Cue combination

Lecture 4

Why study cue combination?

- Very common, within and between modalities
- Simple computation, but still a computation
- Illustrates key notions of Bayesian optimality
- Can be linked to neural basis

Humans integrate visual and haptic information in a statistically optimal fashion

Marc O. Ernst* & Martin S. Banks

Vision Science Program/School of Optometry, University of Califo
94720-2020, USA

Robert J. van Beers · Anne C. Sittig
Jan J. Denier van der Gon

How humans combine simultaneous proprioceptive and visual position information

Current Biology, Vol. 14, 257–262, February 3, 2004,

The Ventriloquist Effect Results from Near-Optimal Bimodal Integratio

David Alais[1,2] and David Burr[1,3,*]
[1]Istituto di Neuroscienze del CNR
56127 Pisa
Italy
[2]Auditory Neuroscience Laboratory
Department of Physiology
University of Sydney
New South Wales 2006
Australia
[3]Department of Psychology
University of Florence
50125 Florence
Italy

Optimal integration of texture and motion cues to depth

Robert A. Jacobs *

Center for Visual Science, University of Rochester, Rochester, NY 14627, USA

urch 1999

Motion illusions as optimal percepts

Yair Weiss[1], Eero P. Simoncelli[2] and Edward H. Adelson[3]

[1] *School of Computer Science and Engineering, Hebrew University of Jerusalem, Givat Ram Campus, Jerusalem 91904, Israel*
[2] *Howard Hughes Medical Institute, Center for Neural Science and Courant Institute of Mathematical Sciences, New York University, 4 Washington Place, New York, New York 10003, USA*
[3] *Brain and Cognitive Sciences Department, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, Massachusetts 02139, USA*
Correspondence should be addressed t

Lip-Reading Aids Word Recognition Most in Moderate Noise: A Bayesian Explanation Using High-Dimensional Feature Space

Wei Ji Ma[1,9]*, Xiang Zhou[2,9], Lars A. Ross[3,4], John J. Foxe[3,4,5], Lucas C. Parra[2]

1 Department of Neuroscience, Baylor College of Medicine, Houston, Texas, United States of America, **2** Department of Biomedical Engineering, The City College of New York, New York, New York, United States of America, **3** Program in Cognitive Neuroscience, Department of Psychology, The City College of New York, New York, New York, United States of America, **4** The Cognitive Neuroscience Laboratory, Nathan S. Kline Institute for Psychiatric Research, Program in Cognitive Neuroscience and Schizophrenia, Orangeburg, New York, United States of America, **5** Program in Neuropsychology, Department of Psychology, Queens College of the City University of New York, Flushing, New York, United States of America

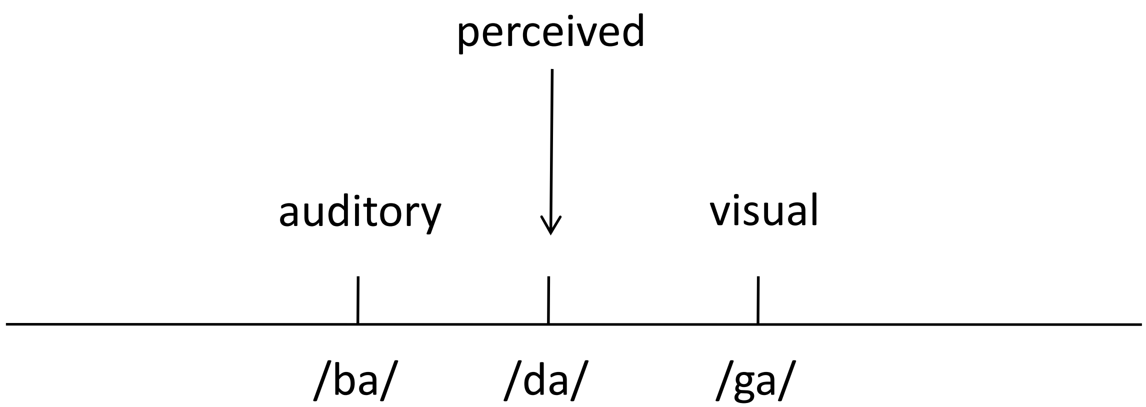
ıally integrate stereo and texture information ır judgments of surface slant?

Ɔavid C. Knill *, Jeffrey A. Saunders

ːiences, University of Rochester, 274 Meliora Hall, Rochester, NY 14627, USA

ːived 2 December 2002; received in revised form 22 April 2003

What is he saying?

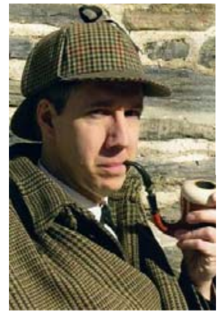


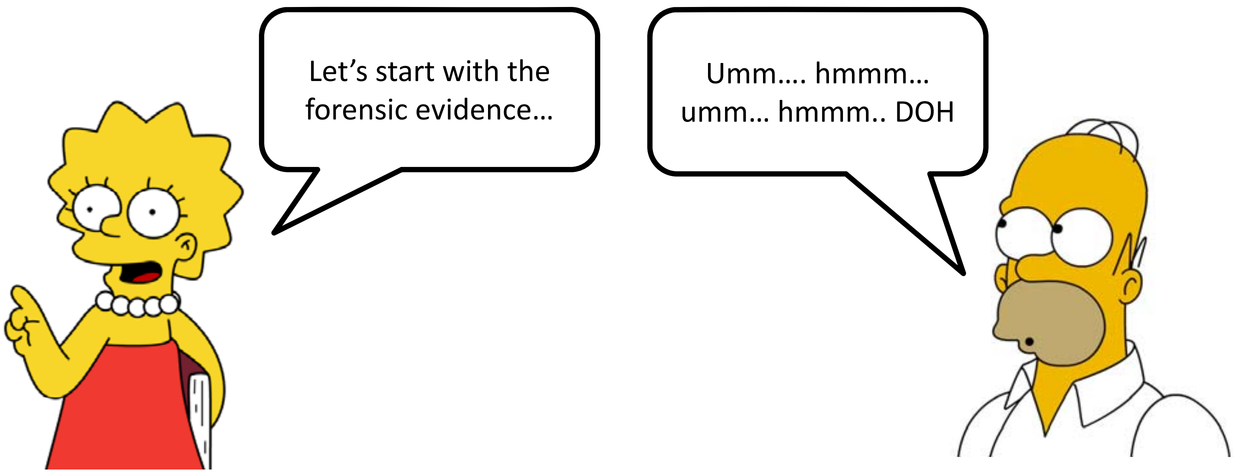
McGurk and MacDonald, Nature 1976

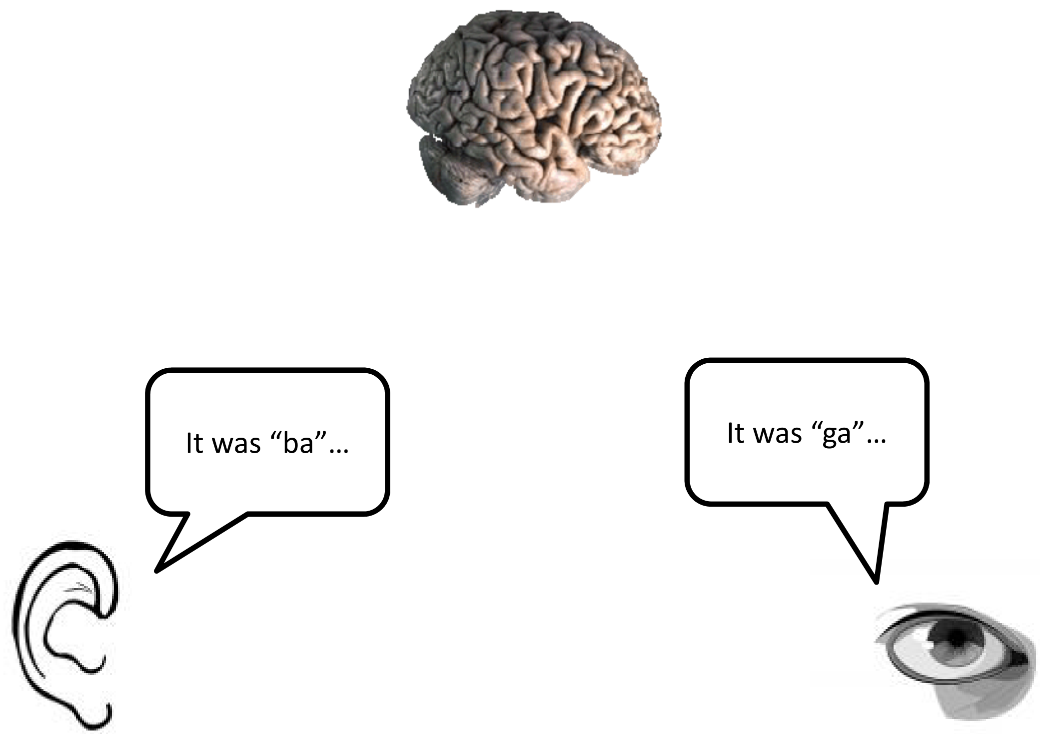
Demo from http://www.media.uio.no/personer/arntm/McGurk_english.html

Why does this happen?

- Syllables very similar → conflict not noticed
- Both stimuli come with uncertainty
- Integrating sound and vision is normally useful.
- The brain interprets observations in terms of their cause(s): perception as inference

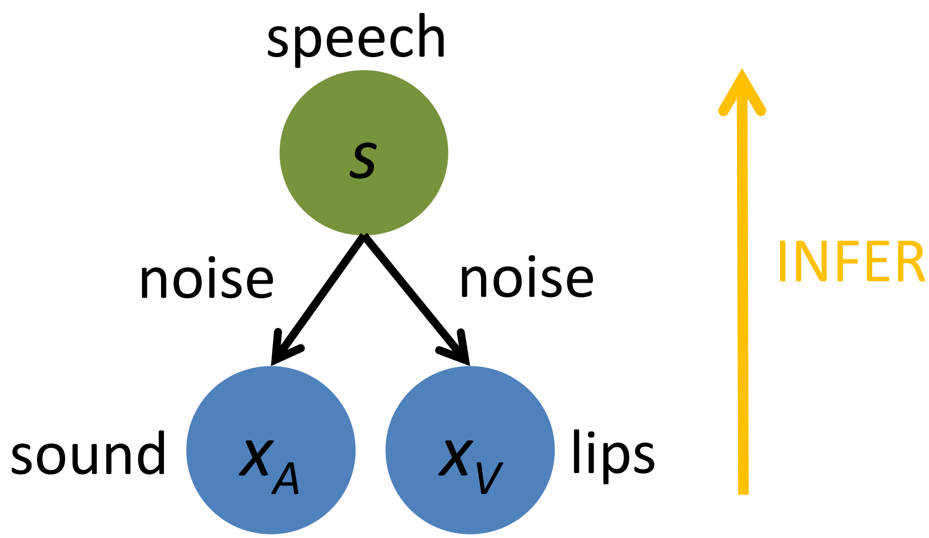



Let’s start with the forensic evidence…
Umm…. hmmm… umm… hmmm.. DOH



It was “ba”…
It was “ga”…

Generative model



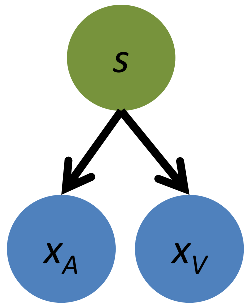
Exercise: what is the posterior over *s*, given this generative model?

$$p(s \mid x_A, x_V) \propto p(x_A, x_V \mid s)\, p(s)$$
$$= p(x_A \mid s)\, p(x_V \mid s)\, p(s)$$

Conditional independence → *multiplying* likelihood functions

Single source or two sources?

- This generative model assumes that there is a single source.
- In most cue integration experiments, there are in fact two sources.
- However, these are kept close enough for the subject to believe that the conflict is due to noise and that there is really one source.
- Later, we will examine the case when there can be one or two sources.



$$p(s \mid x_A, x_V) \propto p(x_A, x_V \mid s)\, p(s)$$
$$= p(x_A \mid s)\, p(x_V \mid s)\, p(s)$$

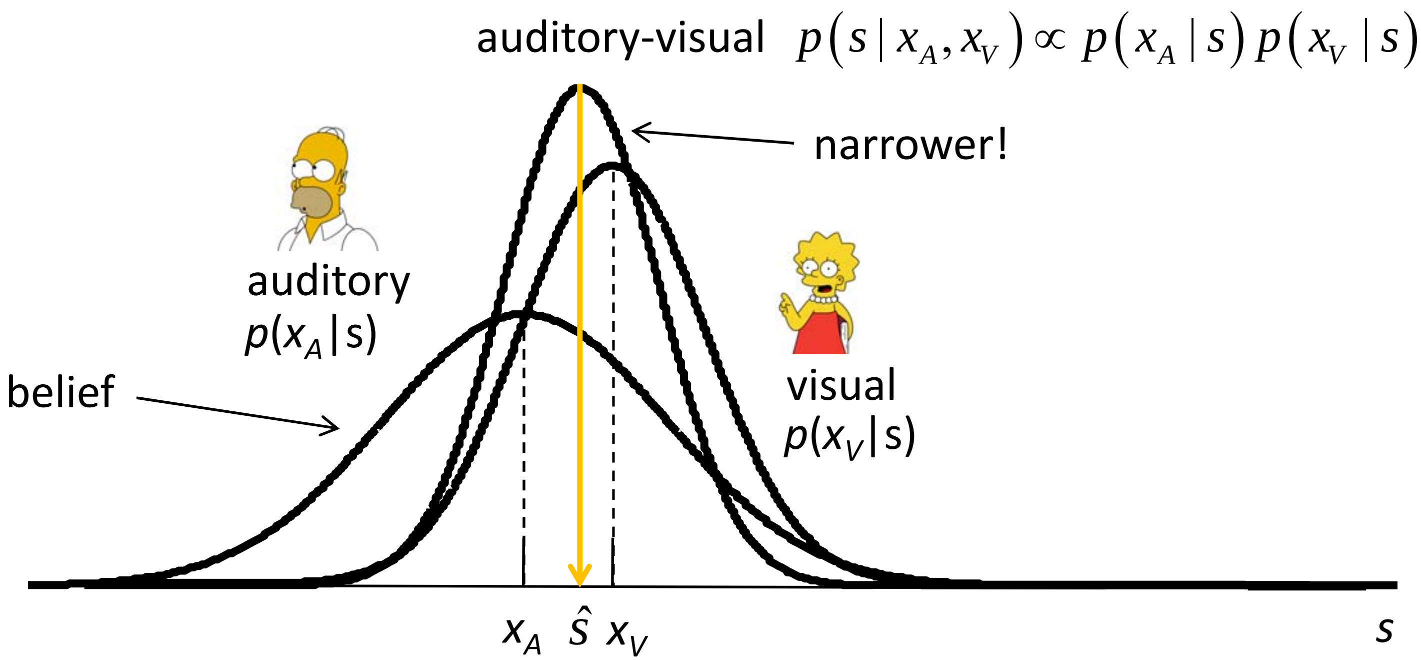
Assumptions about these distributions:

$$p(x_A \mid s) = \frac{1}{\sqrt{2\pi\sigma_A^2}} e^{-\frac{(x_A - s)^2}{2\sigma_A^2}}$$

$$p(x_V \mid s) = \frac{1}{\sqrt{2\pi\sigma_V^2}} e^{-\frac{(x_V - s)^2}{2\sigma_V^2}}$$

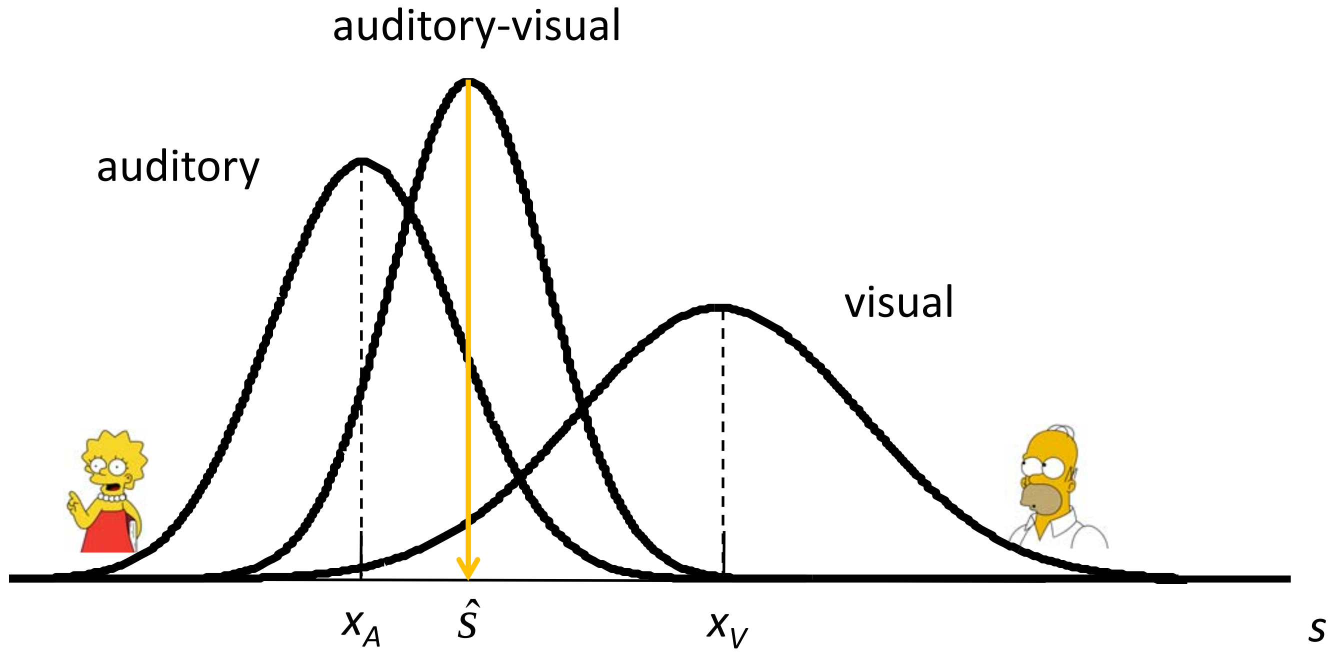
$$p(s) = \text{constant}$$

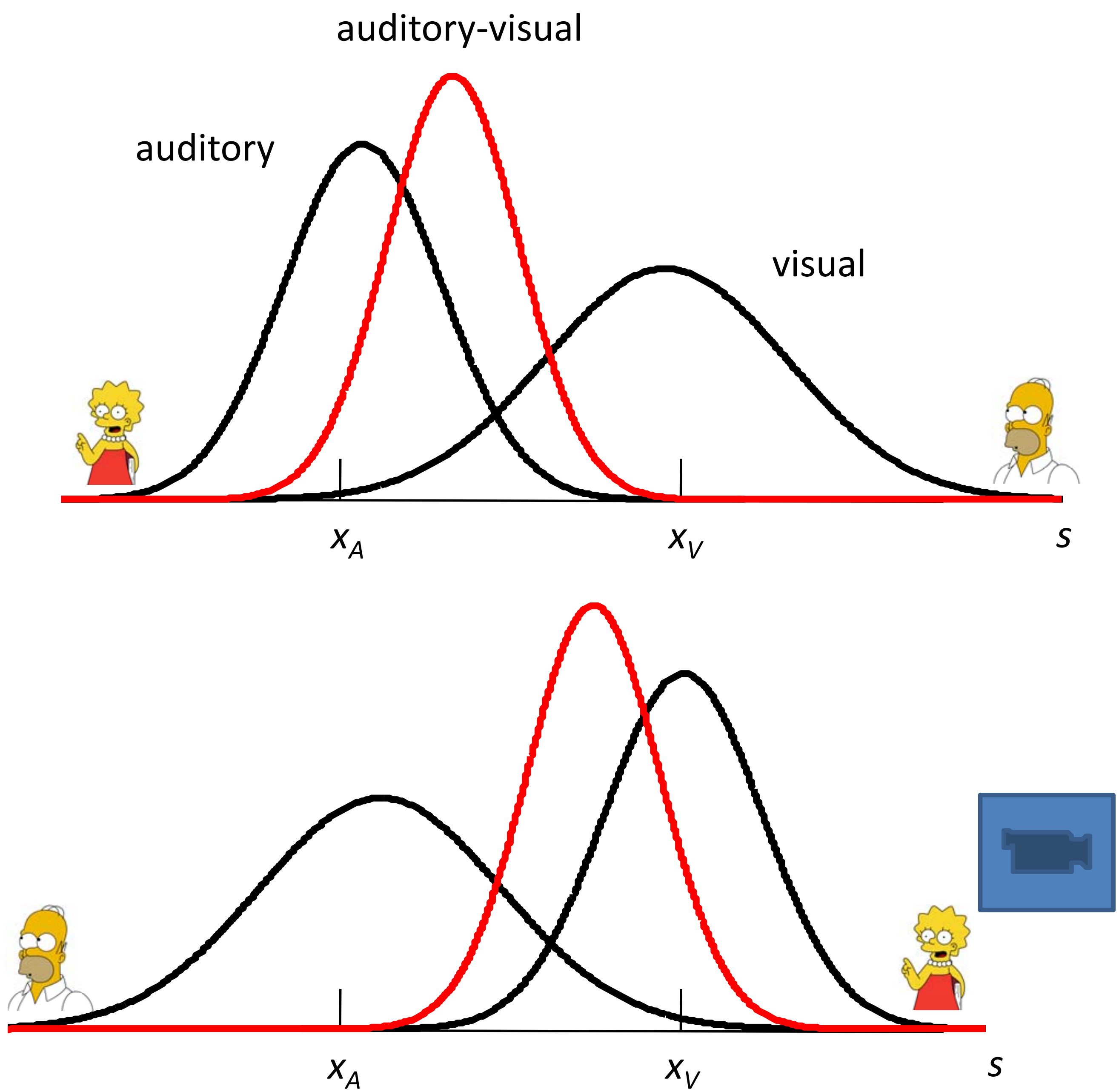
Cue integration without artificial conflict



Cue integration with artificial conflict

(not really different)




auditory-visual
auditory
visual
x_A
x_V
s
x_A
x_V
s

Exercise

Given $$p(s \mid x_A, x_V) \propto p(x_A \mid s)\, p(x_V \mid s)$$

$$p(x_A \mid s) = \frac{1}{\sqrt{2\pi\sigma_A^2}} e^{-\frac{(x_A - s)^2}{2\sigma_A^2}} \qquad p(x_V \mid s) = \frac{1}{\sqrt{2\pi\sigma_V^2}} e^{-\frac{(x_V - s)^2}{2\sigma_V^2}}$$

show that $p(s|x_A, x_V)$ is a normal distribution over s, with mean

$$\hat{s} = \frac{w_A x_A + w_V x_V}{w_A + w_V}$$ where $w_A = \frac{1}{\sigma_A^2}$ and $w_V = \frac{1}{\sigma_V^2}$

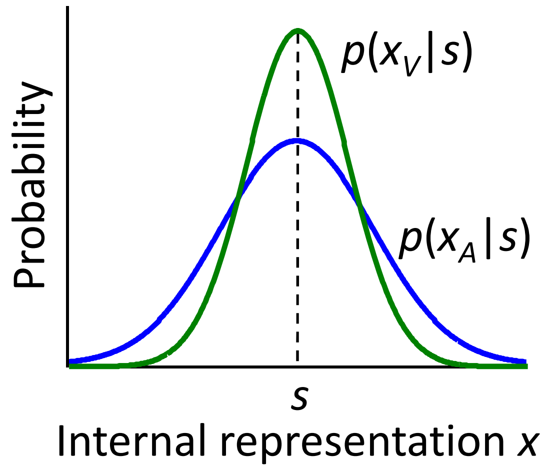
and standard deviation

$$\sigma_{AV} = \frac{\sigma_A \sigma_V}{\sqrt{\sigma_A^2 + \sigma_V^2}}$$ (or equivalently, $\frac{1}{\sigma_{AV}^2} = \frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2}$)

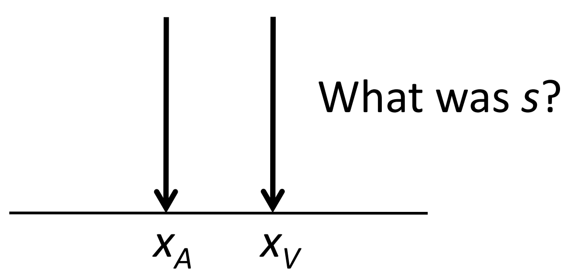
Weighting by reliability

- Urban legend about Bayesian inference: all about the prior
- Here: assumed flat prior, still Bayesian inference
- Key: taking into account uncertainty (σ_A, σ_V) on a single trial → allows weighting by reliability
- Requires knowledge of uncertainty
- Automatic in Bayesian coding: posterior distribution $p(s|\mathbf{r})$
- Bayesian inference is about keeping track of probability distributions over stimuli, instead of just single values.
- Another urban legend: Bayesian inference is the same as Bayesian decoding

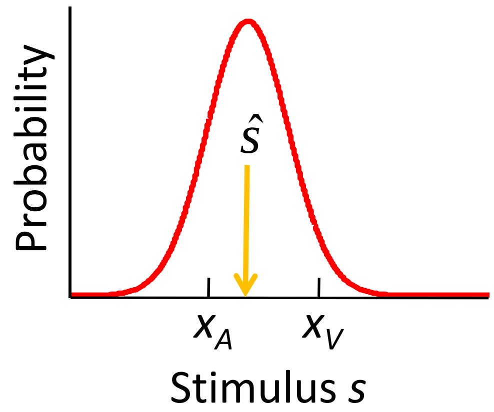
Over many trials, the internal representation follows a distribution when conditioned on a particular stimulus value.



However, **on a single trial**, the brain has to perform inference over the stimulus based on a single set of noisy internal representations, (x_A, x_V).



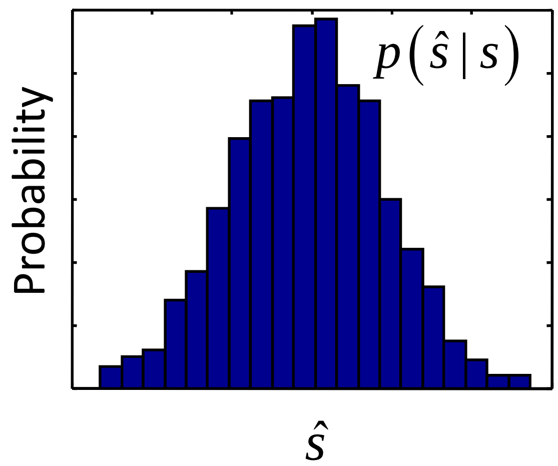
On a single trial, we have a posterior distribution over s, $p(s|x_A, x_V)$. However, this is a *belief*, not an empirical distribution.



On a single trial, the posterior produces a single response, $\hat{s}$.

$$x_A, x_V \longrightarrow \hat{s} = \frac{w_A x_A + w_V x_V}{w_A + w_V}$$

Across many repetitions of the same stimulus *s*, the responses form a response distribution. This distribution can be measured experimentally.



Experimental techniques:

- Estimation
- Discrimination → psychometric curve
  (can also be regarded as extra step in generative model)

Exercise

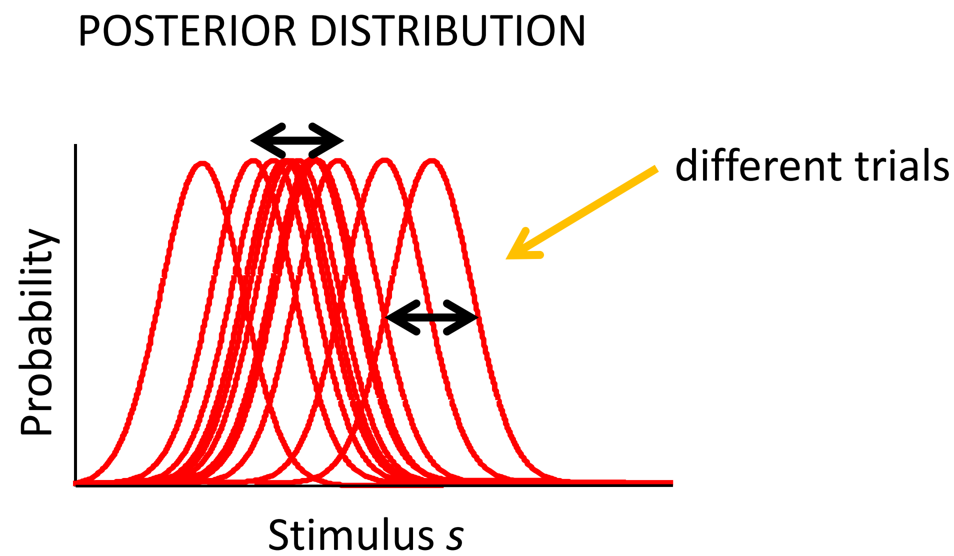
Given $\hat{s} = \frac{w_A x_A + w_V x_V}{w_A + w_V}$, calculate mean and variance of $p(\hat{s} \mid s)$

$w_A = \frac{1}{\sigma_A^2} \qquad w_V = \frac{1}{\sigma_V^2}$

What if x_A is drawn from $p(x_A \mid s_A) = \frac{1}{\sqrt{2\pi\sigma_A^2}} e^{-\frac{(x_A - s_A)^2}{2\sigma_A^2}}$

and x_V from $p(x_V \mid s_V) = \frac{1}{\sqrt{2\pi\sigma_V^2}} e^{-\frac{(x_V - s_V)^2}{2\sigma_V^2}}$?

The posterior wiggles around from trial to trial



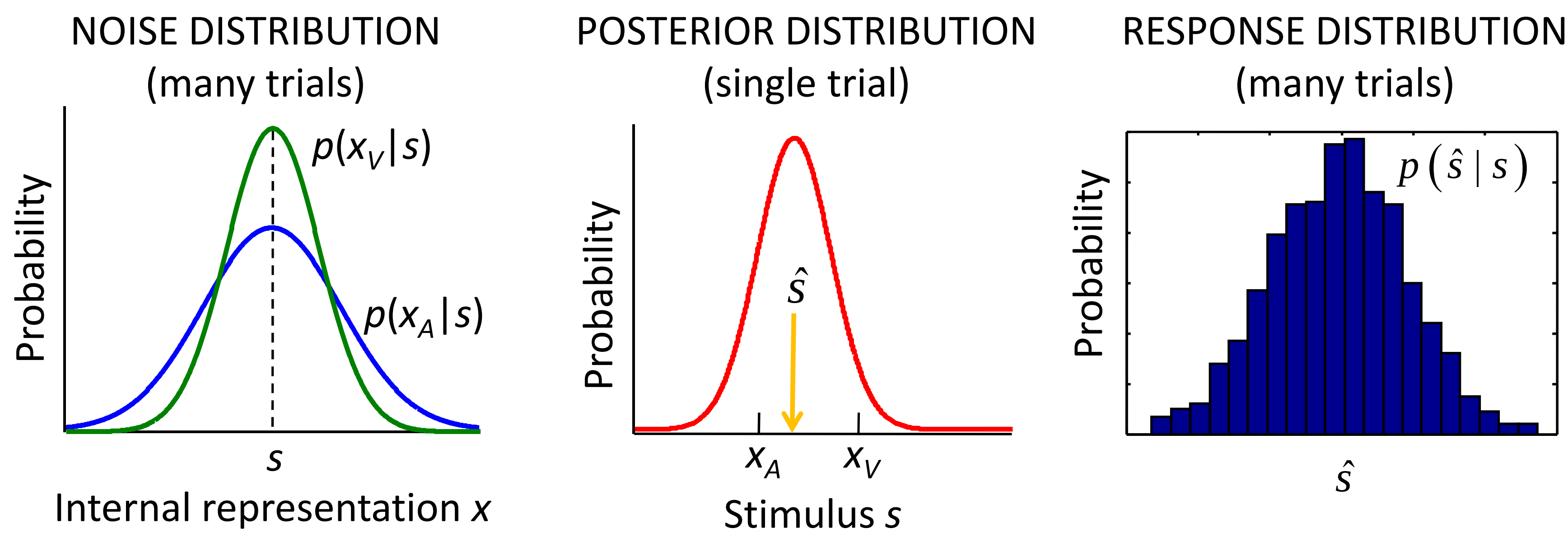
$$\sigma^2_{\text{response}} = \sigma^2_{\text{posterior}}$$

What?!

- The variance of the response distribution is equal to the variance of the posterior…
- The relation between a *single-trial estimate* and the *observations* is the same as that between the *mean estimate* and the *true stimuli*…
- Is this generally true?!

No!

Consequence of Gaussian distributions and multiplicative operation
In general there are three **completely different distributions**:

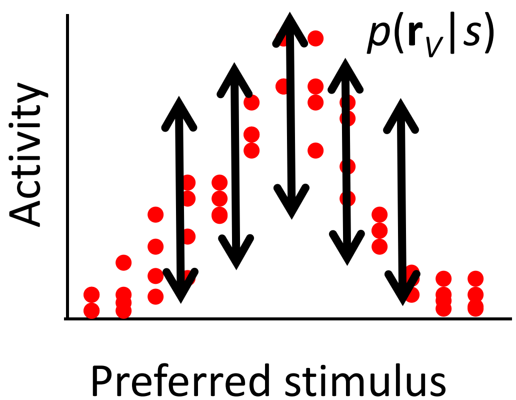


Do not confuse them! Common mistake in Bayesian modeling

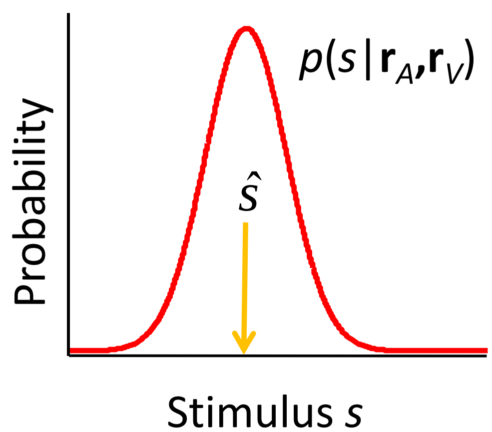
There is no direct way to measure the posterior (on a single trial)

Neural version

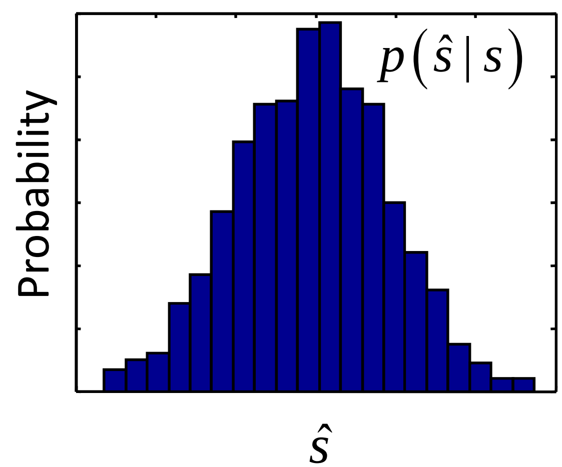
NOISE DISTRIBUTION
(many trials)



POSTERIOR DISTRIBUTION
(single trial)



RESPONSE DISTRIBUTION
(many trials)



Exercise

What is the general equation for the response distribution (assuming some decoder) in terms of the posterior distribution?

$$p(\hat{s} \mid s) = \ldots$$

Exercise

Work out a case where the posterior distribution and the response distribution are both continuous but very different from each other. (For example, choose non-Gaussian distributions and/or a more complex generative model.)

Bonus: make as general as possible the conditions under which the variance of posterior and response distribution are the same.

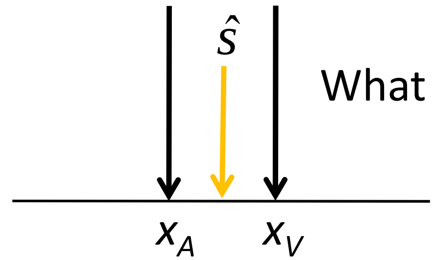
Fisher information

$$
\begin{aligned}
I_{AV}(s) &= -\left\langle \frac{\partial^2}{\partial s^2} \log p(x_A, x_V \mid s) \right\rangle \\
&= -\left\langle \frac{\partial^2}{\partial s^2} \log\left(p(x_A \mid s)\, p(x_V \mid s) \right) \right\rangle \\
&= -\left\langle \frac{\partial^2}{\partial s^2} \log p(x_A \mid s) \right\rangle - \left\langle \frac{\partial^2}{\partial s^2} \log p(x_V \mid s) \right\rangle \\
&= I_A(s) + I_V(s)
\end{aligned}
$$

Optimal cue integration preserves Fisher information.

What does this mean in the Gaussian cue integration case?

Non-optimal cue integration



$$\hat{s} = \frac{x_A + x_V}{2}$$

Example: suppose $\sigma_A^2 = 100$ and $\sigma_V^2 = 1$

Then the variance of the estimate is 101/4 ≈ 25

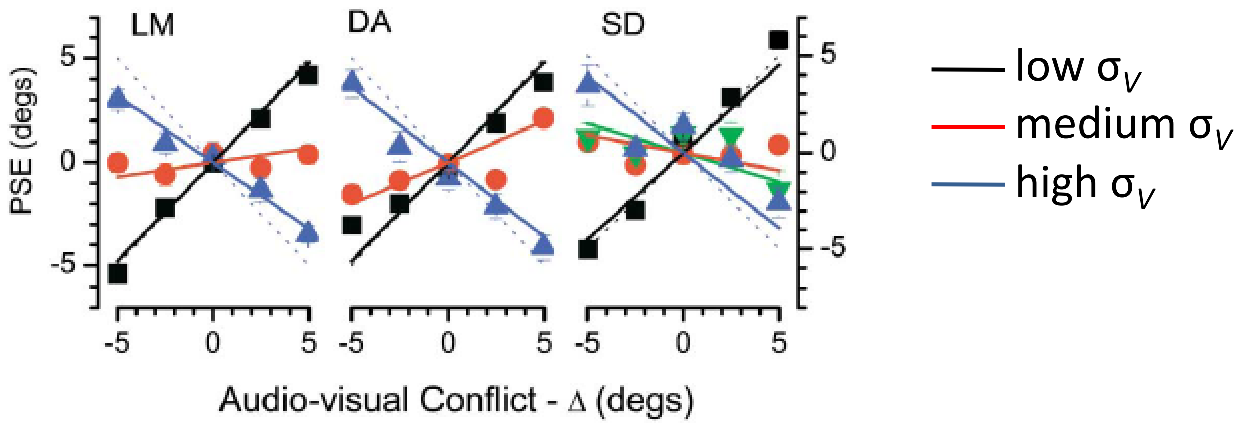
Optimal estimate: $\hat{s} = \frac{0.01 x_A + x_V}{1.01}$

Variance of optimal estimate: 1/(0.01 + 1) ≈ 0.99

Multisensory bias

In the presence of a cue conflict, $s_V = \Delta$, $s_A = -\Delta$, what is the mean multisensory estimate?

$$\langle \hat{s} \rangle = \frac{\sigma_A^2 - \sigma_V^2}{\sigma_A^2 + \sigma_V^2}\Delta$$



Lines are predicted slopes from unisensory experiment

Alais and Burr, 2004

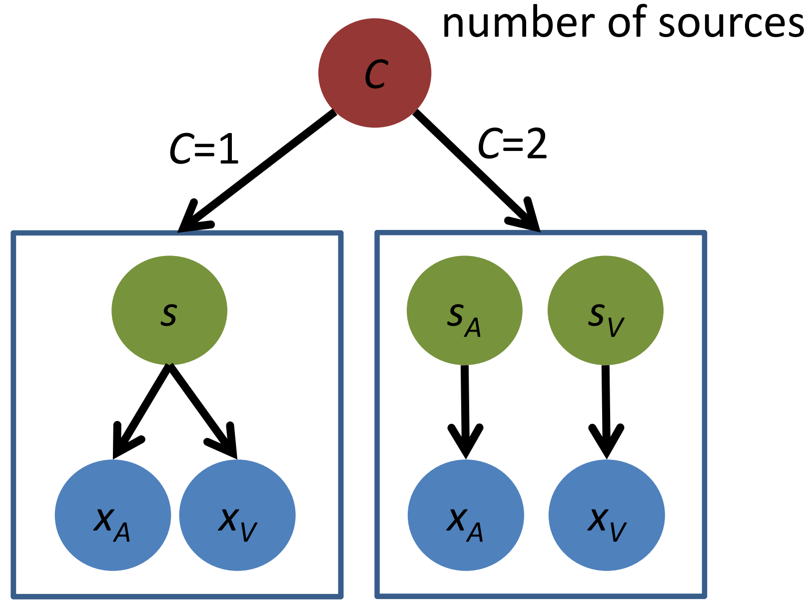
Exercise

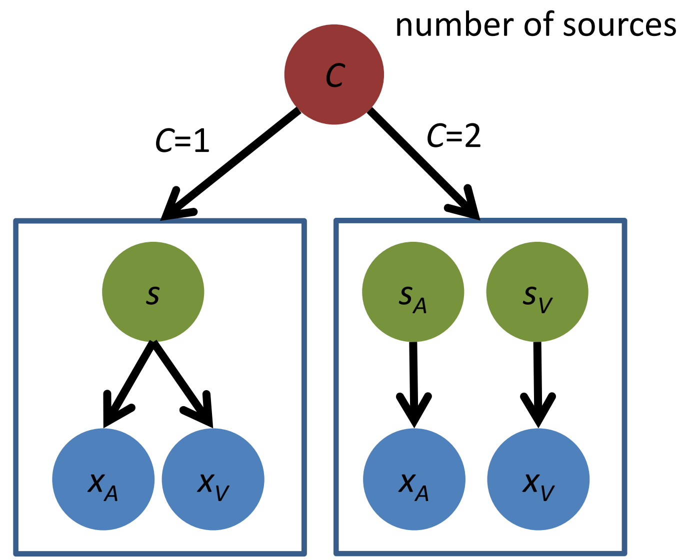
Even when a single stimulus has to be inferred from a single cue, a bias can arise due to a prior. Assuming a Gaussian noise model and a Gaussian prior (with specified mean and variance), compute the bias as a function of the stimulus.

Causal inference

- You don’t always integrate cues
- Two cues often have two different sources
- How to decide whether there are one or two sources?
- Bayesian inference on number of sources!

Generative model





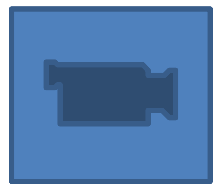
$$p\left(s_A \mid x_A, x_V\right) = \sum_{C=1}^{2} p\left(s_A \mid x_A, x_V, C\right) p\left(C \mid x_A, x_V\right)$$

$$\begin{aligned} p\left(C \mid x_A, x_V\right) &\propto p\left(x_A, x_V \mid C\right) p\left(C\right) \\ &= p\left(C\right) \iint p\left(x_A, x_V \mid s_A, s_V\right) p\left(s_A, s_V \mid C\right) ds_A ds_V \\ &= p\left(C\right) \iint p\left(x_A \mid s_A\right) p\left(x_V \mid s_V\right) p\left(s_A, s_V \mid C\right) ds_A ds_V \end{aligned}$$

$$p\left(s_A, s_V \mid C=1\right) = k\delta\left(s_A - s_V\right)$$

$$\begin{aligned} p\left(C=1 \mid x_A, x_V\right) &= kp\left(C=1\right) \iint p\left(x_A, x_V \mid s_A, s_V\right) \delta\left(s_A - s_V\right) ds_A ds_V \\ &= kp\left(C=1\right) \int p\left(x_A \mid s_A\right) p\left(x_V \mid s_A\right) ds_A \\ &= kp\left(C=1\right) \frac{1}{\sqrt{2\pi\left(\sigma_A^2 + \sigma_V^2\right)}} e^{-\frac{\left(x_A - x_V\right)^2}{2\left(\sigma_A^2 + \sigma_V^2\right)}} \end{aligned}$$

Ventriloquist effect



Bayesian explanation?

Small project 1

- Auditory-visual speech perception data
- Identify a syllable as /ba/ or /da/
- Factorial design
- In each condition, % responses "/ba/" and "/da/"
- Predict responses using a Bayesian model
- *Compare predictions* with those of established model (FLMP)

		Auditory					
		BA	2	3	4	DA	none
Visual	BA						
	2						
	3						
	4						
	DA						
	none						

Massaro et al., 1993
http://mambo.ucsc.edu/psl/data/mass93a.html