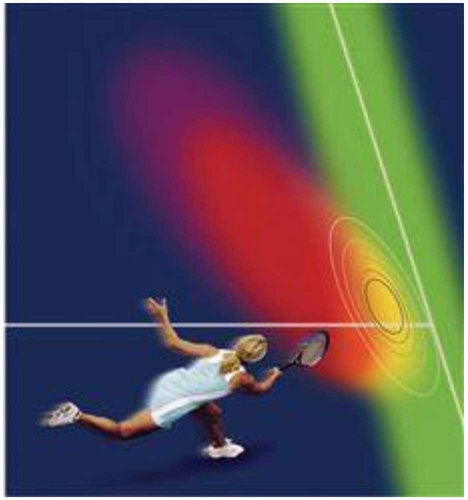


www.wolpertlab.com

Bayesian statistics

Computational Approaches to Neuroscience (NSCI 850)
Gunnar Blohm

Outline

- Introduction to Bayesian problems
- Bayes' theorem
  - Probabilities primer
  - Conditional probabilities
- Population codes
  - Coding and decoding
  - Representing uncertainty with population codes
- Bayesian integration
  - Cue combination
  - Estimation of priors
  - Causality and inference
- Discussion

Introduction to Bayesian problems

- The world is highly variable
  - Sensory uncertainties
  - Noisy neural codes
  - Conflicting sensory cues (e.g. illusions)
- Questions:
  - How does the brain generate a perceptual experience despite all this uncertainty?
  - How can we infer a state (i.e. code in the brain, attribute of an object, sensory state, etc)?
  - What is the optimal way to act in this noisy world?
  - How do we decide what cue to trust and/or how much?

Bayes' theorem

Bayes' theorem

- Probabilities
  - Probability of X=x: $0 \leq P(X = x) \leq 1$
  - Sum (or integral) of P of all outcomes: $\sum_x P(X = x) = 1$
  - Also true for **X** = vector, i.e. combined states
  - For statistically independent variables x and y: $P(x, y) = P(x)P(y)$

"and"

Bayes' theorem

- Conditional probabilities
  - Interaction:, i.e. probability of x given y: $P(x|y) = \frac{P(x,y)}{P(y)}$
  - This works both ways:

$$P(x \mid y) \cdot P(y) = P(y \mid x) \cdot P(x)$$

Bayes' theorem

- Thomas Bayes (1702-1761)

Conditional probability of B given A
or: likelihood function

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Conditional probability of A given B
or: posterior probability

Prior probability

Marginal probability:

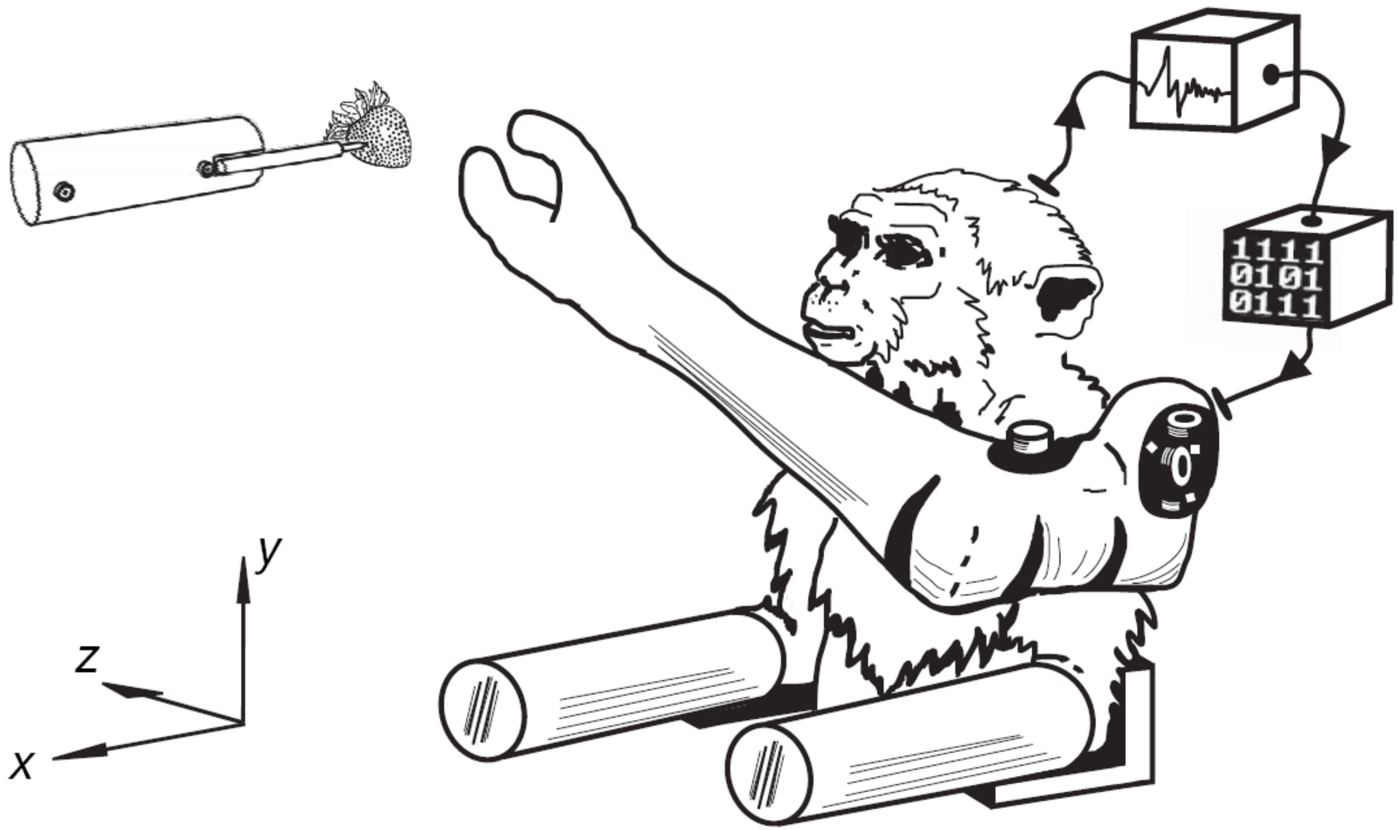
$$P(B) = \sum_i P(B|A_i) \cdot P(A_i)$$

- A: represents a specific hypothesis
- B: represents a certain event

Population codes

Population codes

- Example: cortical prosthetics



Velliste et al. Nature 2008

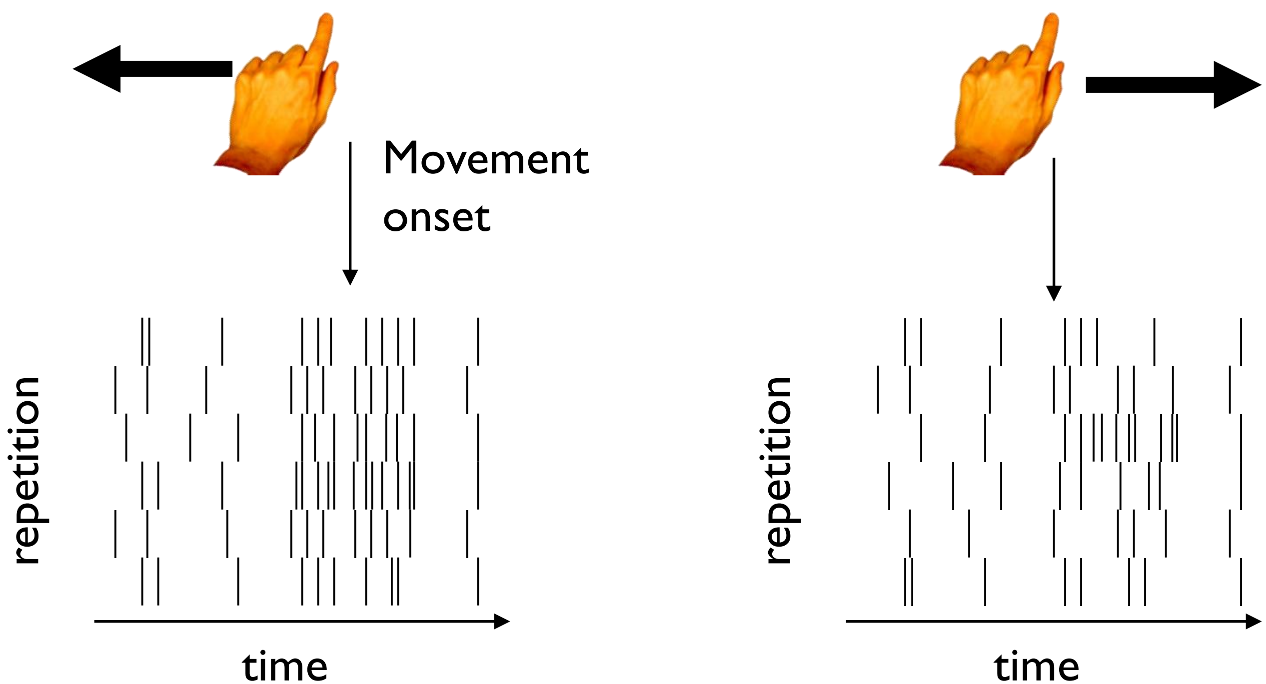
Population codes

- Example: cortical prosthetics

Courtesy Dr. Andy Schwartz

Population codes

- How do we decode the neural code?
  - Simplest situation: left or right?



Population codes

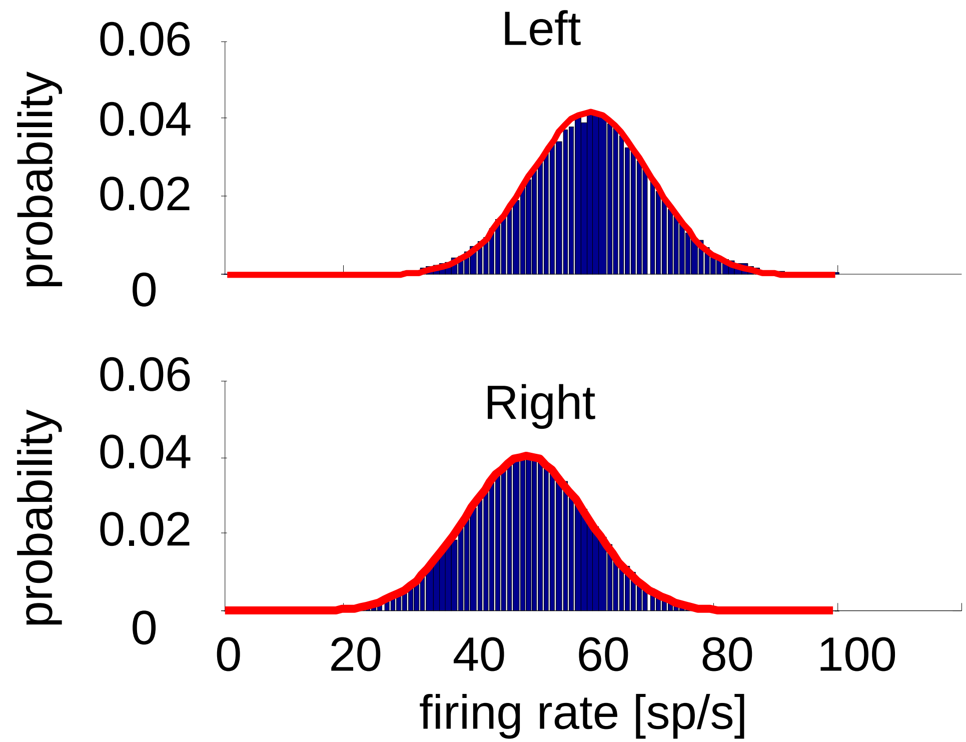
- Decoding based on firing rates
  - Suppose you have a neuron and record firing rates for left and right movements, such that:

Firing rate during left trial i	Firing rate during right trial i
53.9781	58.3077
57.6395	56.3932
56.1187	38.9440
38.0109	39.9052
67.3739	50.7743
…	…

Population codes

- Decoding based on firing rates
  - How to decide based on the firing rate of the neuron which direction to go?

$$p(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



Population codes

- Decoding based on firing rates
  - How to decide based on the firing rate of the neuron which direction to go?
    - Bayesian rule for decoding:

      probability of left, given spike rate S: $p(L|S) = \frac{p(S|L)p(L)}{p(S)}$

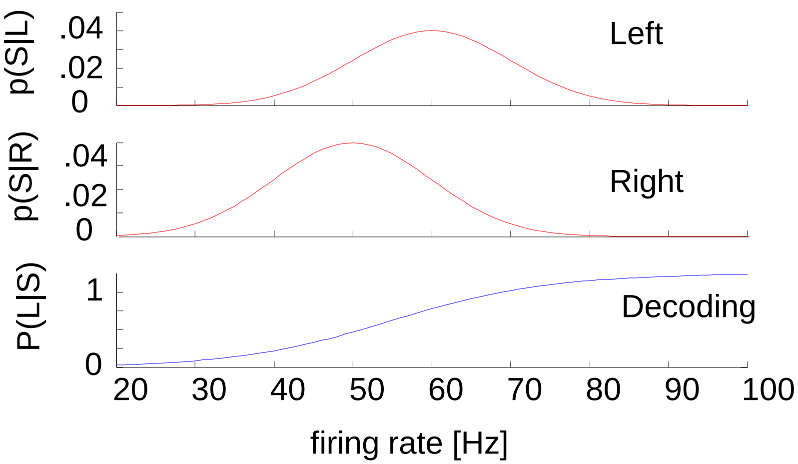
$$p(L|S) = \frac{p(S|L)p(L)}{p(S|L)p(L)+p(S|R)p(R)}$$

    - For equal proportions left and right movements, p(L)=p(R)=0.5

$$p(L|S) = \frac{p(S|L)}{p(S|L)+p(S|R)}$$

Population codes

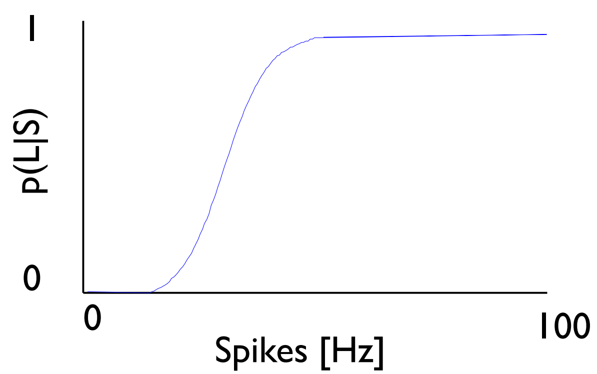
- Decoding based on firing rates
  - How to decide based on the firing rate of the neuron which direction to go?
    - Bayesian rule for decoding



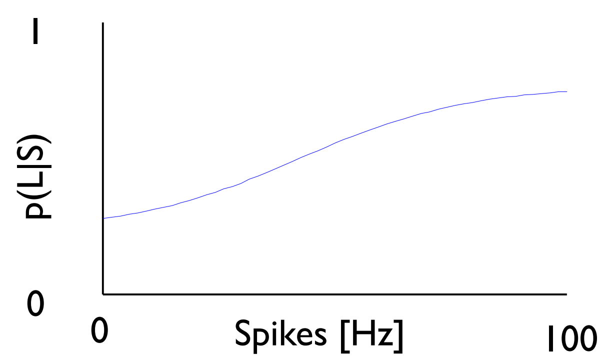
$$p(L|S) = \frac{p(S|L)}{p(S|L)+p(S|R)}$$

Population codes

- Decoding based on firing rates
  - Typical posterior probabilities in the brain



Very informative



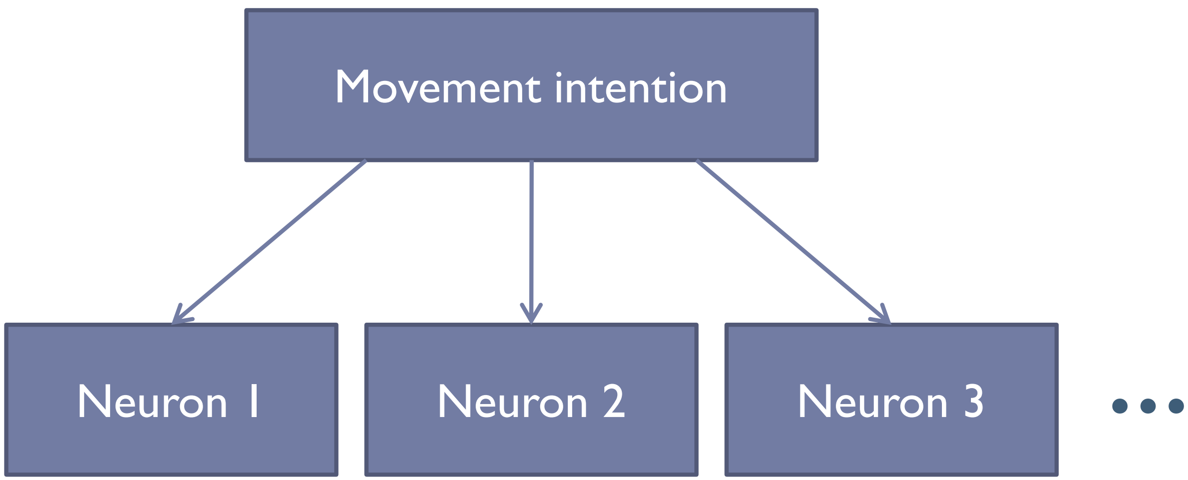
Not very informative

↓

This is the typical case in the CNS

Population codes

- Decoding based on firing rates
  - How can we get a better understanding of the neuronal code?
    - Record from multiple neurons!



    - Naive Bayesian assumption: neurons are independent

Population codes

- Decoding based on firing rates

  - Combination of 2 neurons:

$$p(L \mid S_1, S_2) = \frac{p(S_1, S_2 \mid L) \cdot p(L)}{p(S_1, S_2)}$$

$$= \frac{p(S_1 \mid L) \cdot p(S_2 \mid L) \cdot p(L)}{p(S_1 \mid L) \cdot p(S_2 \mid L) \cdot p(L) + p(S_1 \mid R) \cdot p(S_2 \mid R) \cdot p(R)}$$

  - More generally:

$$p(L \mid S_1, S_2, \ldots, S_N) = \frac{p(L) \cdot \prod_i p(S_i \mid L)}{p(S_1, S_2, \ldots, S_N)}$$

Bayesian integration

Bayesian integration

- Cue combination

  - E.g. Multi-sensory integration
    - McGurk effect, ventriloquism, …
  - The study of how different sensory modalities (vision, touch, sound, etc) get combined into a perceptual experience that is coherent and unified.
  - The brain always uses all available useful information.
  - Information from different sources is combined in a statistically optimal fashion

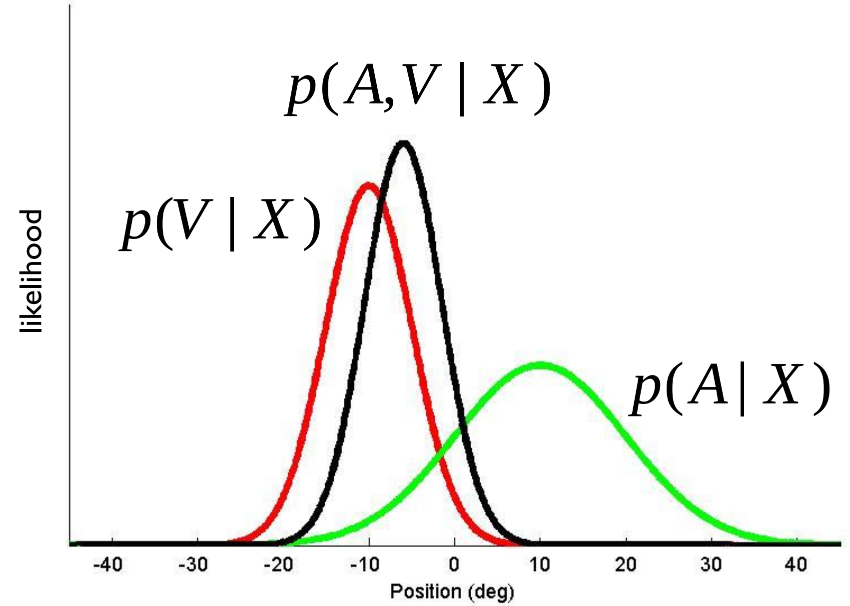
Bayesian integration

- Cue combination
  - Optimal Bayesian observer $$p(X \mid A,V) = \frac{p(A,V \mid X) \cdot p(X)}{p(A,V)}$$
  - Independent observations A, V

$$p(A,V \mid X) = p(V \mid X) \cdot p(A \mid X)$$

  - If uniform priors, then

$$p(X \mid A,V) \propto p(V \mid X) \cdot p(A \mid X)$$



Bayesian integration

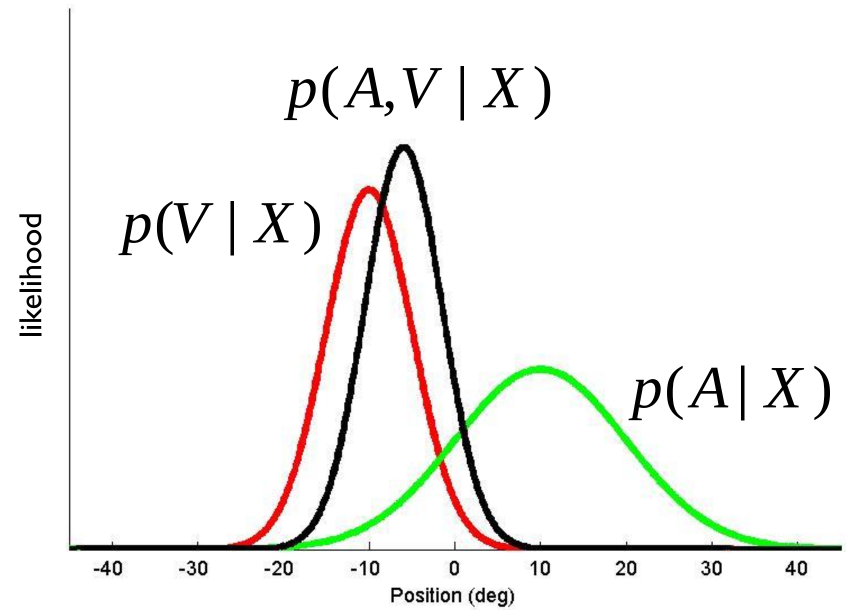
- Cue combination
  - Gaussian likelihood functions

$$\frac{1}{\sigma^2} = \frac{1}{\sigma^2{}_1} + \frac{1}{\sigma^2{}_2}$$

$$\Longleftrightarrow$$

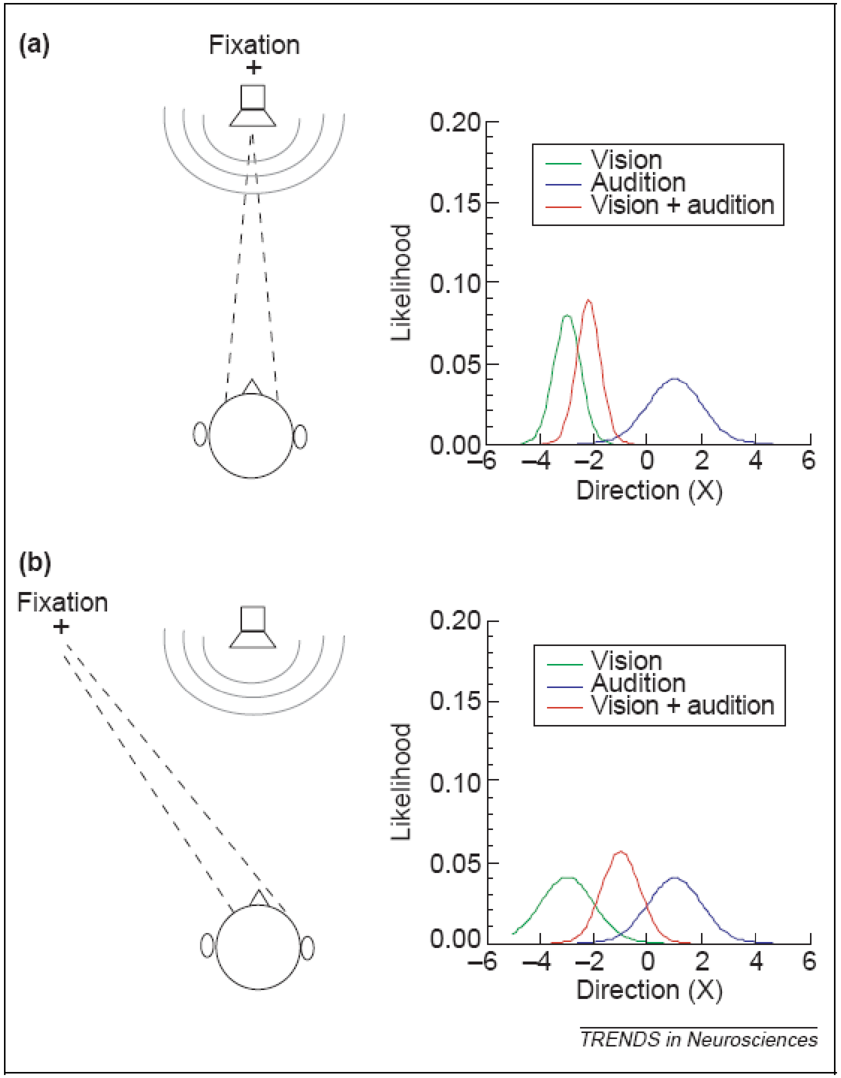
$$\sigma^2 = \frac{\sigma^2{}_1 \cdot \sigma^2{}_2}{\sigma^2{}_1 + \sigma^2{}_2}$$

$$\mu = \sigma^2 \cdot \left(\frac{\mu_1}{\sigma^2{}_1} + \frac{\mu_2}{\sigma^2{}_2} \right)$$



Bayesian integration

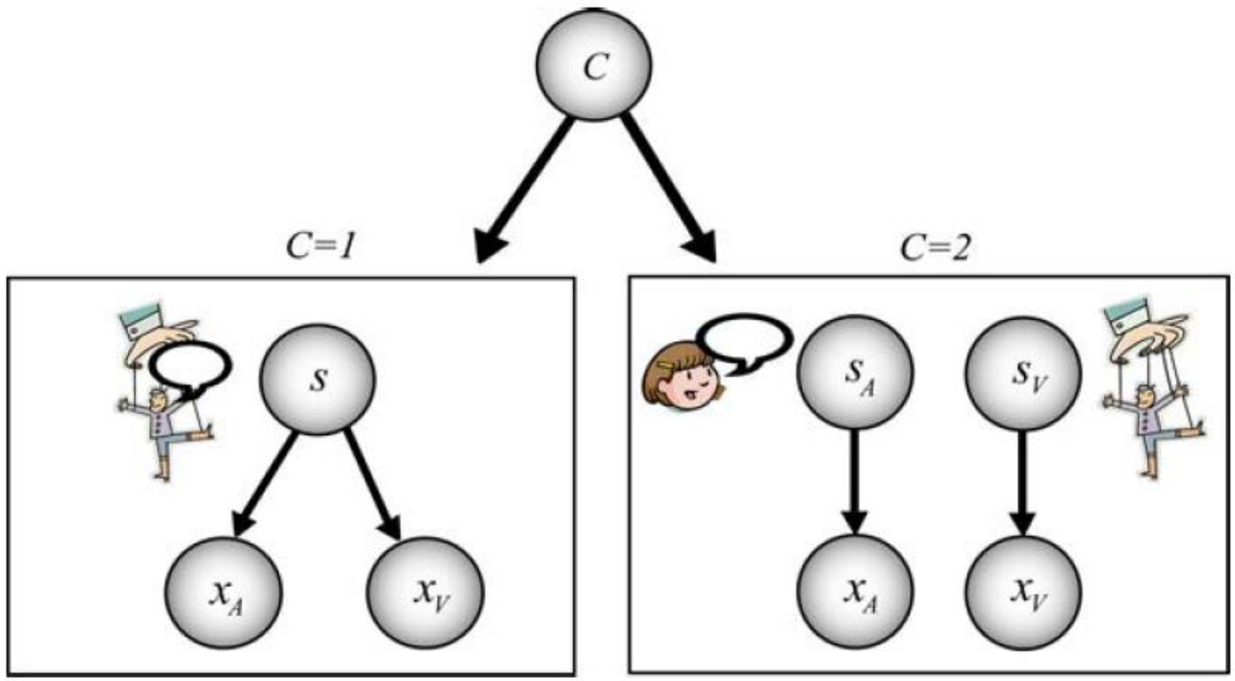
- Cue combination
  - E.g. Audio-visual integration



Knill & Pouget, TINS 2004

Bayesian integration

- Causality and inference
  - Until now, we have considered Bayesian integration of multiple cues related to a unique *cause* (measurable state s)
  - But what if sensory inputs came from different causes?



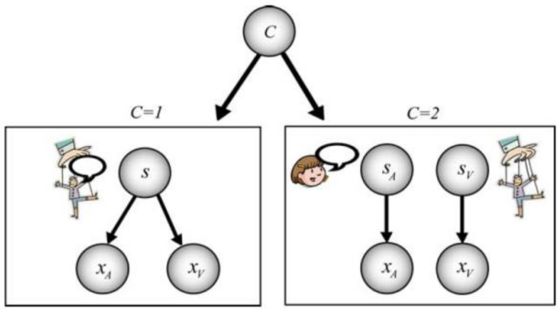
Körding et al. PLoSOne 2007

Bayesian integration

- Causality and inference
  - For 2 cues (A,V), Bayes' rule: $p(C \mid x_V, x_A) = \dfrac{p(x_V, x_A \mid C) \cdot p(C)}{p(x_V, x_A)}$
  - Choose $p(x_V, x_A)$ so that $p(x_V, x_A \mid C=1) + p(x_V, x_A \mid C=2) = 1$

$$p(C=1 \mid x_V, x_A) = \frac{p(x_V, x_A \mid C=1) \cdot p(C=1)}{p(x_V, x_A \mid C=1) \cdot p(C=1) + p(x_V, x_A \mid C=2) \cdot (1 - p(C=1))}$$

  - In most cases, the normalization factors can be neglected and we can thus concentrate on $p(x_V, x_A \mid C=1)$



Körding et al. PLoSOne 2007

Bayesian integration

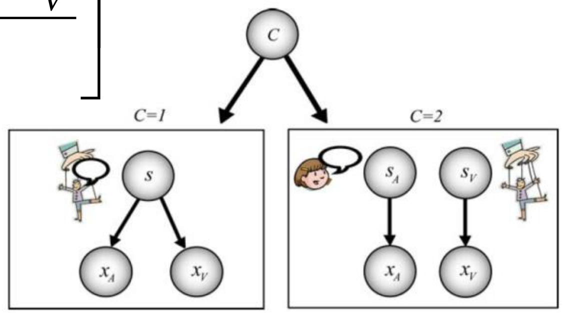
- Causality and inference

$$p(x_V, x_A \mid C=1) = \int p(x_V, x_A \mid s)\,p(s)\,ds$$
$$= \int p(x_V \mid s)\,p(x_A \mid s)\,p(s)\,ds$$

- If all 3 factors in this integral are Gaussian, then

$$p(x_V, x_A \mid C=1) = \frac{1}{2\pi\sqrt{\sigma_V^2\sigma_A^2 + \sigma_V^2\sigma_P^2 + \sigma_A^2\sigma_P^2}}$$
$$\cdot \exp\left[-\frac{1}{2}\frac{(x_V - x_A)^2\sigma_P^2 + (x_V - \mu_P)^2\sigma_A^2 + (\mu_P - x_A)^2\sigma_V^2}{\sigma_V^2\sigma_A^2 + \sigma_V^2\sigma_P^2 + \sigma_A^2\sigma_P^2}\right]$$

- with prior $p(s) = N(\mu_P, \sigma_P^2)$



Körding et al. PLoSOne 2007

Bayesian integration

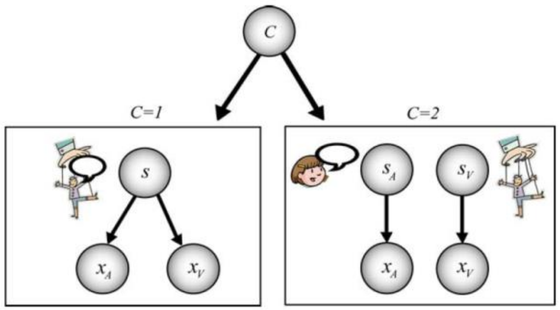
- Causality and inference
  - Similarly for C=2:

$$p(x_V, x_A \mid C=2) = \int p(x_V, x_A \mid s_V, s_A) p(s_V, s_A) ds$$
$$= \left(\int p(x_V \mid s_V) p(s_V) ds_V\right) \cdot \left(\int p(x_A \mid s_A) p(s_A) ds_A\right)$$

  - If all terms are Gaussian:

$$p(x_V, x_A \mid C=2) = \frac{1}{2\pi\sqrt{(\sigma_V^2 + \sigma_P^2)\cdot(\sigma_A^2 + \sigma_P^2)}} \exp\left[-\frac{1}{2}\left(\frac{(x_V - \mu_P)^2}{\sigma_V^2 + \sigma_P^2} + \frac{(x_A - \mu_P)^2}{\sigma_A^2 + \sigma_P^2}\right)\right]$$

  - Since $p(x_V, x_A \mid C=1) + p(x_V, x_A \mid C=2) = 1$
  - Common cause if: $p(x_V, x_A \mid C=1) > \frac{1}{2}$

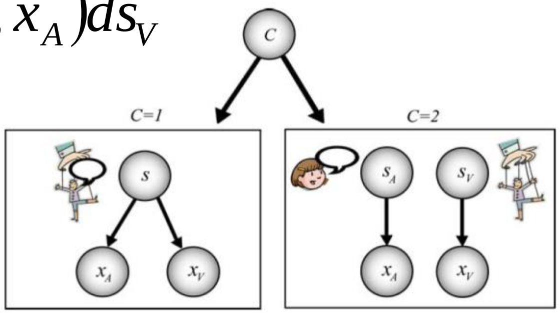


Körding et al. PLoSOne 2007

Bayesian integration

- Optimal estimation
  - How do we optimally estimate a state s if the number of causes are uncertain?
  - Minimize a cost function: $Cost = (\hat{s}_V - s_V)^2 + (\hat{s}_A - s_A)^2$
  - An optimal estimate leads to the lowest expected cost under the subject's posterior belief:

$$Cost_V = p(C=1 \mid x_V, x_A) \int (\hat{s}_V - s_V)^2 \, p(s_V \mid C=1, x_V, x_A) \, ds_V$$
$$+ p(C=2 \mid x_V, x_A) \int (\hat{s}_V - s_V)^2 \, p(s_V \mid C=2, x_V, x_A) \, ds_V$$



Körding et al. PLoSOne 2007

Bayesian integration

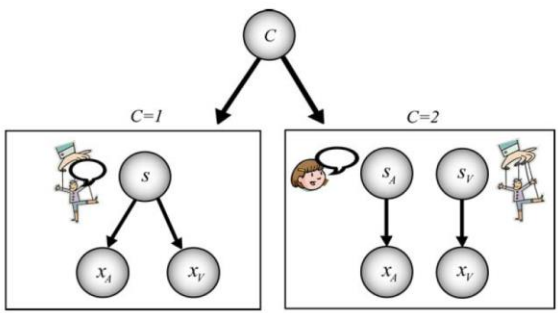
- Optimal estimation
  - For quadratic costs, optimality = find mean of posterior distribution. Thus:

$$\hat{s}_V = p(C=1 \mid x_V, x_A) \cdot \hat{s}_{V,C=1} + p(C=2 \mid x_V, x_A) \cdot \hat{s}_{V,C=2}$$

Estimates obtained if we were certain about the number of causes.

- Gaussians:

$$\hat{s}_{V,C=2} = \frac{\frac{x_V}{\sigma_V^2} + \frac{x_P}{\sigma_P^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_P^2}} \qquad \hat{s}_{V,C=1} = \frac{\frac{x_V}{\sigma_V^2} + \frac{x_P}{\sigma_P^2} + \frac{x_A}{\sigma_A^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_P^2} + \frac{1}{\sigma_A^2}}$$

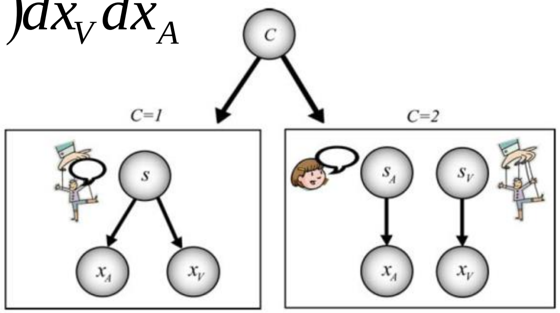


Körding et al. PLoSOne 2007

Bayesian integration

- Optimal estimation
  - While $\hat{s}_{V,C=1}$, $\hat{s}_{V,C=2}$ are linear combinations of x_V and x_A, the factors $p(C \mid x_V, x_A)$ are non-linear.
  - Therefore, the overall optimal estimates are non-linear!
  - To describe a system that performs causal inference, the linear approximation is not sufficient.
  - Finally, the distribution of visual positions is obtained through marginalization:

$$p(\hat{s}_V \mid s_V, s_A) = \iint p(\hat{s}_V \mid x_V, x_A)\, p(x_V \mid s_V)\, p(x_A \mid s_A)\, dx_V\, dx_A$$



Körding et al. PLoSOne 2007

Further readings

- Doya et al., Bayesian Brain: Probabilistic approaches to neural coding, MIT Press 2007
- Abbott & Sejnowski, Neural codes and distributed representations, MIT Press 1999
- Trommershäuser, Körding, Landy (eds.), Sensory Cue Integration, Oxford University Press 2011
- https://drive.google.com/file/d/0B7BtxrHIZHgpUmUxXzI2WHVsUEk/view