

G89.2223 Perception

Cue Integration

Laurence T. Maloney



Visual Tasks (Estimation)



Constantine Brancusi

Size?
Shape?
Distance?

Material?
Weight?
Color?

Visual Tasks (Estimation)



Constantine Brancusi

Size?
Shape? *Depth*
Distance?

~~*Material?*~~
~~*Weight?*~~
~~*Color?*~~

We can develop cue combination models of these but not today!

There are typically multiple **cues** to depth/distance in a scene

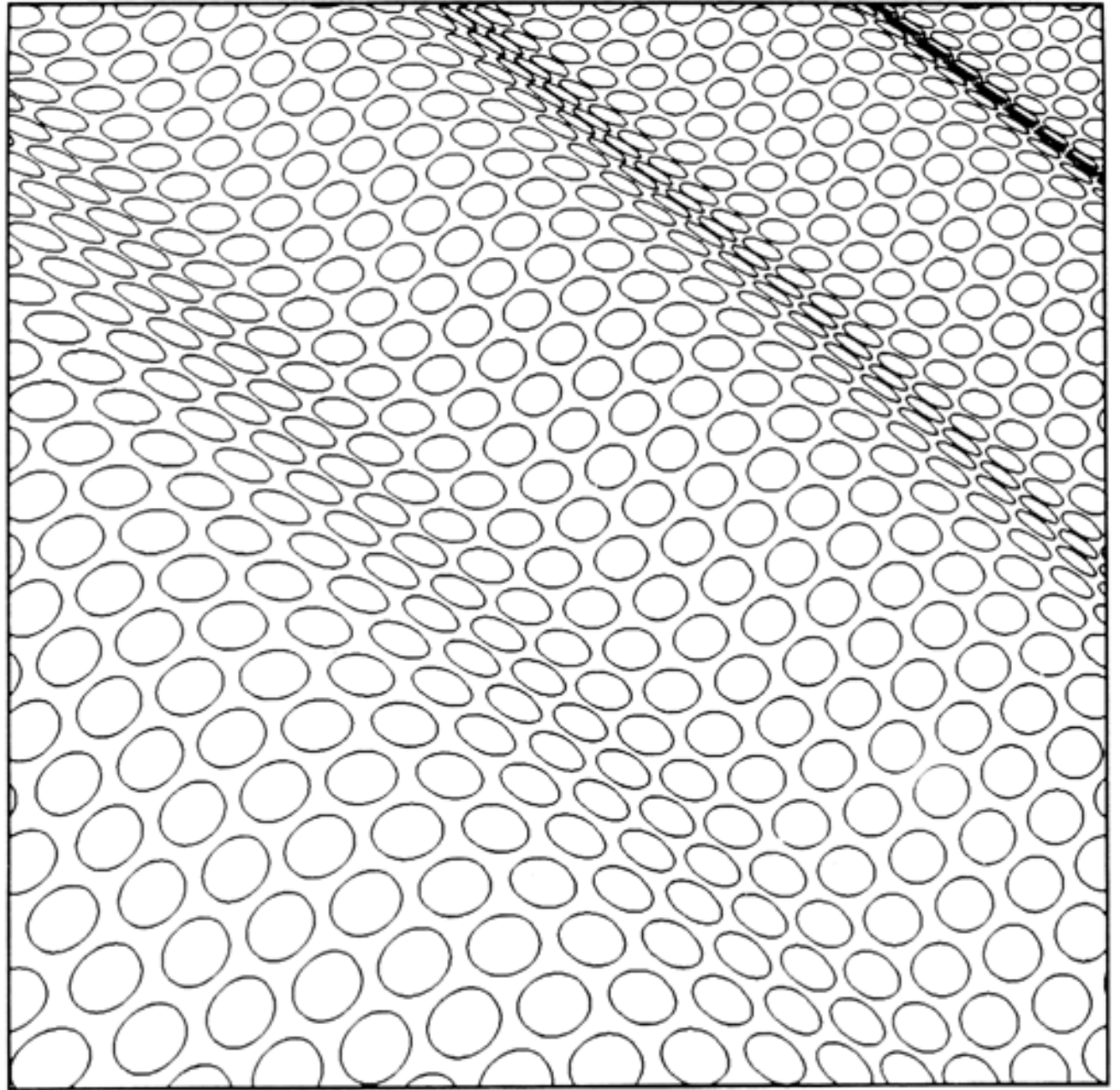


Carlo Crivelli c. 1486

Some examples of cues to
depth or shape

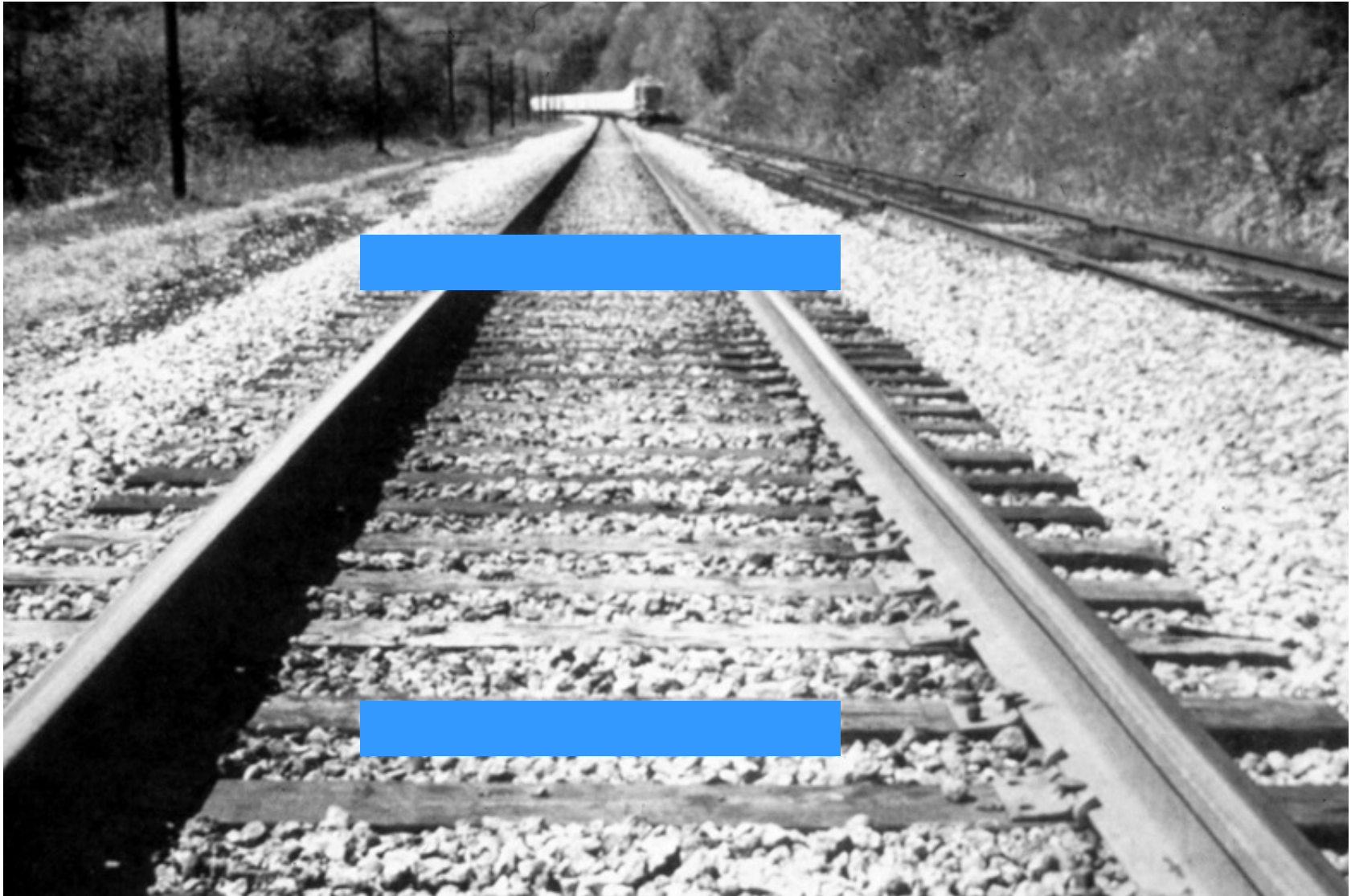
Texture Gradients

1. Density
2. Foreshortening
3. Size

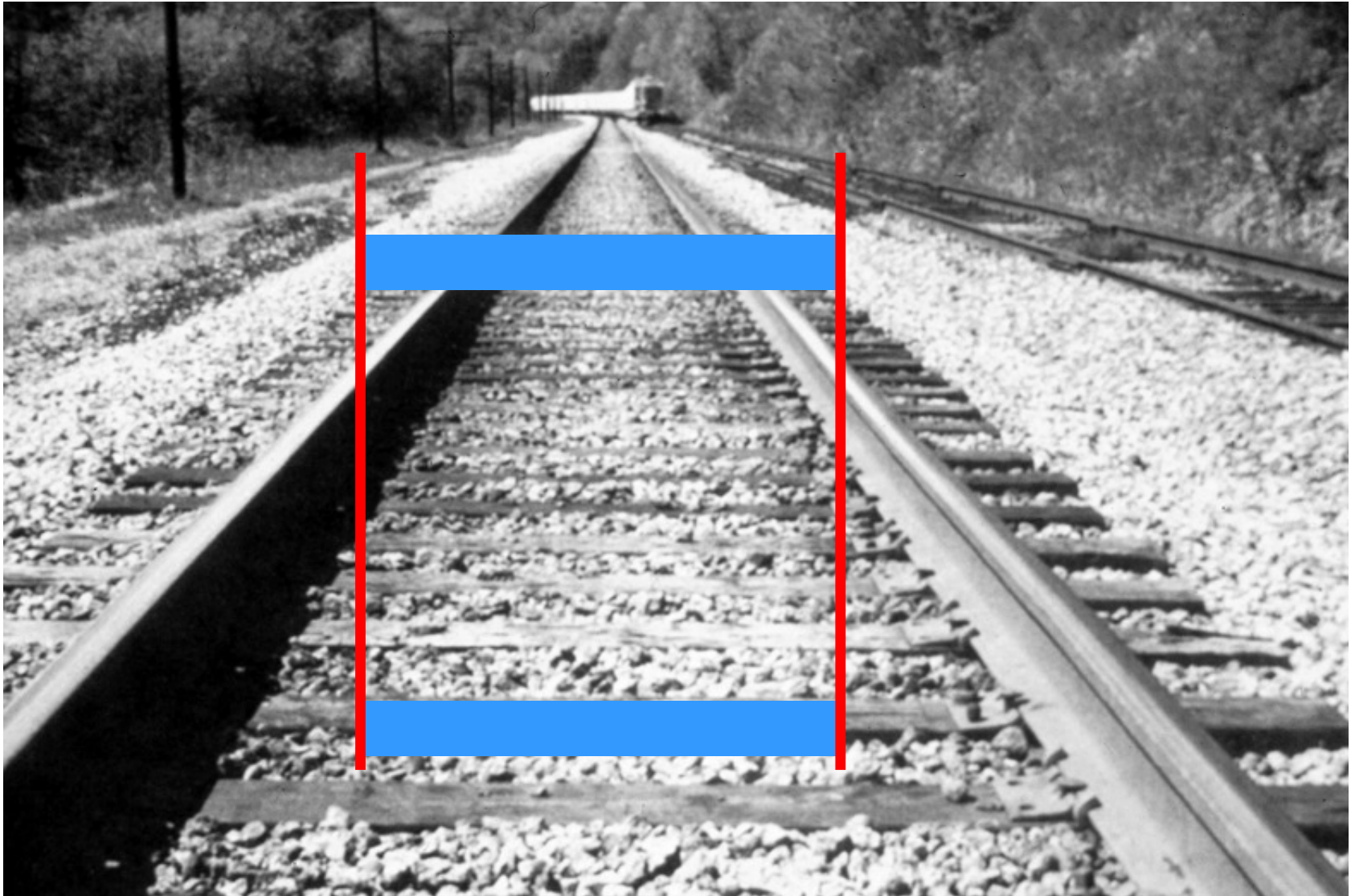


Linear perspective

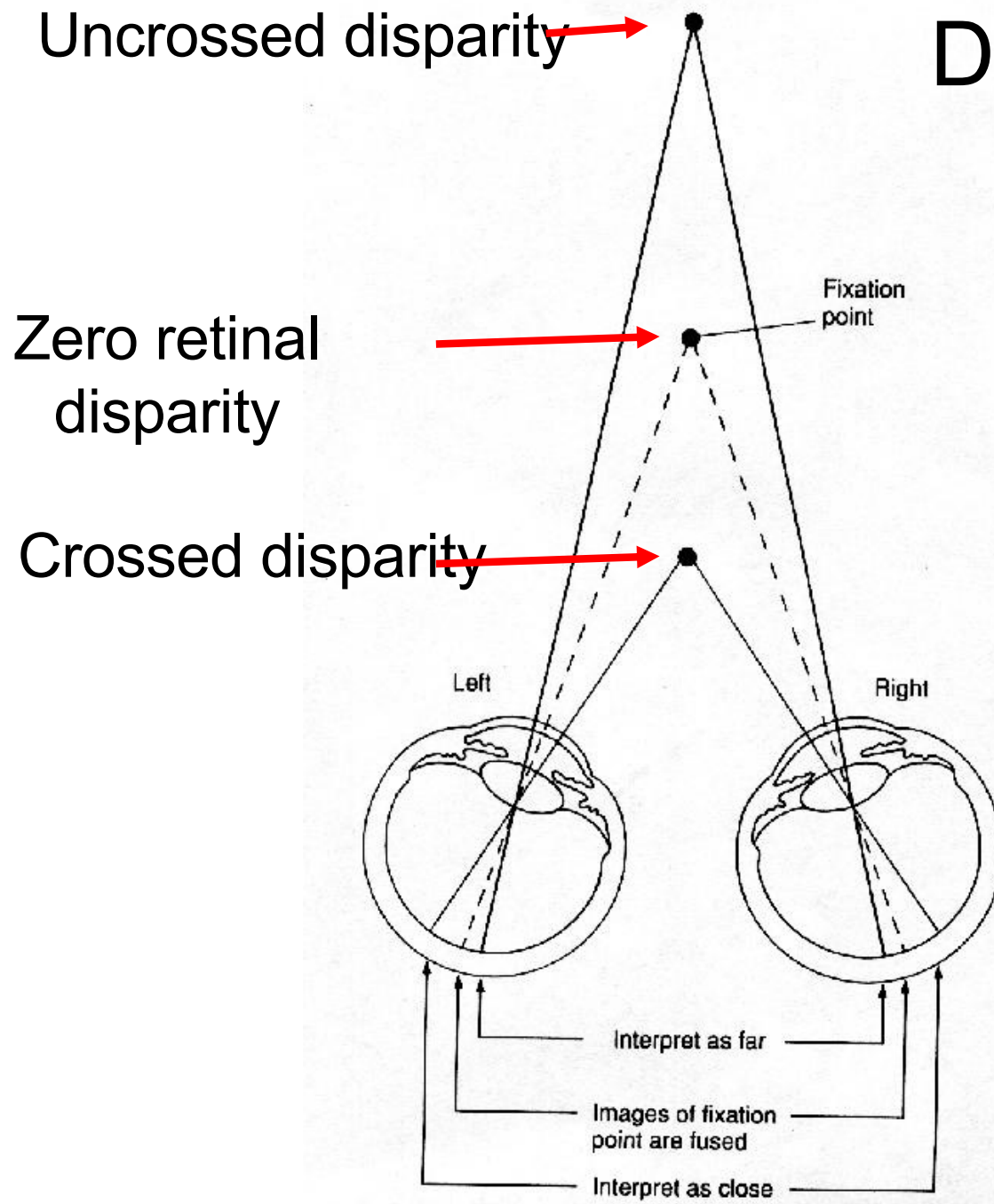
Aside: illusion



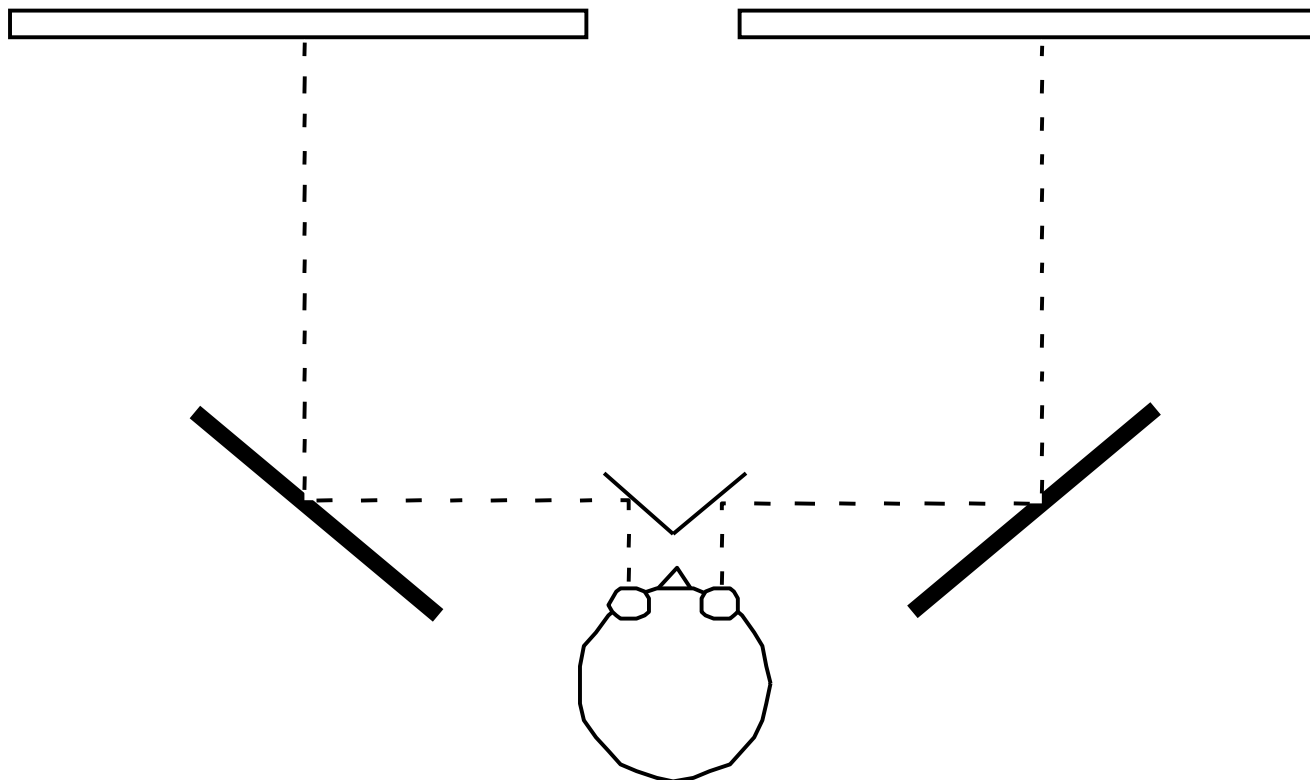
Linear perspective



Binocular Disparity



Dual mirror stereoscope



Cues in Conflict

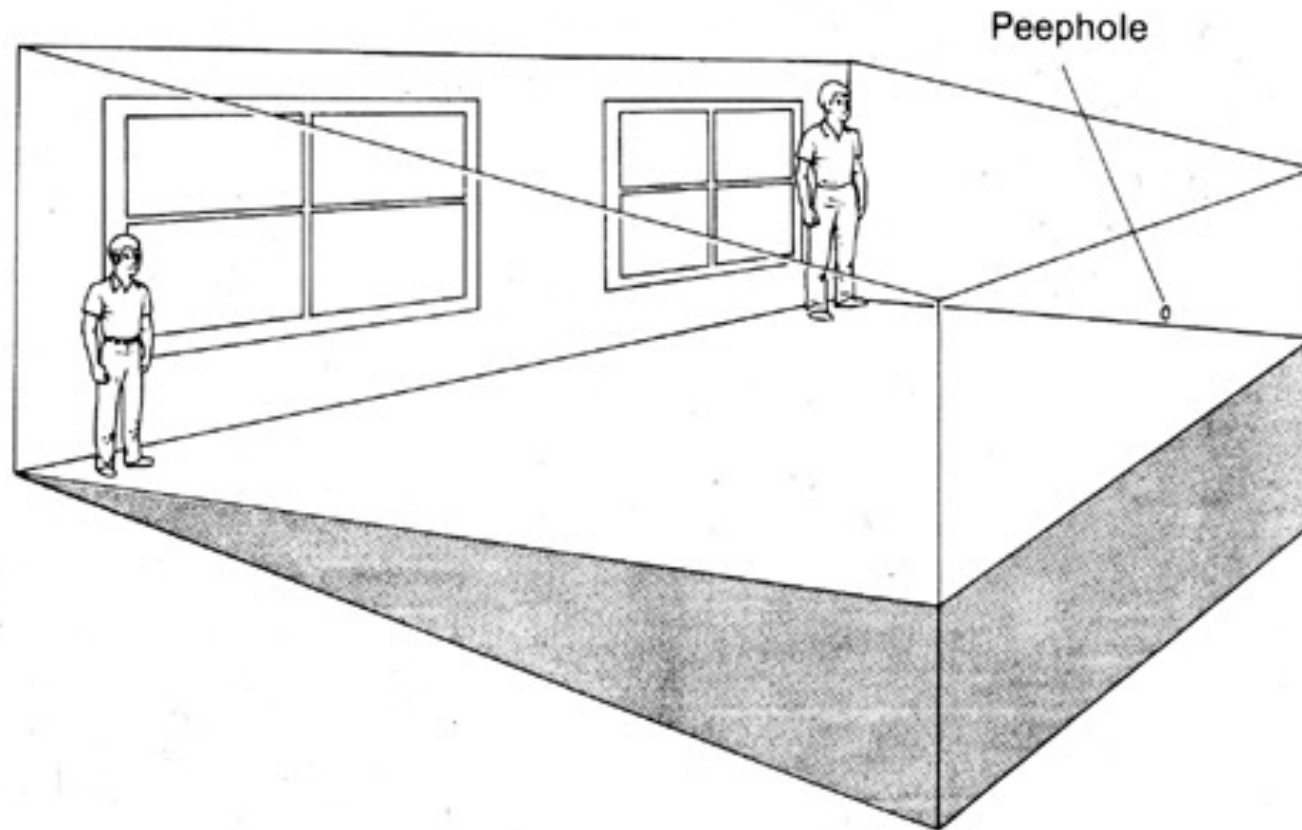
When we have multiple cues to depth or shape the cues may conflict

*When two cues disagree
(a lot) what do we do?*

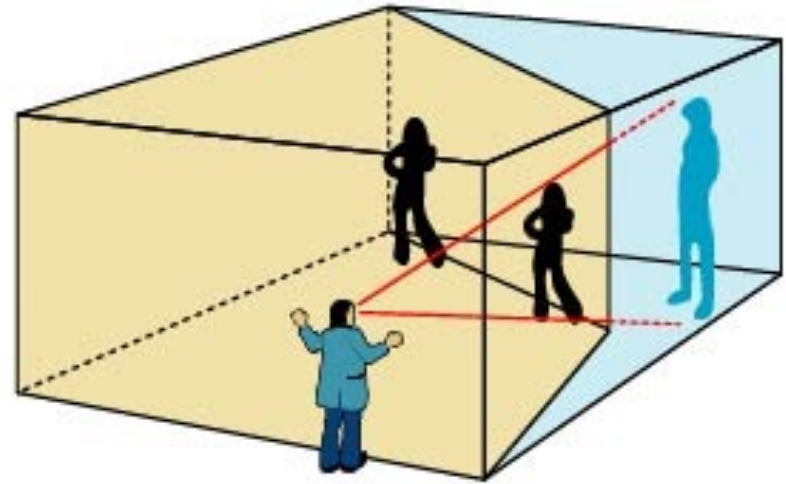
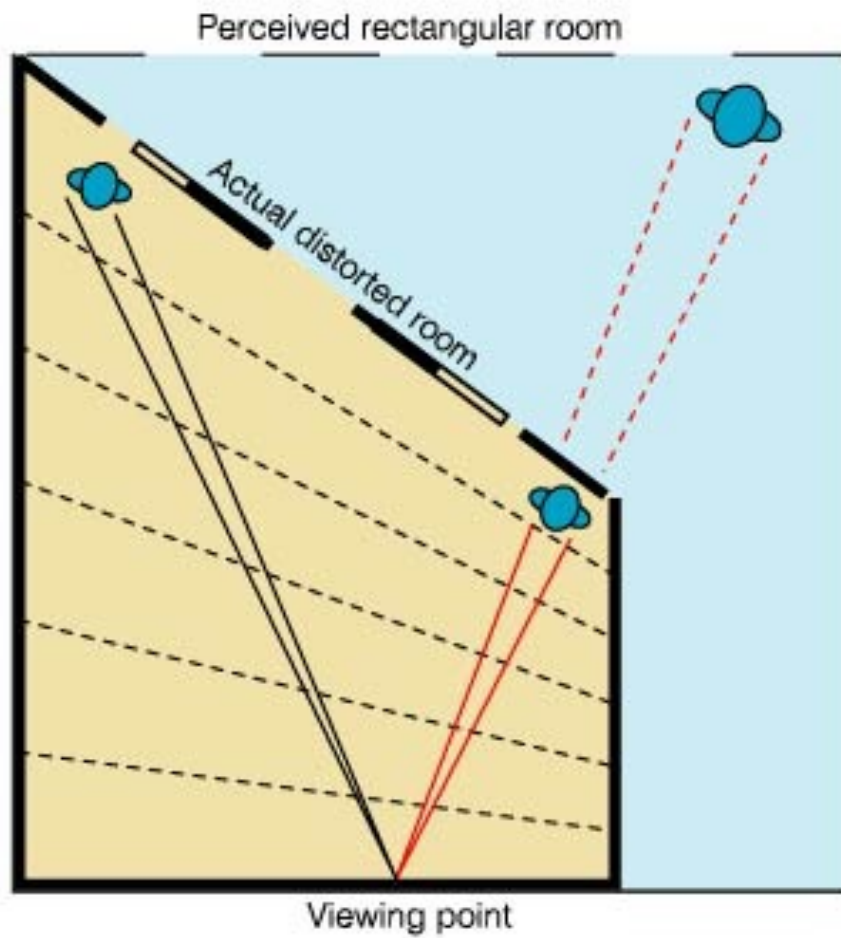
Ames Room



Ames Room: Real Layout



Ames room



Cue Integration

When we have multiple cues to the same thing:

What should we do?

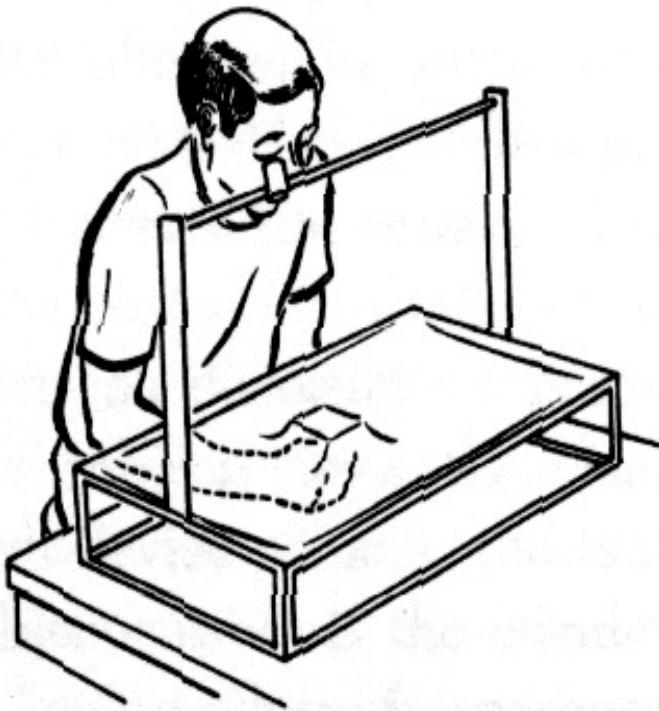
Pick the best?

Somehow integrate them?

What does the brain do?

Rock & Victor (1964)

**View object through distorting lens
while exploring object haptically**



Irv Rock

**Visually and haptically specified shapes differed.
What shape is perceived?**

Rock & Victor (1964)

Experimental Design

Stimulus Presentation		
Vision alone	Haptic alone	Conflict
<i>V</i> 	<i>H</i> 	<i>V</i> <i>H</i>  
<i>V</i> 	<i>H</i> 	<i>V</i> <i>H</i>  

Rock & Victor (1964)

Results

Stimulus Presentation		
Vision alone	Haptic alone	Conflict

<i>V</i>  13.4	<i>H</i>  23.1	<i>V</i> <i>H</i>   14.1 mm
<i>V</i>  14.1	<i>H</i>  20.5	<i>V</i> <i>H</i>   14.5 mm

Rock & Victor (1964) Results

Stimulus Presentation		
Vision alone	Haptic alone	Conflict
<i>V</i>  1.90	<i>H</i>  0.98	<i>V</i> <i>H</i>   1.85
<i>V</i>  13.4	<i>H</i>  23.1	<i>V</i> <i>H</i>   14.1 mm
<i>V</i>  14.1	<i>H</i>  20.5	<i>V</i> <i>H</i>   14.5 mm

***Vision seems to dominate haptic.
“Visual Capture”***

Cue Integration

When we have multiple cues to the same thing:

What should we do?

Pick the best?

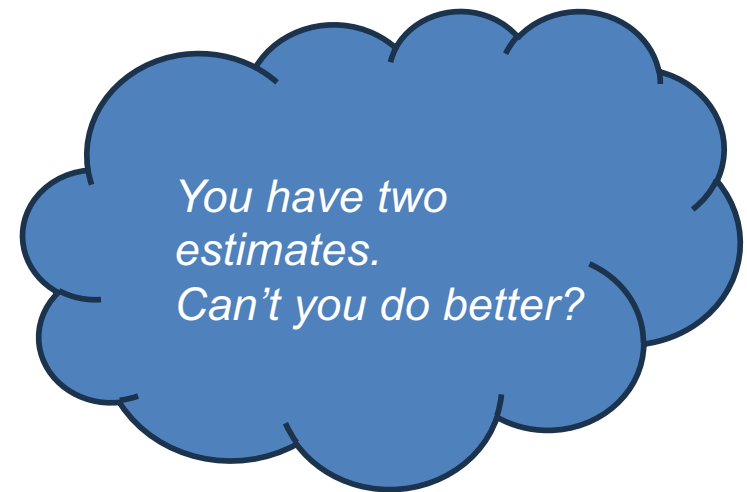
Average the cues?

Cue Integration

When we have multiple cues to the same thing:

What should we do?

Pick the best?



*Your inner
statistician*

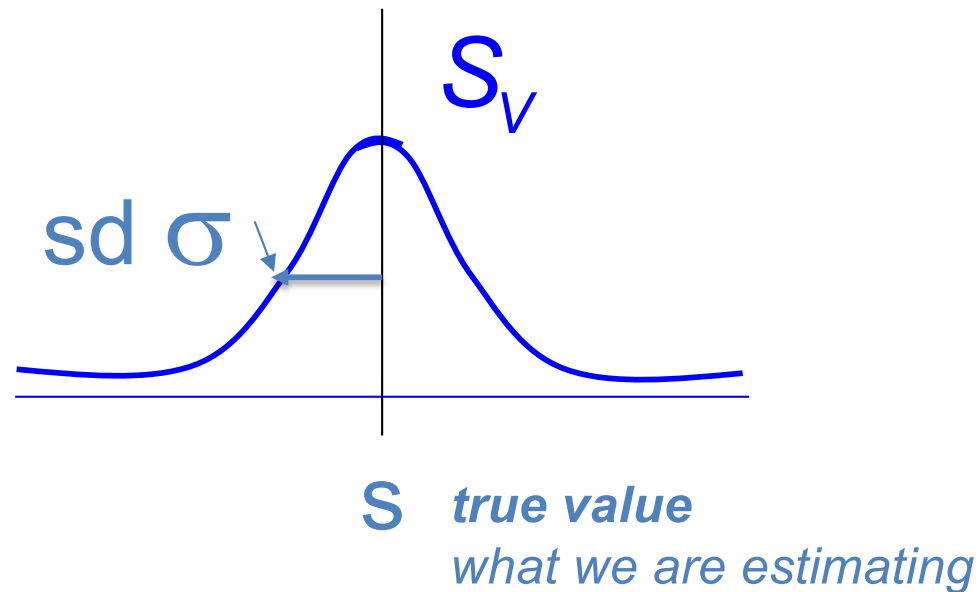
Modeling Cue Combination

Modeling Cue Integration

**Different people use different terminology.*

(How) should we model cues?

Gaussian random variables



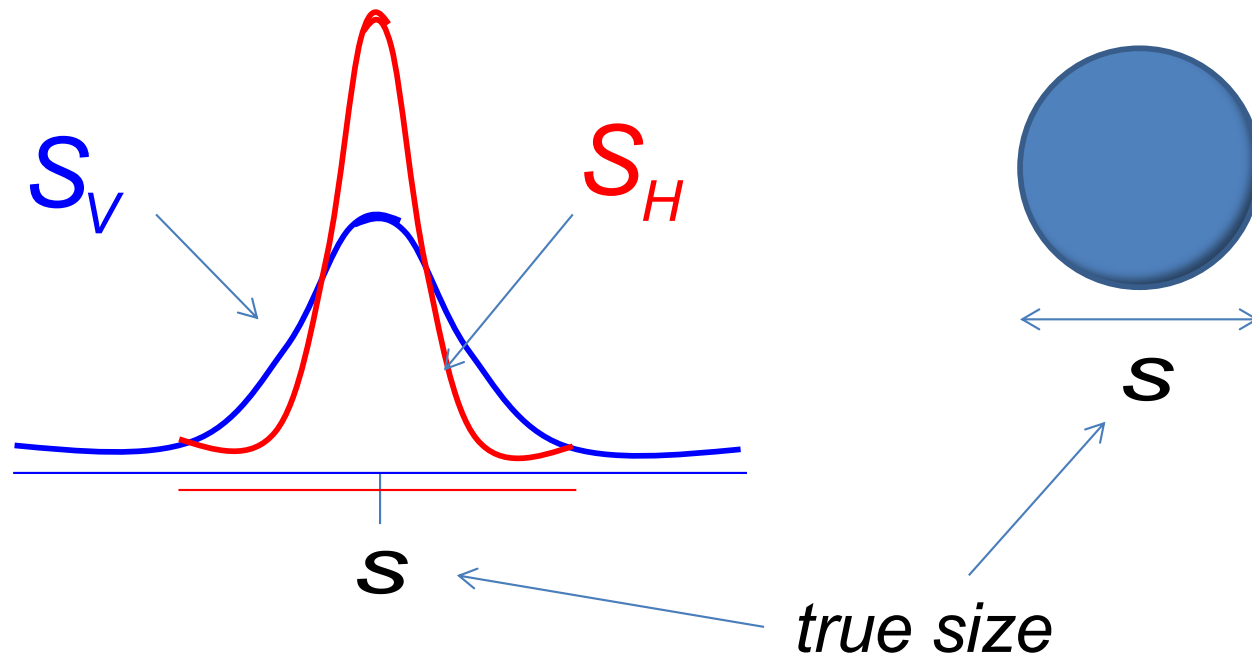
Two Cues

true size

$$S_H \sim \text{Gaussian}(s, \sigma_H)$$

$$S_V \sim \text{Gaussian}(s, \sigma_V)$$

some important
assumptions
Gaussian
Unbiased
Independent

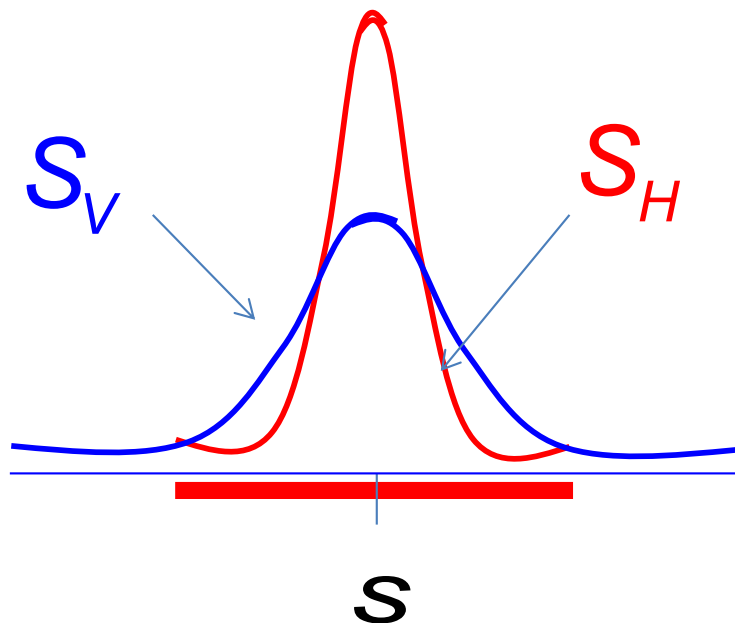


Thought Example

$$S_H \sim \text{Gaussian}(10 \text{ cm}, 1 \text{ cm})$$

$$S_V \sim \text{Gaussian}(10 \text{ cm}, 2 \text{ cm})$$

*haptic cue is
less variable*



A Thought Game

Pick a cue

\$10 if you are within
1 cm of s

Which cue?

Chances of winning?

Can we do better than using just S_H ?

Notation

σ standard deviation

σ^2 variance

Exercise

Some Possible Rules

1. Fixed hierarchy rule

Vision > Auditory > Haptic > TikTok hypothetical order
Use first available cue in order above.

2. Best single cue

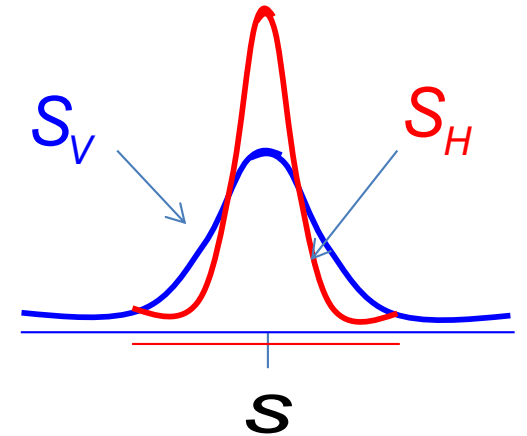
Use cue with lowest variance.
Discard others.

How does the visual system get an estimate of variance?

3. Weighted average of the cues ...

How choose weights?

Can we do better by
combining cues?

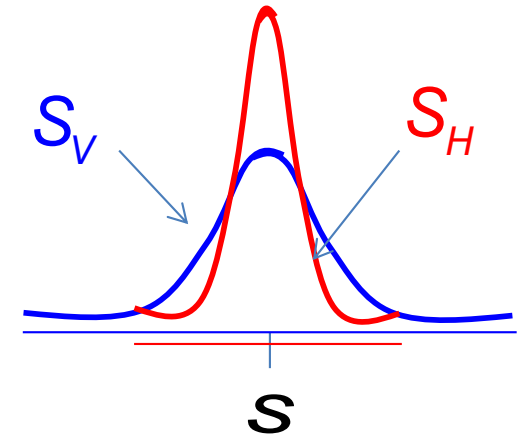


$$S = wS_H + (1-w)S_V$$

$$0 \leq w \leq 1$$

“weighted linear combination”

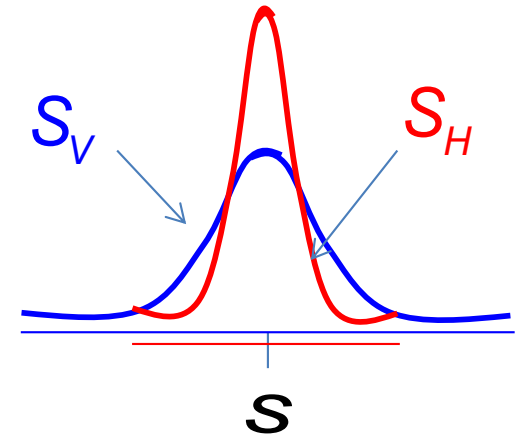
All Three Rules are weighted
linear combinations



$$S = wS_H + (1-w)S_V$$

1. Fixed hierarchy rule $w = 0$ (Vision)
2. Best single cue $w = 1$ (Haptic)
3. Combine the cues somehow ... $w = ?$

Can we do better by combining cues?



$$S = wS_H + (1-w)S_V$$

$$0 \leq w \leq 1$$

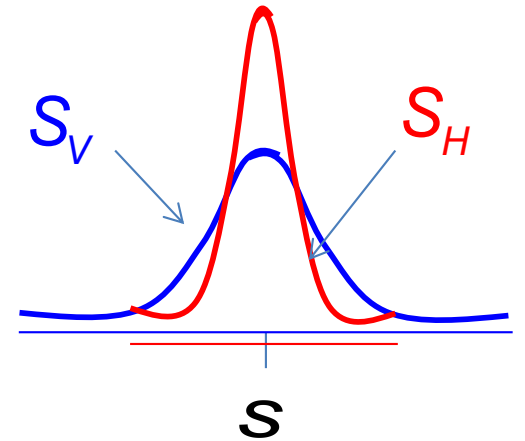
What is the “best” value of w ?

We have to decide what our goal is

We combine two cue S_H and S_V to get a new cue S .

Goal: We want S to be *unbiased* and to have *minimum variance*.

ASSUMPTION: UMVUE



$$\begin{aligned} E[S] &= wE[S_H] + (1-w)E[S_V] \\ &= ws + (1-w)s = s \end{aligned}$$

unbiased

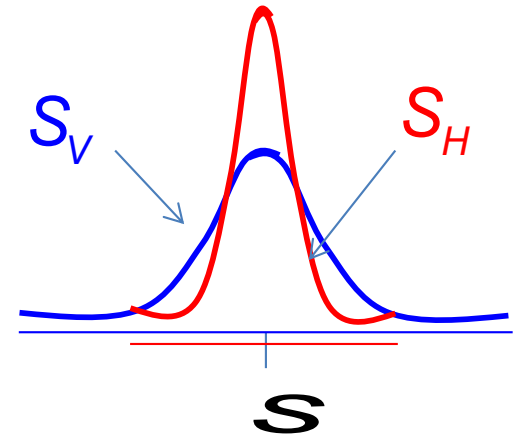
Review on variance σ^2

$$\text{Var}[sX] = s^2 \text{Var}[X]$$

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$$

X, Y *independent* variables

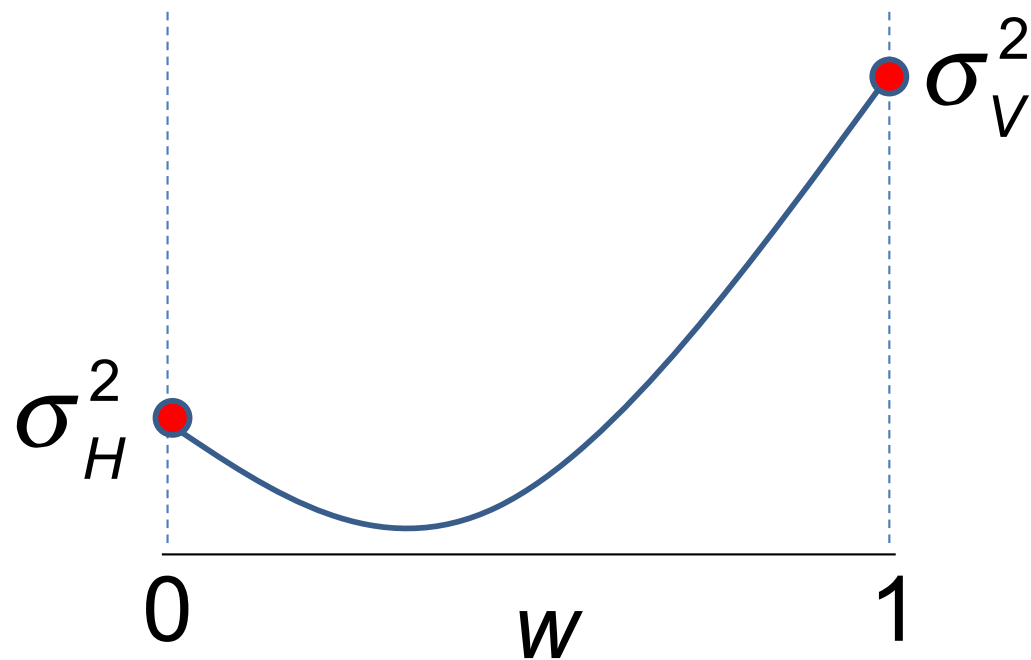
*ASSUMPTION: S_H and S_V independent
See Oruc, Maloney & Landy (2003)*



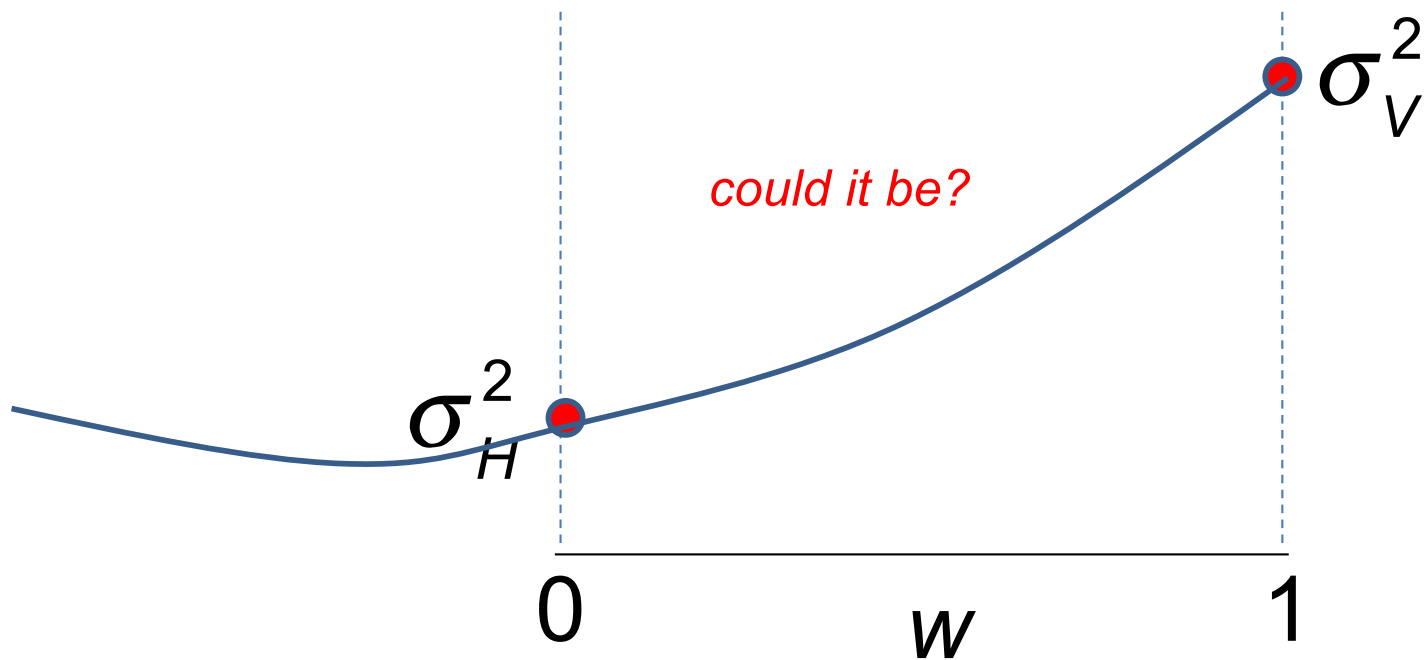
$$\begin{aligned} \text{Var}[S] &= w^2 \text{Var}[S_H] + (1-w)^2 \text{Var}[S_V] \\ &= w^2 \sigma_V^2 + (1-w)^2 \sigma_H^2 \end{aligned}$$

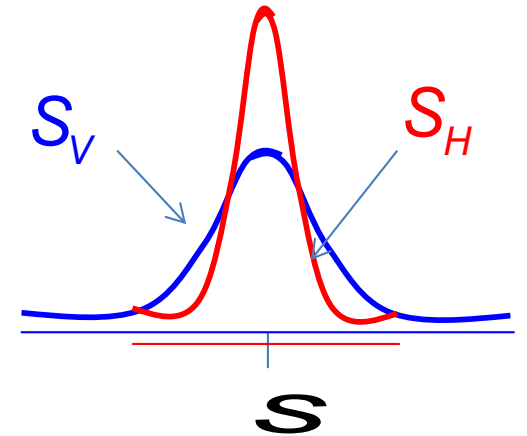
a parabola in w
up-facing or down?

$$\text{Var}[S] = w^2 \sigma_H^2 + (1-w)^2 \sigma_V^2$$



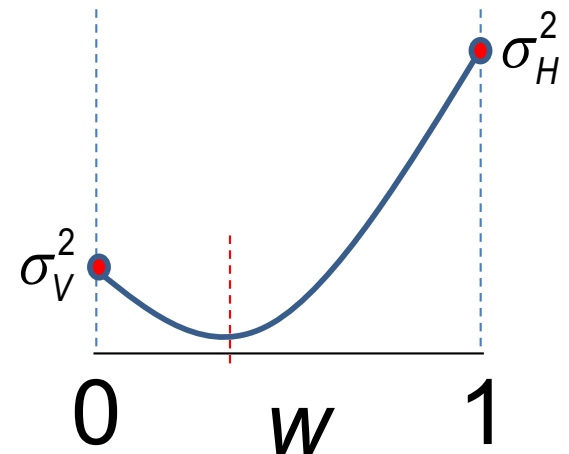
$$\text{Var}[S] = w^2 \sigma_H^2 + (1-w)^2 \sigma_V^2$$



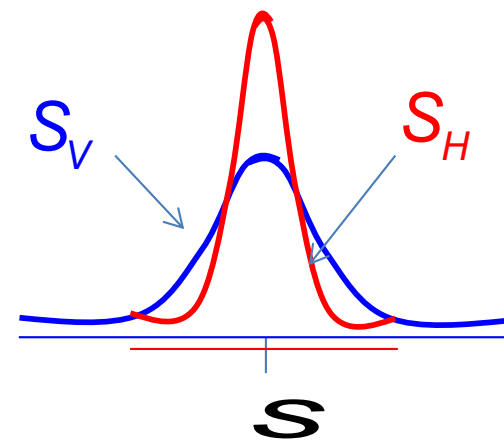


$$\frac{\partial \text{Var}[S]}{\partial w} = 2w\sigma_H^2 - 2(1-w)\sigma_V^2 \stackrel{!}{=} 0$$

$$w = \frac{\sigma_V^2}{\sigma_V^2 + \sigma_H^2}$$

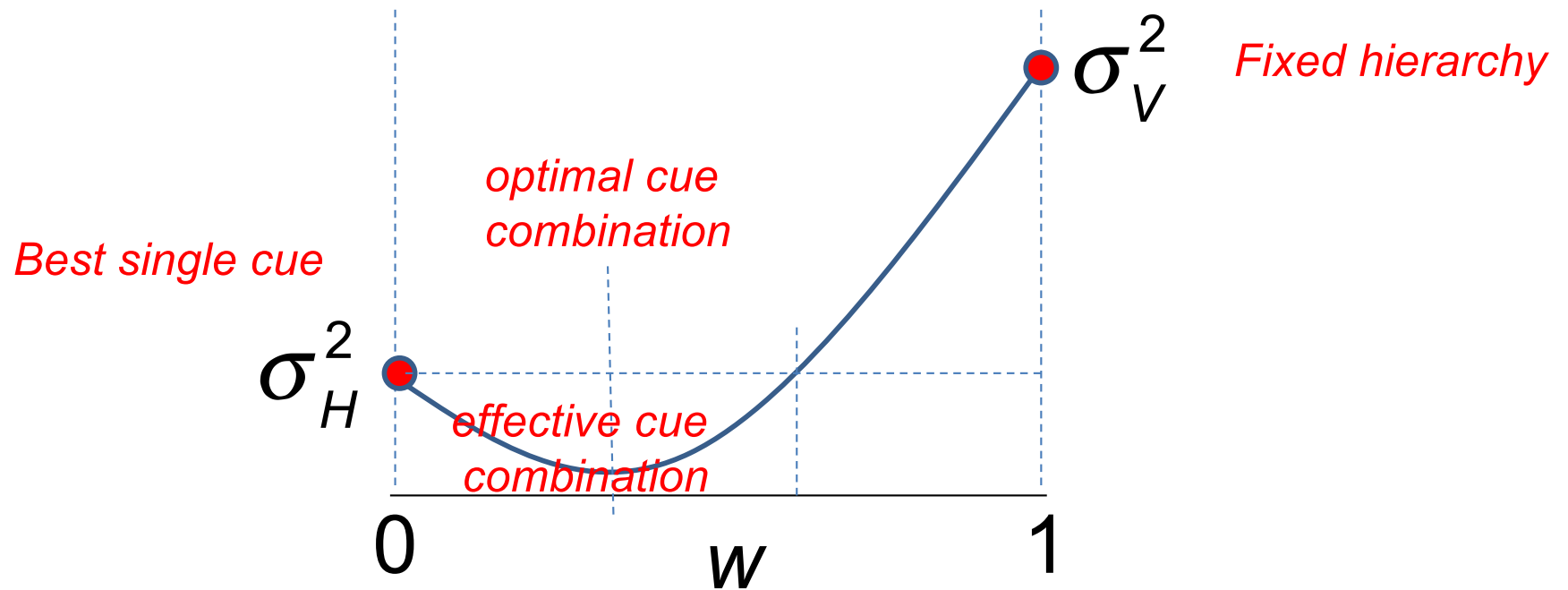


minimum or maximum?



$$W = \frac{\sigma_V^2}{\sigma_V^2 + \sigma_H^2}$$

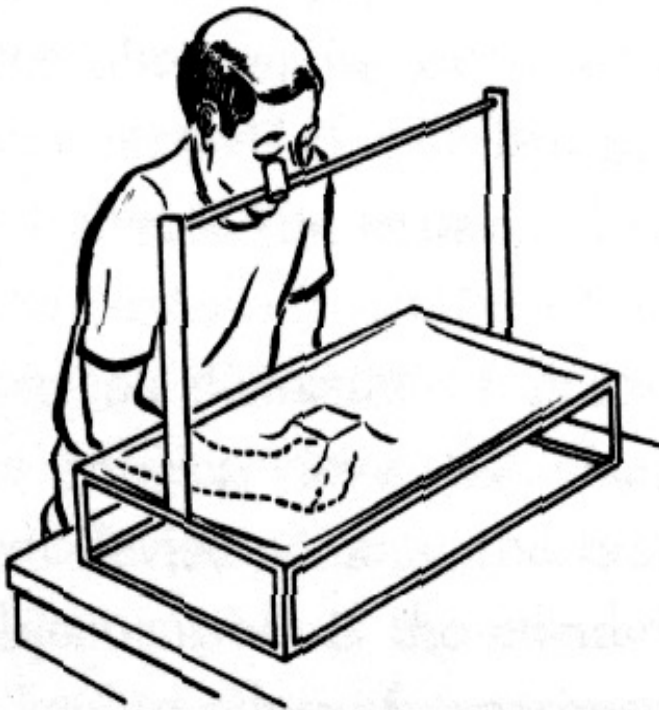
$$\text{Var}[S] = w^2 \sigma_H^2 + (1-w)^2 \sigma_V^2$$



It is always better to use all available cues – wisely.

Rock & Victor (1964)

**View object through distorting lens
while exploring object haptically**



Irv Rock

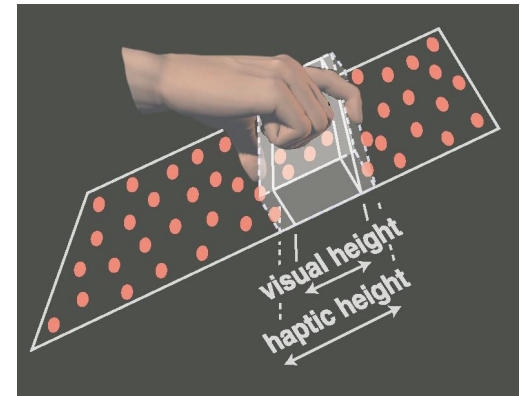
Why visual capture?

**Visually and haptically specified shapes differed.
What shape is perceived?**

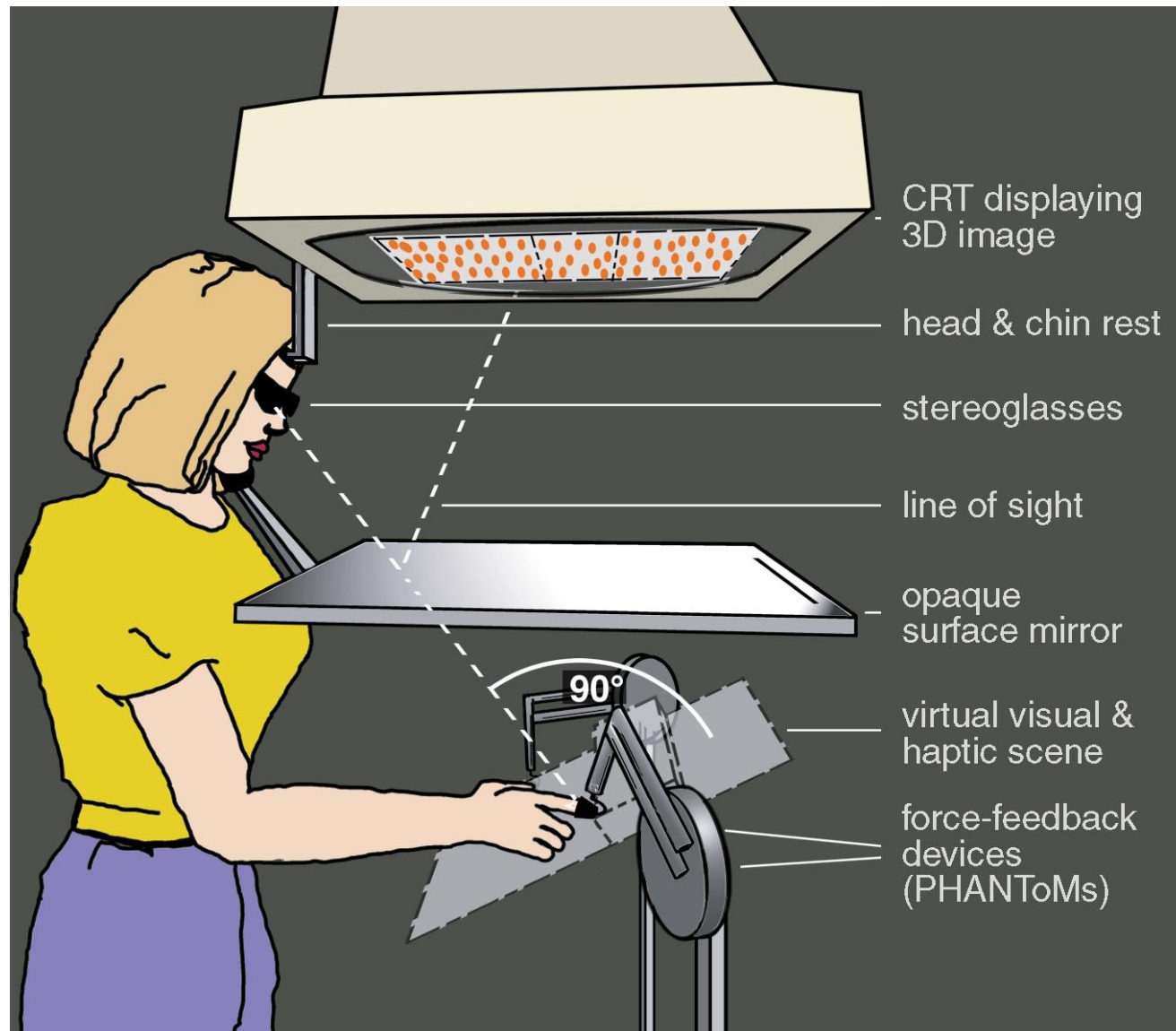
Humans integrate visual and haptic information in a statistically optimal fashion

Marc O. Ernst^{*} & Martin S. Banks

*Vision Science Program/School of Optometry, University of California, Berkeley
94720-2020, USA*



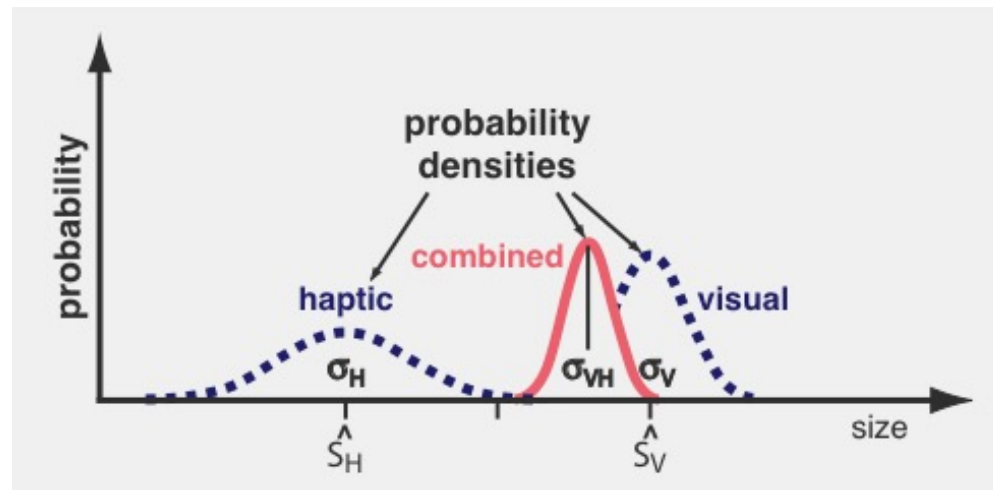
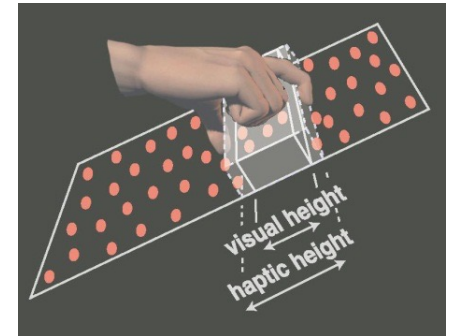
Visual/Haptic Setup



Visual Capture ?

Why should vision be the “gold standard”
all other modalities are compared to?

NOT MINIMUM VARIANCE



$$S_{VH} = w_V S_V + w_H S_H$$

Weights

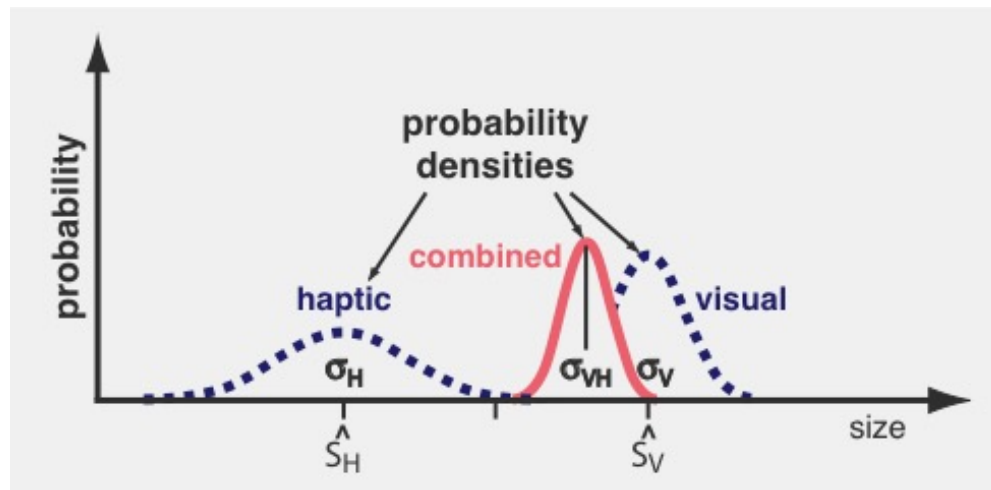
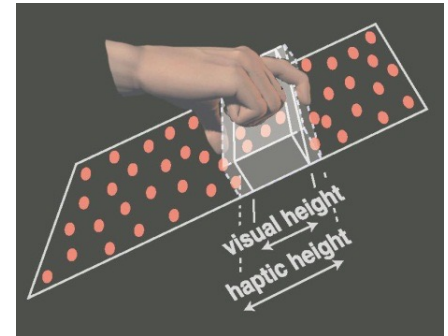
$$w_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2}$$

Variance

$$\frac{1}{\sigma_{VH}^2} = \frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2}$$

Visual Capture ?

Why should vision be the “gold standard” all other modalities are compared to?



$$S_{VH} = w_V S_V + w_H S_H$$

Weights

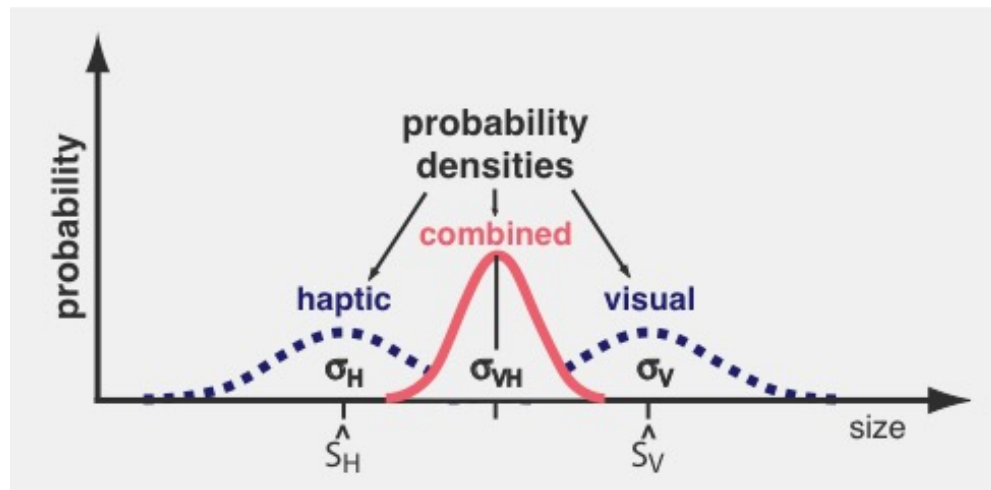
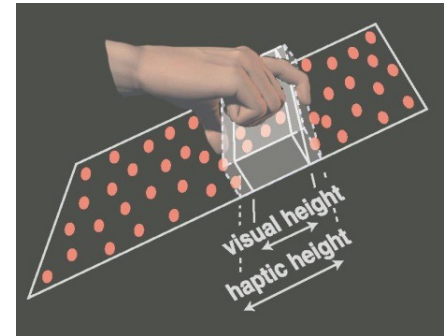
$$w_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2}$$

Variance

$$\frac{1}{\sigma_{VH}^2} = \frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2}$$

Visual Capture ?

Why should vision be the “gold standard” all other modalities are compared to?



$$S_{VH} = w_V S_V + w_H S_H$$

Weights

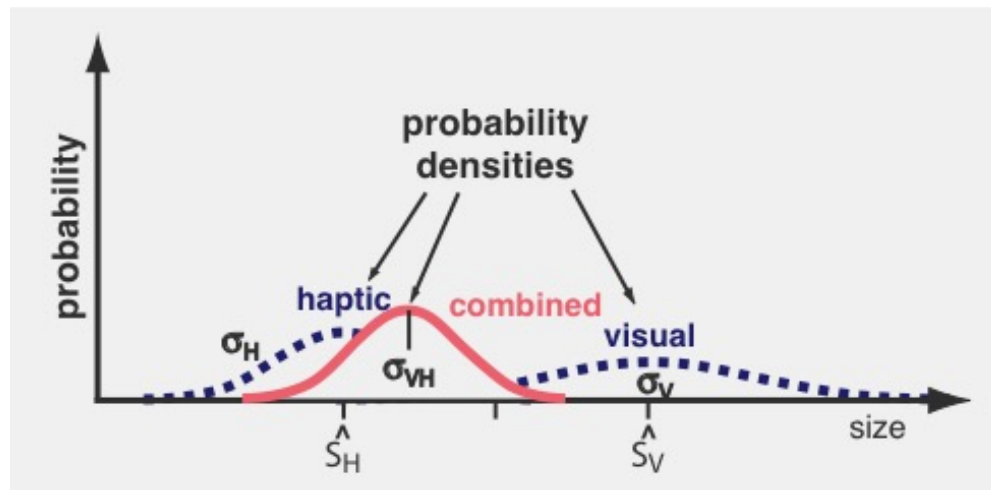
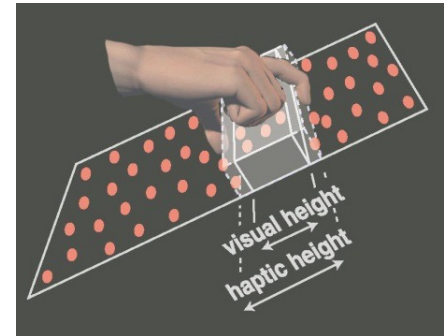
$$w_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2}$$

Variance

$$\frac{1}{\sigma_{VH}^2} = \frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2}$$

Visual Capture ?

Why should vision be the “gold standard” all other modalities are compared to?



$$S_{VH} = w_V S_V + w_H S_H$$

Weights

$$w_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2}$$

Variance

$$\frac{1}{\sigma_{VH}^2} = \frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2}$$

Weights

$$w_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2}$$

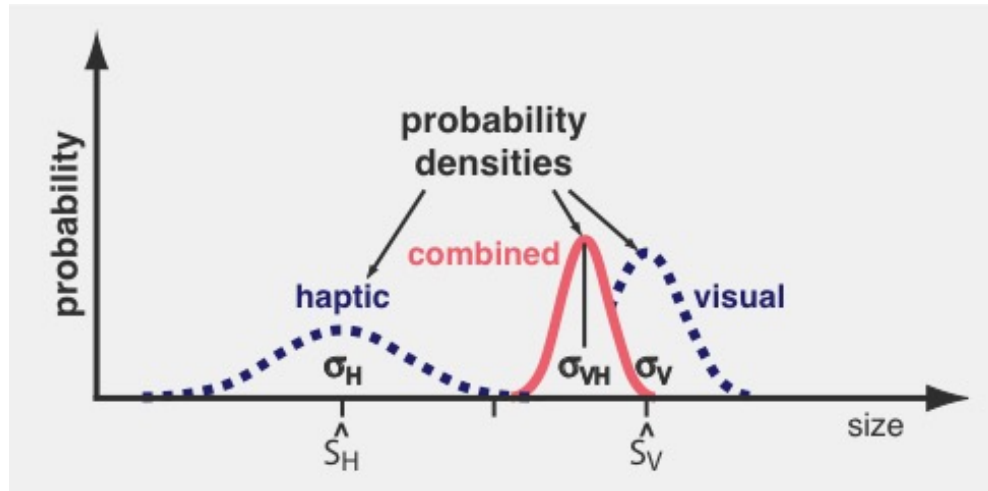
Variance

$$\frac{1}{\sigma_{VH}^2} = \frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2}$$

More cues always
lead to higher
precision

[on board]

Experimental Outline



1) manipulate & determine within-modality variances

- discrimination thresholds (2-IFC, constant stimuli)

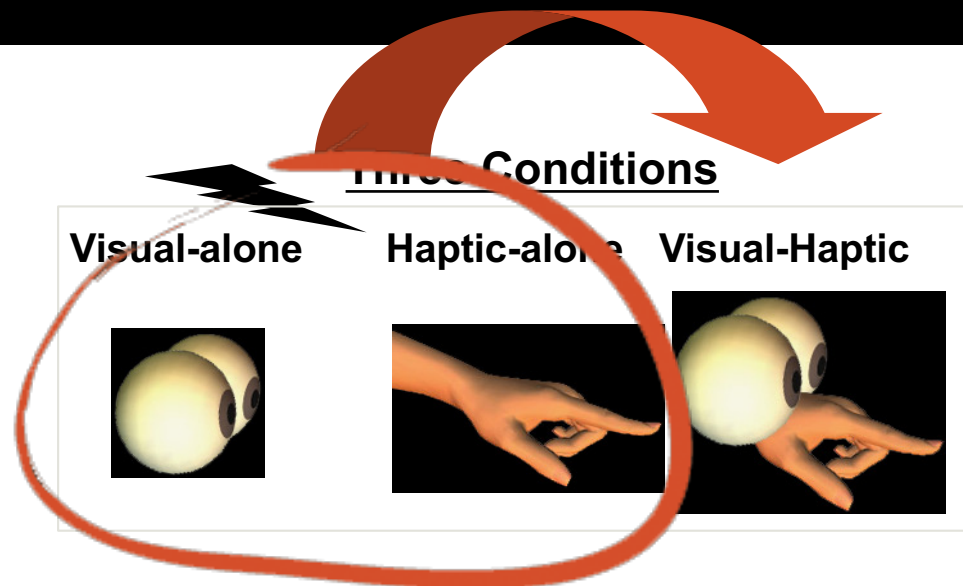
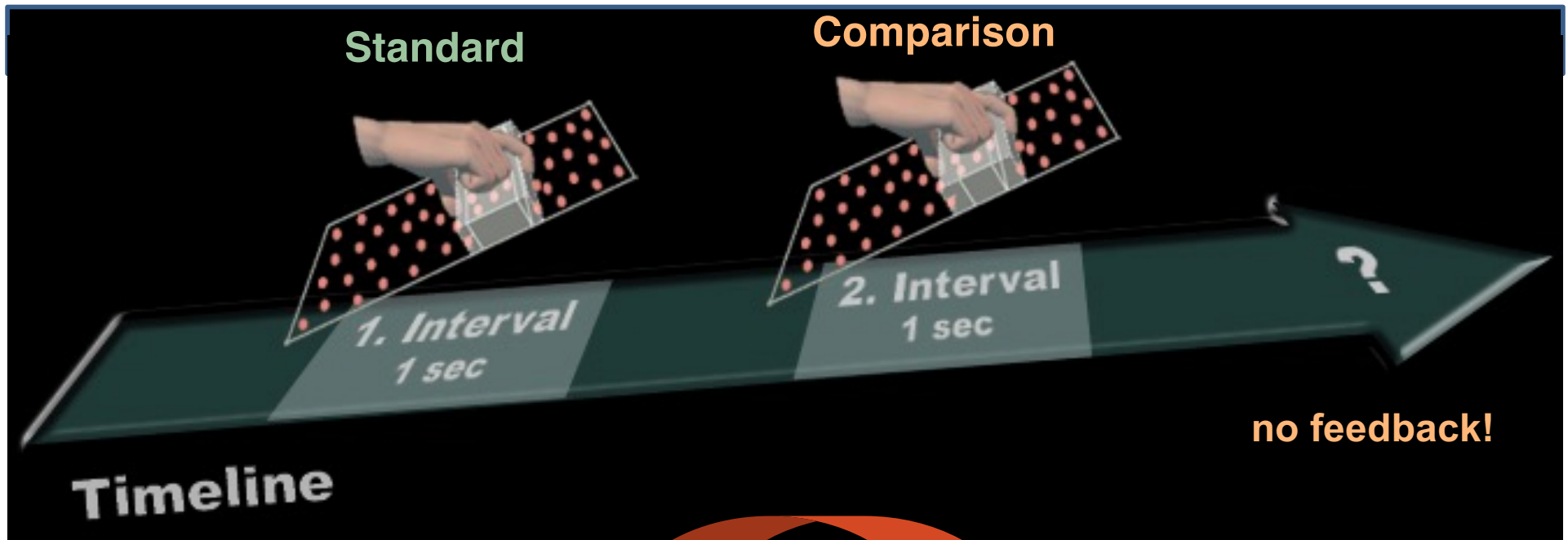
2) make predictions for combined performance

- using MLE model to predict weights & combined variance.

3) measure combined performance & compare to prediction

- similar to within-modality 2-IFC discrimination task (get PSE and thresholds)

2-IFC Task



STOP: How do we estimate the
variance (or SD) of a cue?

$$X \sim \textit{Gaussian}(s, \sigma)$$

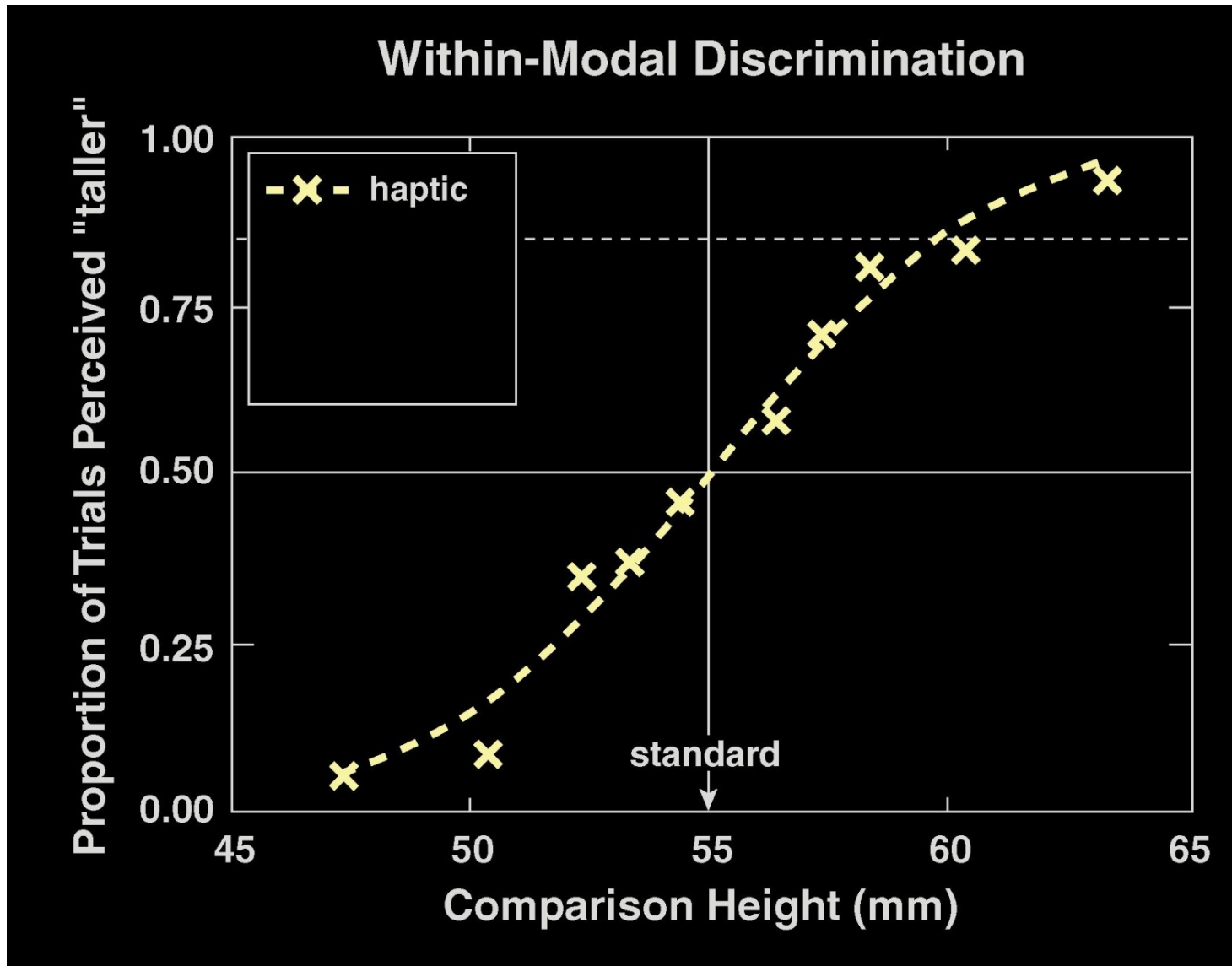
cdf cumulative distribution function
psychometric function

[whiteboard]

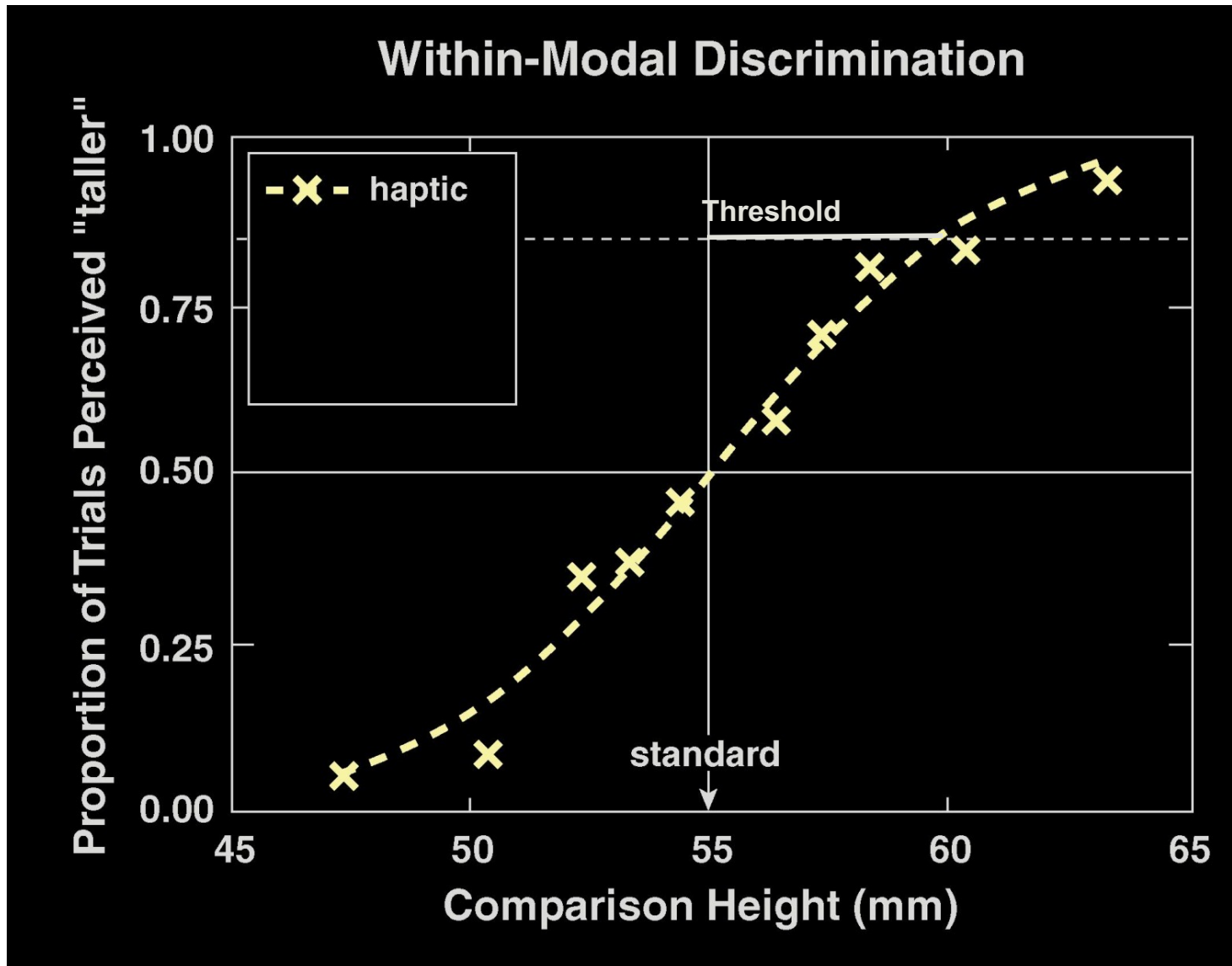
Effect of varying σ ?

location μ scale σ

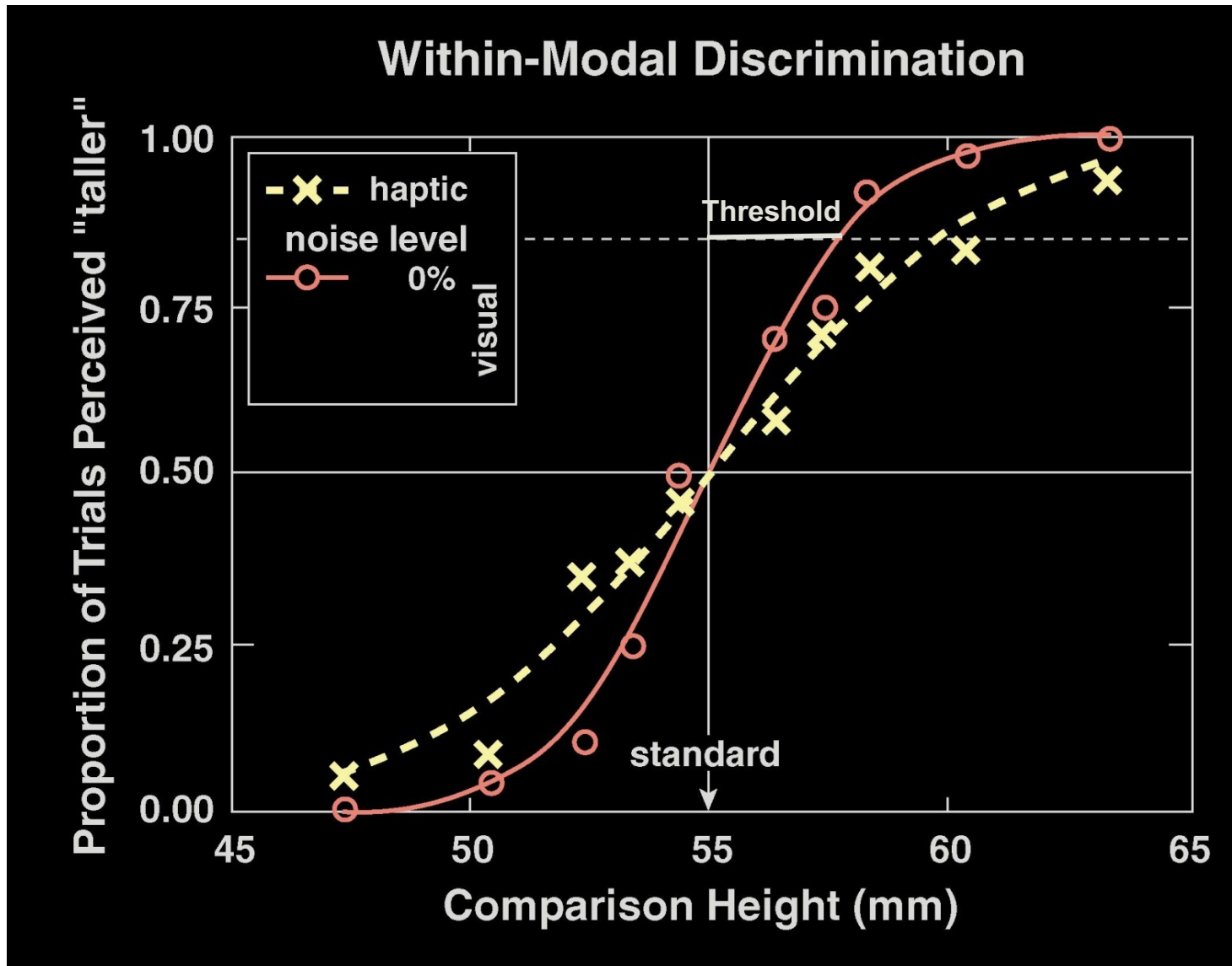
Determining Within-Modality Variance



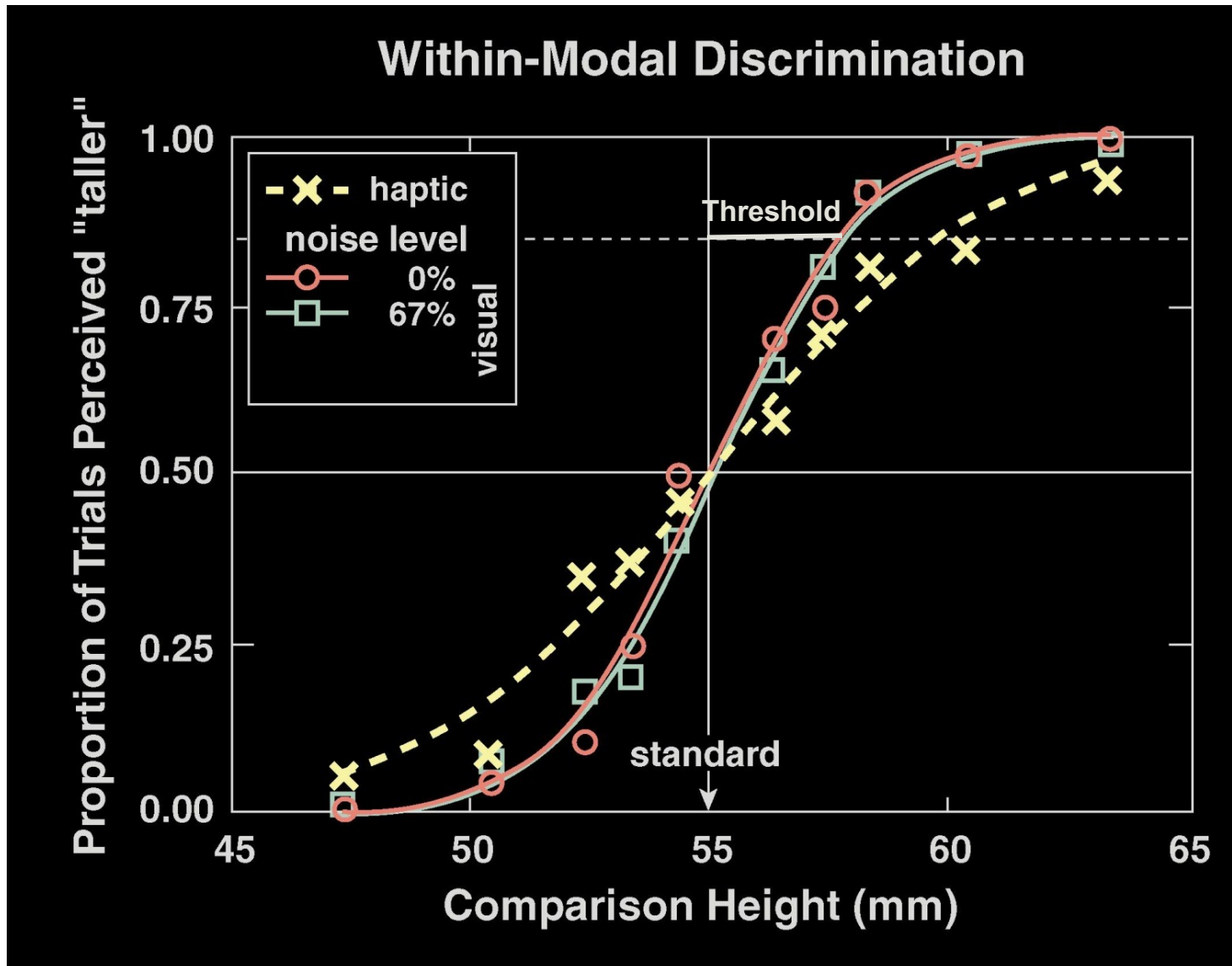
Determining Within-Modality Variance



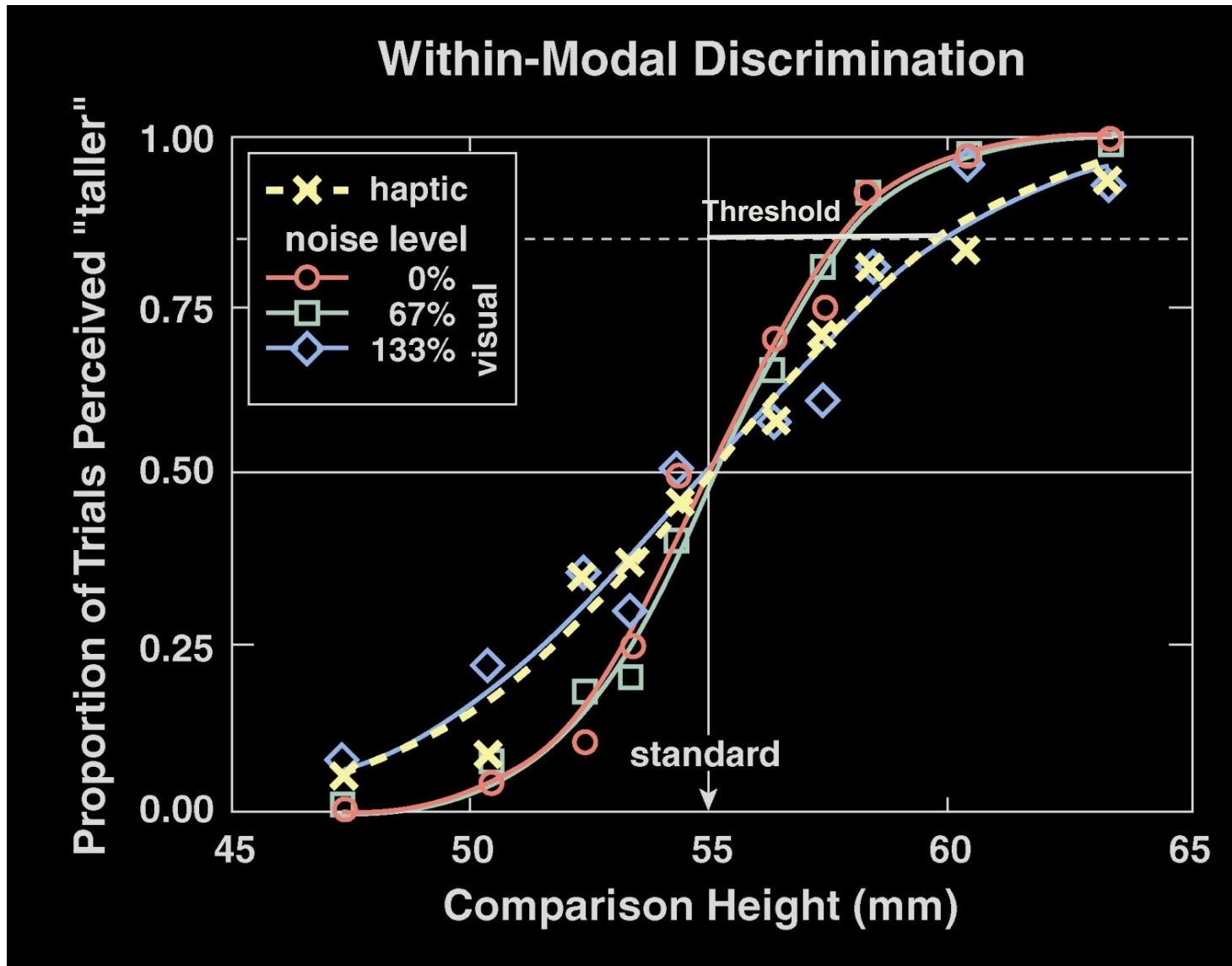
Determining Within-Modality Variance



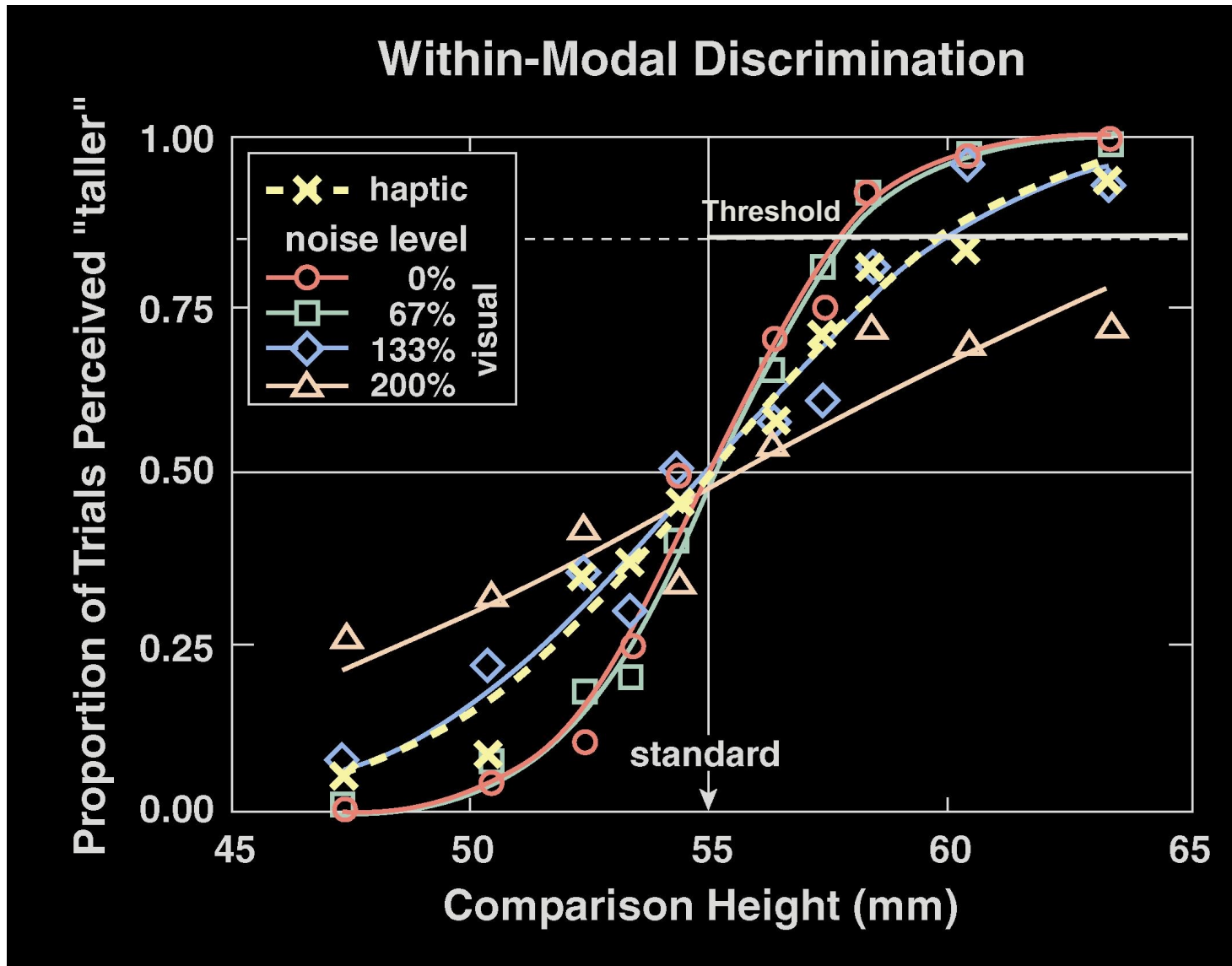
Determining Within-Modality Variance



Determining Within-Modality Variance



Determining Within-Modality Variance



From Variance to Threshold

Predicted weights for combined performance from within-modal data

estimators weights

$$w_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2}$$

$$JND_i = \sqrt{2} \cdot \sigma_i$$

estimators weights

$$w_V = \frac{JND_H^2}{JND_V^2 + JND_H^2}$$

Predicted combined threshold from within-modal data

visual-haptic variance

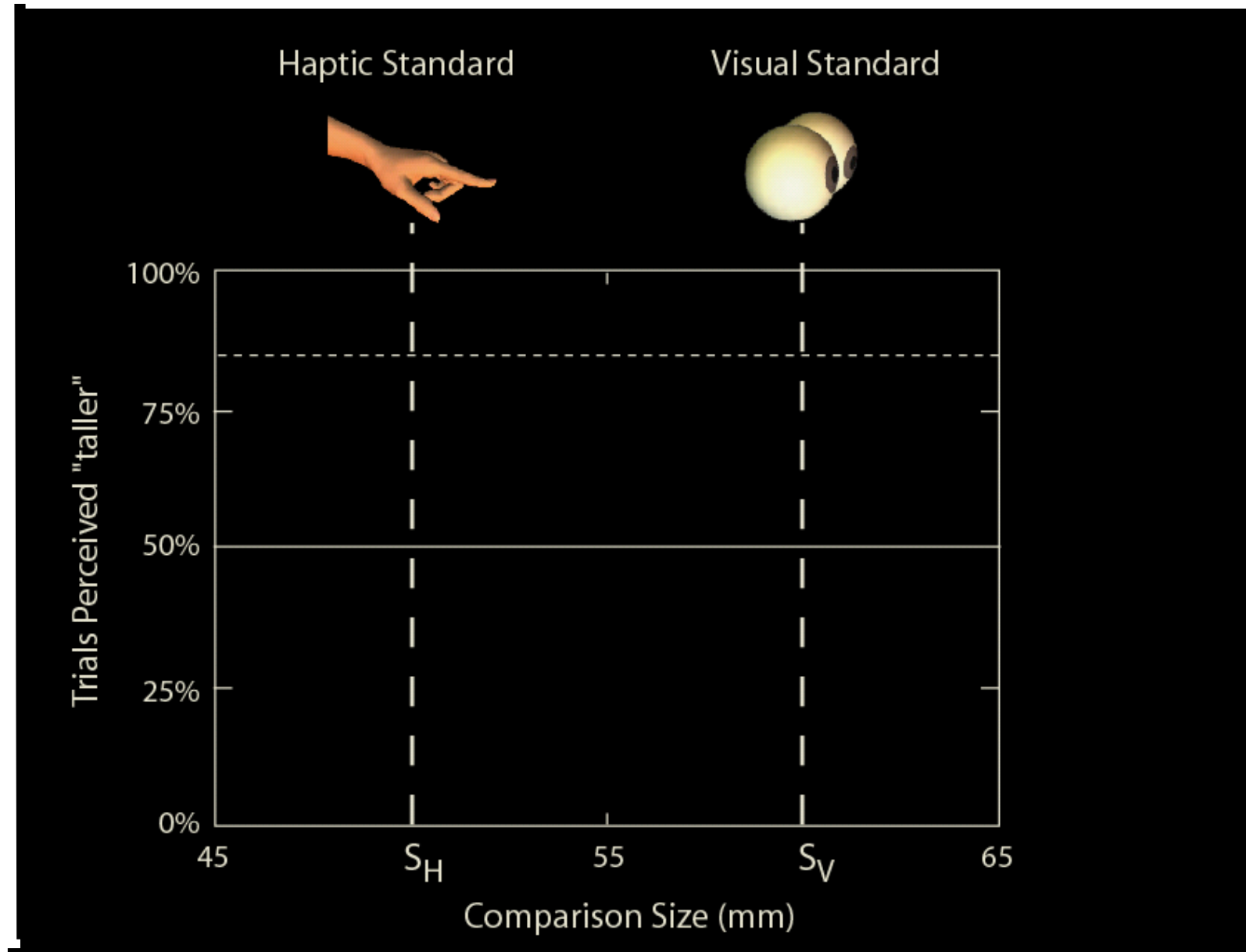
$$\frac{1}{\sigma_{VH}^2} = \frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2}$$

$$JND_i = \sqrt{2} \cdot \sigma_i$$

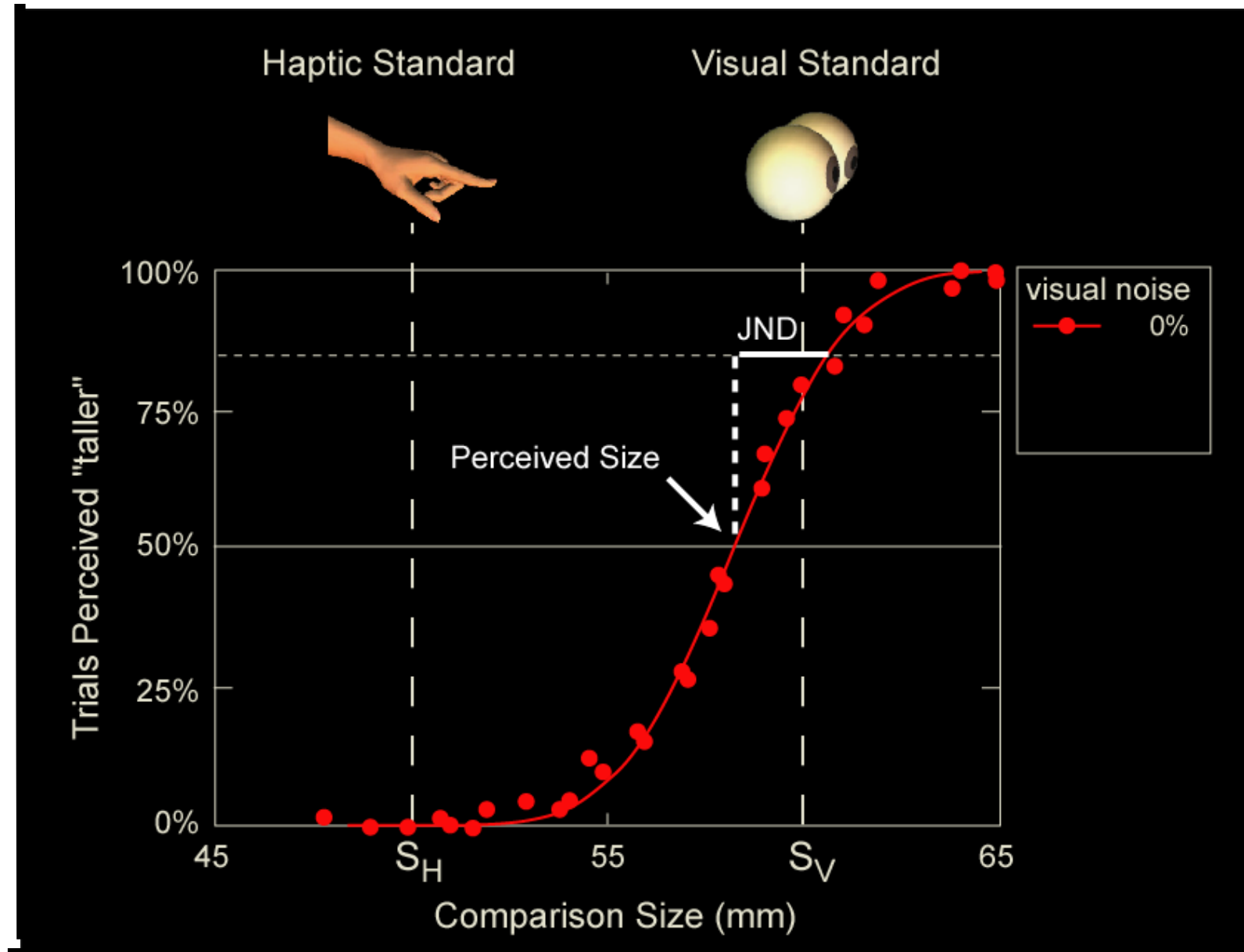
visual-haptic threshold

$$\frac{1}{JND_{VH}^2} = \frac{1}{JND_V^2} + \frac{1}{JND_H^2}$$

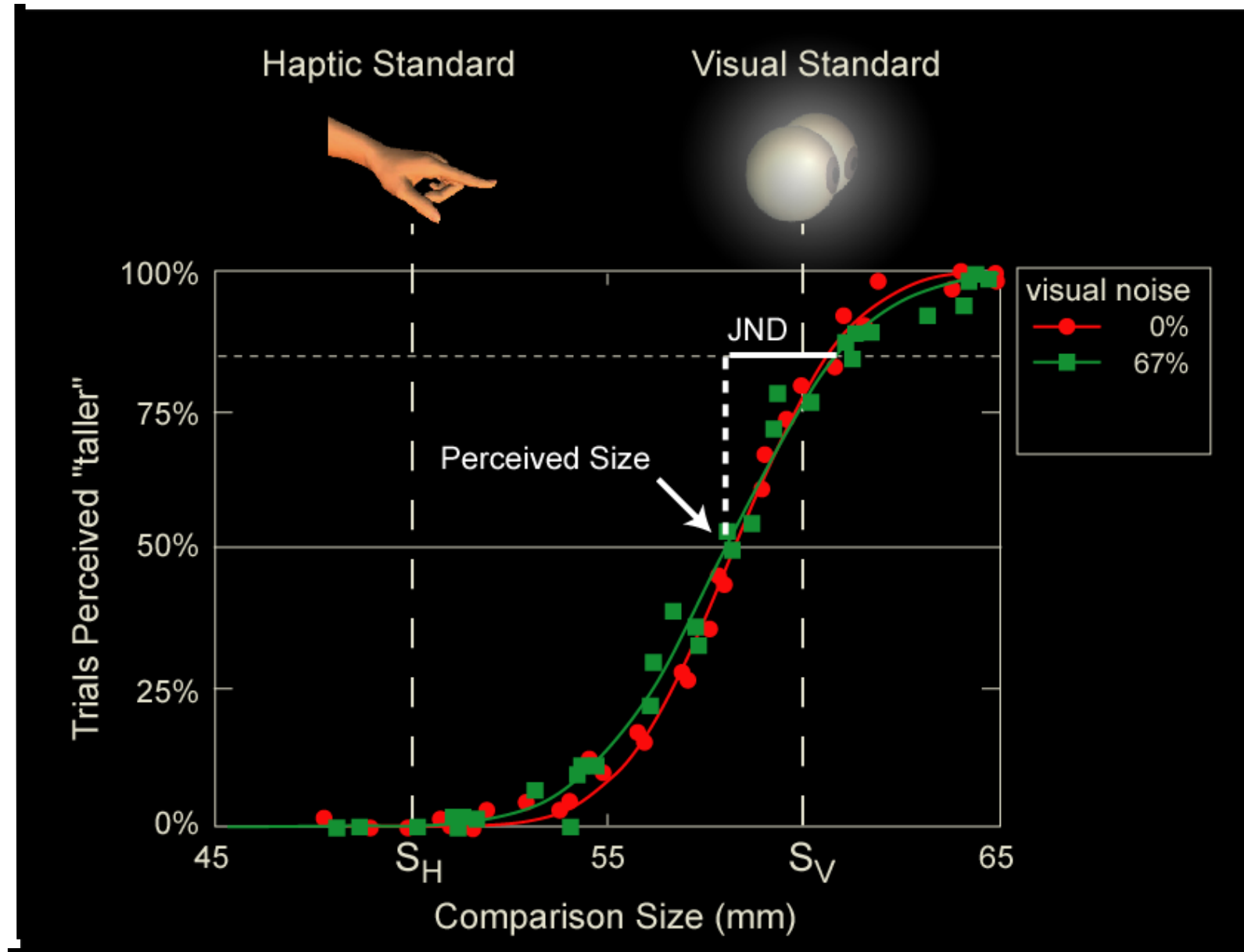
Visual-Haptic Discrimination



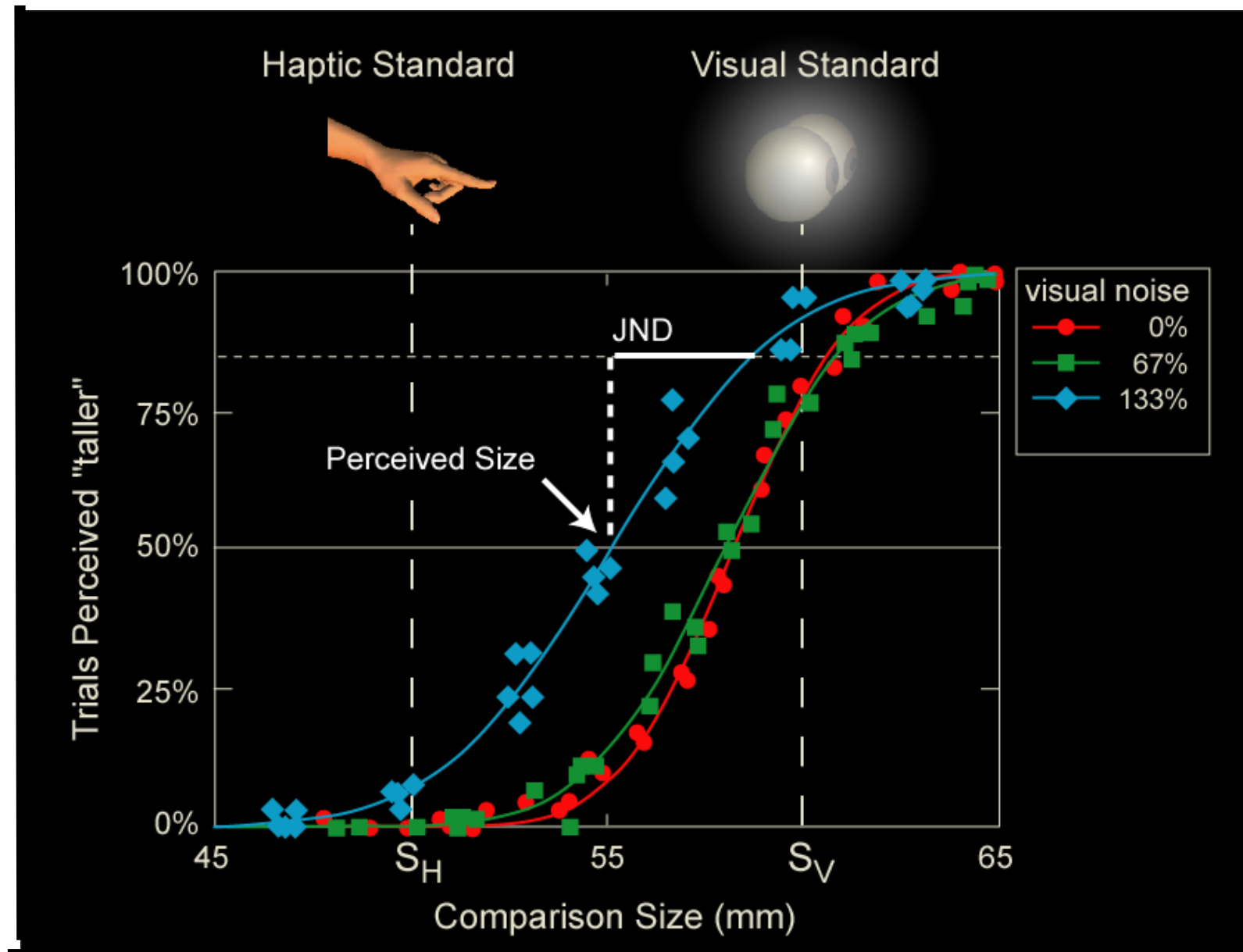
Visual-Haptic Discrimination



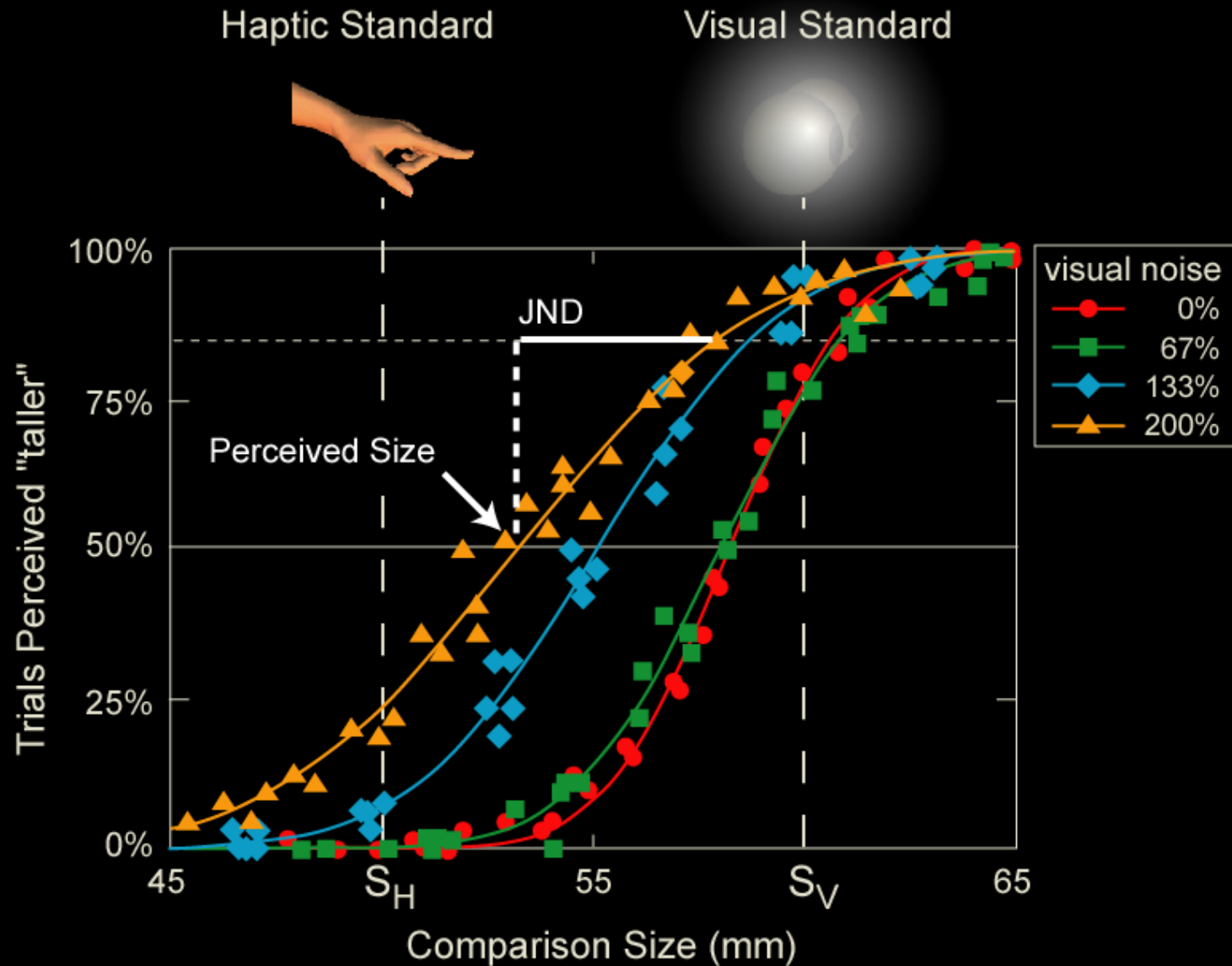
Visual-Haptic Discrimination



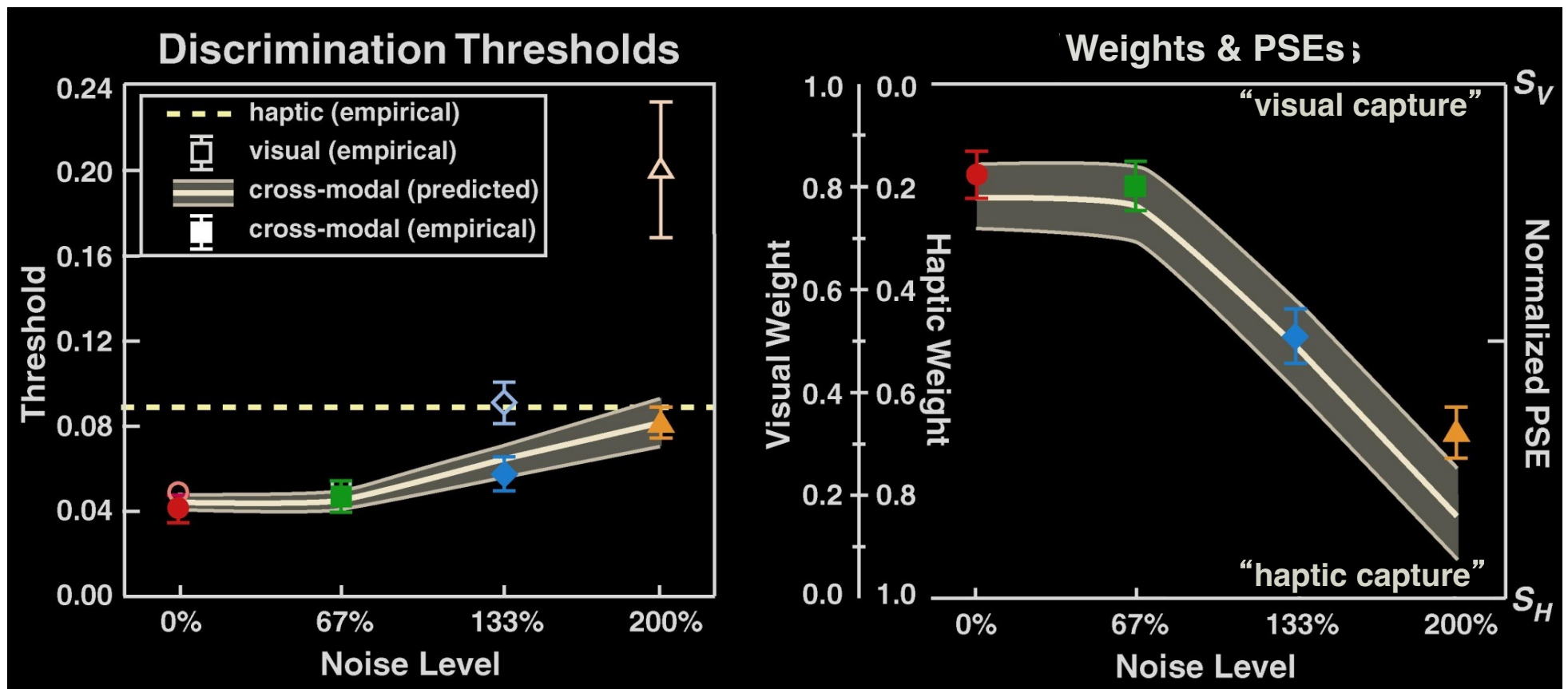
Visual-Haptic Discrimination



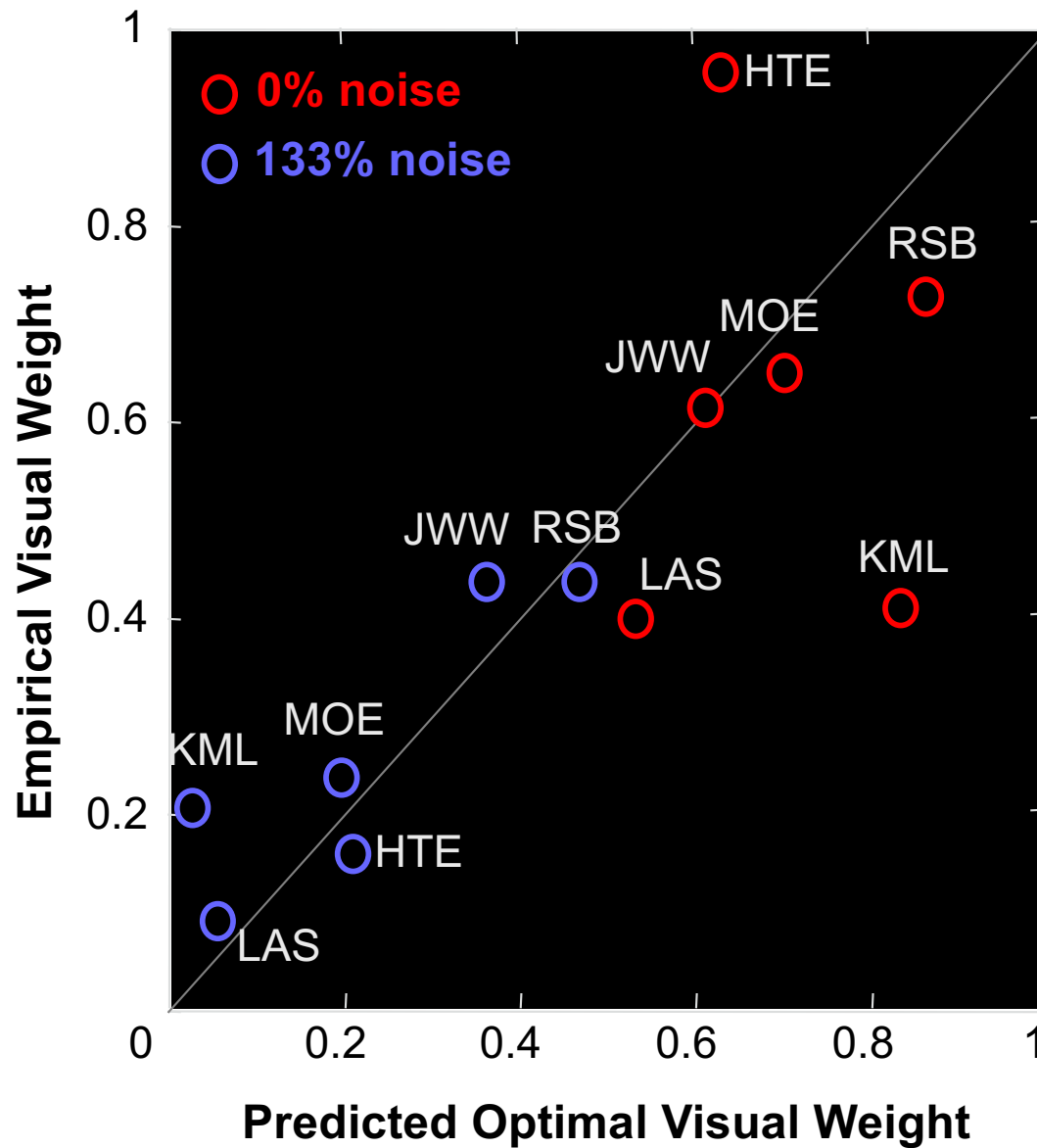
Visual-Haptic Discrimination



Empirical Thresholds and Weights



Individual Differences

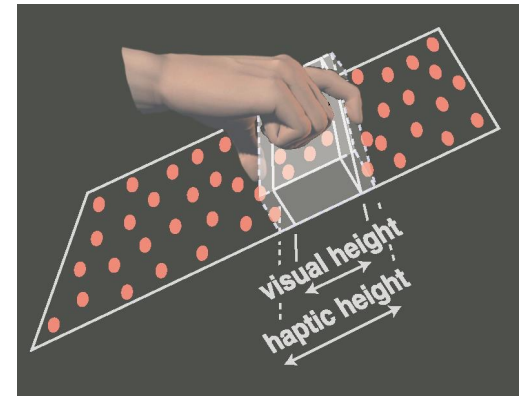


Conclusions

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 California, Berkeley
94720-2020, USA*



- **Combination reduces variance. below that of either cue.**
- **Linear weighting scheme for visual-haptic perception.**
- **Explains behavior like apparent “visual capture” or visual dominance.**

ASSUMPTIONS

- Cues are

Gaussian
Independent

What if they are not?

Oruç, I, Maloney, L. T., & Landy, M. S. (2003), Weighted linear cue combination with possibly correlated error, *Vision Research*, 43, 2451-2468.

- Goal: UMVUE

*What if we have other goals?
next lecture*