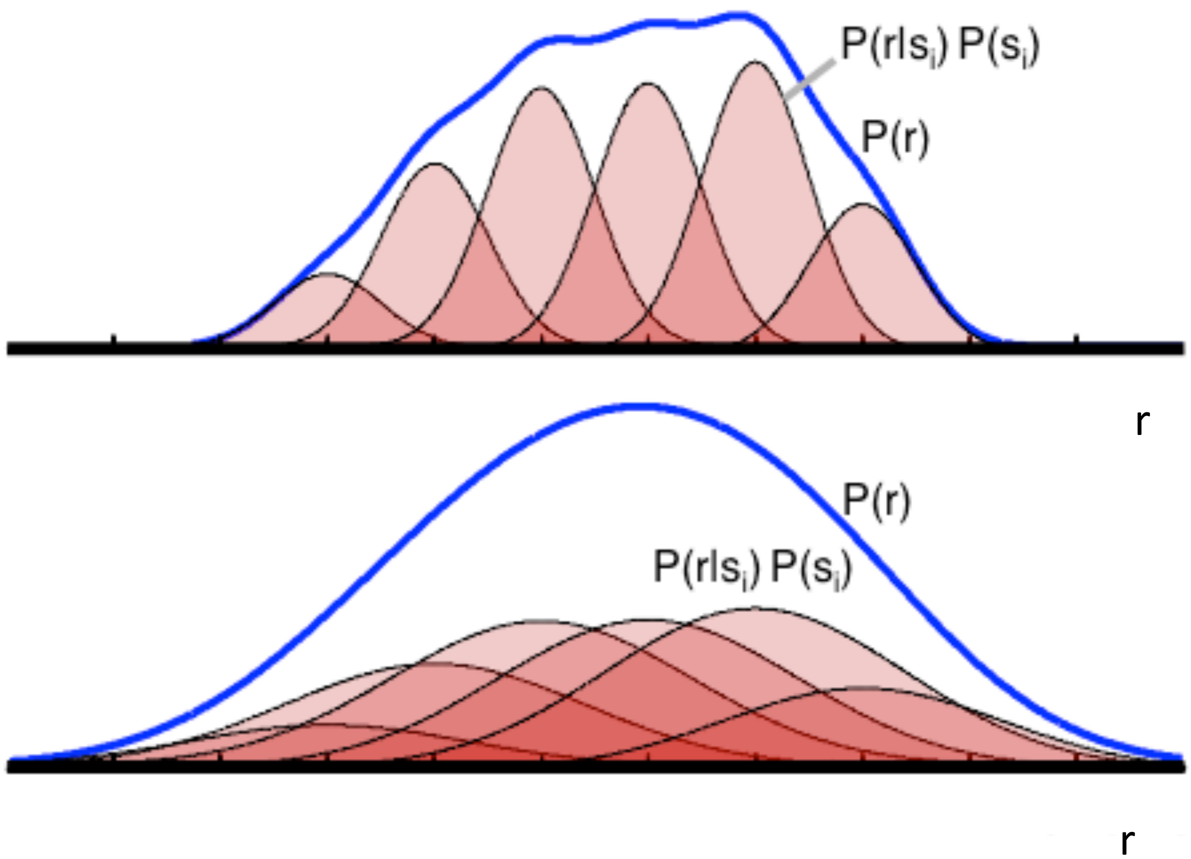
r



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

Response is perfectly predicted by stimulus

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:

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

Mutual information

Information quantifies how *independent* R and S are:

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

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

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

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** $H[R]$