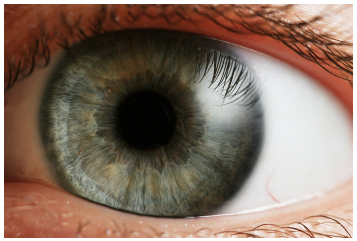
Bayesian modeling in cognitive science

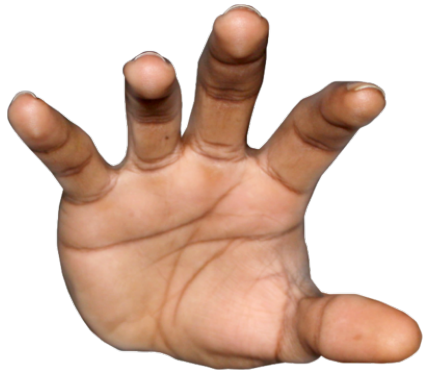
Dave Kleinschmidt
University of Rochester

6cm ± 2cm



size?

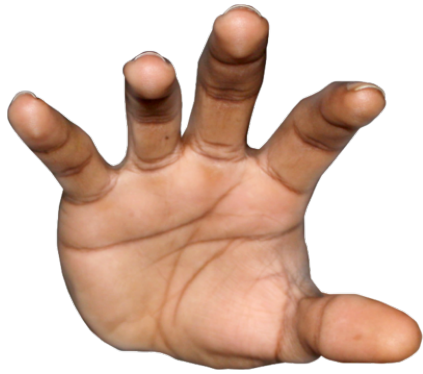
4cm ± 1cm

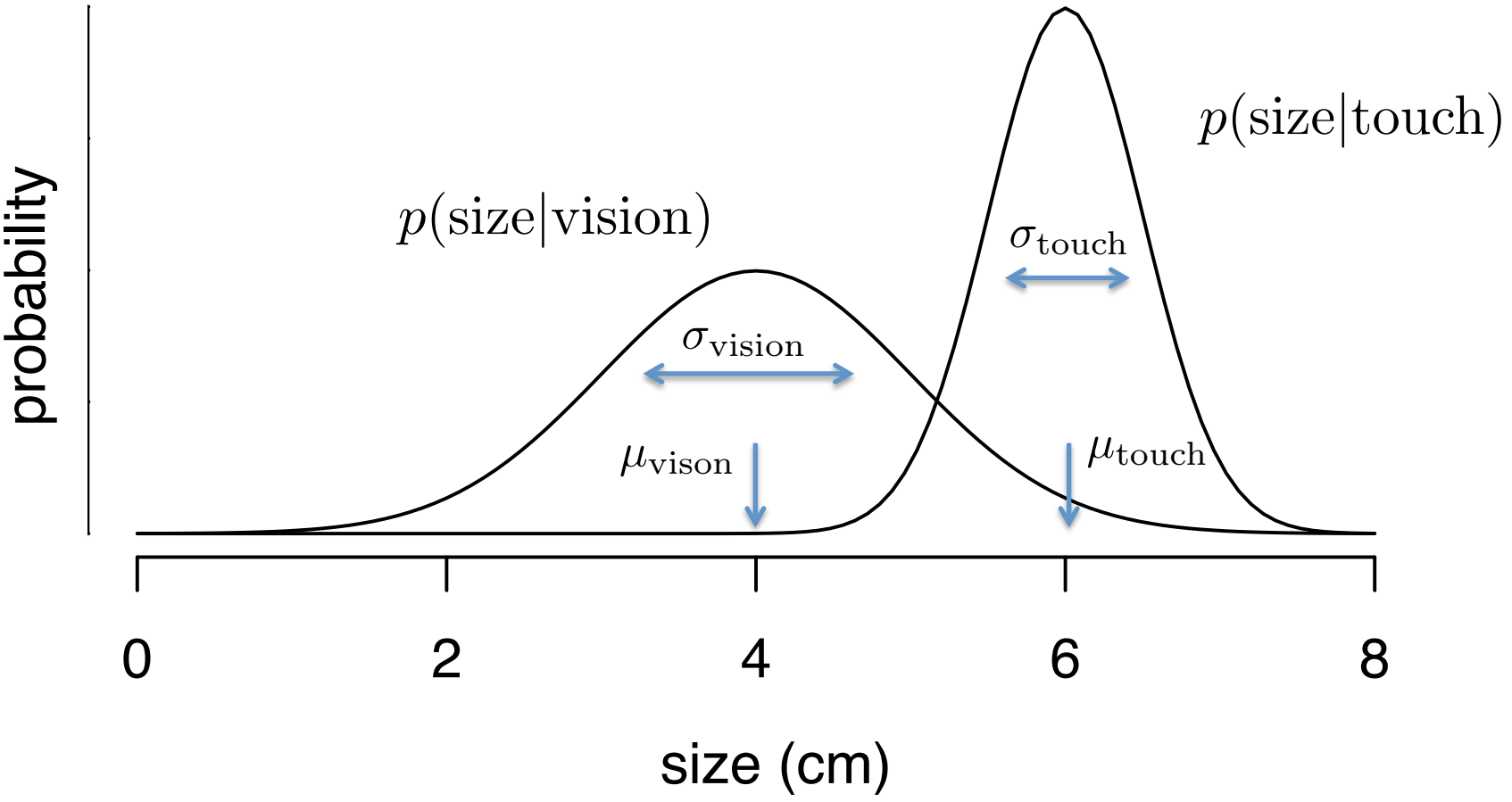


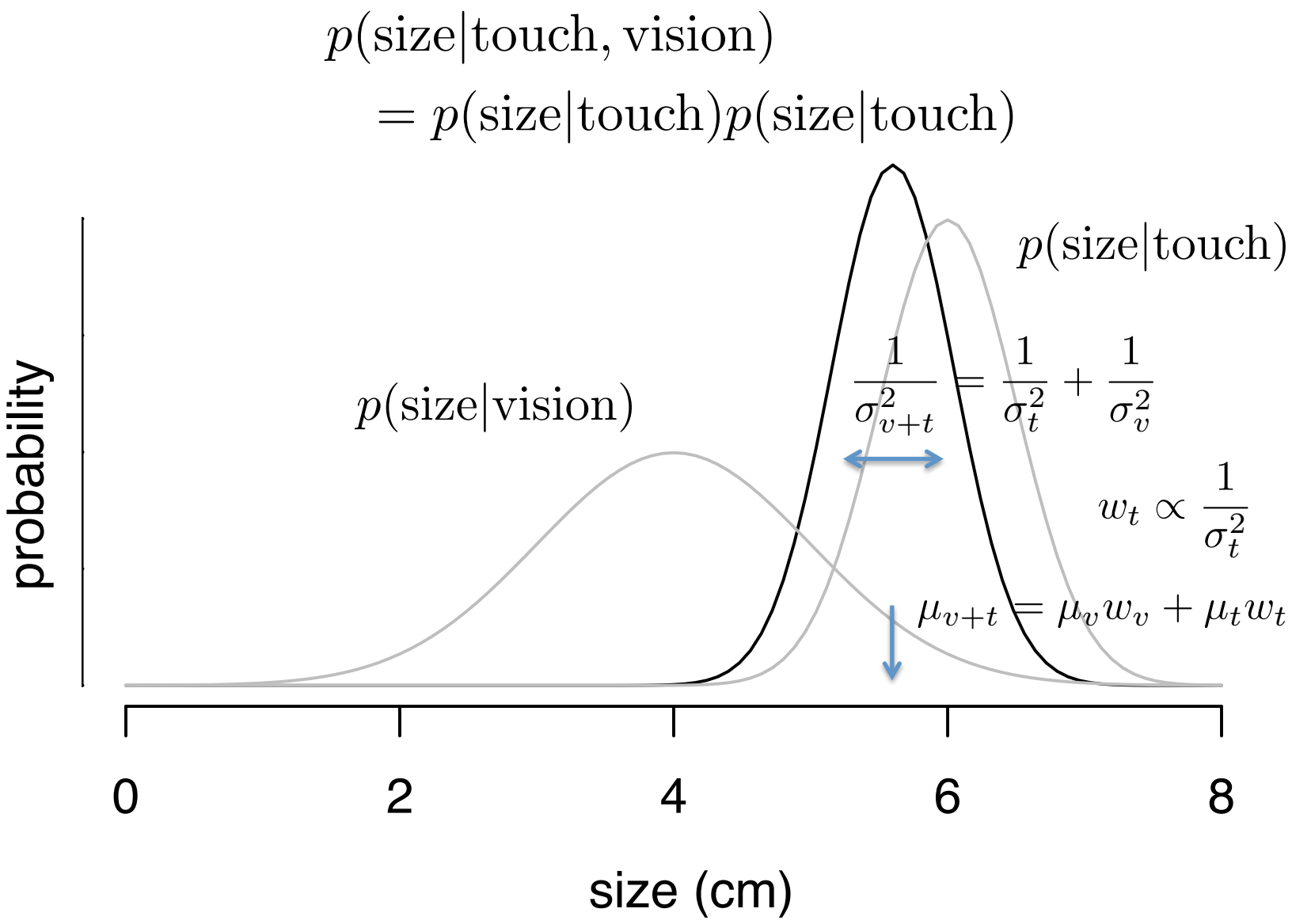
6cm ± 0.5cm

size?

4cm ± 1cm



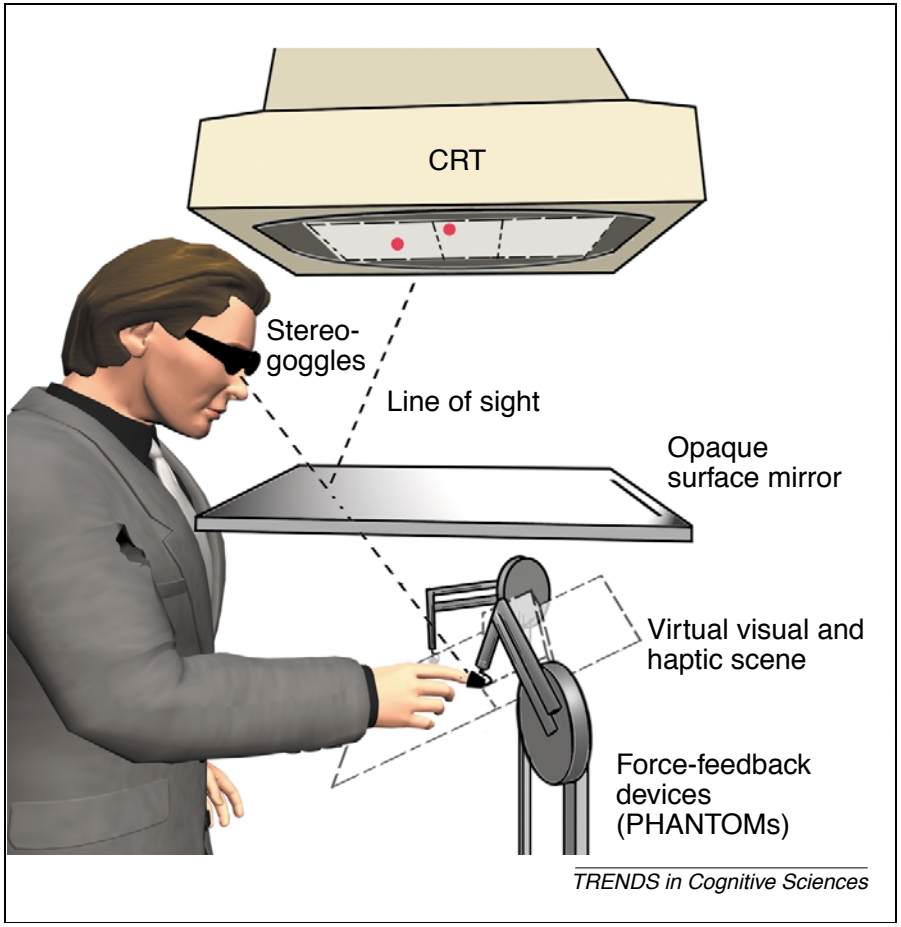

probability
$p(\text{size}|\text{vision})$
$p(\text{size}|\text{touch})$
σ_{vision}
σ_{touch}
μ_{vison}
μ_{touch}
0
2
4
6
8
size (cm)


$p(\text{size}|\text{touch}, \text{vision})$
$= p(\text{size}|\text{touch})p(\text{size}|\text{touch})$
$p(\text{size}|\text{touch})$
$p(\text{size}|\text{vision})$
$\frac{1}{\sigma^2_{v+t}} = \frac{1}{\sigma^2_t} + \frac{1}{\sigma^2_v}$
$w_t \propto \frac{1}{\sigma^2_t}$
$\mu_{v+t} = \mu_v w_v + \mu_t w_t$
probability
0
2
4
6
8
size (cm)

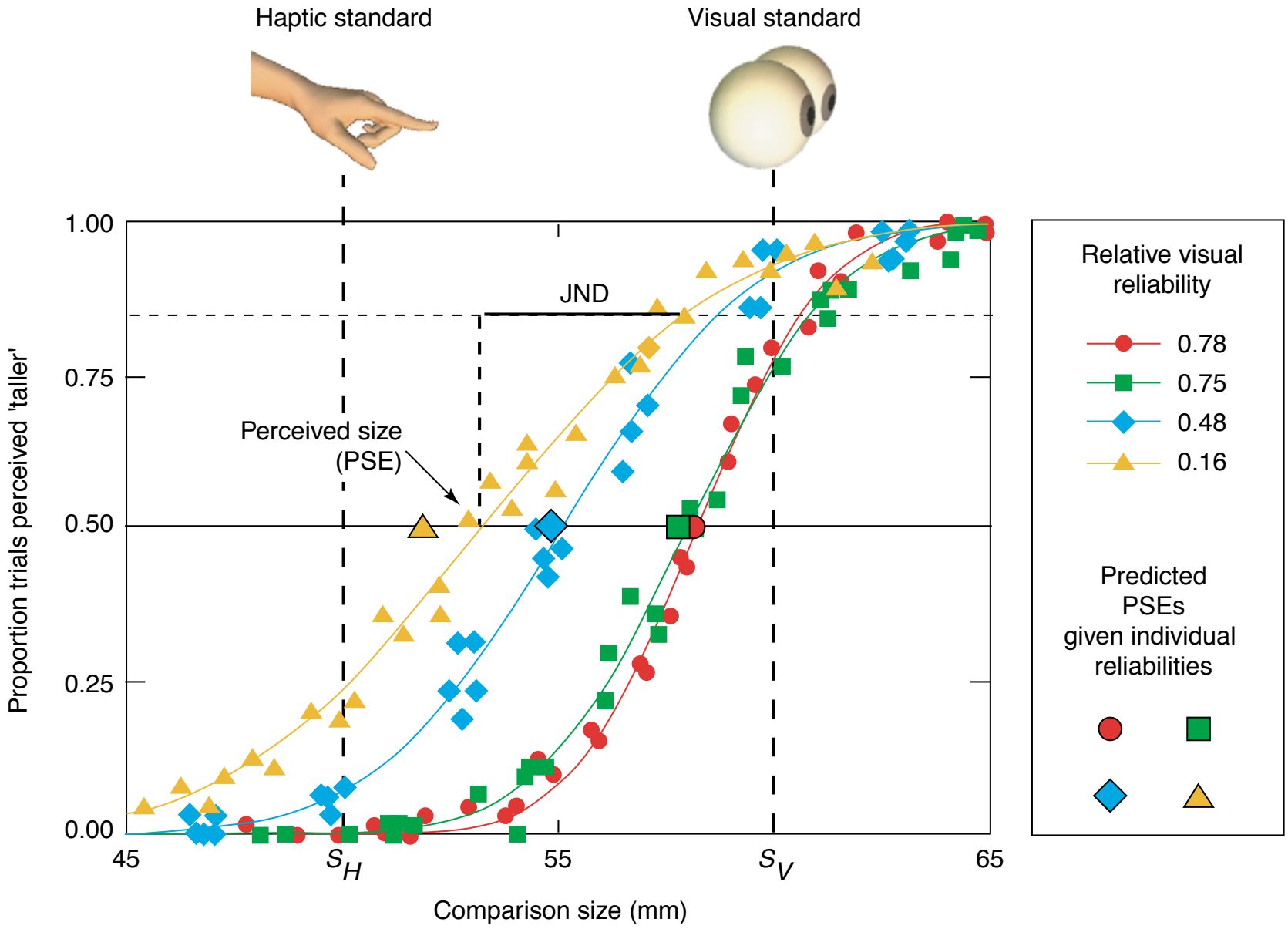
Does the brain really do this?

- Important to make **predictions** about behavior based on model.
- **Qualitative** predictions: “when I increase *x*, *y* should also increase”
- **Quantitative** predictions: “if I measure this thing, it will have value *x*”

An experiment!



- Decide which of two objects are bigger.
- One object has a **cue conflict**: felt and seen size are mismatched.
- Cue combination predicts:
  - More reliable cue is weighted more
  - Weights are directly proportional to reliability
  - Reliability of two cues sums.


Haptic standard
Visual standard
1.00
0.75
0.50
0.25
0.00
JND
Perceived size
(PSE)
Proportion trials perceived 'taller'
45
S_H
55
S_V
65
Comparison size (mm)
Relative visual
reliability
0.78
0.75
0.48
0.16
Predicted
PSEs
given individual
reliabilities

Bayesian modeling

- Represent uncertain information
- Everything is a probability distribution
- Usually describe **optimal** performance for a task, given certain information and assumptions about task structure.
- Not always directly related to actual cognitive processing or neural activity.

Bayesian modeling: levels of analysis

- David Marr (1982), *Vision*
- **Computational:** what is the problem that the system solves?
- **Algorithmic**: what are the **representations** and **processes** used to carry out the computations?
- **Implementation**: what is the hardware?

Bayesian modeling

- Generally at the **computational level**:
  - What is the **task** to be performed?
  - What is the **information** available to do it?
  - What is the **model** (relationship) of information and task?
- Predicts **best possible performance.**

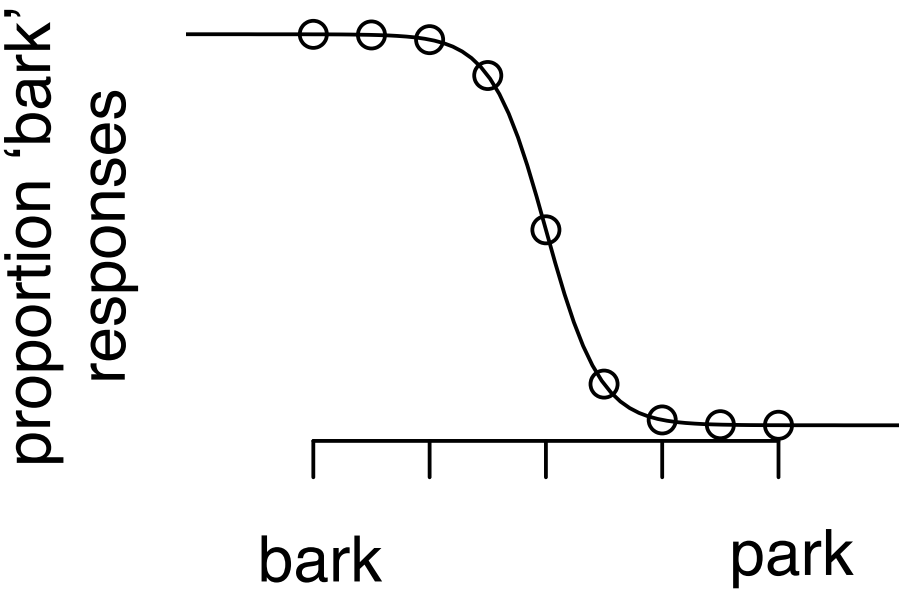
Bayesian modeling: Cue combination

- **task**: figure out how big a thing is
- **information**: multiple noisy/uncertain estimates
- **model**: all estimates generated by same actual size in the world

How to make a model

- Start with a theory
  - "People combine information about size from different senses by weighting each sense according to its reliability"
- Express the theory quantitatively
  - Information is a probability distribution
  - Reliability is the inverse variance of that distribution.

Categorical speech perception



Categorical speech perception

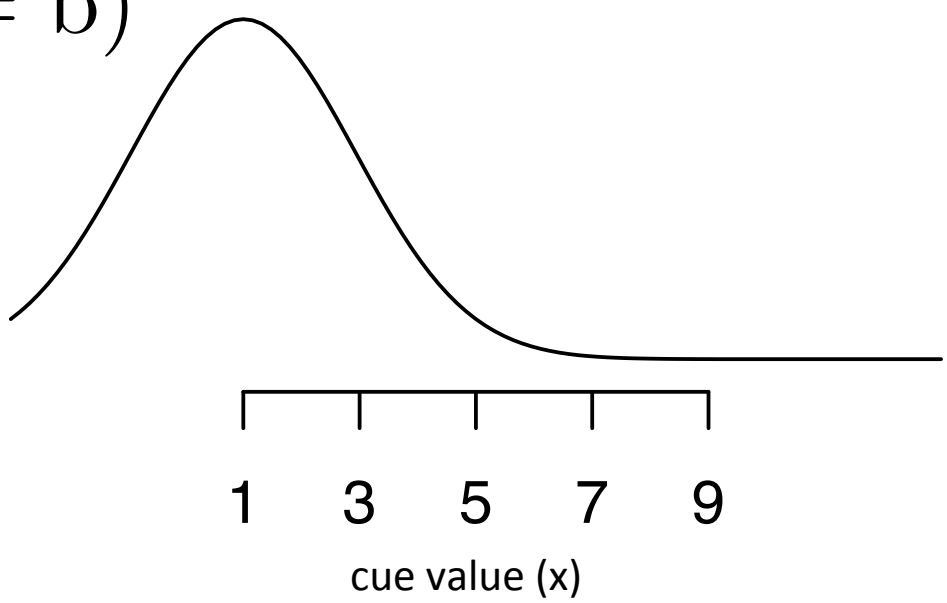
- Bayesian model: combination of
  - The task (infer word/phonetic category)
  - Information available (distributions of sounds)

$$p(\text{cue value}|\text{category}) = p(x|c)$$

Categorical speech perception

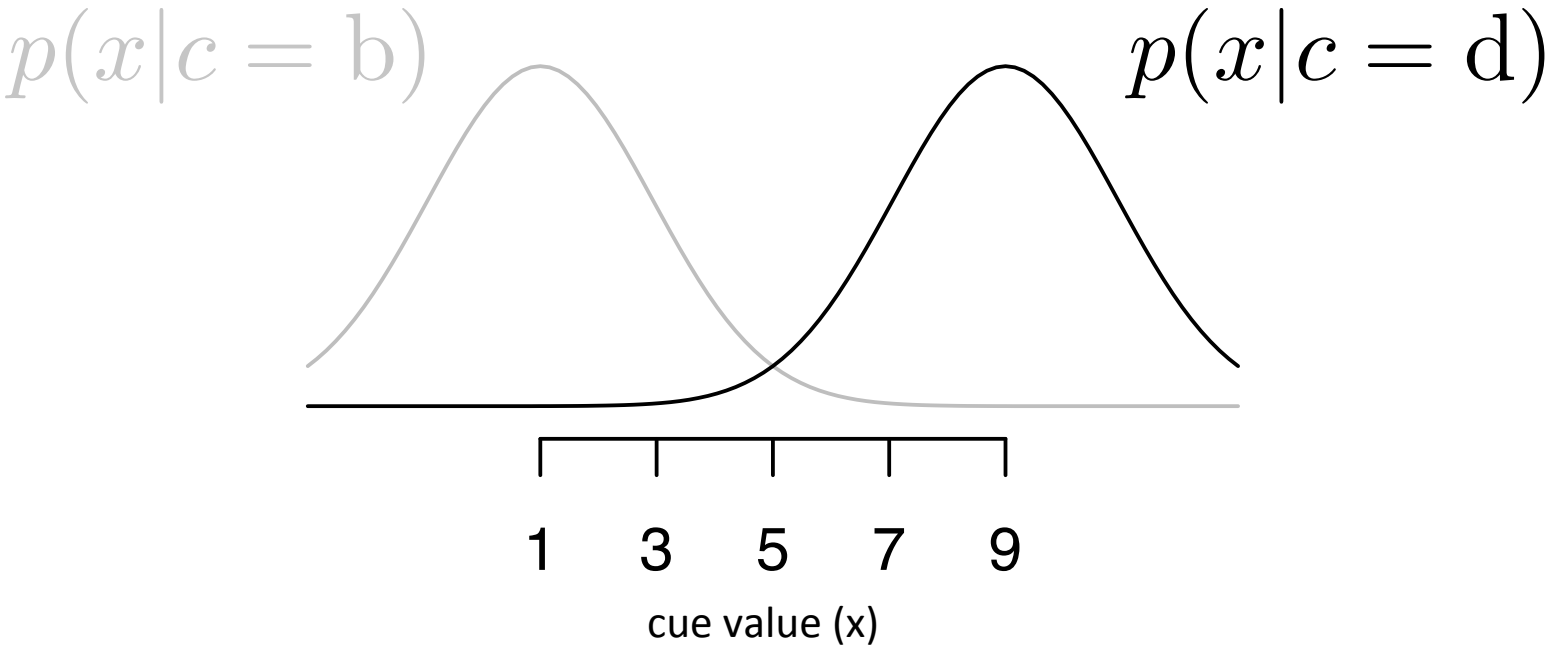
$$p(\text{cue value}|\text{category}) = p(x|c)$$

$p(x|c = \text{b})$



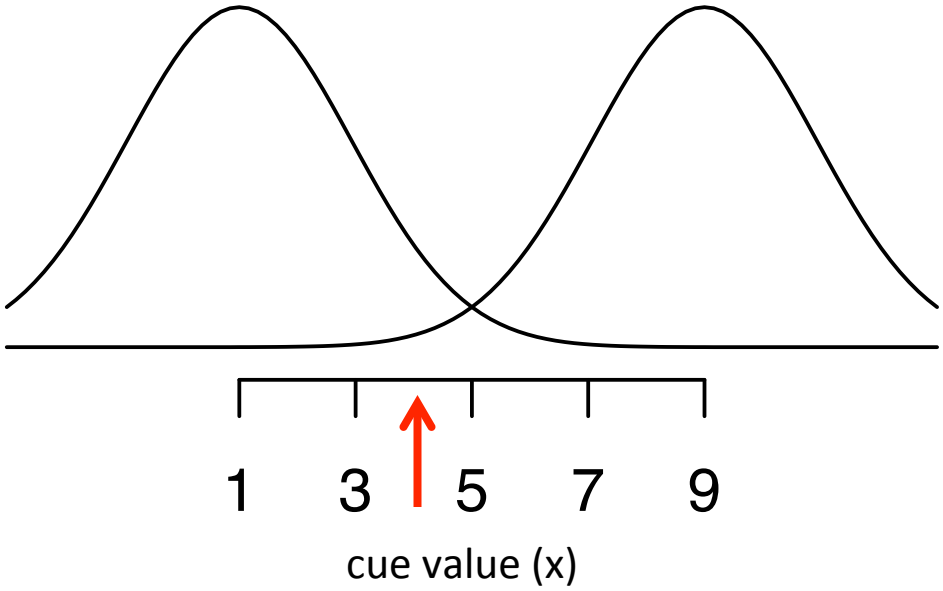
Categorical speech perception

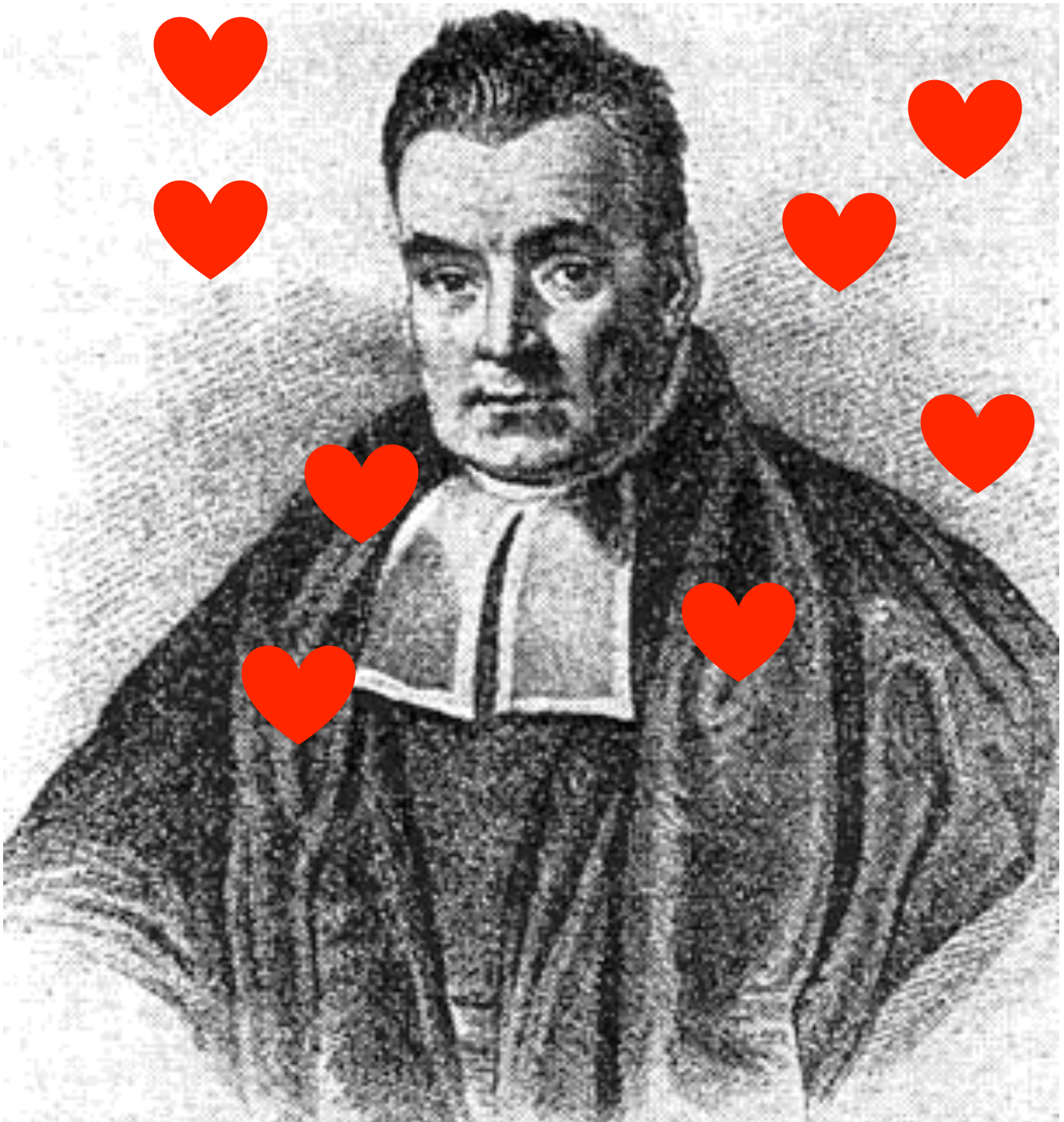
$$p(\text{cue value}|\text{category}) = p(x|c)$$



Categorical speech perception

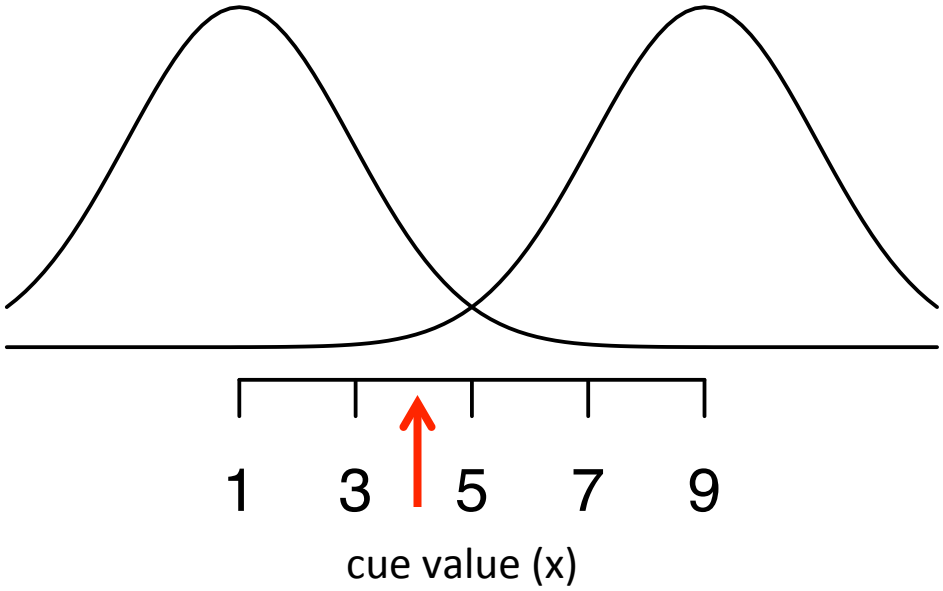
$$p(c = \mathrm{b}|x = 4) = ?$$





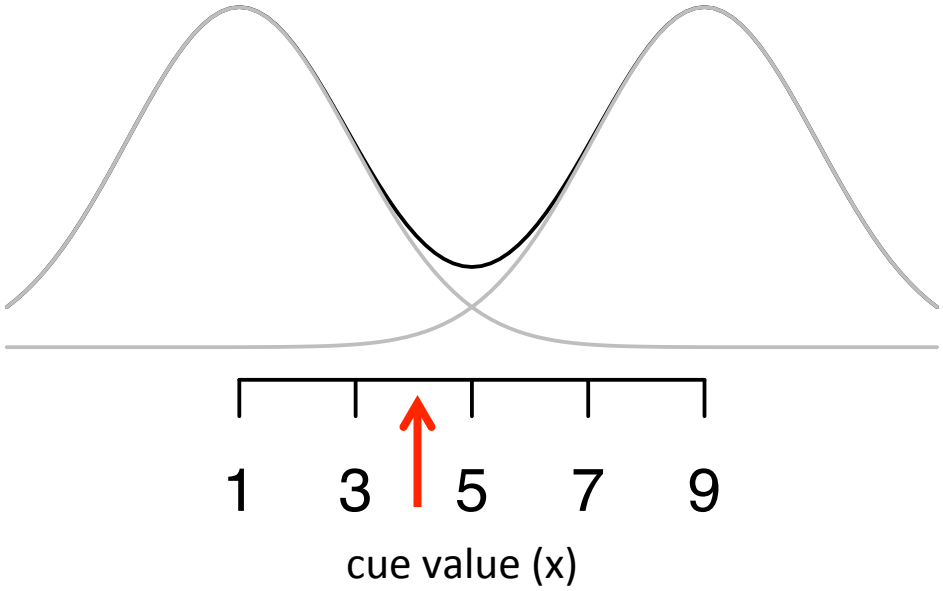
Categorical speech perception

$$p(c = \mathrm{b}|x = 4) = \frac{p(x = 4|c = \mathrm{b})p(c = \mathrm{b})}{p(x = 4)}$$



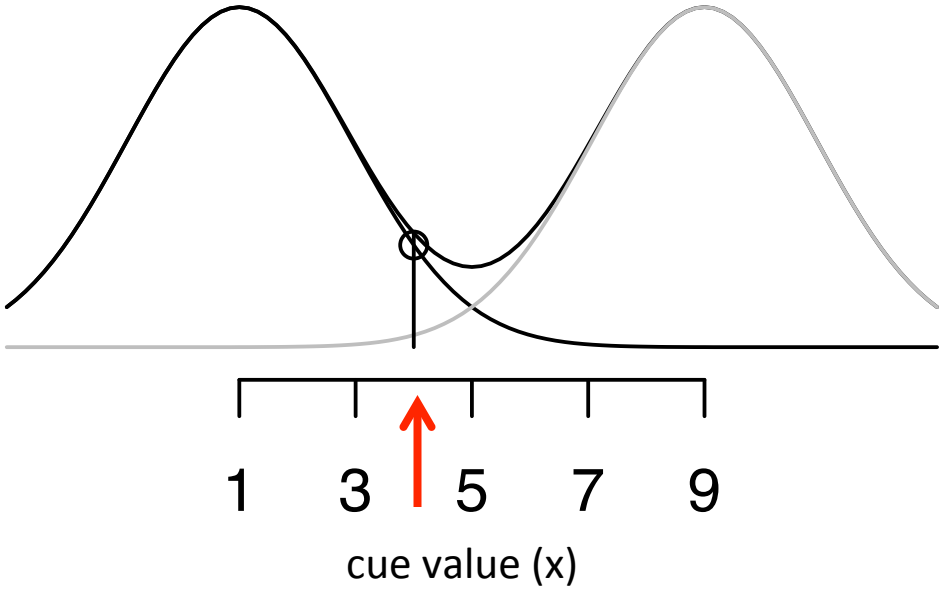
Categorical speech perception

$$p(c = \mathrm{b}|x = 4) = \frac{p(x = 4|c = \mathrm{b})p(c = \mathrm{b})}{p(x = 4)}$$



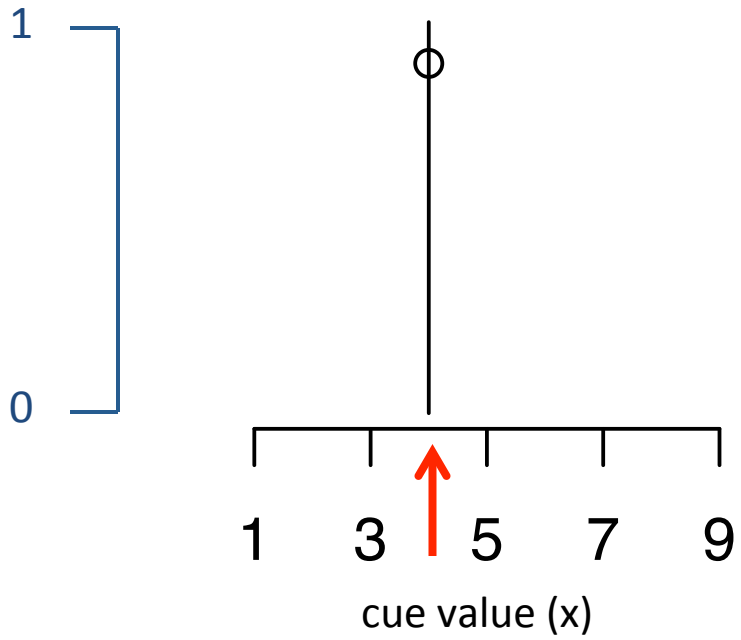
Categorical speech perception

$$p(c = \mathrm{b}|x = 4) = \frac{p(x = 4|c = \mathrm{b})p(c = \mathrm{b})}{p(x = 4)}$$



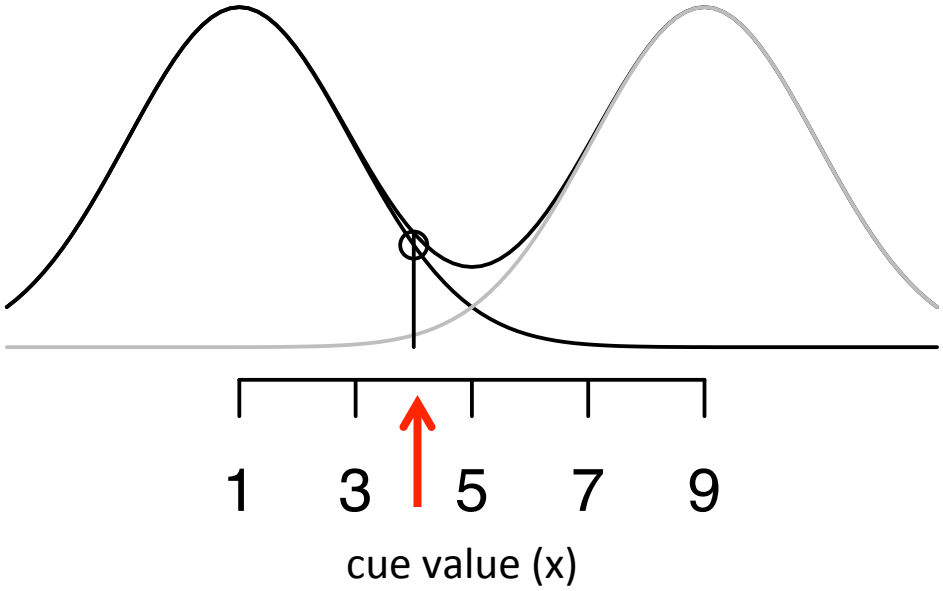
Categorical speech perception

$$p(c = \mathrm{b}|x = 4) = \frac{p(x = 4|c = \mathrm{b})p(c = \mathrm{b})}{p(x = 4)}$$



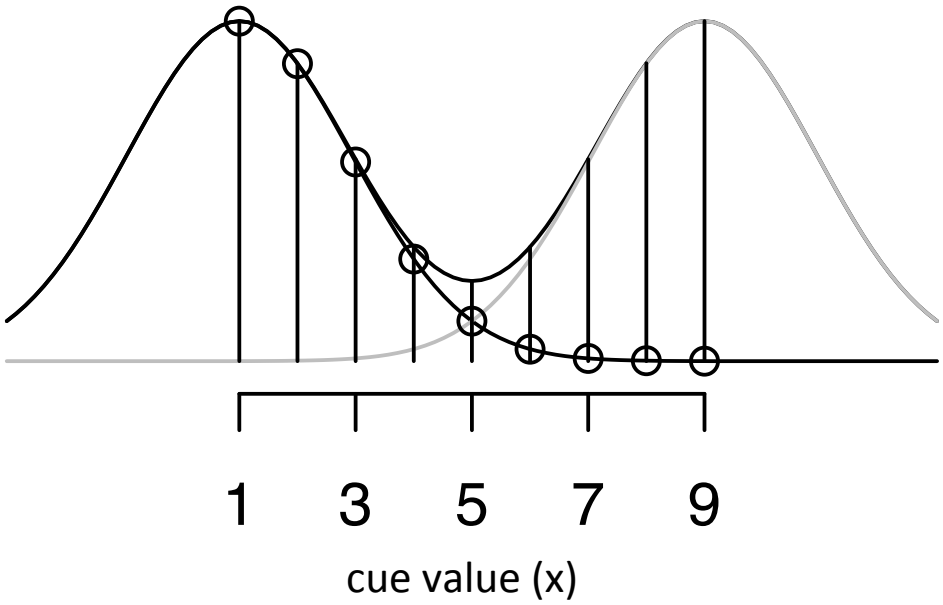
Categorical speech perception

$$p(c = \mathrm{b}|x = 4) = \frac{p(x = 4|c = \mathrm{b})p(c = \mathrm{b})}{p(x = 4)}$$



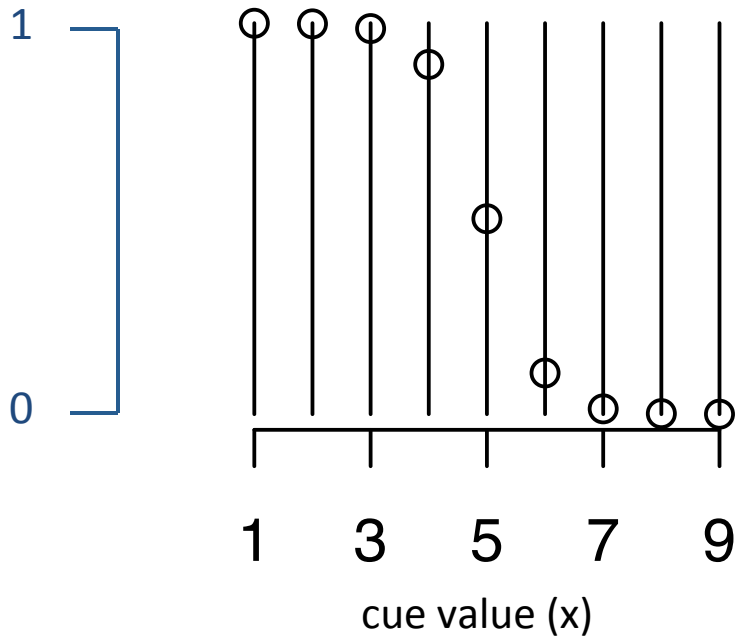
Categorical speech perception

$$p(c = \mathrm{b}|x) = \frac{p(x|c = \mathrm{b})p(c = \mathrm{b})}{p(x)}$$



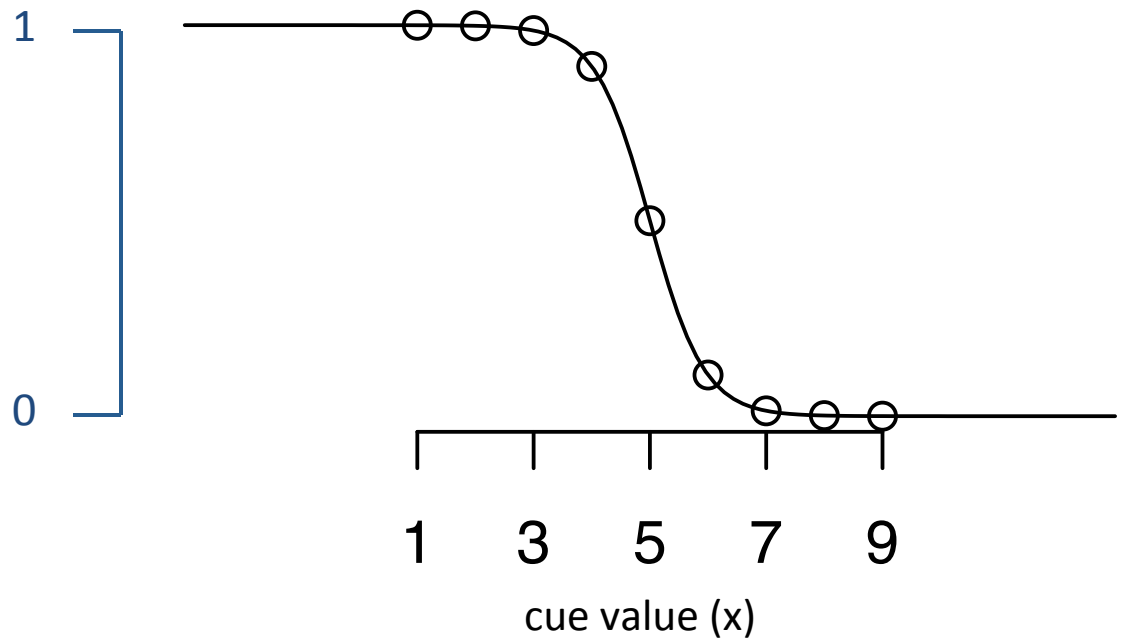
Categorical speech perception

$$p(c = \mathrm{b}|x) = \frac{p(x|c = \mathrm{b})p(c = \mathrm{b})}{p(x)}$$



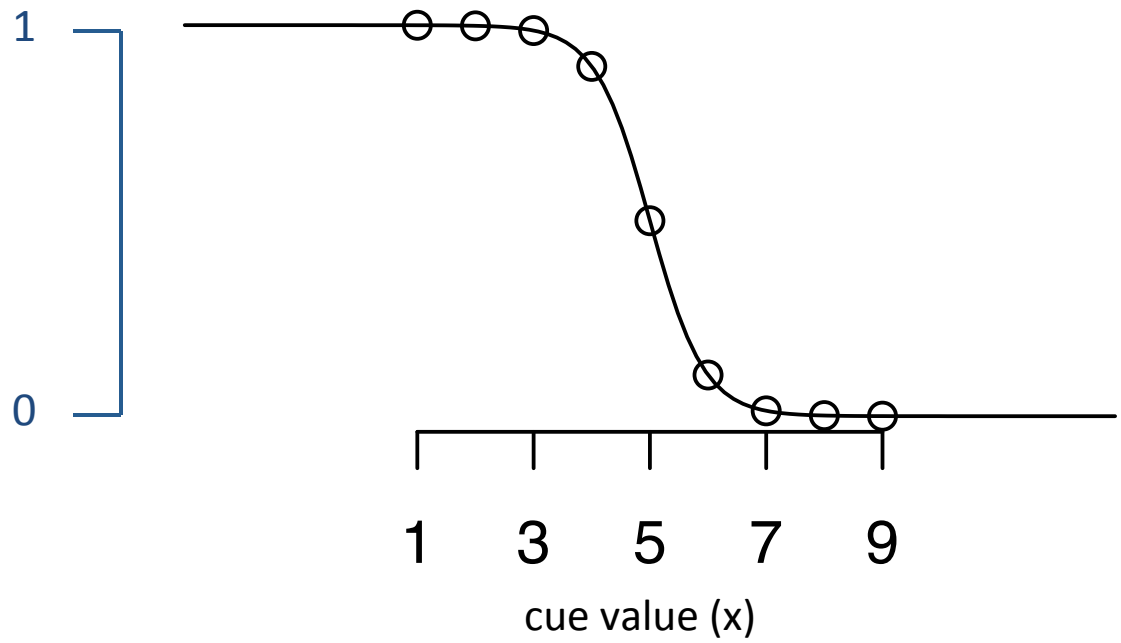
Categorical speech perception

$$p(c = \mathrm{b}|x) = \frac{p(x|c = \mathrm{b})p(c = \mathrm{b})}{p(x)}$$



Categorical speech perception

$$p(c = \mathrm{b}|x) = \frac{p(x|c = \mathrm{b})p(c = \mathrm{b})}{p(x)}$$



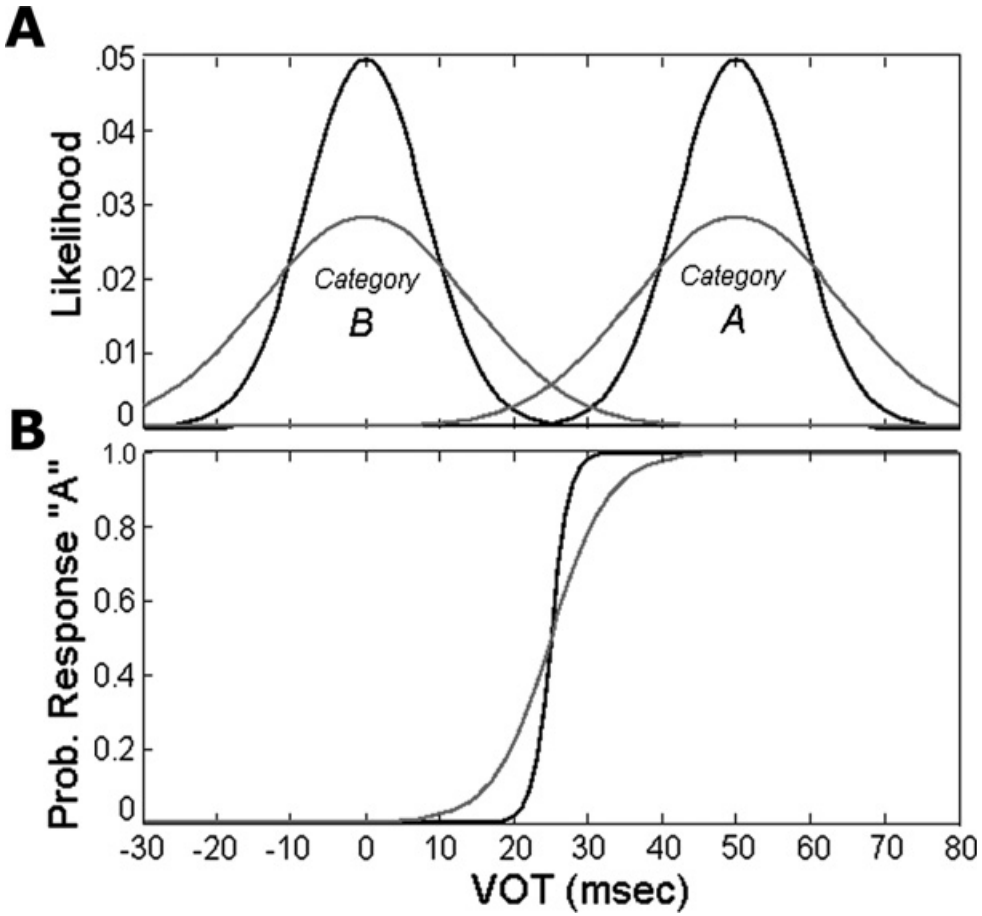
PREDICTIONS

- If people are behaving like an ideal listener, they should be sensitive to cue statistics
- Clayards et al. (2008) set out to test for sensitivity to **variance**

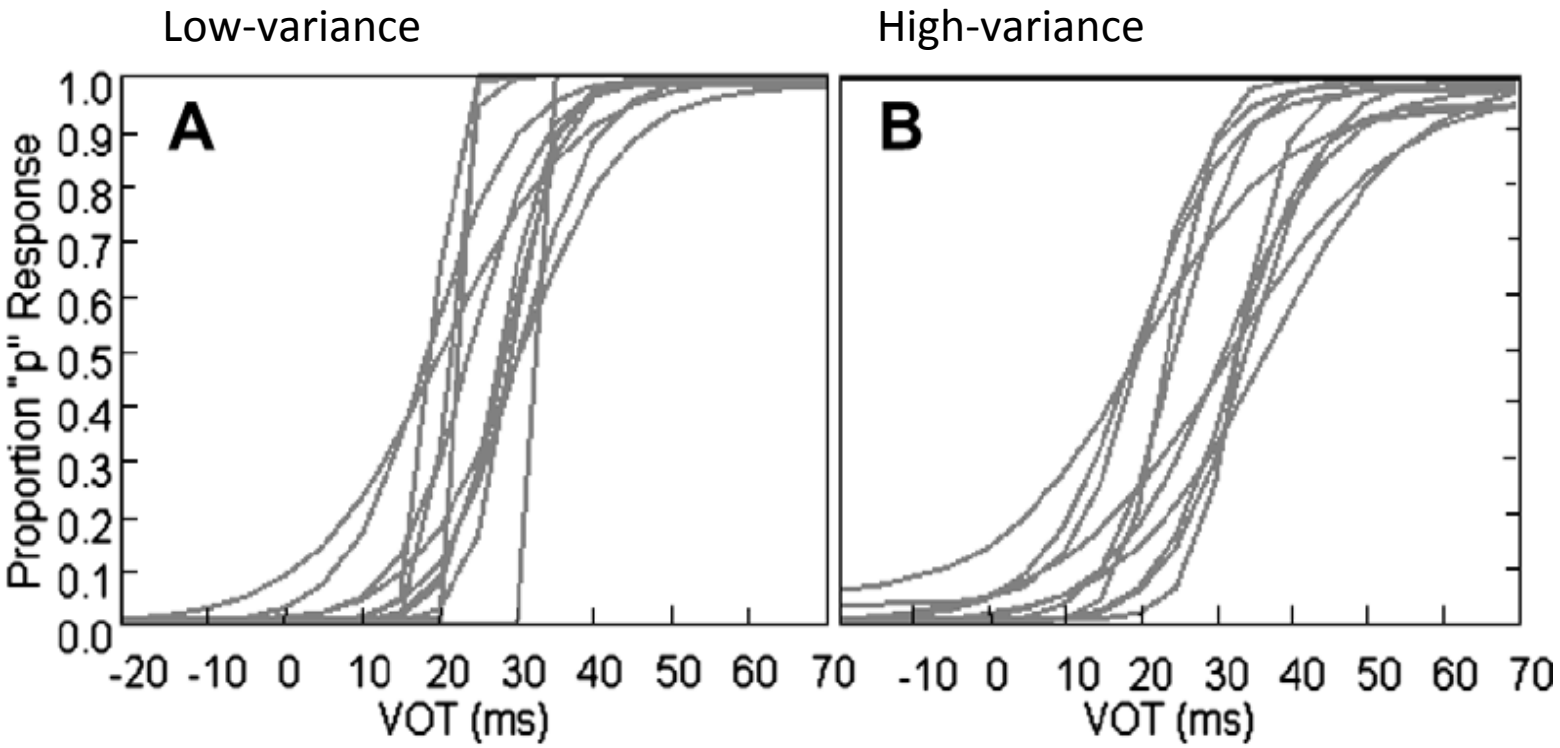
CLAYARDS ET AL. (2008)

- /b/-/p/ minimal pairs (VOT continnum)
- Two conditions: high variance and low variance VOT distributions.

CLAYARDS ET AL. (2008) PREDICTIONS



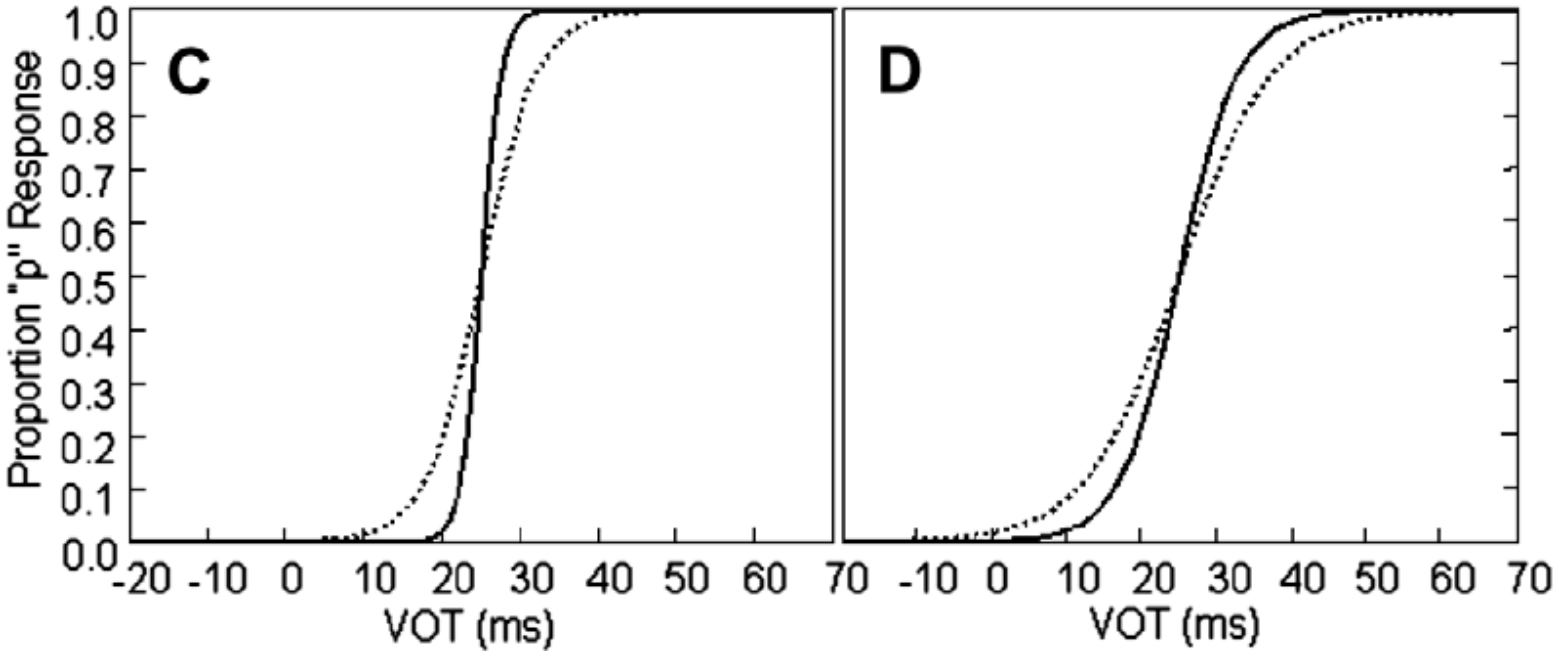
CLAYARDS ET AL. (2008) RESULTS



CLAYARDS ET AL. (2008) RESULTS

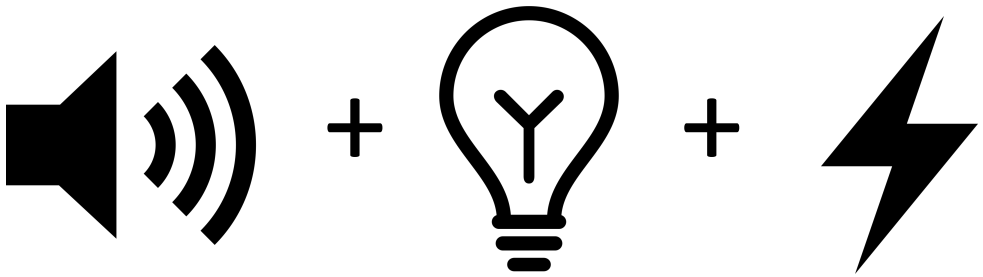
Low-variance

High-variance

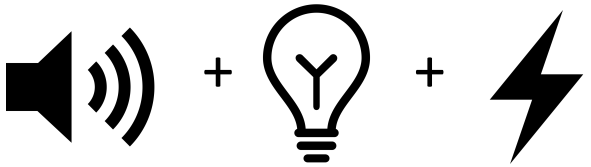


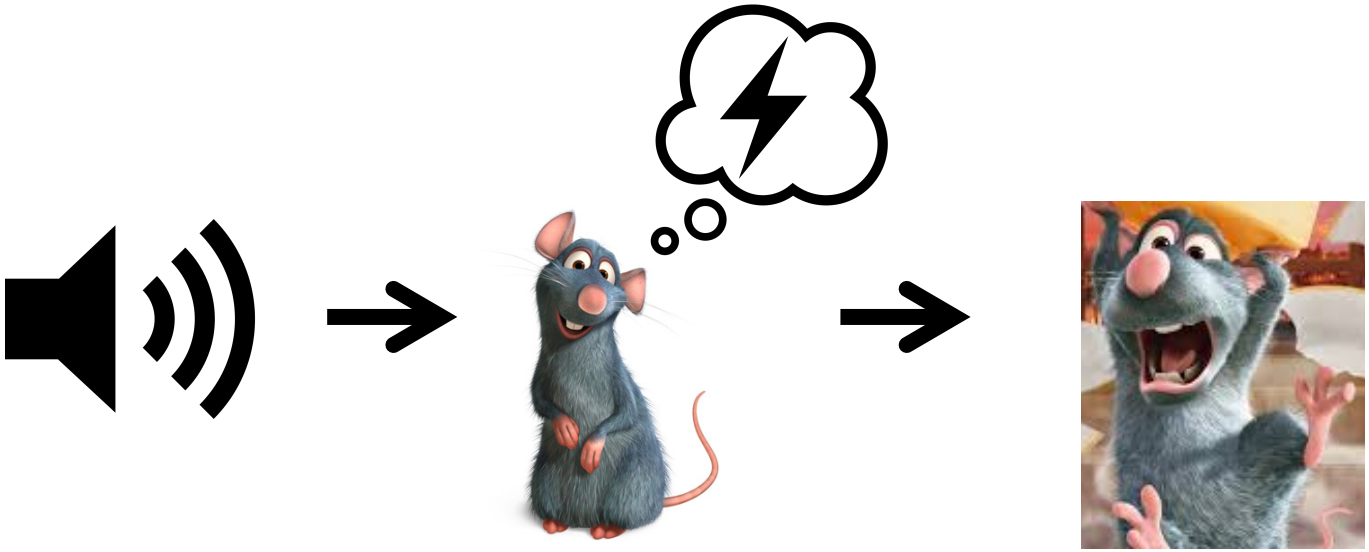
Backwards blocking/ Explaining away

Phase 1

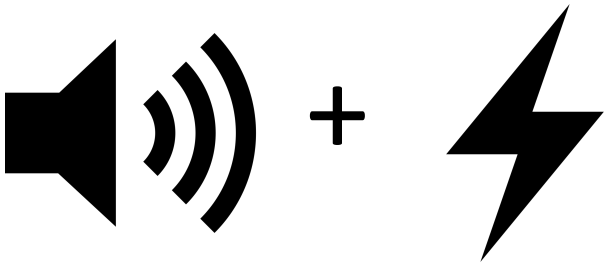


Phase 1: test

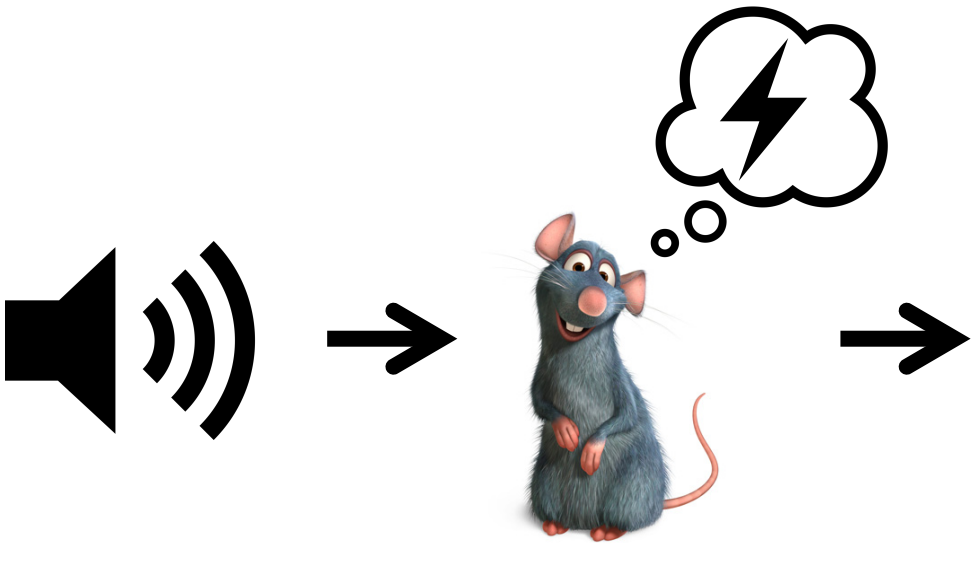


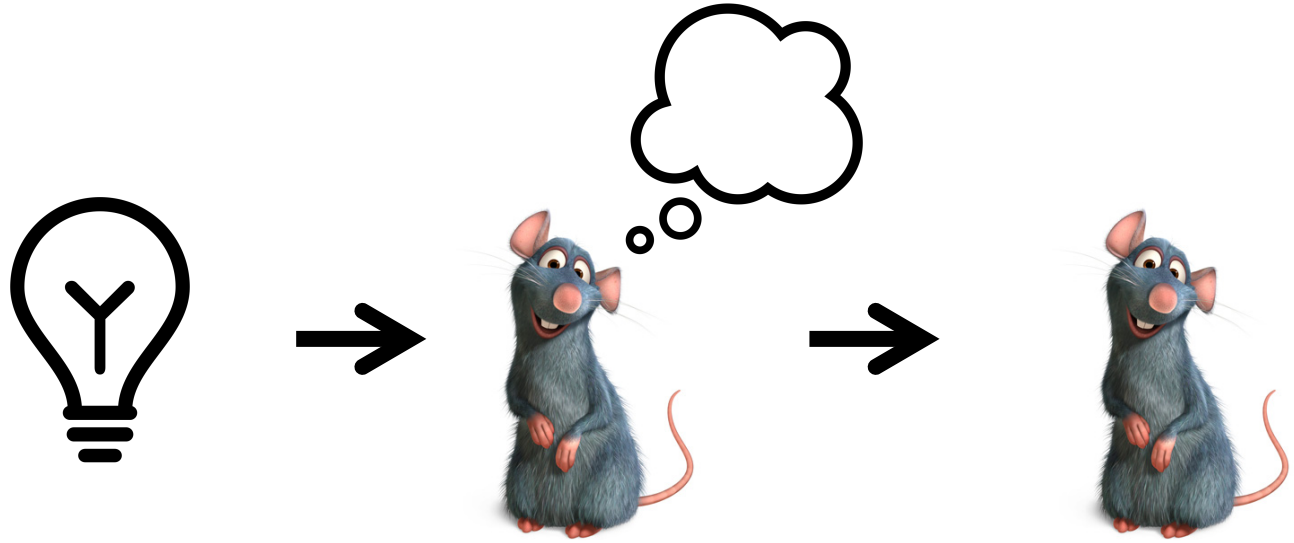


Phase 2



Phase 2: test





Explaining away & backward blocking

- Bayesian model:
- Cue weights (how well cue predicts shock), sum to 1.
- After exposure to light+sound+shock, could be anywhere from all light to all sound and mixtures in between.
- After exposure to sound+shock, high weight for sound is supported, and hence weight for light drops.