The neuroscience of probability

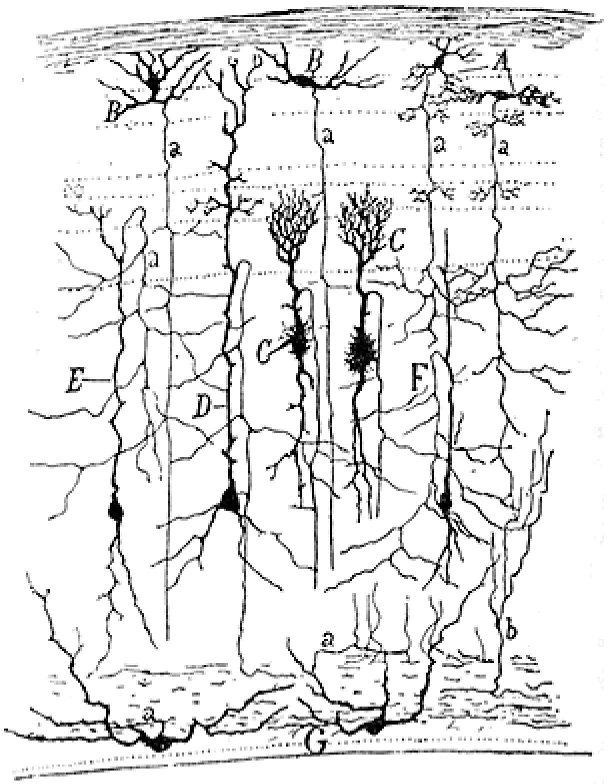
Wei Ji Ma

Baylor College of Medicine

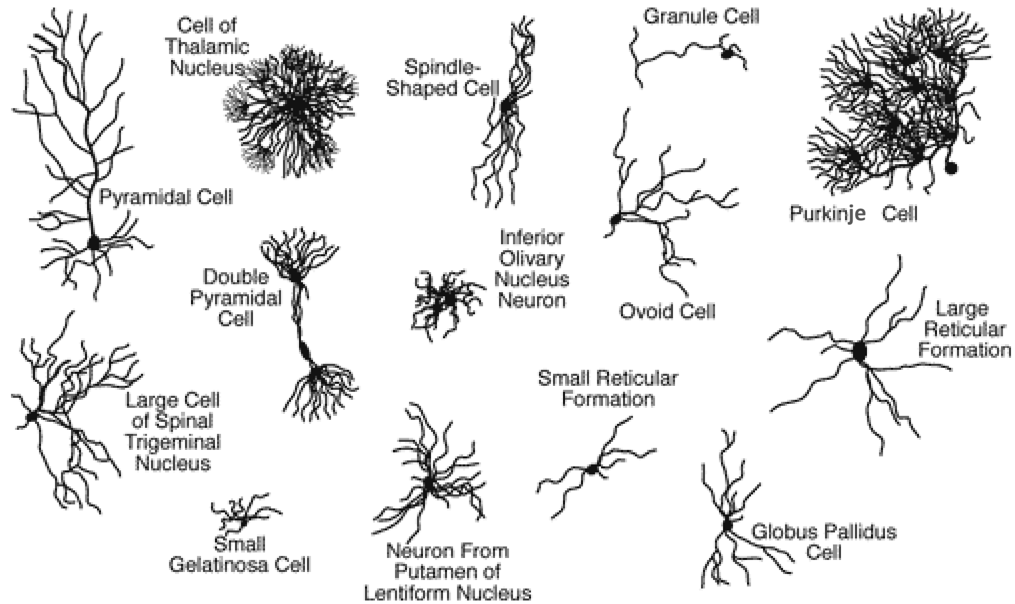
10^{10}-10^{11} neurons

Each neuron has 10^{3}-10^{4} synaptic connections

Neuron diversity

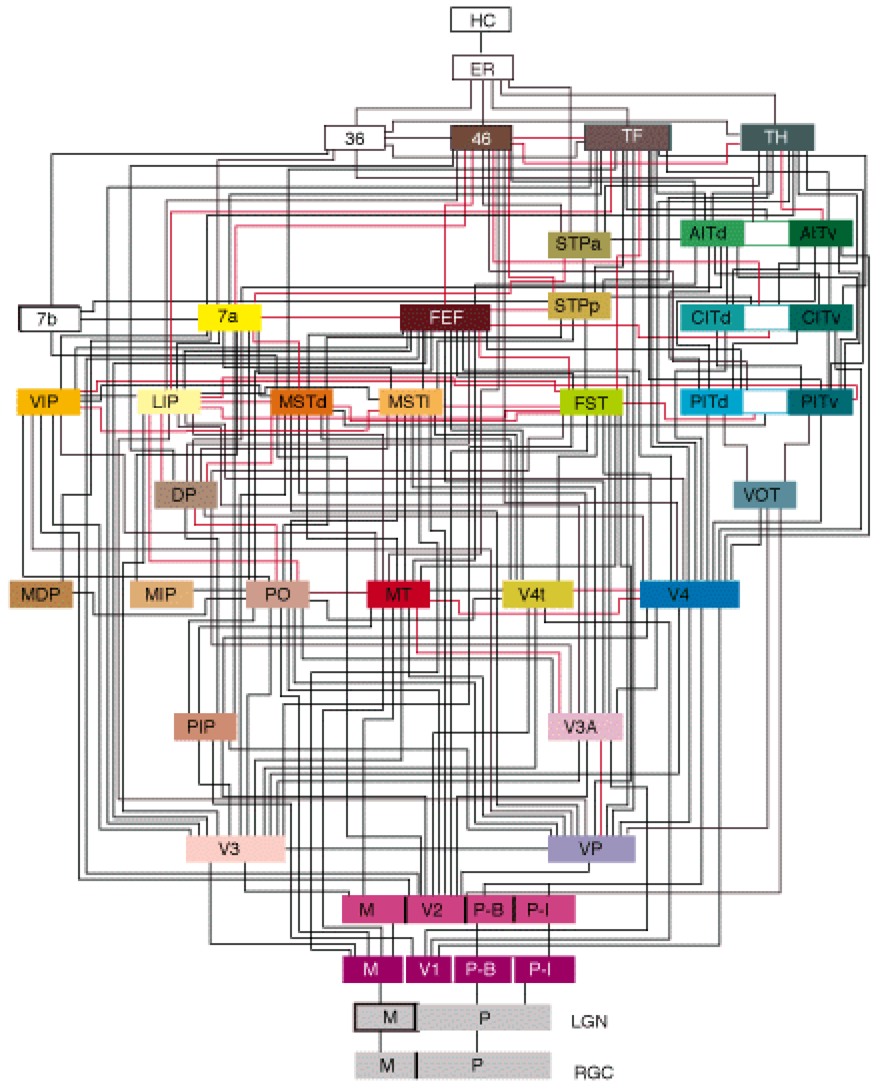


Optic tectum of the sparrow
Ramón y Cajal, 1905



R. Stufflebeam, The Mind Project
Based on drawings by Cajal

Circuit diagram of macaque visual areas



Felleman and Van Essen, 1991

Levels of organization

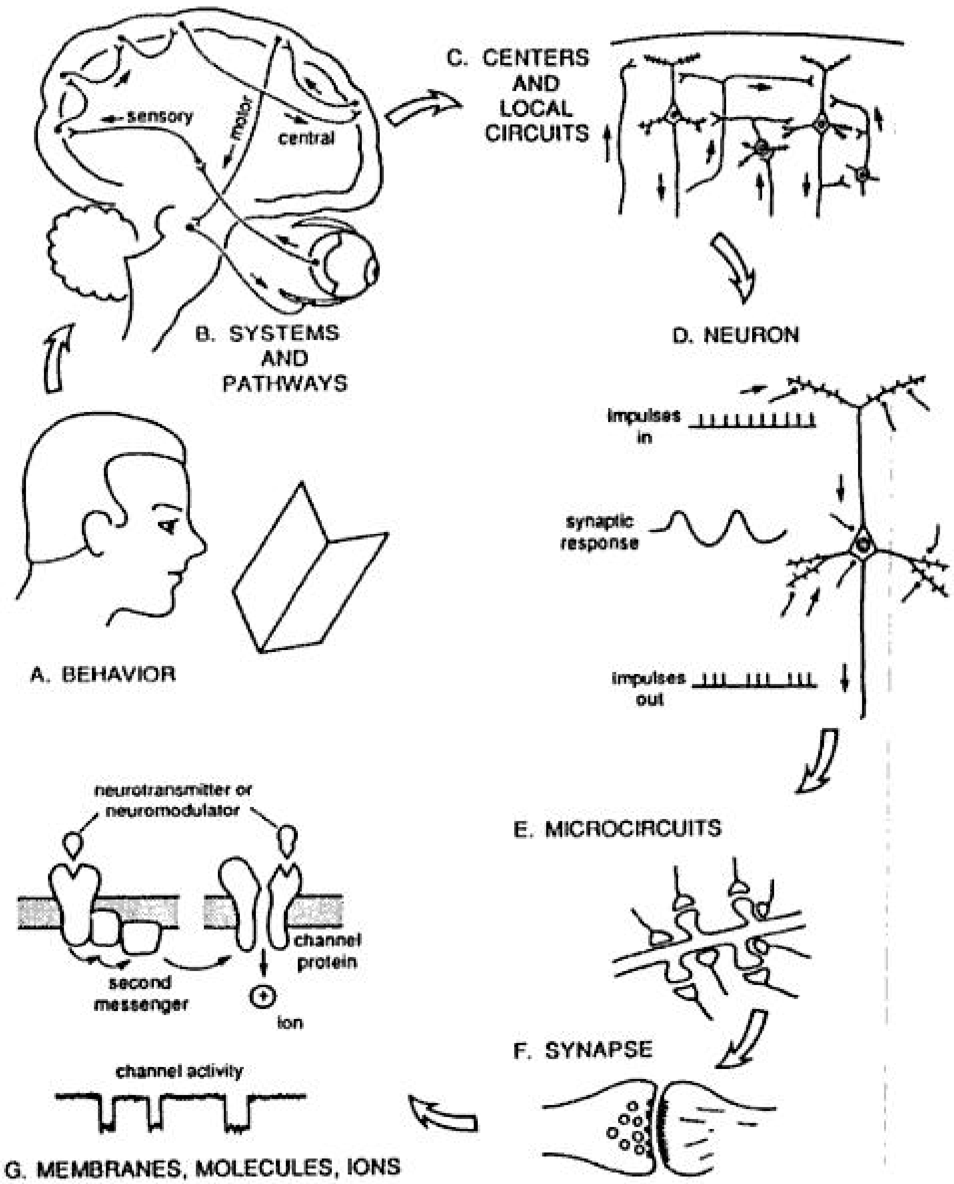


Figure 2.1 Levels of organization in the nervous system, as characterized by Gordon Shepherd (1988a).

Difficult question:

How to model the brain?

Simulating the brain bottom-up

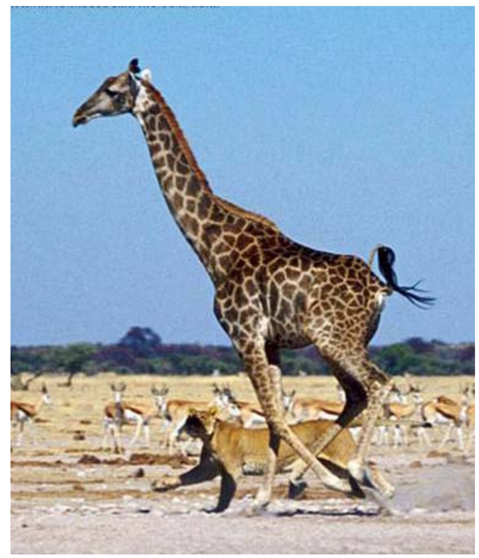
Blue
Brain
Project

Model neocortical microcircuit

What?

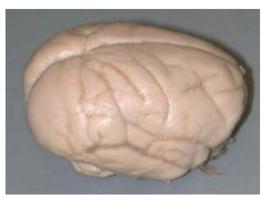
Why?

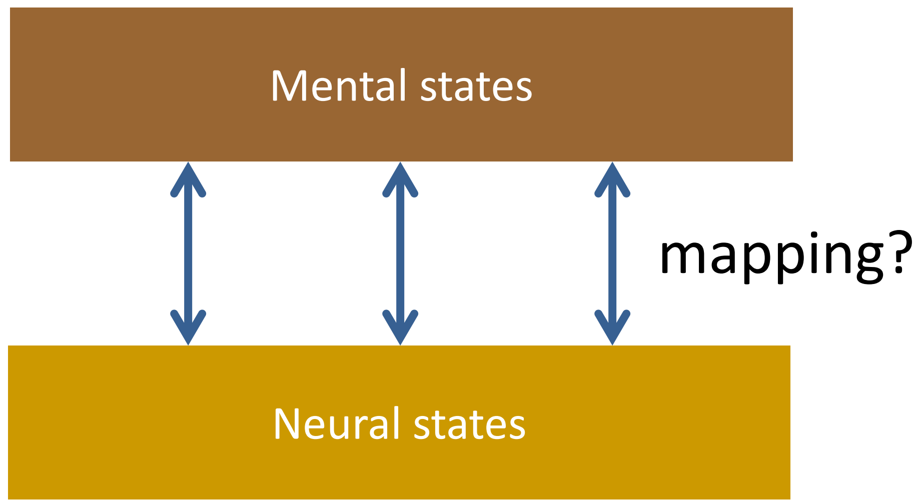
Why do we have a brain?



Essential function: using sensory information to guide behavior

Top-down view: Start with behavior





Goal: to understand the relationship between neural and mental states in quantitative terms

Approach: normative/optimality: what *should* the brain be doing? → is it really doing that?

System: perception in humans

Gareth Oliver – Britain's Got Talent 2009

How is our brain fooled into thinking that the puppet is talking?

Hypothesis 1: The puppet is talking.

Support: - We see that the puppet's movements match the speech.
- We see that the human's face isn't moving.

Hypothesis 2: The human is talking.

Support: - We know that most puppets don't come with sound.
- We (kind of) hear the sound coming from the human.

How does our brain decide between these two hypotheses?

- Optimal: compute the *probability* of each from the observations and prior knowledge.

  probability that the puppet is talking given observations

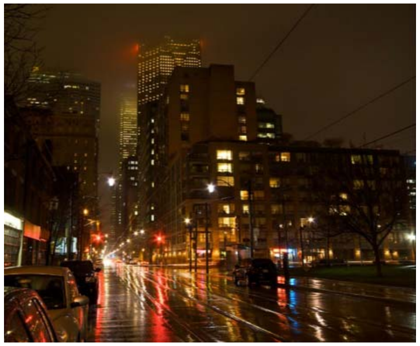
  probability that the human is talking given observations

- What is perceived is the hypothesis with the highest probability.
- Claim: *all perception* consists of computing probabilities

Computing probabilities means we are uncertain.

Why can’t we be certain?

Low-quality input

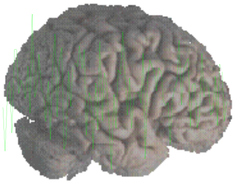


It's dark

Visibility is low

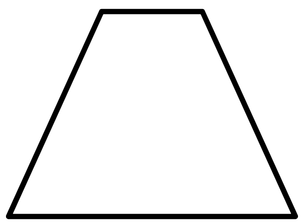
Stuff is far away

Stuff happens in the periphery

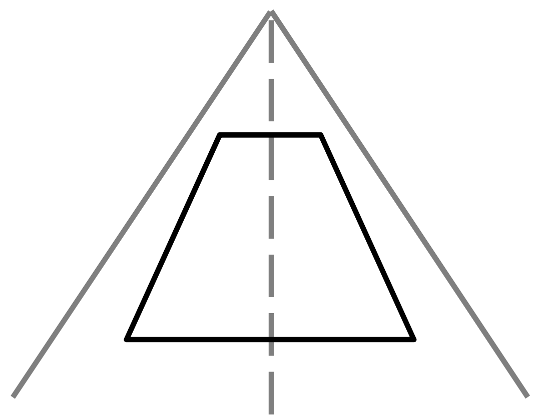


Noise in the brain

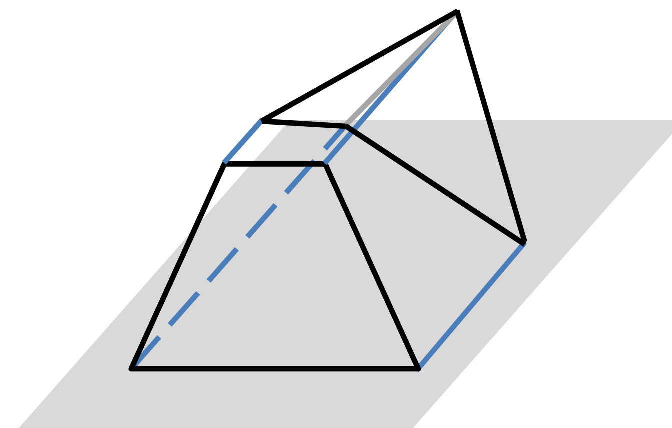
Ambiguity



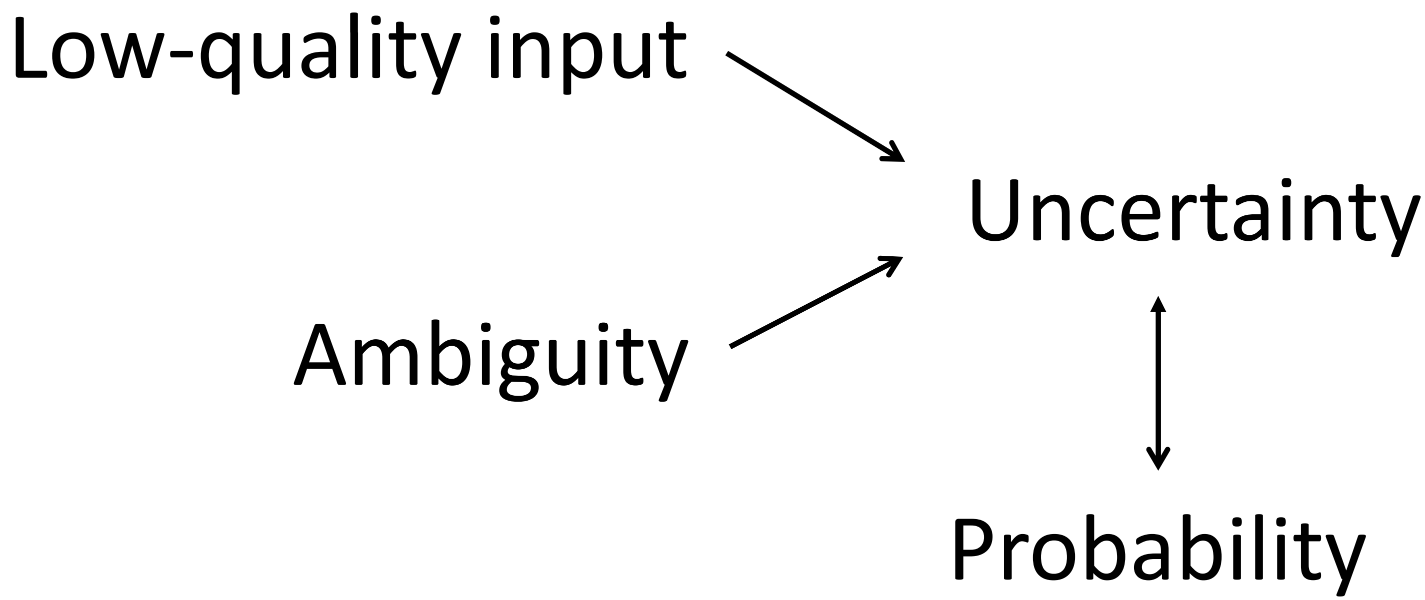
a trapezoid



a rectangle on a road



a weird wire frame


Low-quality input
Ambiguity
Uncertainty
Probability

Perception as inference

- The brain, forced to interpret *low-quality* and *ambiguous* observations, computes probabilities: (probabilistic) inference.

probability of hypothesis given observations

p **(hypothesis | observations)**

- The brain is not a recording device!

Interpretation

↕

Inference

Al Hazen (Ibn al-Haytham), 965-1040
"Perception requires unnoticed judgments."

Pierre-Simon Laplace, 1749-1827
"One may even say, strictly speaking, that almost all our knowledge is only probable."

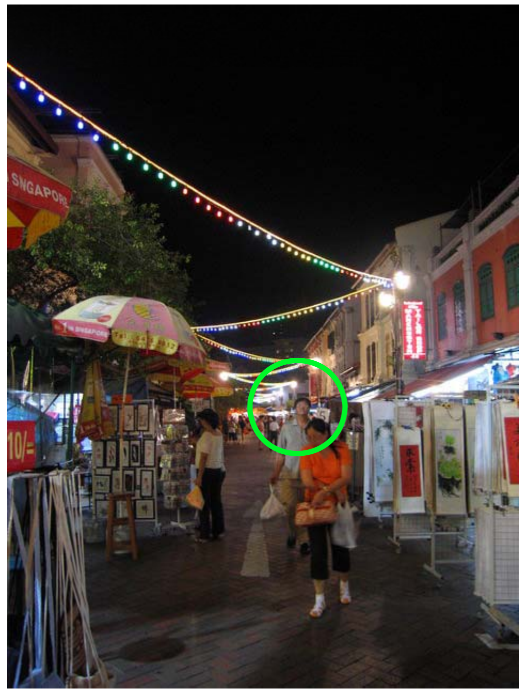
Hermann von Helmholtz, 1821-1894
"Perception is unconscious inference."

Is almost all our knowledge only probable?

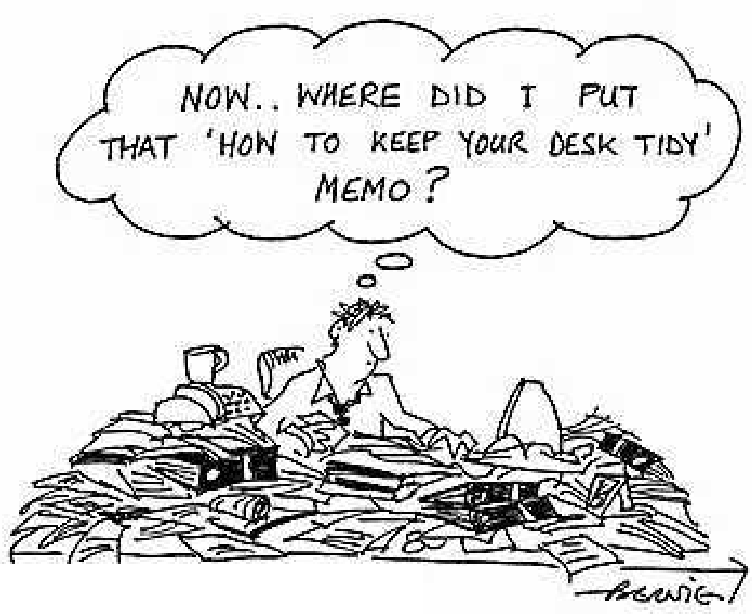
Do we compute p (hypothesis | observations) ?

Let’s look at daily life!

Perception…



p(that is my friend | visual information)



p(memo is present on desk | messy visual information)

Prediction…

p(it will rain | atmospheric data)

p(I will get sick if I eat this apple | look, smell)

p(my teammate will catch the ball | peripheral visual information)

p(I can jump over the stream | visual information, jumping ability)

Complex decision-making…

How did you decide to come here?

p(this is an interesting REU | announcement)

p(this is a nice place to work | first impression)

p(this is the guy we need| first impression)

p(he’s the one| his behavior)

p(it will rain | atmospheric data)

p(I will get sick if I eat this apple | look, smell)

p(my teammate will catch the ball | peripheral visual information)

p(I can jump over the stream | visual information, jumping ability)

p(that is my friend | visual information)

p(this is a nice place to work | first impression)

p(this is the guy we need| first impression)

p(he's the one| his behavior)

↑

***p*(hypothesis| observations)**

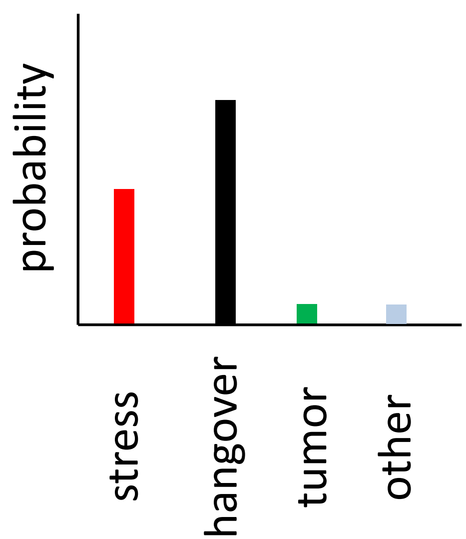
The brain interprets.

Probabilities are everywhere.

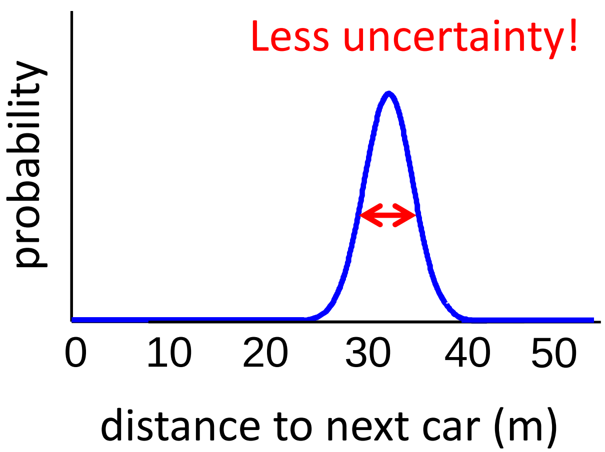
All yes/no variables!

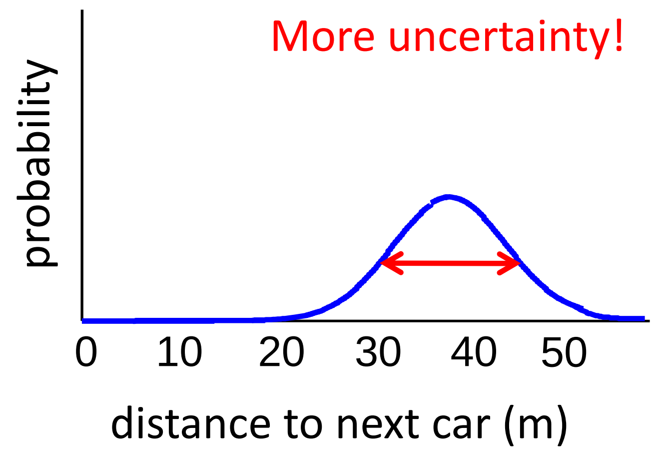
… but easily extends to variables with >2 possible values

p(stress | headache)
p(hangover | headache)
p(brain tumor | headache)
p(other | headache)



... and to *continuous* variables





"We (kind of) hear the sound coming from the human."

A small experiment…

How does the brain decide who's talking?

Prior knowledge:

p(puppet talking) = 0.20

p(human talking) = 0.80

Probabilities given observations:

p(puppet talking | visual observations) = 0.95

p(human talking | visual observations) = 0.05

p(puppet talking | auditory observations) = 0.40

p(human talking | auditory observations) = 0.60

Optimally combined probabilities:

$$\frac{p(\text{puppet talking} \mid \text{observations})}{p(\text{human talking} \mid \text{observations})} = \frac{0.20}{0.80} \times \frac{0.95}{0.05} \times \frac{0.40}{0.60} = 3.2$$

→ It is more probable that the puppet is talking!

Bayes' rule

posterior probability

$$p(\text{hypothesis} \mid \text{observations}) \propto$$

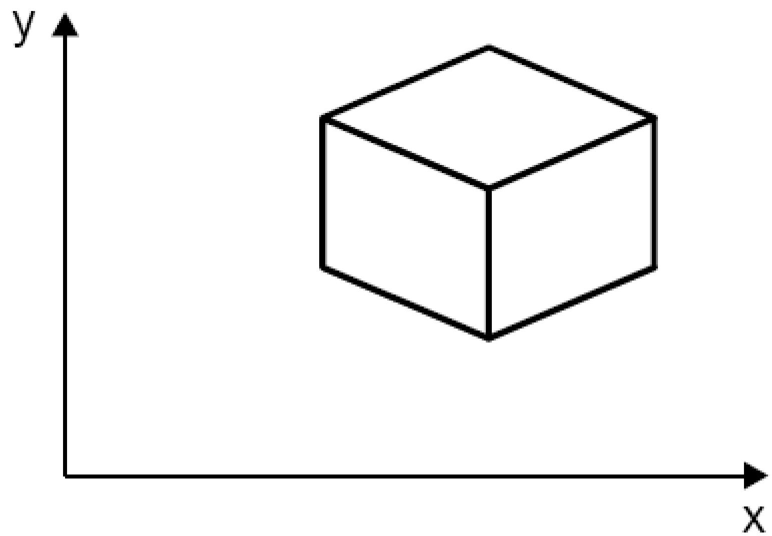
$$p(\text{observations} \mid \text{hypothesis}) \; p(\text{hypothesis})$$

likelihood of hypothesis　　prior probability

Example: object recognition

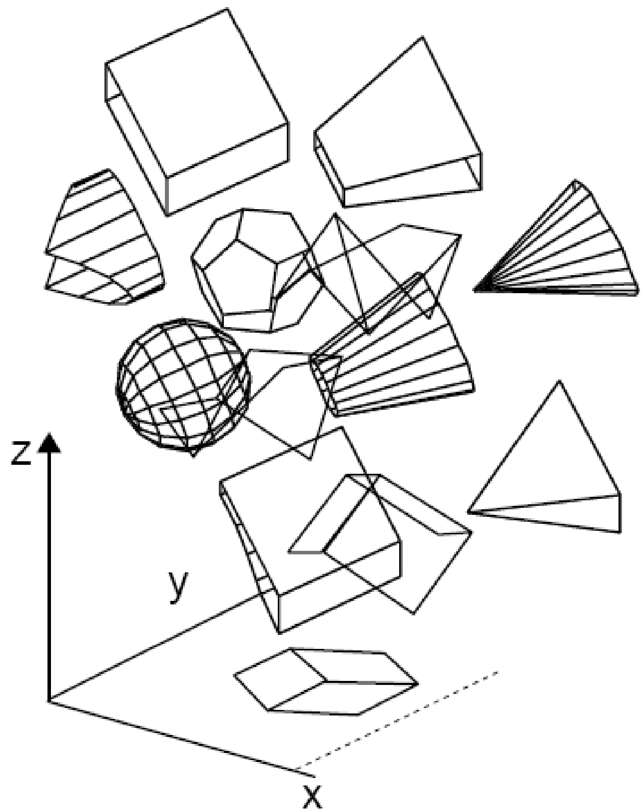
s: object identity

image data *I*

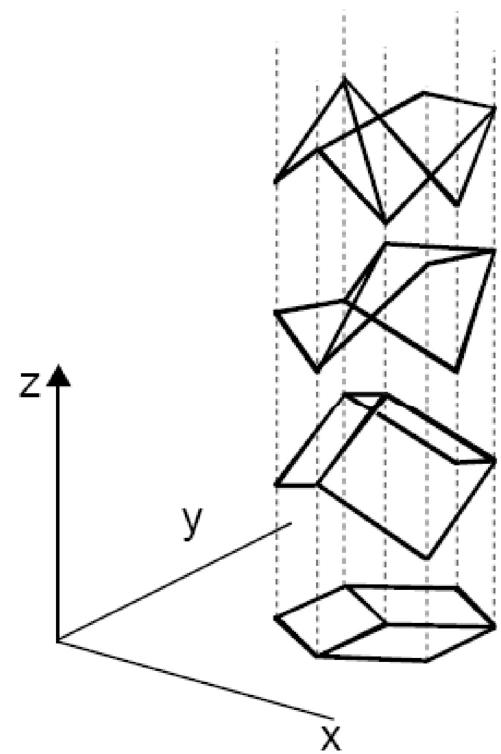


Kersten and Yuille, 2003

Prior over objects
$p(s)$

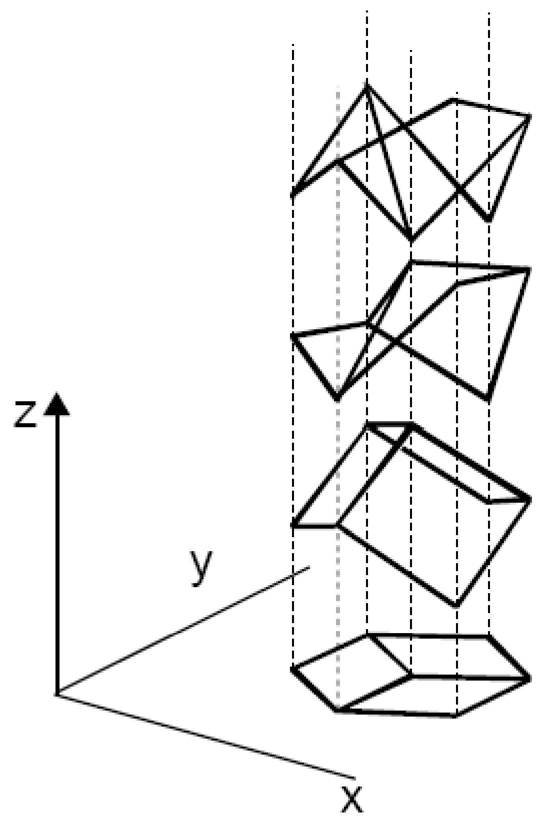


Likelihood over objects given 2D image
$L(s) = p(I|s)$



Kersten and Yuille, 2003

Posterior over objects
$p(s|I)$



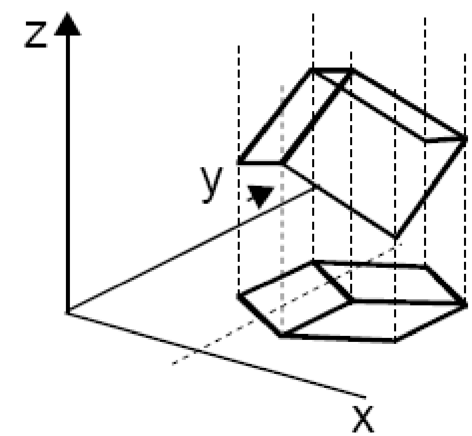
$p(s_1|I) = p_1$

$p(s_2|I) = p_2$

$p(s_3|I) = p_3$

Perceive the object with the highest posterior probability

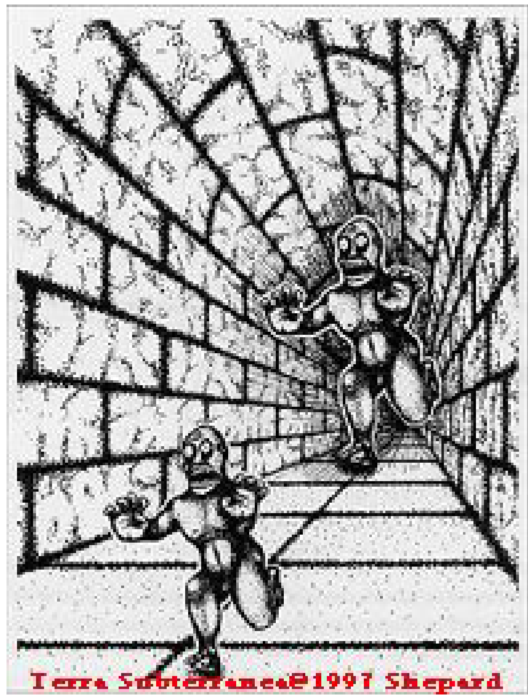
$p(s_3|I) = p_3$ is biggest



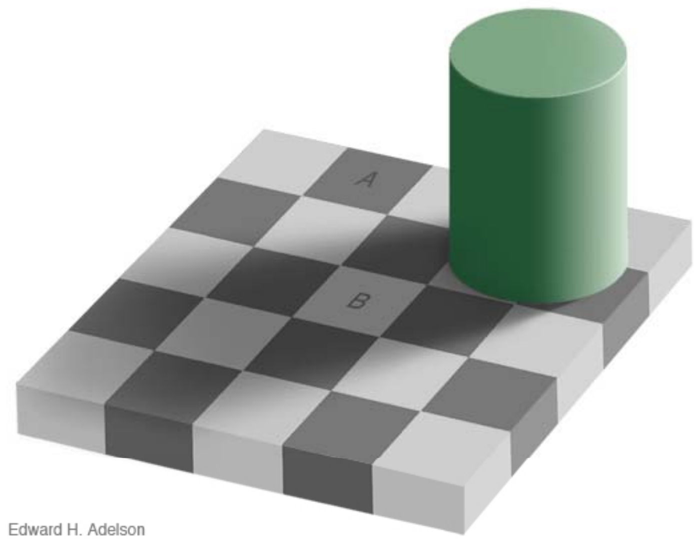
Kersten and Yuille, 2003

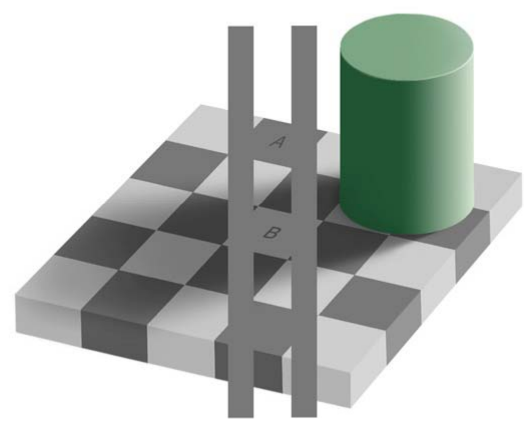
Many perceptual effects can be explained as consequences of Bayesian inference.

Ponzo illusion

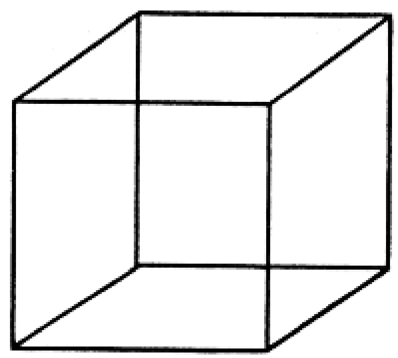


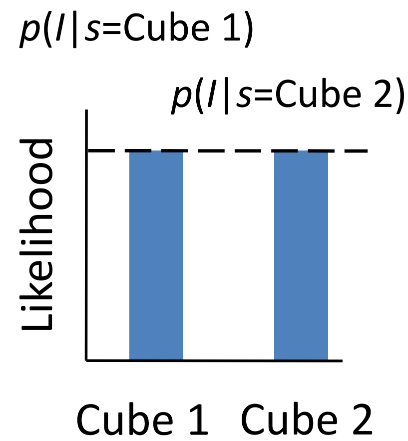
Checkerboard shadow illusion





Uninformative likelihood: Necker cube





Prior probability comes to the rescue:

$$\frac{p\left(s=\text{Cube 1}\mid I\right)}{p\left(s=\text{Cube 2}\mid I\right)}=\frac{p\left(I\mid s=\text{Cube 1}\right)p\left(s=\text{Cube 1}\right)}{p\left(I\mid s=\text{Cube 2}\right)p\left(s=\text{Cube 2}\right)}\approx\frac{p\left(s=\text{Cube 1}\right)}{p\left(s=\text{Cube 2}\right)}$$

Uninformative likelihoods in nature

Peppered moth caterpillar
Biston betularia
Noor et al., PLOS ONE 2008

Uninformative likelihood: hidden messages?

- Led Zeppelin, *Stairway to heaven* (1971)
- Played backwards
- Satanism?

Or the Story of Aidan?

Daniel Goldreich and lab, McMaster University

Aidan and the Quicksand: *In a faraway land, a little boy named Aidan wanders lost in the wilderness. Desperate, his father consults a shaman who can hear the "broken music" of the spirit world. The shaman reports that Aidan made a campfire for warmth, but that the fire burned out as the wood turned to ash. Ash ("quicksand") is an evil power that, unchecked by the heat of fire, escapes to engulf its victims. Aidan's father sings this lament: …*

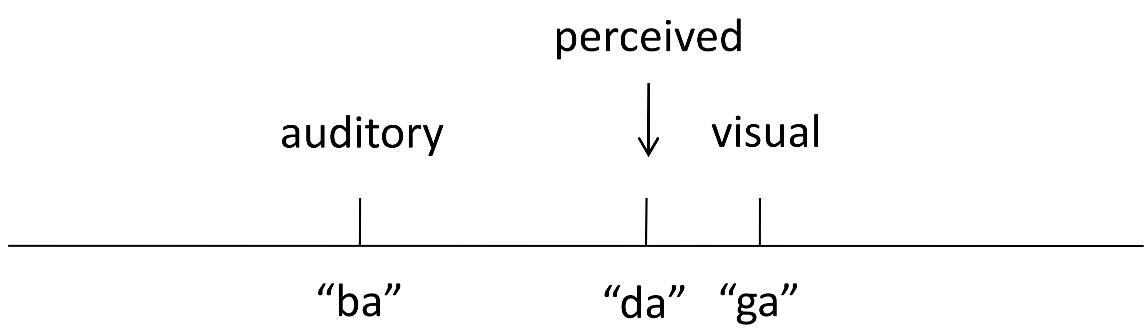
How to do science on this?

- Theory: the human brain optimally computes probabilities in making perceptual judgments
- Experimental test: measure human responses in a controlled behavioral task (psychophysics)
- Compare with alternative theories

Speech recognition gone wild

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

McGurk effect

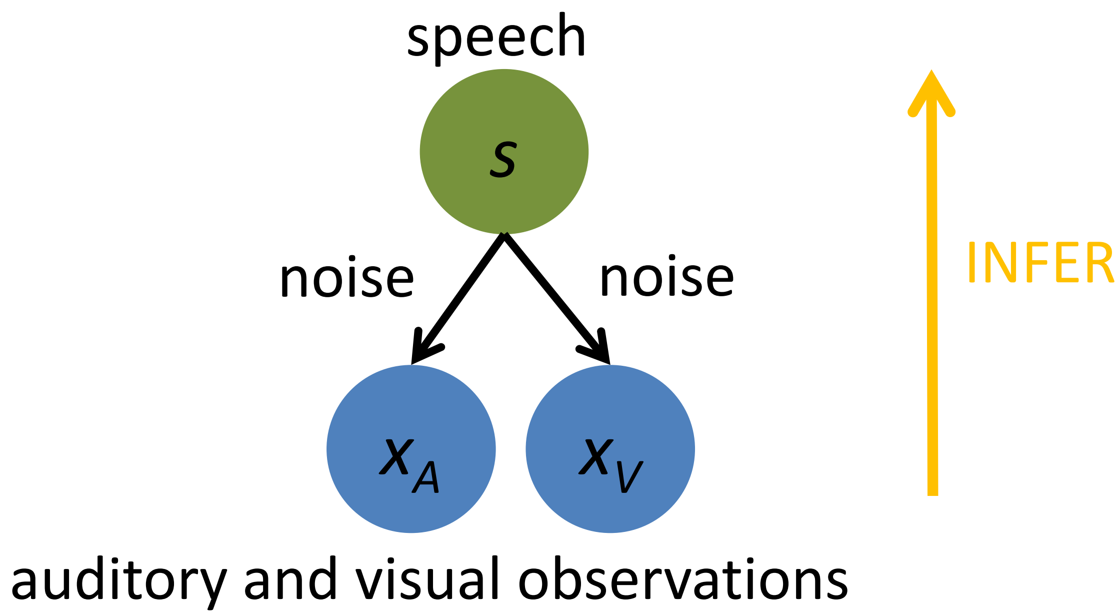


McGurk and MacDonald, Nature 1976

Why does this happen?

- Hypotheses: “ba”, “ga”, ”da”, other syllables
- Noisy auditory (A) evidence for “ba”
- Noisy visual (V) evidence for “ga”
- The brain computes p(syllable | A,V): *cue combination*

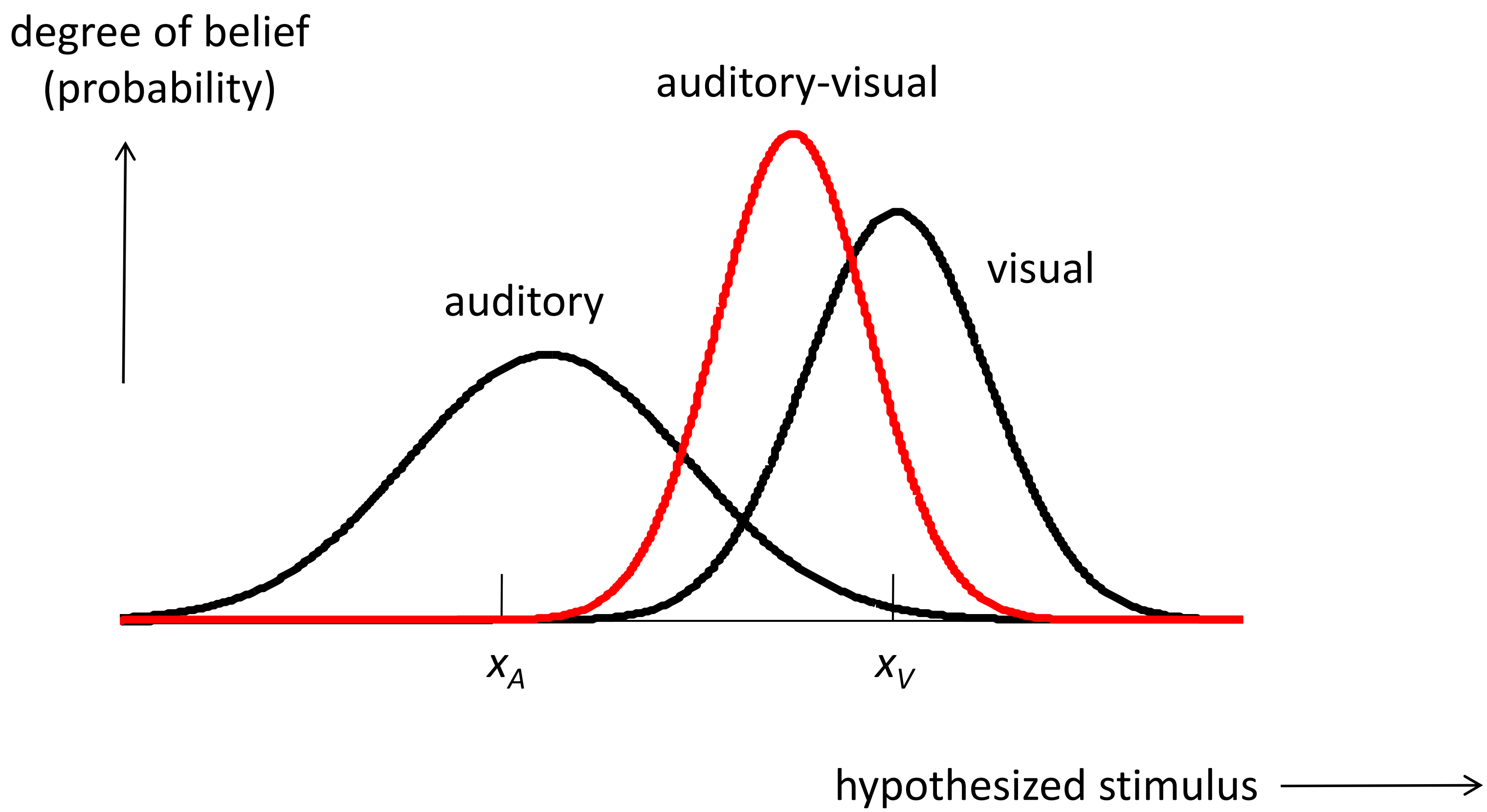
Generative model



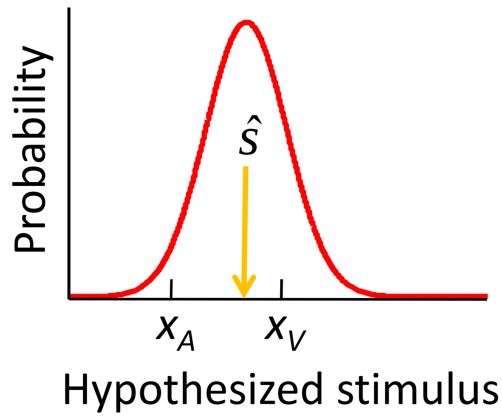
Bayes' rule →

$$p\left(s \mid x_A, x_V\right) \propto p\left(x_A \mid s\right) p\left(x_V \mid s\right)$$

Computing the posterior



What was s?



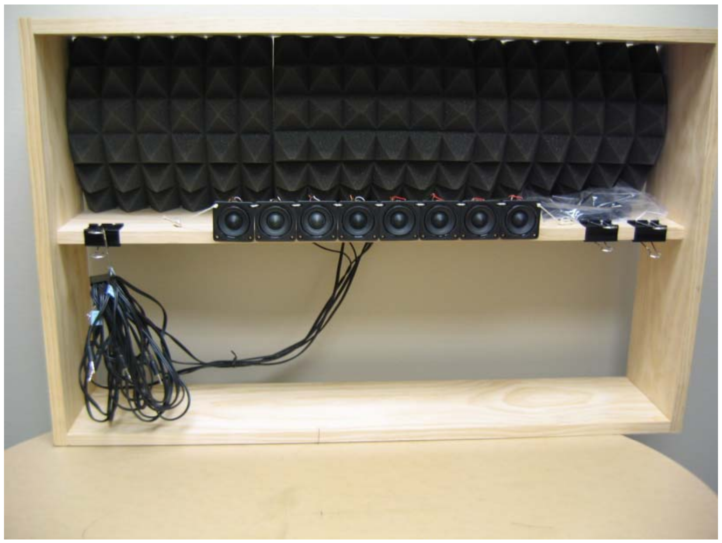
Weighted average:

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

The Bayesian observer weighs cues by their reliabilities, on a single trial.

Do humans do this?

Example: ventriloquist effect



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 Integration

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 M Noise: A Bayesian Explanation Using Feature Space

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

1 Department of Neuroscience, Baylor College of Medicine, Houston, Texas, United States of America, **2** Departi York, New York, New York, United States of America, **3** Program in Cognitive Neuroscience, Department of Psych 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

Do humans optimally integrate stereo and texture information for judgments of surface slant?

David C. Knill *, Jeffrey A. Saunders

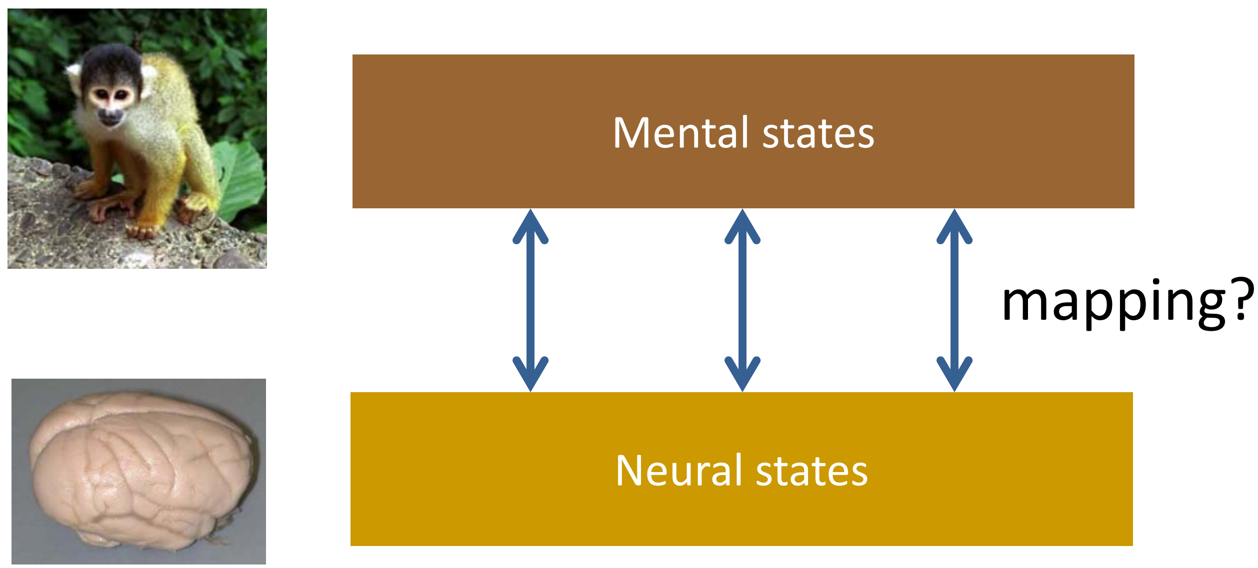
Center for Visual Sciences, University of Rochester, 274 Meliora Hall, Rochester, NY 14627, USA

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

Is something wrong with the brain?

- We perceive a puppet to be talking…
  We hear speech that is not there…

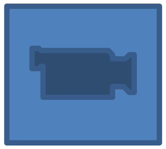
  Are we delusional?!
- Nothing wrong! The brain uses probabilities that are correct based on *normally occurring stimuli*.
- In illusions, the stimuli/task are artificially created to make those probabilities misleading.


Mental states
mapping?
Neural states

Two questions about neural states

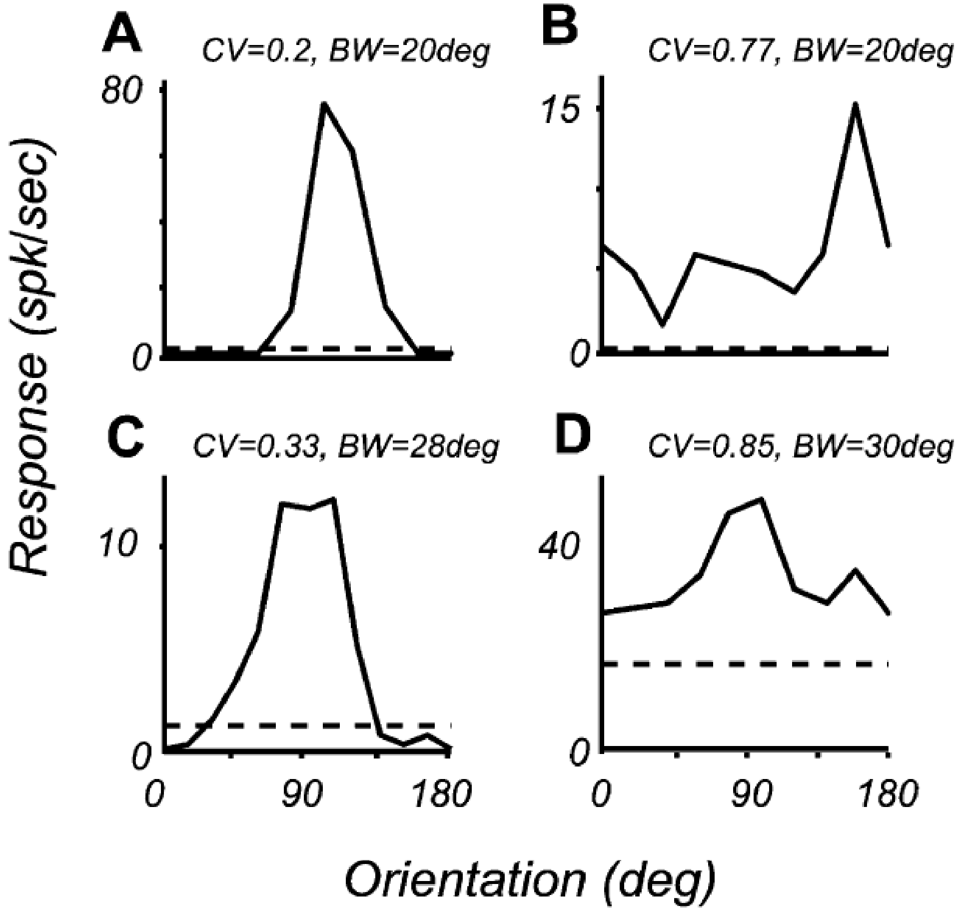
- How do neurons encode probability distributions? (*representation*)
- How do neurons perform Bayesian inference? (*computation*)

Hubel and Wiesel



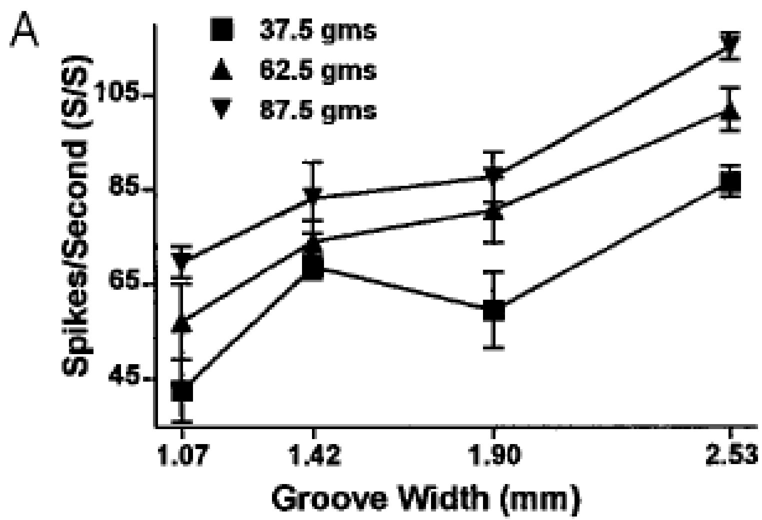
4:10

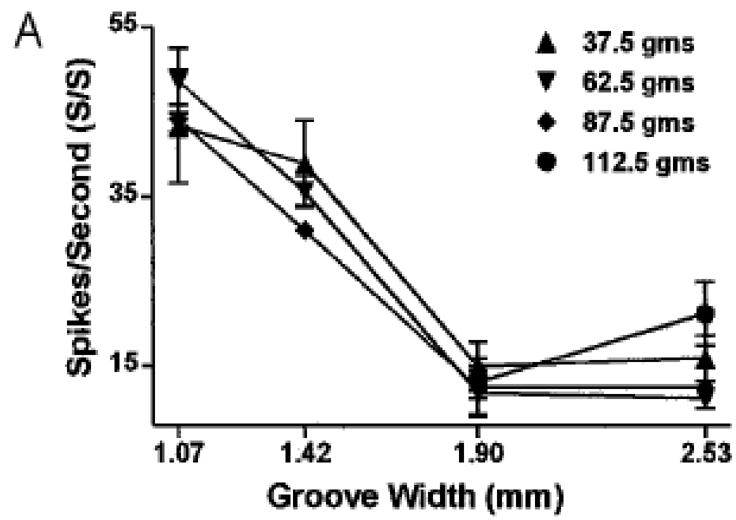
Tuning curve of a single neuron



Macaque V1
Shapley et al., 2003

Tuning curve of a single neuron

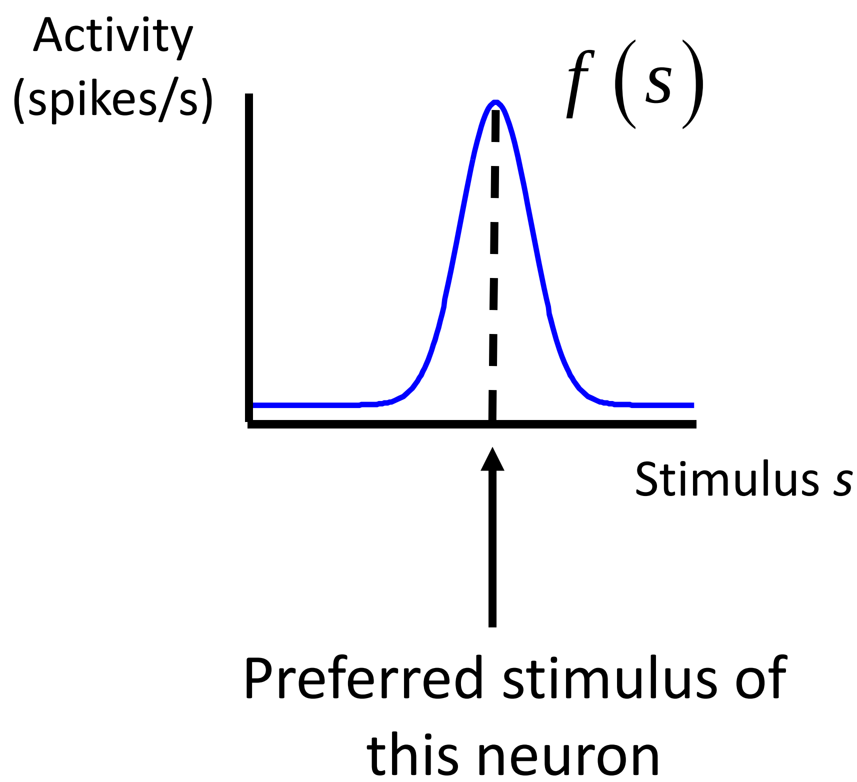




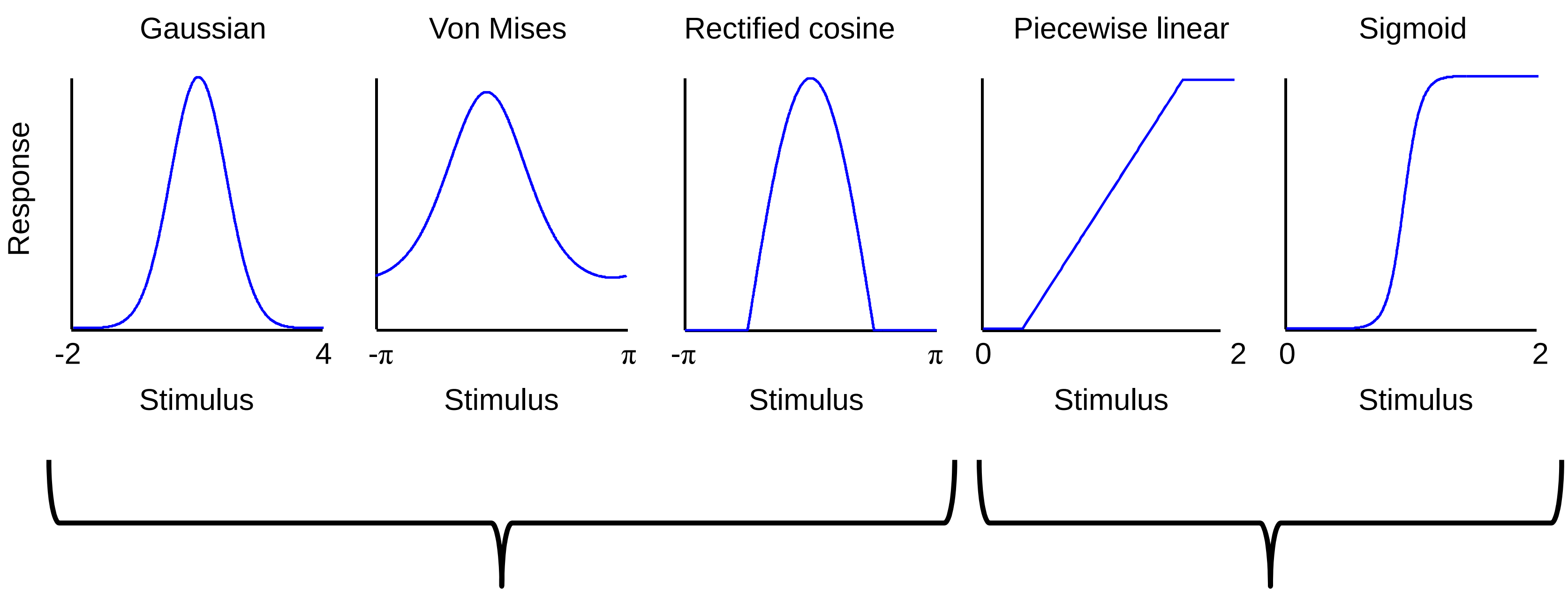
Macaque S2
Pruett et al., 2000

Idealized tuning curve

Mean response as a function of the stimulus



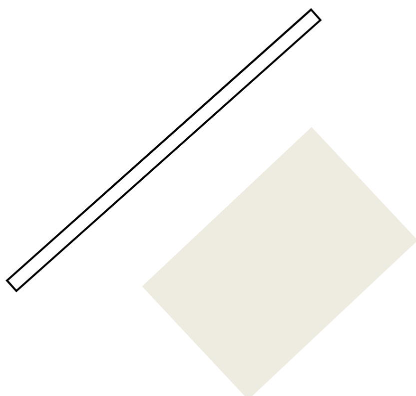
Models of tuning curves



These have a preferred stimulus.

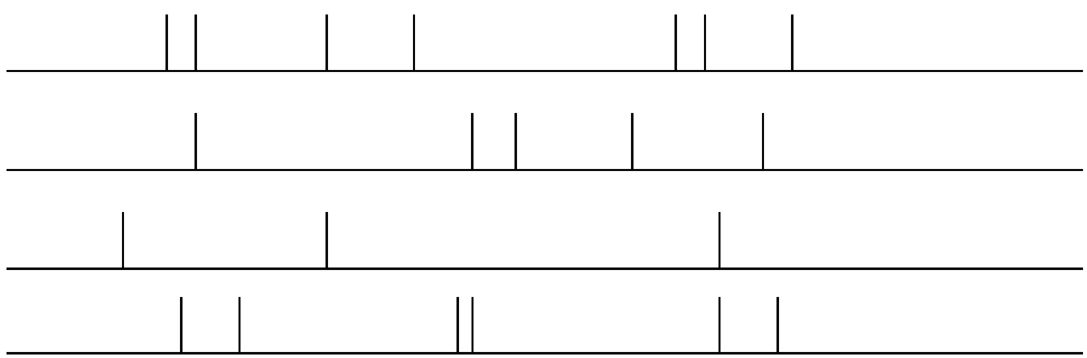
These don't.

Variability around the mean response



Response distribution: $p(r \mid s)$

What functional form?



Trial 1: 7 spikes

Trial 2: 5 spikes

Trial 3: 3 spikes

Trial 4: 6 spikes

Poisson variability

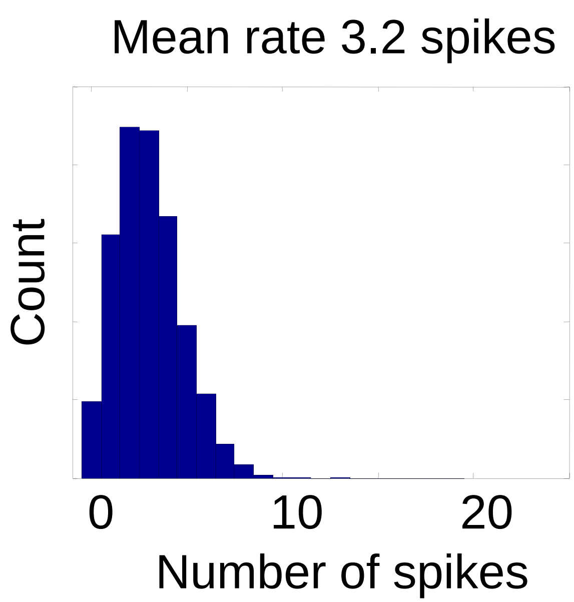
- Discrete distribution (spike counts)

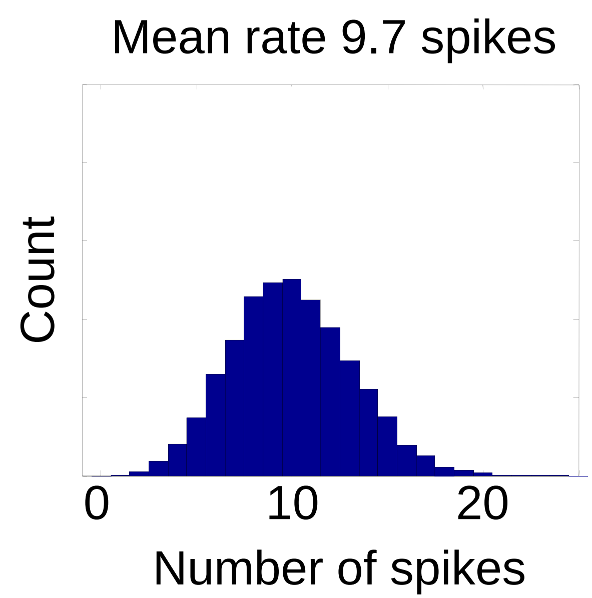
$$p(r \mid s) = \frac{e^{-f(s)} f(s)^r}{r!}$$

- *r* is an integer, *f*(*s*) not necessarily
- Mean of *r*:

$$\langle r \rangle = \sum_{r=0}^{\infty} r p(r \mid s) = f(s)$$

Histograms of a Poisson random variable

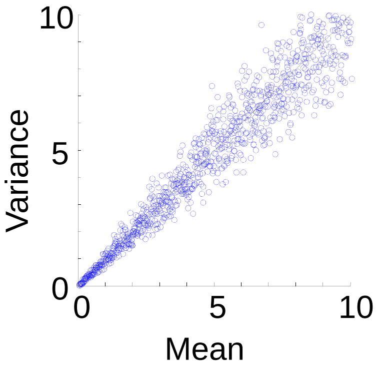




Fano factor

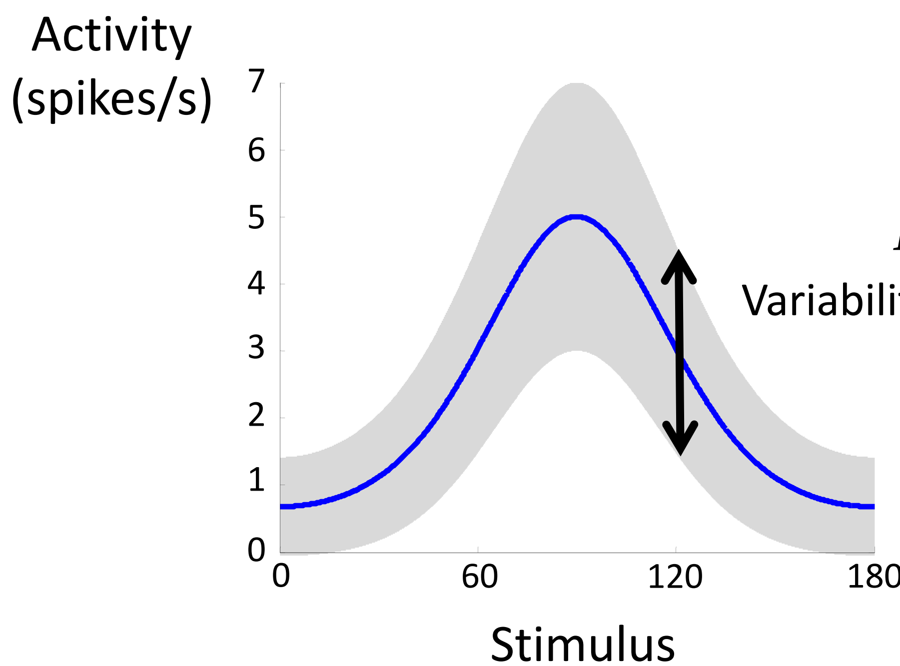
$$\text{Fano factor} = \frac{\text{variance of spike count}}{\text{mean spike count}}$$

Poisson process: Fano factor = 1



Physiology: Fano factor in range 0.3 to 1.8

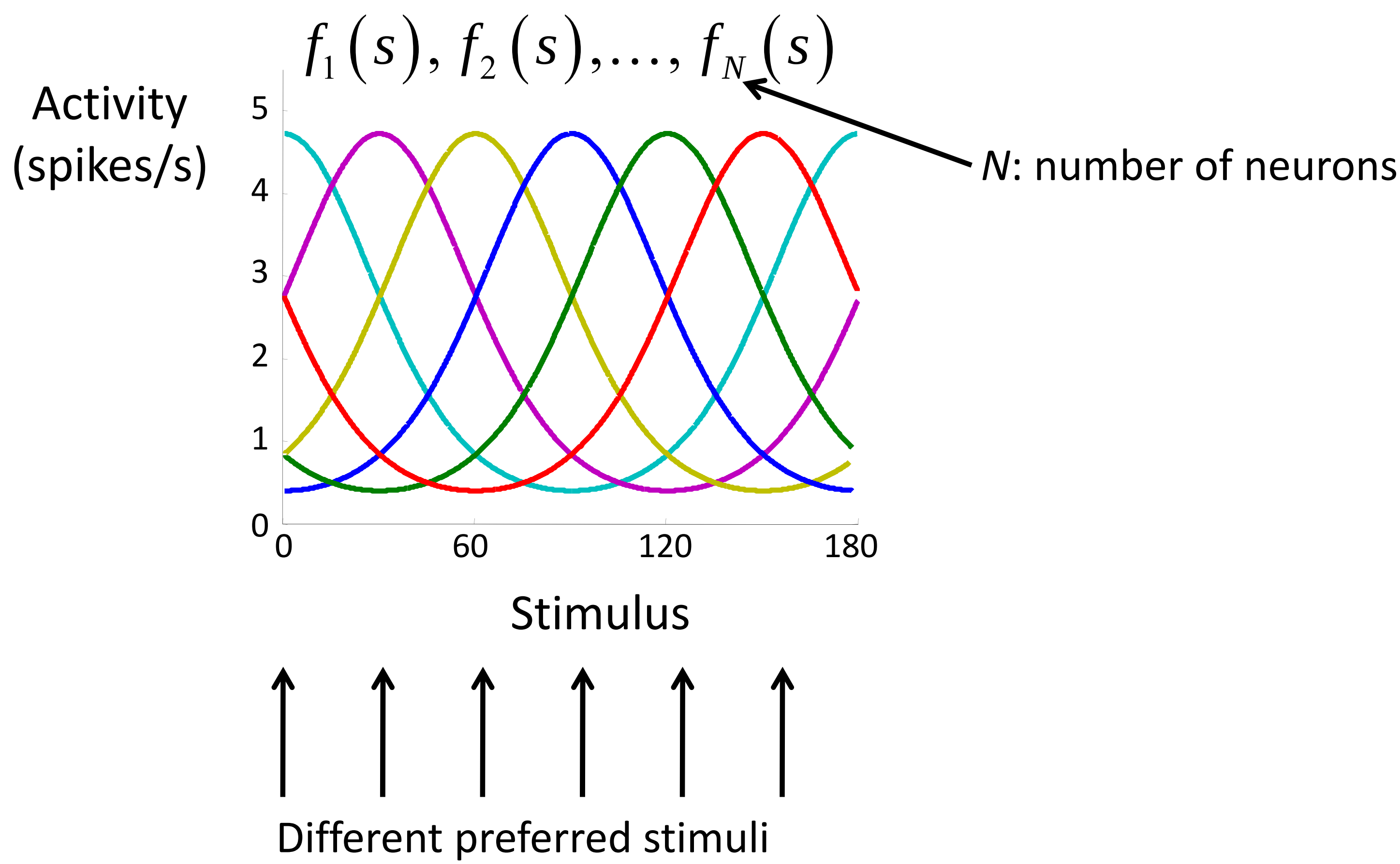
Single neuron – response variability



$$p(r \mid s)$$

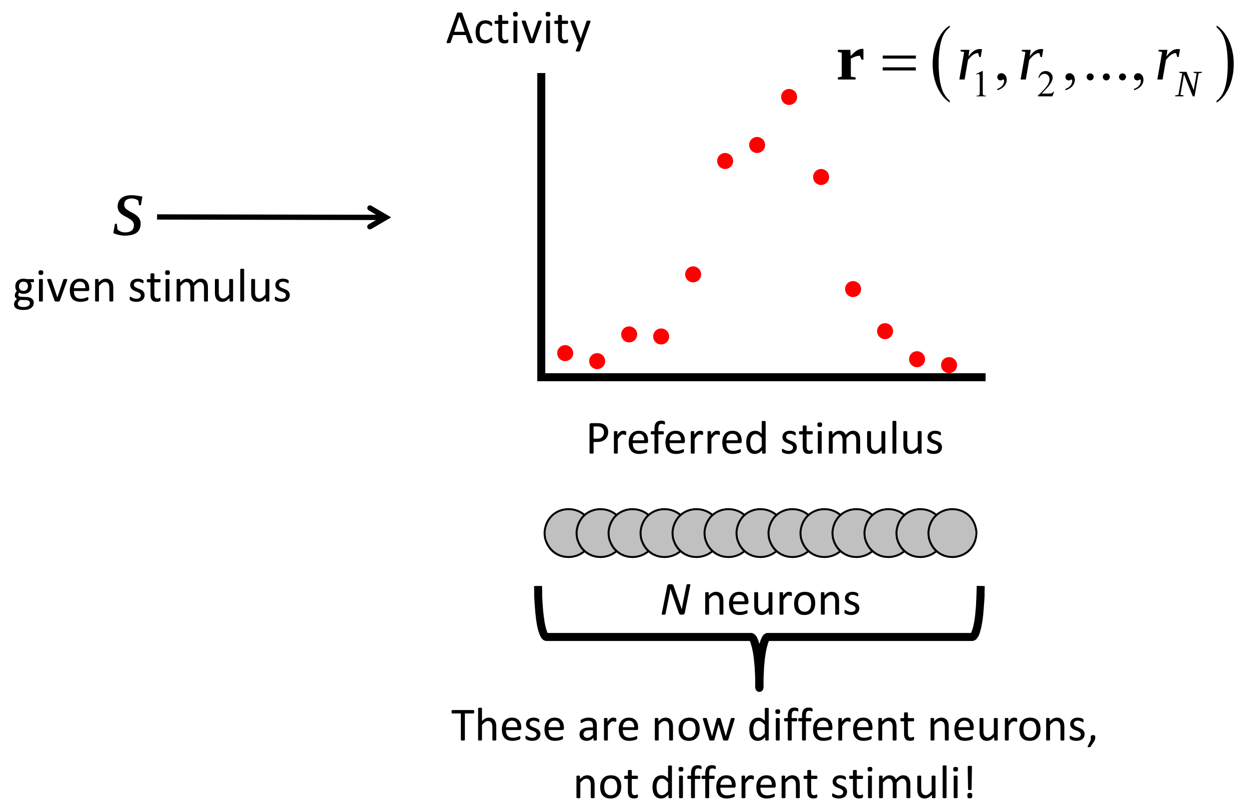
Variability depends on stimulus

Population of neurons

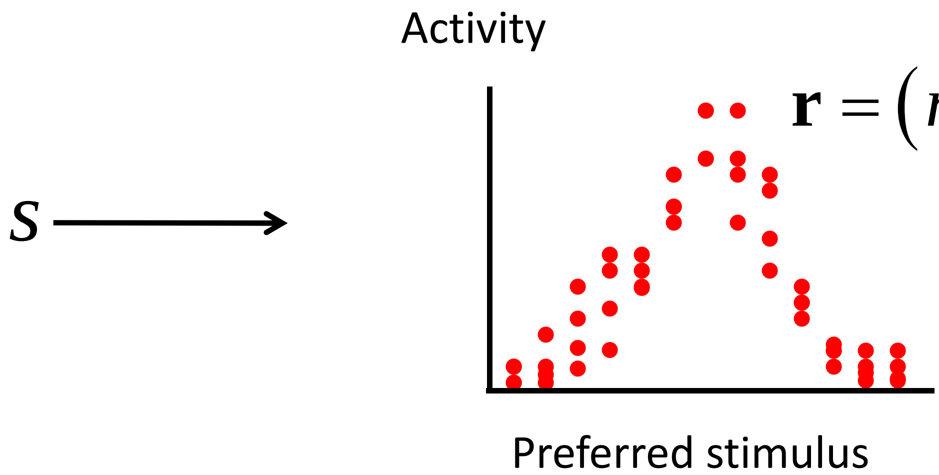


What is unrealistic about this picture?

Population activity on a single trial



Population activity – variability



$$\mathbf{r} = \left(r_1, r_2, \ldots, r_N\right)$$

Response distribution
(noise distribution): $p\left(\mathbf{r} \mid s\right)$

Independent Poisson variability

One neuron: $$p(r \mid s) = \frac{e^{-f(s)} f(s)^r}{r!}$$

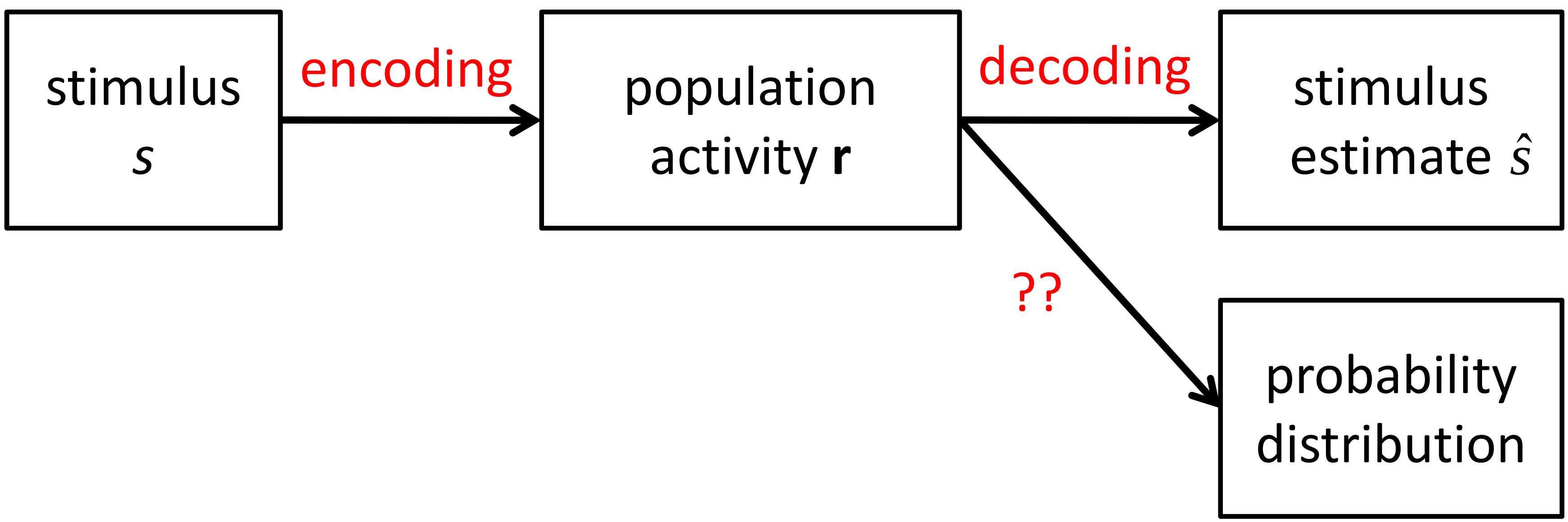
Population: $$p(\mathbf{r} \mid s) = \prod_{i=1}^{N} \frac{e^{-f_i(s)} f_i(s)^{r_i}}{r_i!}$$

If not independent, then *noise correlations*

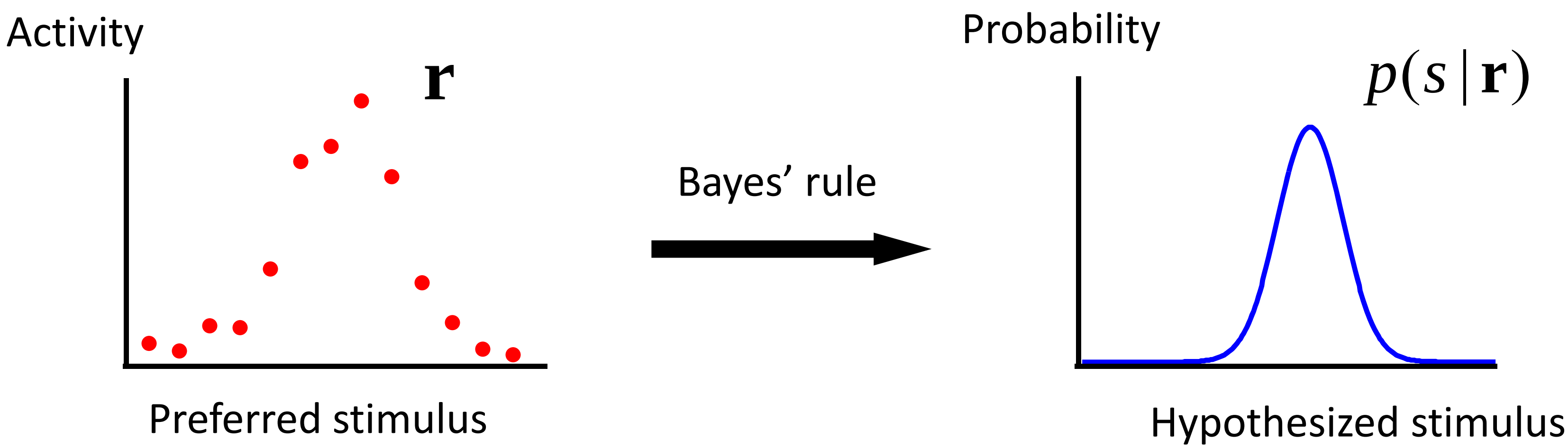
Population codes in the brain

- Primary visual cortex (orientation, spatial frequency)
- MT (motion direction, velocity)
- IT (human faces, objects)
- SC (saccade direction)
- Primary motor cortex (arm movement direction)
- Hippocampus in rat (self location)
- Cercal interneurons in cricket (wind direction)
- Prefrontal cortex (numerosity)

Why population coding and not single-neuron coding?


stimulus
s
encoding
population
activity $\mathbf{r}$
decoding
stimulus
estimate $\hat{s}$
??
probability
distribution

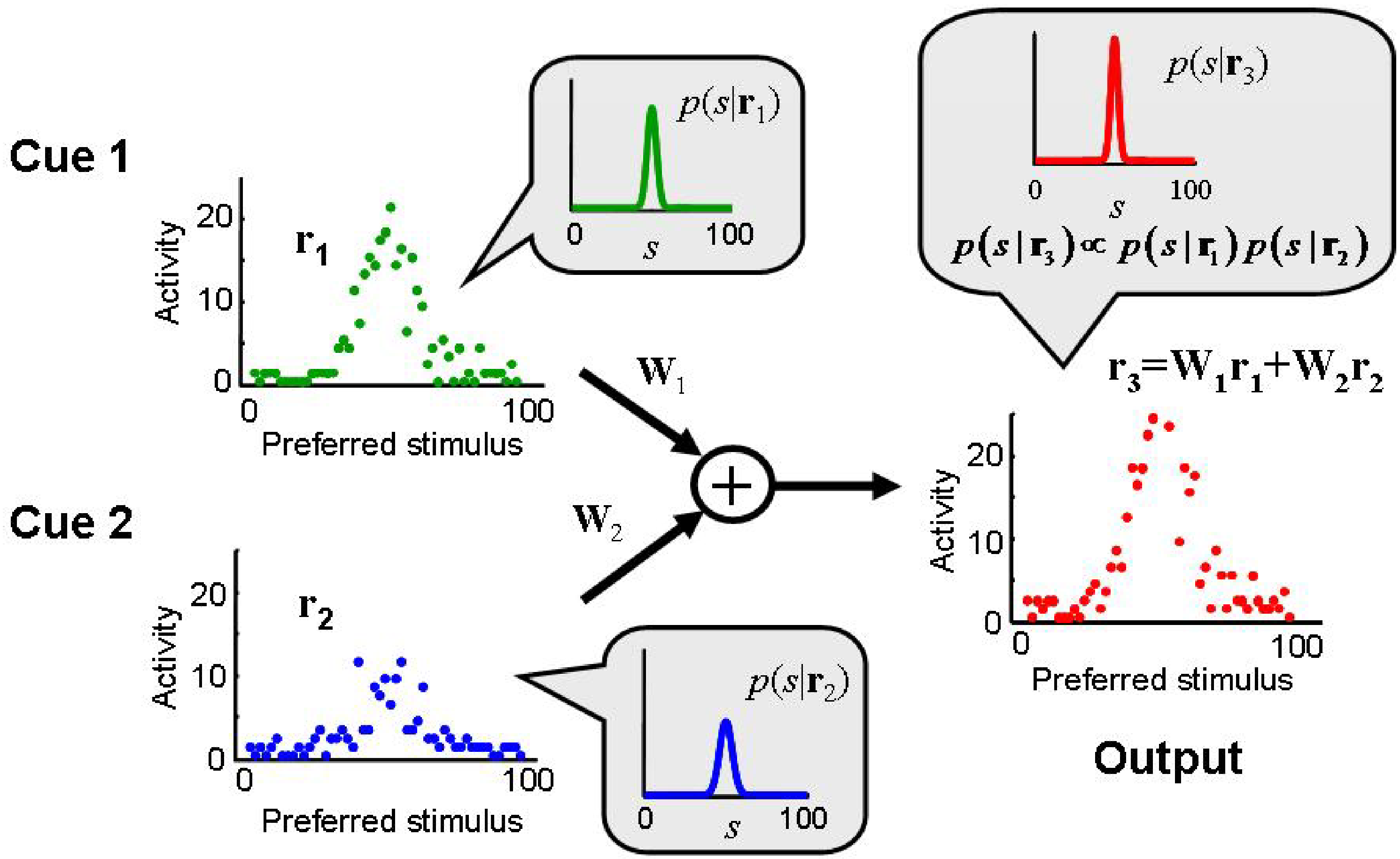
Decoding a probability distribution



$$p(s \mid \mathbf{r}) \propto p(\mathbf{r} \mid s)\, p(s)$$

population variability

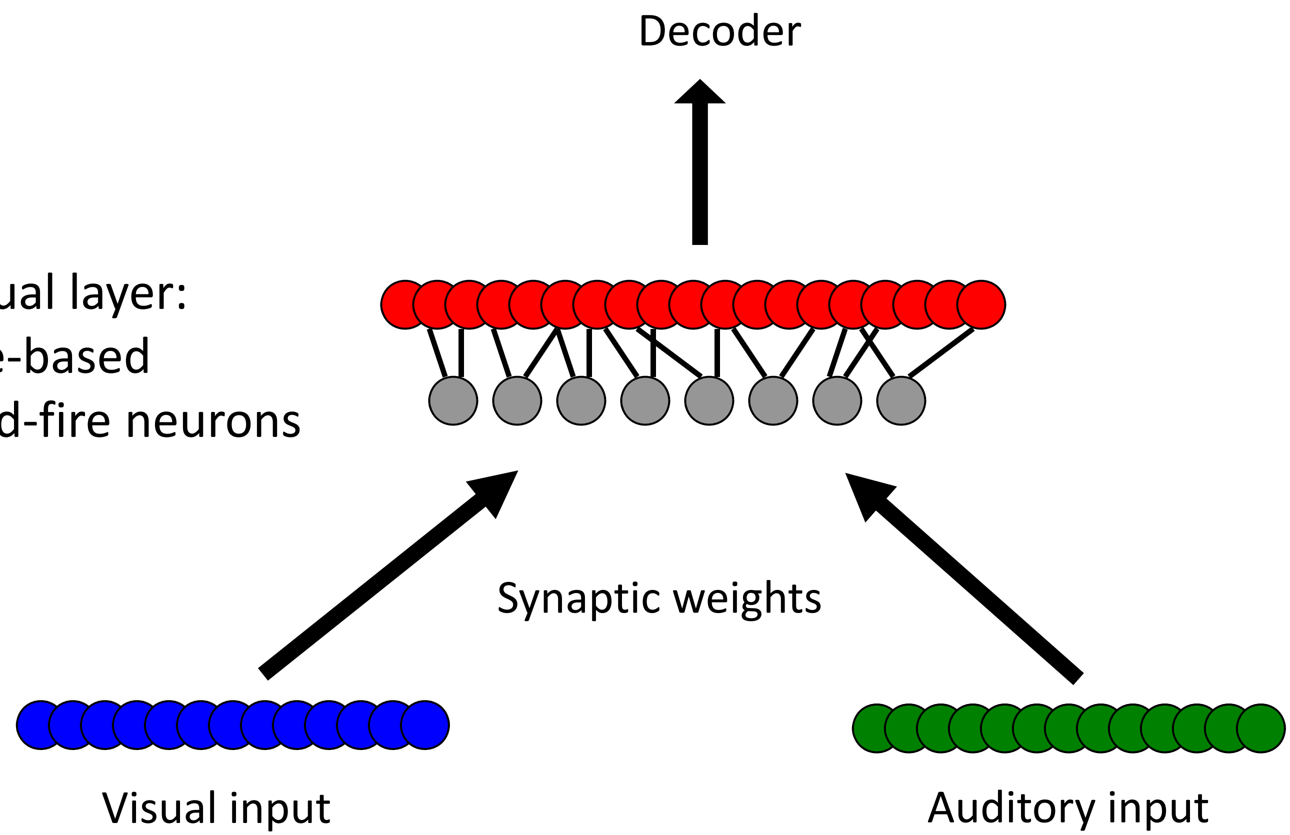
Computation with population codes



Ma et al., Nature Neuroscience 2006

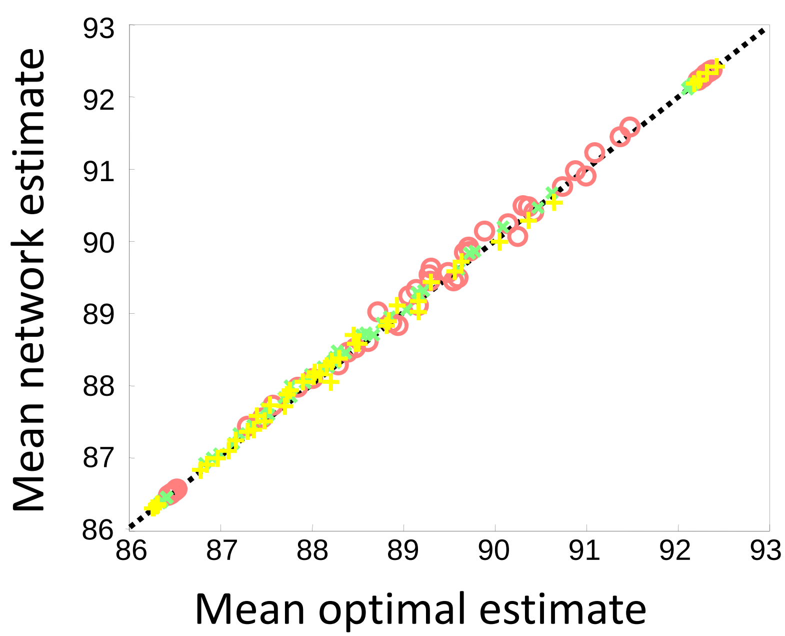
Integrate-and-fire network

Auditory-visual layer: conductance-based integrate-and-fire neurons

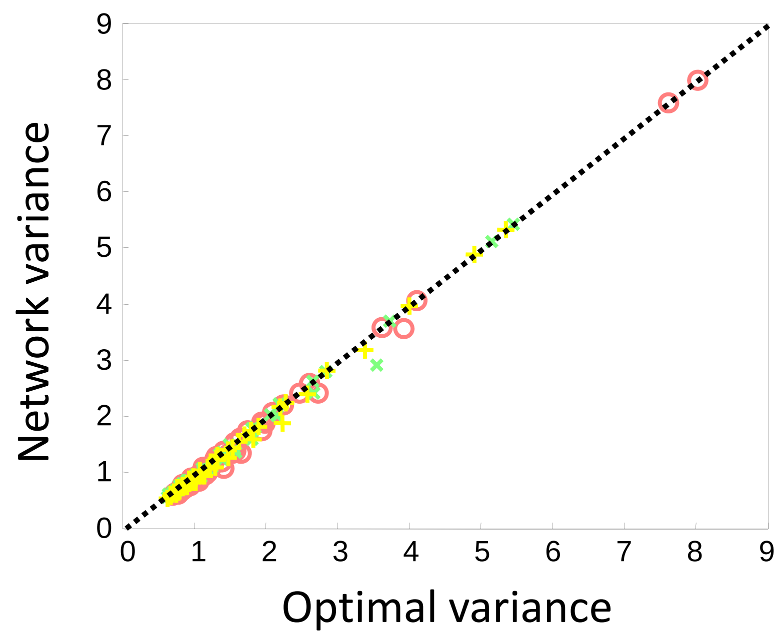


Network performs near-optimally

Mean estimates



Estimate variance



o Same tuning curves, same covariance matrices

\+ Same tuning curves, different covariance matrices

x Different tuning curves, different covariance matrices

Testing predictions…

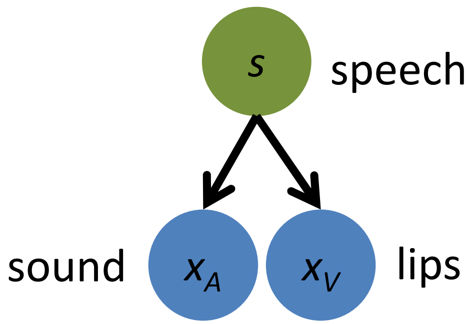
Neural correlates of multisensory cue integration in macaque MSTd

Yong Gu[1], Dora E Angelaki[1,3] & Gregory C DeAngelis[1–3]

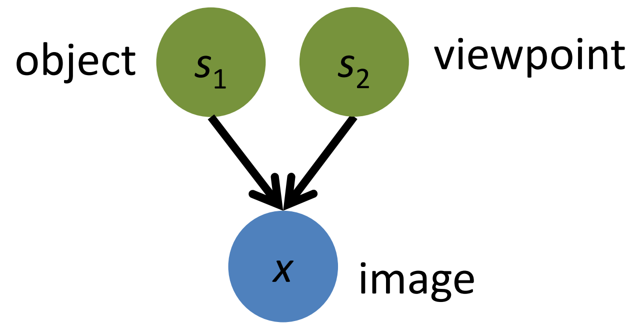
Multisensory Integration in Macaque Visual Cortex Depends on Cue Reliability

Michael L. Morgan,[1] Gregory C. DeAngelis,[2,3] and Dora E. Angelaki[1,3,*]
[1]Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, MO 63110, USA
[2]Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, NY 14627, USA
[3]These authors contributed equally to this work
*Correspondence: angelaki@pcg.wustl.edu
DOI 10.1016/j.neuron.2008.06.024

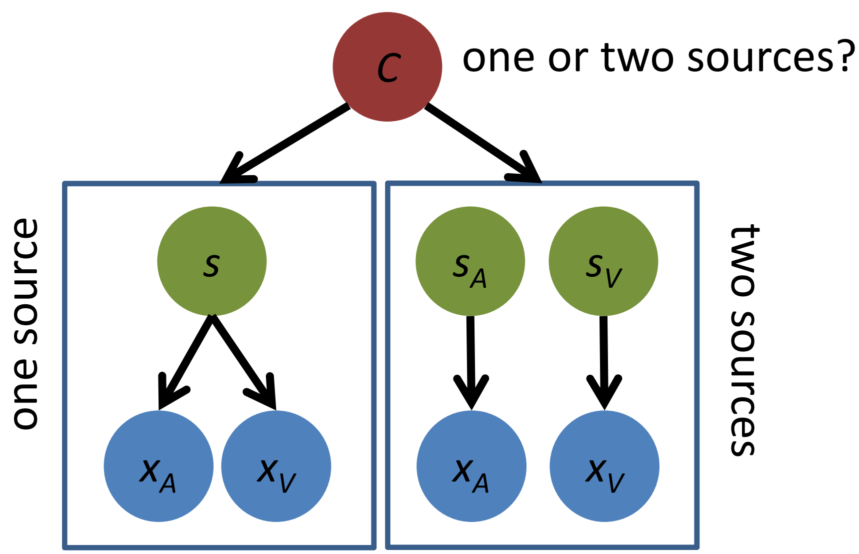
More complex forms of inference



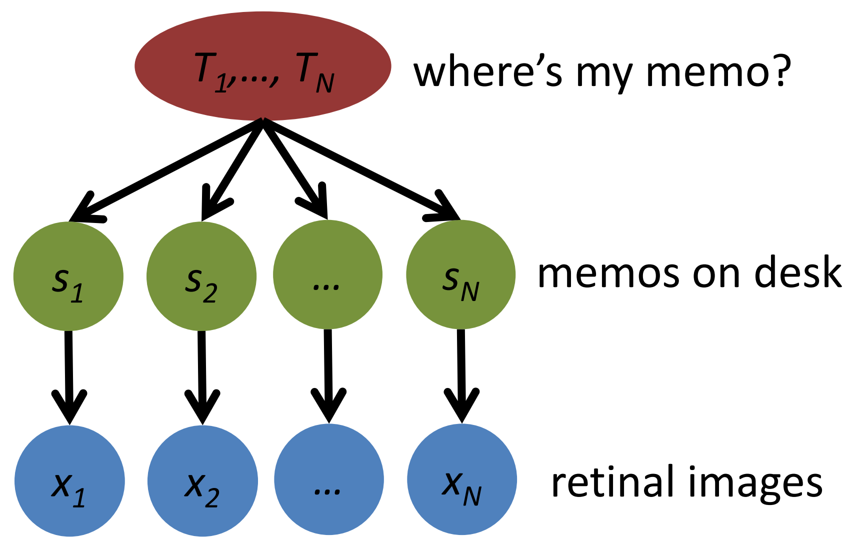
Cue integration



Invariant perception (discounting)



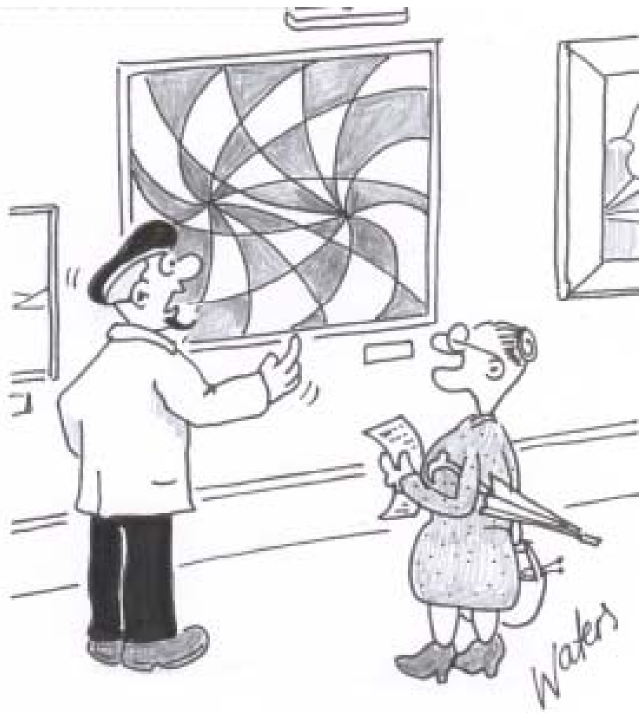
Causal inference



Visual search

Your brain is an inference machine

- The brain *interprets*, not just transmits, sensory input.
- The brain *computes posterior probabilities* to interpret the world: Bayesian inference
- Many *illusions* can be explained as consequences of probabilistic inference.
- *Humans are Bayesian observers* in many psychophysical tasks. They weigh observations by reliability.
- Explaining human *behavior* using a Bayesian model can elucidate underlying *neural processes*.



"It's not an optical illusion, Madame, it just looks that way!"