Integration of cues

- Quick review of depth cues
- Cue combination: Minimum variance
- Cue combination: Bayesian
- Nonlinear cue combination: Causal models
- Statistical decision theory

Distance, depth, and 3D shape cues

- Pictorial depth cues: familiar size, relative size, brightness, occlusion, shading and shadows, aerial/ atmospheric perspective, linear perspective, height within image, texture gradient, contour
- Other static, monocular cues: accommodation, blur, [astigmatic blur, chromatic aberration]
- Motion cues: motion parallax, kinetic depth effect, dynamic occlusion
- Binocular cues: convergence, stereopsis/binocular disparity
- Cue combination

Basic distinctions

- Types of depth cues
  - Monocular vs. binocular
  - Pictorial vs. movement
  - Physiological
- Depth cue information
  - What is the information?
  - How could one compute depth from it?
  - Do we compute depth from it?
  - What is learned: ordinal, relative, absolute depth, depth ambiguities

Definitions

- Distance: Egocentric distance, distance from the observer to the object
- Depth: Relative distance, e.g., distance one object is in front of another or in front of a background
- Surface Orientation: Slant (how much) and tilt (which way)
- Shape: Intrinsic to an object, independent of viewpoint

Distance, depth, and 3D shape cues

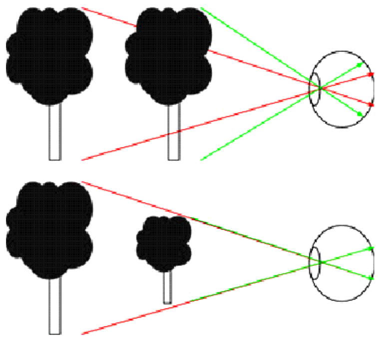
- Pictorial depth cues: familiar size, relative size, [brightness], occlusion, shading and shadows, aerial/ atmospheric perspective, linear perspective, height within image, texture gradient, contour
- Other static, monocular cues: accommodation, blur, [astigmatic blur, chromatic aberration]
- Motion cues: motion parallax, kinetic depth effect, dynamic occlusion
- Binocular cues: convergence, stereopsis/binocular disparity

Epstein (1965) familiar size experiment

How far away
is the coin?

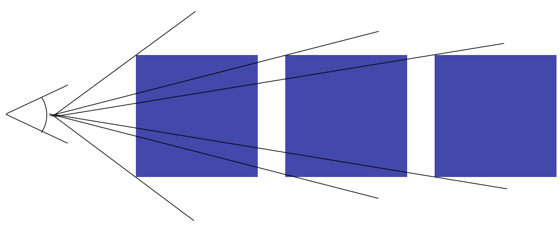
Monocular depth cues

Retinal projection depends on size and distance

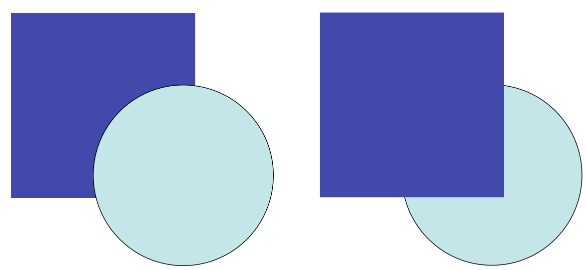


Relative size as a cue to depth

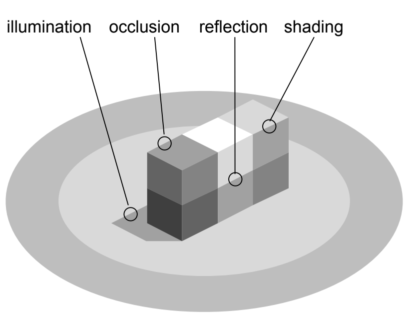
Relative size as a cue to depth



Occlusion as a cue to depth

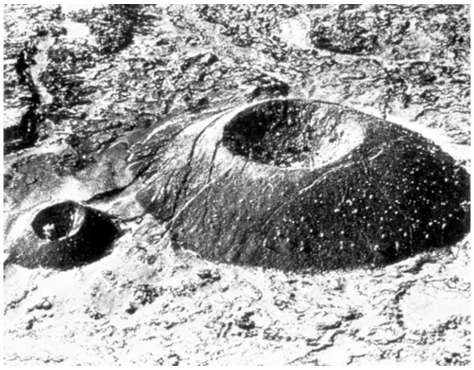


Shading, reflection, and illumination



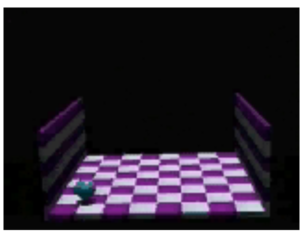
Shading – prior of light-from-above

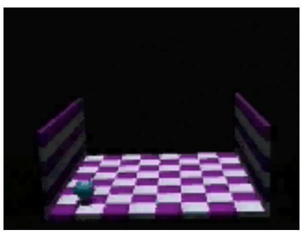
Shading (flip the photo upside-down)



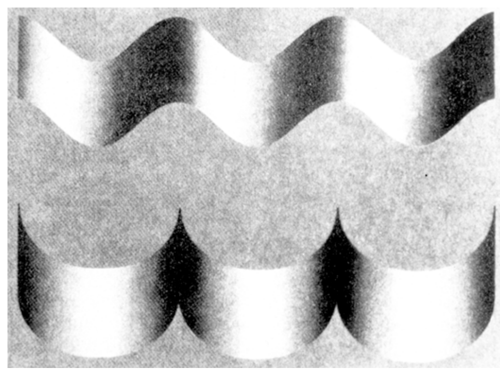
Cast Shadows

Dynamic Cast Shadows





Shading and contour

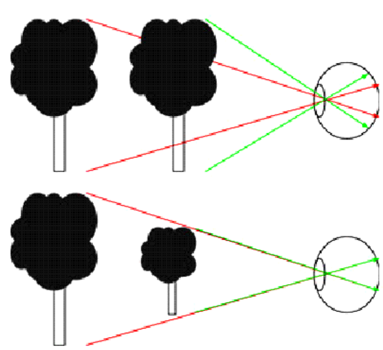


Aerial/Atmospheric Perspective

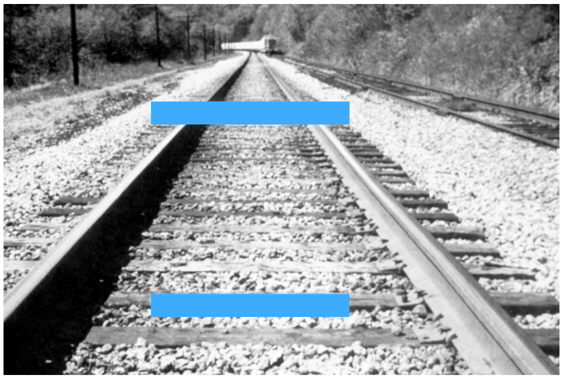
Geometry of Linear Perspective

Retinal projection depends on size and distance:

Size in the world (e.g., in meters) is proportional to size in the retinal image (in degrees) times the distance to the object



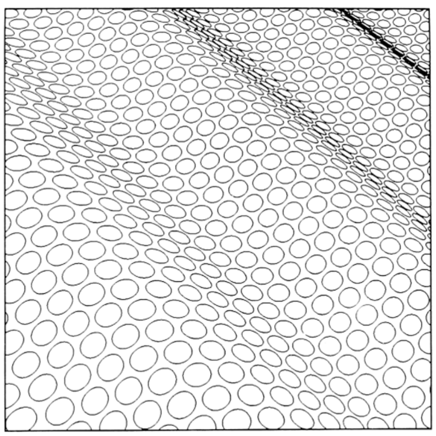
Linear perspective



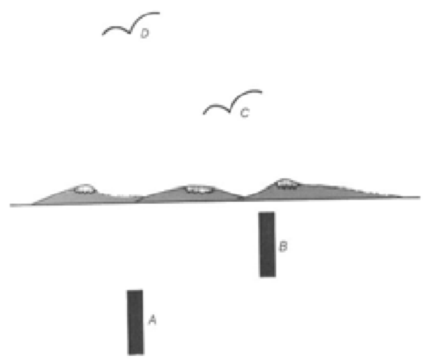
Size constancy

Texture

1. Density
2. Foreshortening
3. Size



Height Within the Image



Distance, depth, and 3D shape cues

- Pictorial depth cues: familiar size, relative size, brightness, occlusion, shading and shadows, aerial/atmospheric perspective, linear perspective, height within image, texture gradient, contour
- Other static, monocular cues: accommodation, blur, [astigmatic blur, chromatic aberration]
- Motion cues: motion parallax, kinetic depth effect, dynamic occlusion
- Binocular cues: convergence, stereopsis/binocular disparity

Monocular Physiological Cues

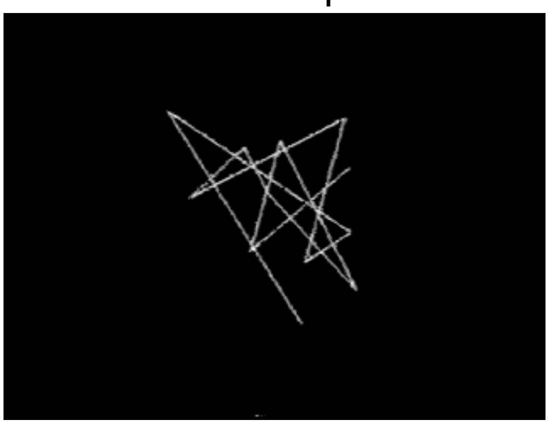
- Accommodation – estimate depth based on state of accommodation (lens shape) required to bring object into focus
- Blur – objects that are further or closer than the accommodative distance are increasingly blur
- Astigmatic blur
- Chromatic aberration

Distance, depth, and 3D shape cues

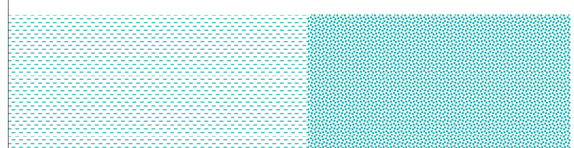
- Pictorial depth cues: familiar size, relative size, brightness, occlusion, shading and shadows, aerial/atmospheric perspective, linear perspective, height within image, texture gradient, contour
- Other static, monocular cues: accommodation, blur, [astigmatic blur, chromatic aberration]
- Motion cues: motion parallax, kinetic depth effect, dynamic occlusion
- Binocular cues: convergence, stereopsis/binocular disparity

Motion Parallax

The Kinetic Depth Effect



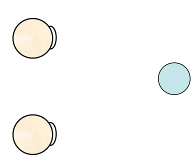
Dynamic (Kinetic) Occlusion



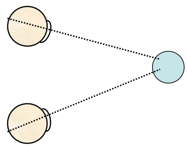
Distance, depth, and 3D shape cues

- Pictorial depth cues: familiar size, relative size, brightness, occlusion, shading and shadows, aerial/atmospheric perspective, linear perspective, height within image, texture gradient, contour
- Other static, monocular cues: accommodation, blur, [astigmatic blur, chromatic aberration]
- Motion cues: motion parallax, kinetic depth effect, dynamic occlusion
- Binocular cues: convergence, stereopsis/binocular disparity

Vergence Angle As One Binocular Source

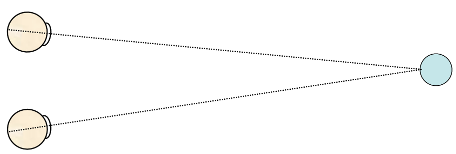


Vergence Angle As One Binocular Source

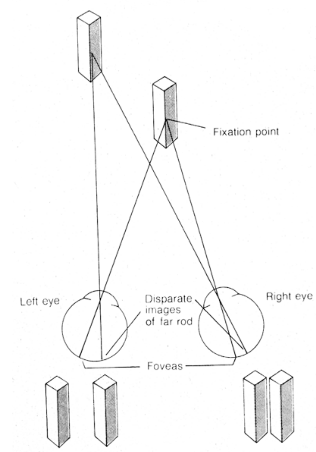


Vergence Angle As One Binocular Source

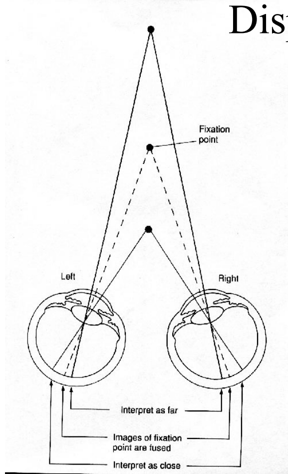
Vergence Angle As One Binocular Source



Binocular disparity



Disparity

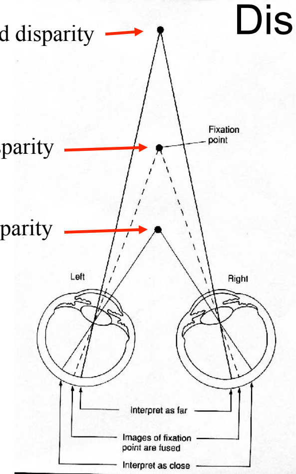


Disparity

Uncrossed disparity

Zero retinal disparity

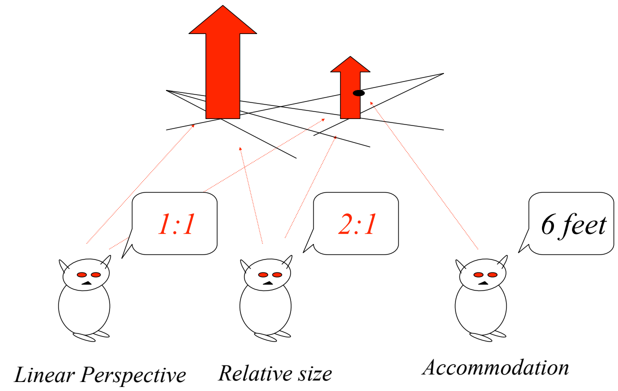
Crossed disparity



Depth Cue Combination: Issues

1. *How do you put all of the depth cue information together?*
2. *What do you do when cues disagree?*
   A little ... ?
   A lot ... ? *errors*
3. *How much weight should each cue get?*

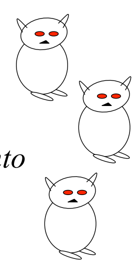
When cues disagree ...



Information Fusion Problem

Multiple sources of information, possibly in error, possibly contradictory

How combine the information into a single judgment?



Rashomon

Optimal Cue Combination: Minimum Variance

$E(X_i) = \mu_1, \quad E(X_2) = \mu_2$

Variances: $\sigma_2^2 \le \sigma_1^2$ Just use one cue?

Suppose we use a linear cue-combination rule:

$X \ = w_1 X_1 + w_2 X_2$ weighted linear combination

$E\left[X\right] = w_1 E\left[X_1\right] + w_2 E\left[X_2\right] = \left(w_1 + w_2\right)\mu$

unbiased?

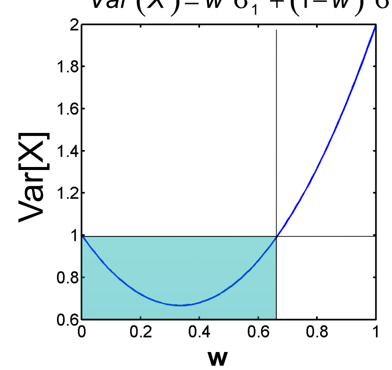
Minimum-Variance Cue Combination

$X \ = wX_1 + \left(1-w\right)X_2$ unbiased

$Var\left(X\right) = w^2 Var\left(X_1\right) + \left(1-w\right)^2 Var\left(X_2\right)$

$= w^2\sigma_1^2 + \left(1-w\right)^2 \sigma_2^2$ minimize

$Var\left(X\right) = w^2\sigma_1^2 + \left(1-w\right)^2 \sigma_2^2$



Minimum-Variance Cue Combination

$$X = wX_1 + (1-w)X_2$$

$$Var(X) = w^2 Var(X_1) + (1-w)^2 Var(X_2)$$

Choose w to minimize variance:

$$w = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2}$$

Reparameterization

Define reliability $r_i = \sigma_i^{-2}$

$$X = w_1 X_1 + w_2 X_2$$

weight proportional to reliability

$$w = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2} = \frac{r_1}{r_1 + r_2}$$

$$r = r_1 + r_2$$

reliabilities add

Cue Combination for Estimation

- Weighted average:

$$D(x,y) = \alpha_s D_s(x,y) + \alpha_m D_m(x,y) + \alpha_t D_t(x,y) + \cdots$$

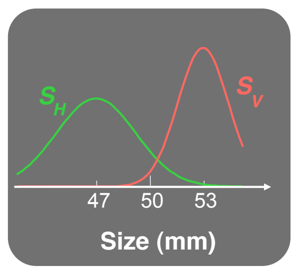
where

$$\sum_i \alpha_i = 1$$

- Optimal weights for independent cues:

$$\alpha_i = \frac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2} = \frac{r_i}{\sum_j r_j}$$

Combining Sensory Estimates

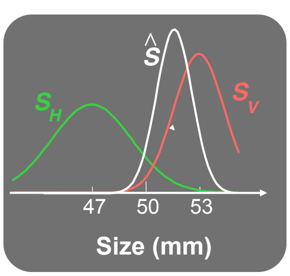


$$\hat{S} = w_H \hat{S}_H + w_V \hat{S}_V$$

$$w_H = \frac{r_H}{r_H + r_V}$$

$$w_V = \frac{r_V}{r_H + r_V}$$

Combining Sensory Estimates

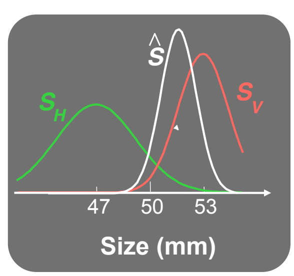


$$\hat{S} = w_H \hat{S}_H + w_V \hat{S}_V$$

$$w_H = \frac{r_H}{r_H + r_V}$$

$$w_V = \frac{r_V}{r_H + r_V}$$

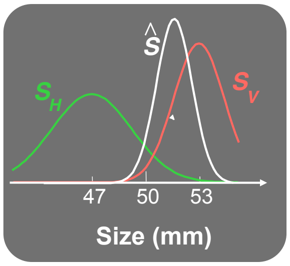
Combining Sensory Estimates



$$\hat{S} = w_H \hat{S}_H + w_V \hat{S}_V$$

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

Combining Sensory Estimates



$$\hat{S} = w_H \hat{S}_H + w_V \hat{S}_V$$

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

$$r_{HV} = r_H + r_V$$

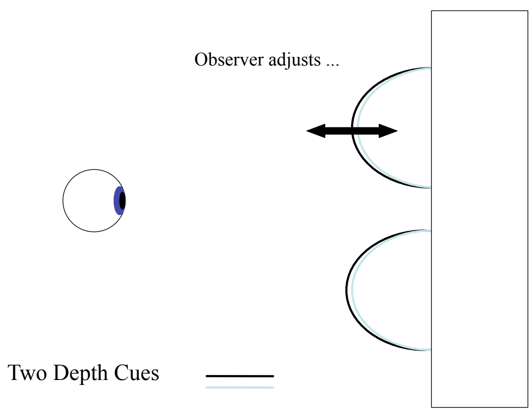
Variance of combined estimate lower than variance of either single-cue estimate

Perturbation Methodology and Influence Measures

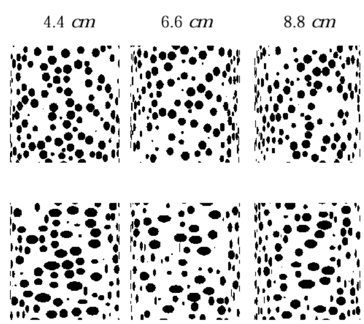
How can we measure the influence of various cues on perceptual judgments in complex scenes?

Goal: Change the stimulus as little as we possibly can.

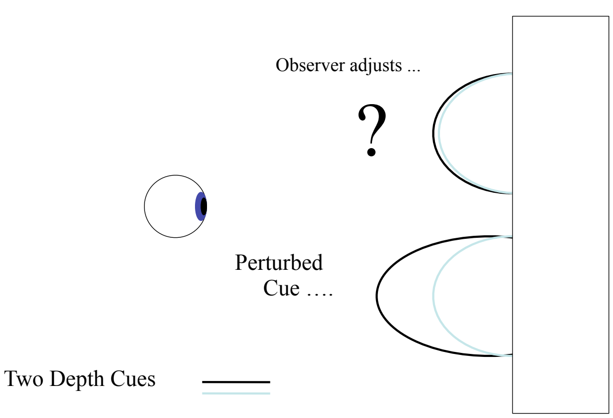
Perturbation Method



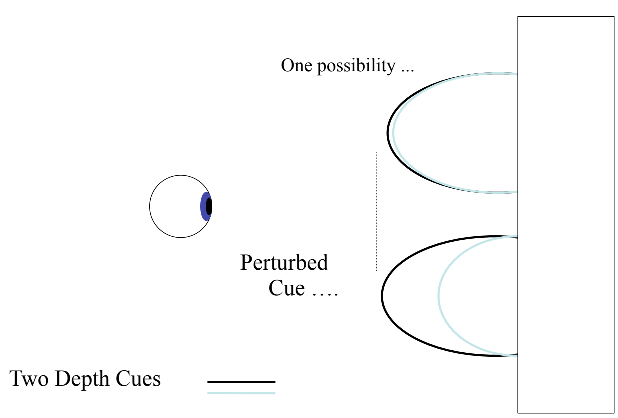
Example: Texture and Motion



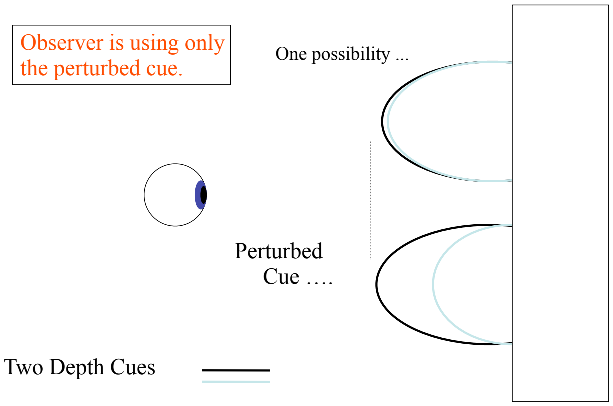
Perturbation Method



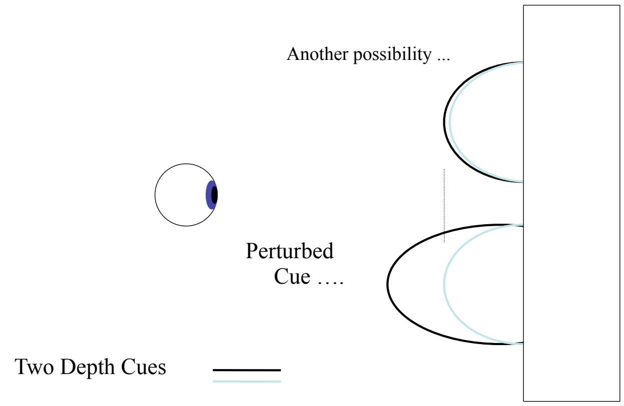
Perturbation Method



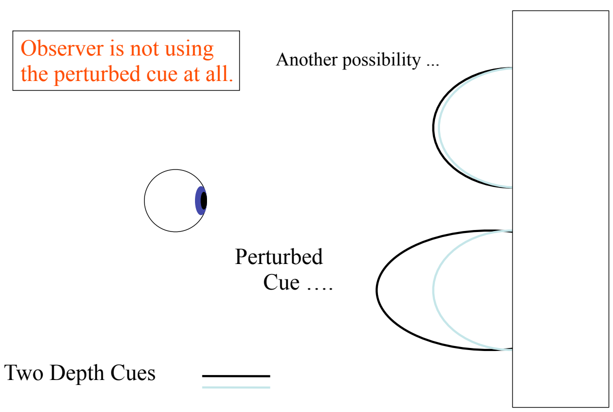
Perturbation Method



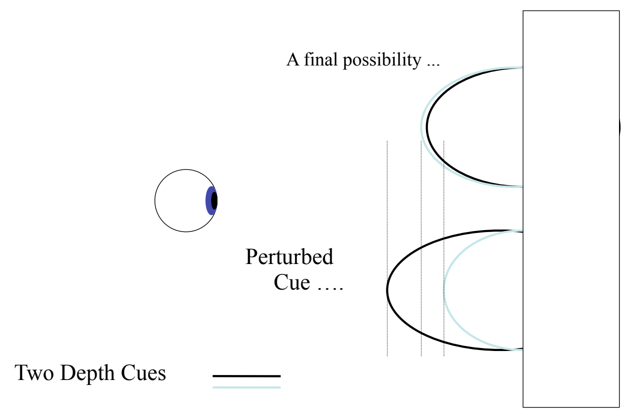
Perturbation Method



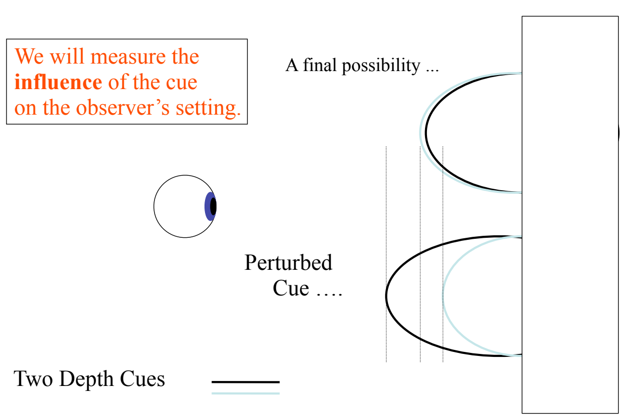
Perturbation Method



Perturbation Method



Perturbation Method



An Experimental Paradigm: Perturbation Analysis

The observer's cue weights can be estimated.

The stimulus comparison:

$$Cue_1 = d \quad Cue_2 = d$$

Matches

$$Cue_1 = d_1 \quad Cue_2 = d_2 = d_1 + \Delta d$$

Therefore

$$\alpha_1 = \frac{d - d_1}{d_2 - d_1} = \frac{\Delta \text{depth}}{\Delta \text{cue}}$$

Influence Measures

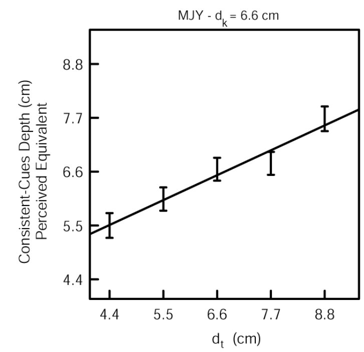
$$I_{cue} = \frac{\Delta_{setting}}{\Delta_{cue}}$$

Change in observer's setting

***Influence** of the cue*

Perturbation of the cue

Texture and Motion: Data



Optimal Cue Combination: Bayesian

Compute posterior:

$$p(depth \mid x_1, x_2) = \frac{p(x_1, x_2 \mid depth)\, p(depth)}{p(x_1, x_2)}$$

Assume conditional independence:

$$p(depth \mid x_1, x_2) \propto p(x_1 \mid depth)\, p(x_2 \mid depth)\, p(depth)$$

If likelihoods and prior are Gaussian, so is posterior, and means and reliabilities are as in minimum-variance case. Prior acts like a static cue.

Optimal Cue Combination: Bayesian

$$p(depth \mid x_1, x_2) \propto p(x_1 \mid depth)\, p(x_2 \mid depth)\, p(depth)$$

Depending on cost function and priors, choose:

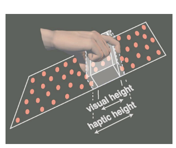
- ML: Maximum-likelihood estimator
- MAP: Maximum a posteriori estimator
- Mean of the posterior
- Median of the posterior
- Etc.

Optimal Cue Combination

Humans integrate visual and haptic information in a statistically optimal fashion

Marc O. Ernst* & Martin S. Banks

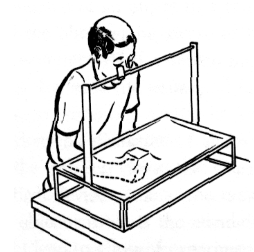
Vision Science Program/School of Optometry, University of California, Berkeley 94720-2020, USA



Rock & Victor (1964)


Irv Rock

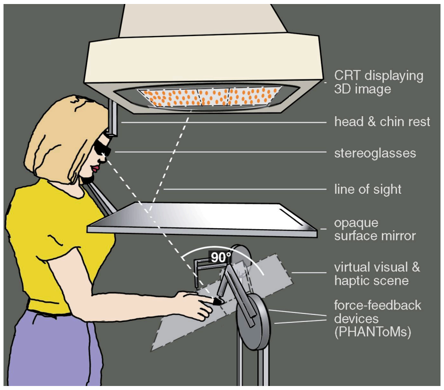
View object through distorting lens while exploring object haptically



Visual capture

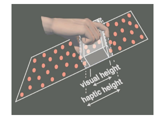
Visually and haptically specified shapes differed. What shape is perceived?

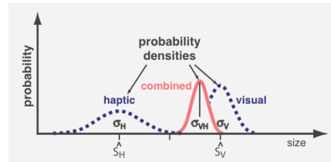
Visual/Haptic Setup



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

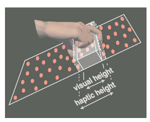
$$w_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2}$$

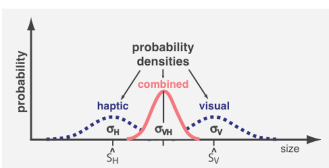
Variance

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

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

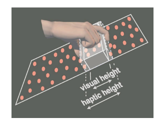
$$w_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2}$$

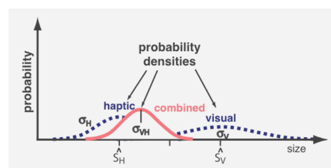
Variance

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

Visual Capture ?

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





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

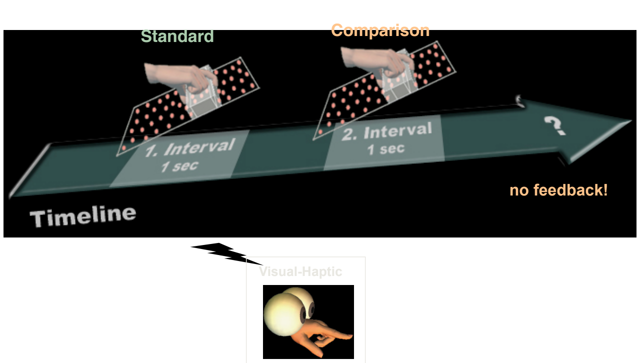
Weights

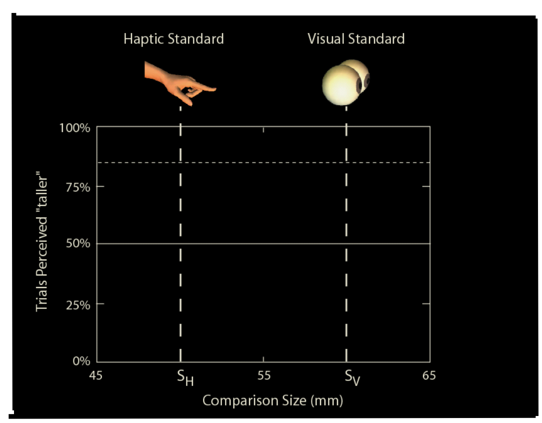
$$w_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2}$$

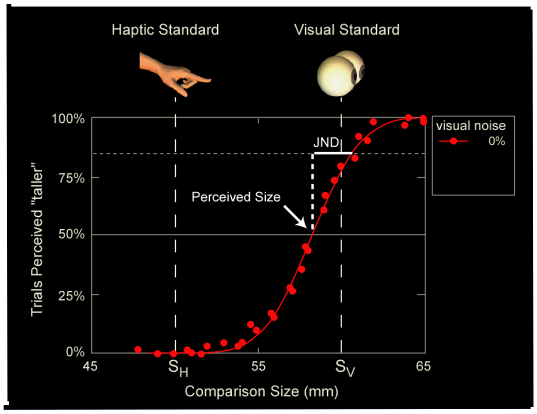
Variance

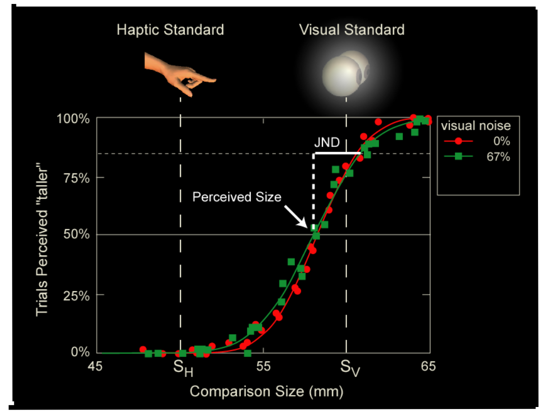
$$\frac{1}{\sigma_{VH}^2} = \frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2}$$

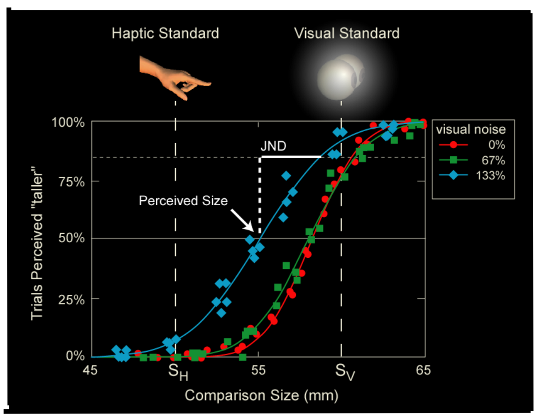
2-IFC Task

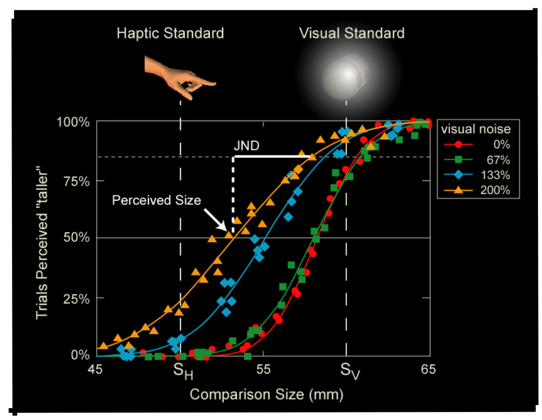




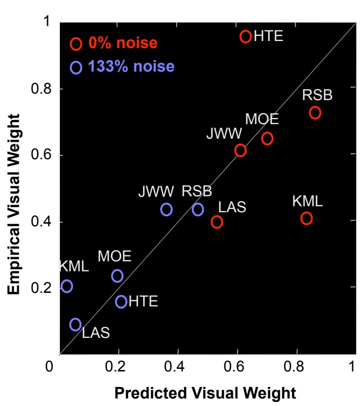




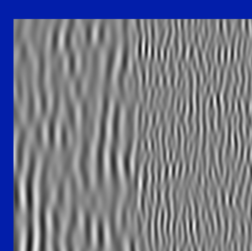


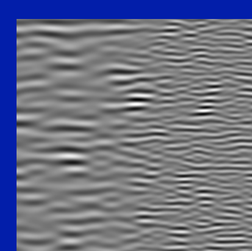


Individual Differences



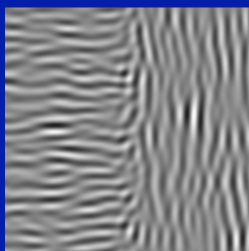
Cue 1: Spatial Frequency

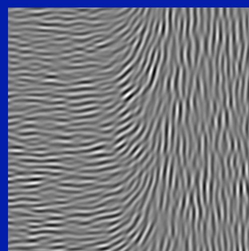




Landy & Kojima (2001)

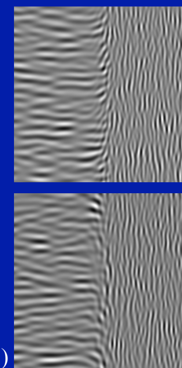
Cue 2: Orientation





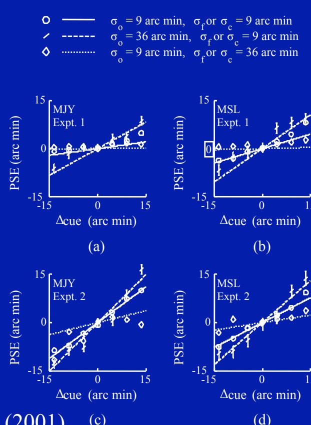
Landy & Kojima (2001)

Task: Vernier



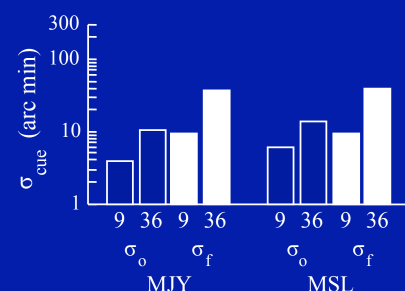
Landy & Kojima (2001)

Texture: Results



Landy & Kojima (2001)

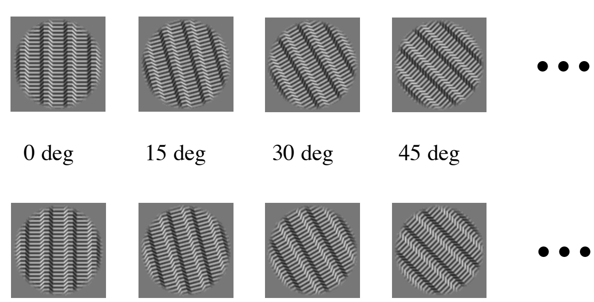
Texture: Fitted Reliabilities

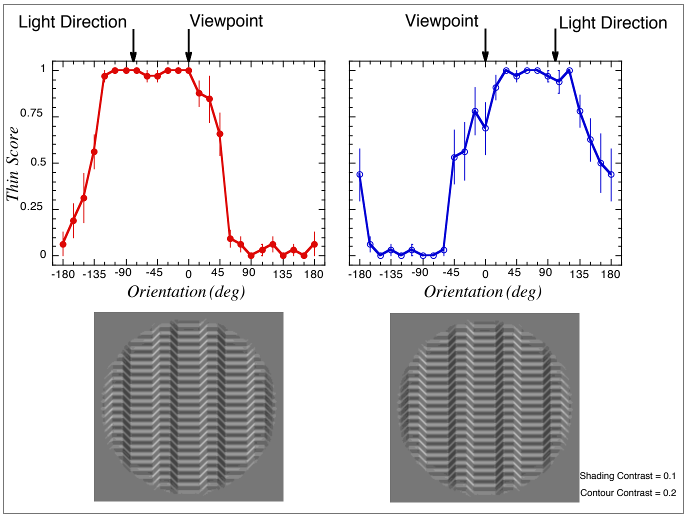


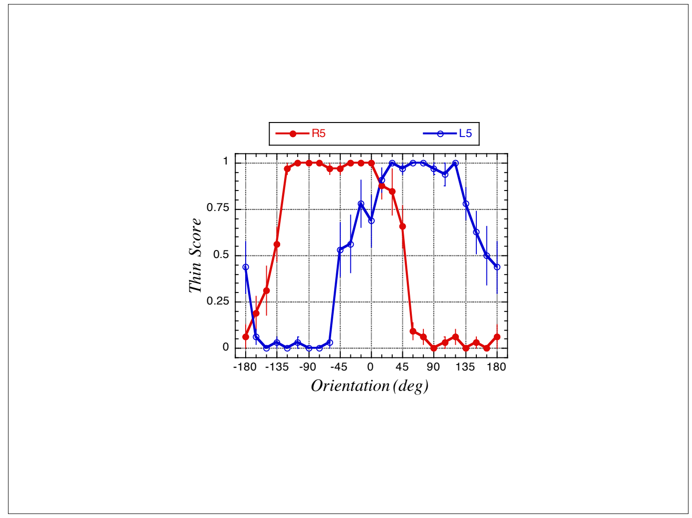
Landy & Kojima (2001)

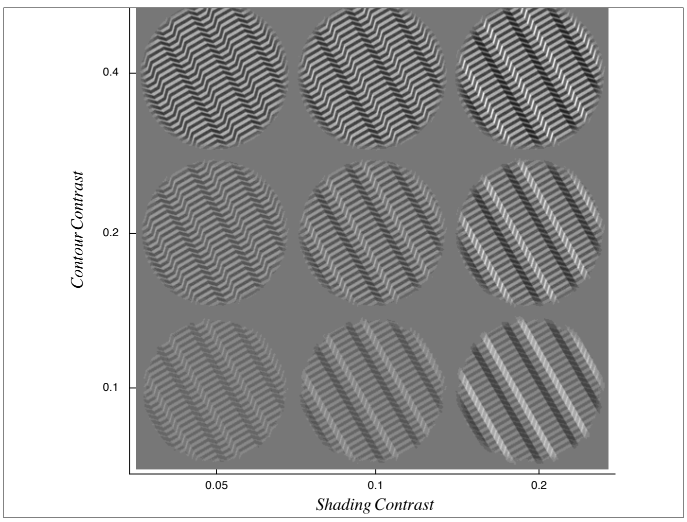
Demo: Landy/Kojima psychophysical task

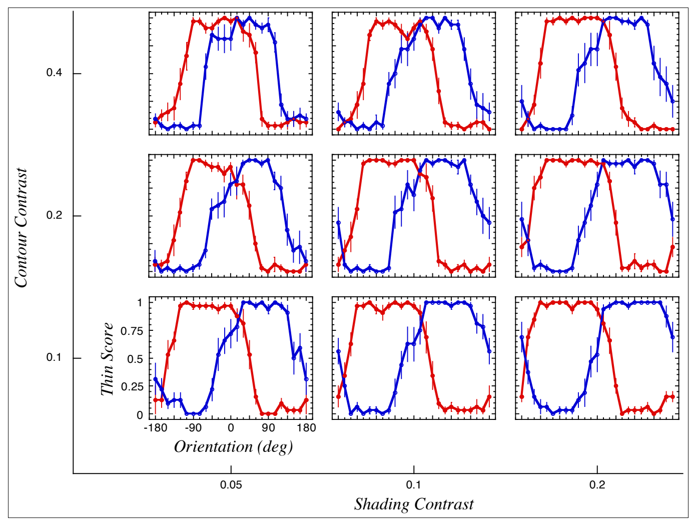
Influence of priors: Mamassian & Landy (1998, 2001)

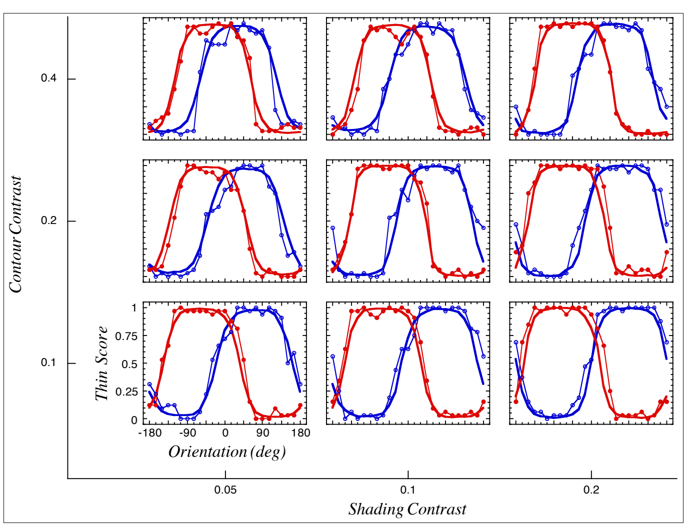


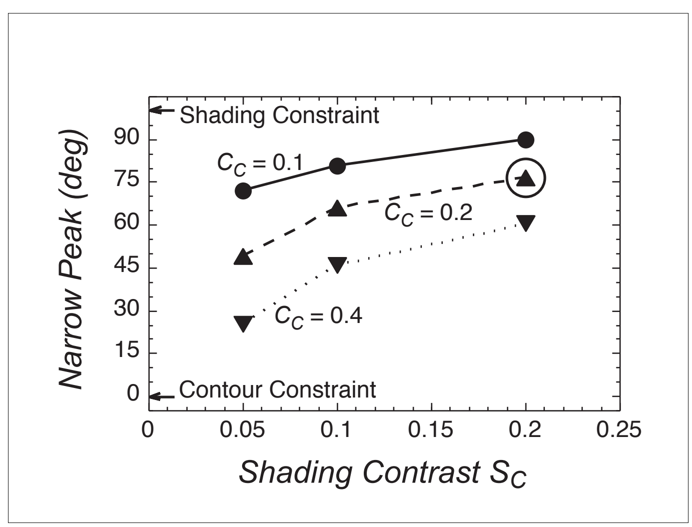

Light Direction
Viewpoint
Viewpoint
Light Direction
Thin Score
Orientation (deg)
Orientation (deg)
Shading Contrast = 0.1
Contour Contrast = 0.2


R5
L5
Thin Score
Orientation (deg)


Contour Contrast
Shading Contrast


Contour Contrast
Thin Score
Orientation (deg)
Shading Contrast


Contour Contrast
Thin Score
Orientation (deg)
Shading Contrast


Shading Constraint
$C_C = 0.1$
$C_C = 0.2$
$C_C = 0.4$
Contour Constraint
Narrow Peak (deg)
Shading Contrast S_C

Cost functions

We've touched on two of the three elements of Bayesian estimation and Bayesian decision-making: the likelihood and the prior. But, what about the third element: the cost function?

Typical Task for Decision-Making Under Risk

Would you rather have

A. $480, or

B. A 50-50 chance for $1,000?

A choice between "lotteries", where a lottery is a list of potential outcomes and their respective probabilities of occurence, e.g.,

(0.5, $0; 0.5, $1,000)

Typical Task for Decision-Making Under Risk

Would you rather have

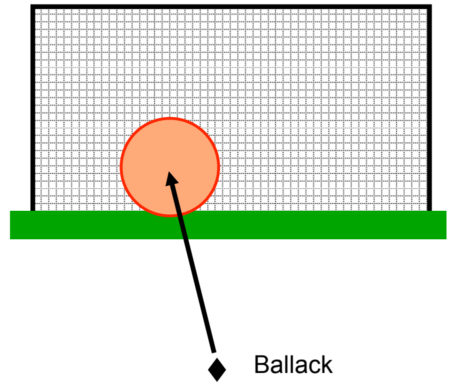
A. $480, or

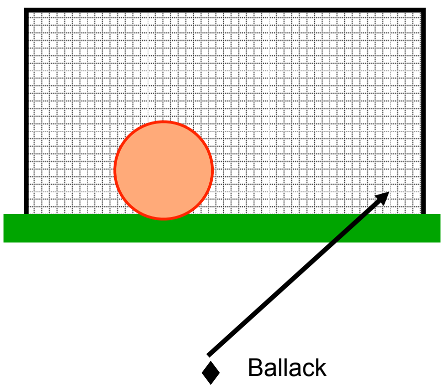
B. A 50-50 chance for $1,000?

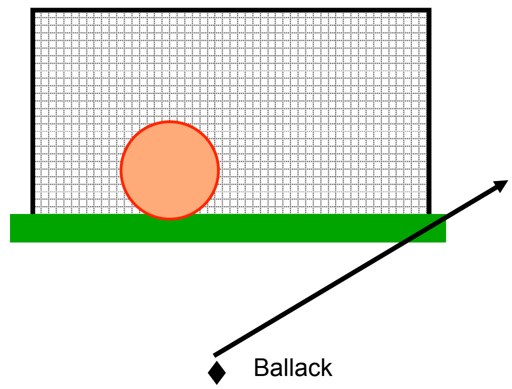
Typically, people choose A, showing risk-aversion for gains, and also show risk-seeking behavior for losses, along with many other "sub-optimal" behaviors, i.e., they don't simply maximize expected gain.

An Implicit (Motor) Decision Task

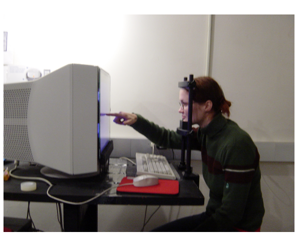
World Cup 2002, semifinal: South Korea vs. Germany

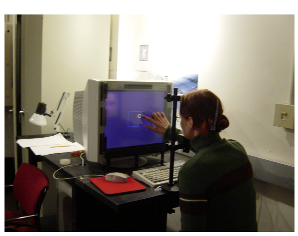




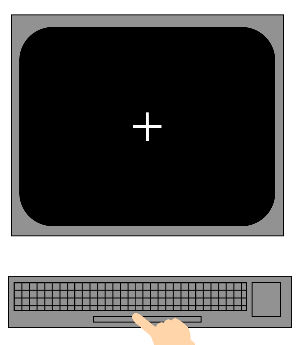


Experimental Task



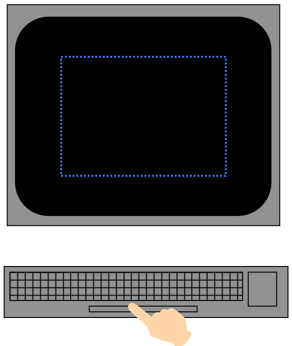


Experimental Task



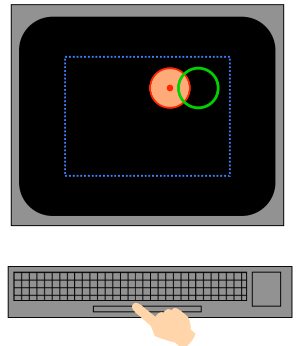
Start of trial:
display of fixation
cross (1.5 s)

Experimental Task



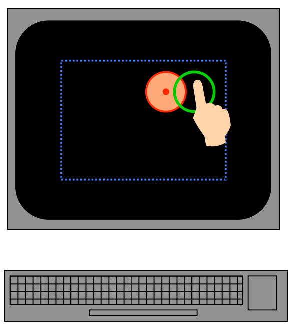
Display of response area,
500 ms before
target onset
(114.2 mm x 80.6 mm)

Experimental Task

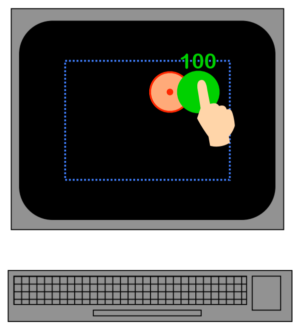


Target display (700 ms)

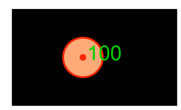
Experimental Task



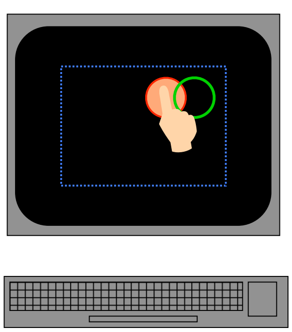
Experimental Task



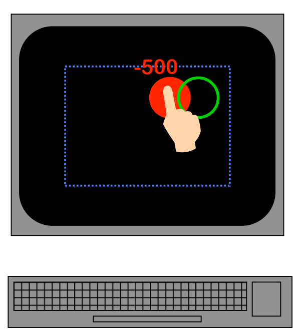
The green target is hit:
+100 points



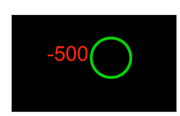
Experimental Task



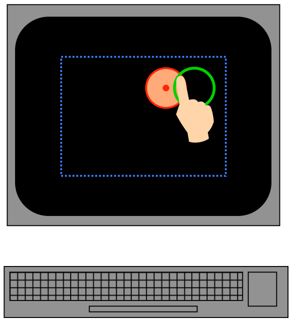
Experimental Task



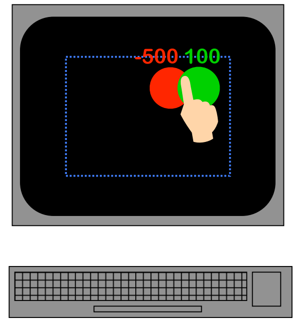
The red target is hit:
-500 points



Experimental Task

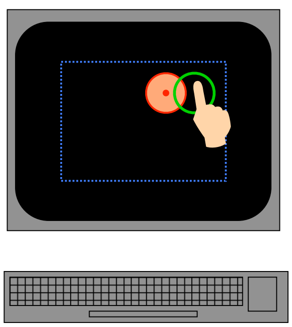


Experimental Task

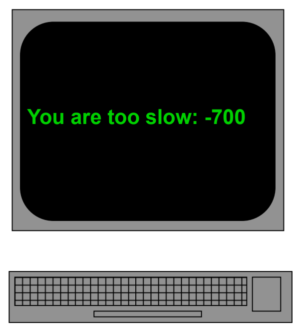


Scores add if both targets are hit:

Experimental Task

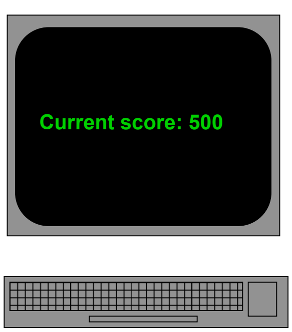


Experimental Task



The screen is hit later than 700 ms after target display: -700 points.

Experimental Task

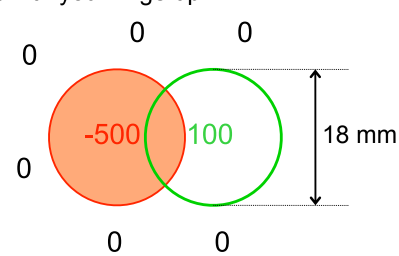


End of trial

Experimental Task

Rapidly touch a point with your fingertip.

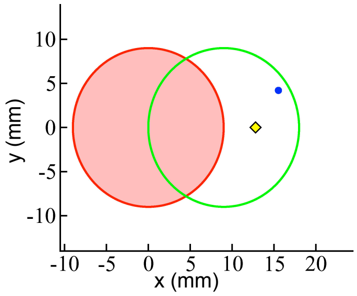
Responding after the time limit:
-700 points



What should you do?

Thought Experiment

: -500 : 100 points (2.5 ¢)

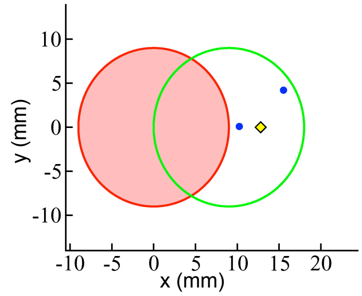


100 points

σ = 4.83 mm

Thought Experiment

: -500 : 100 points (2.5 ¢)



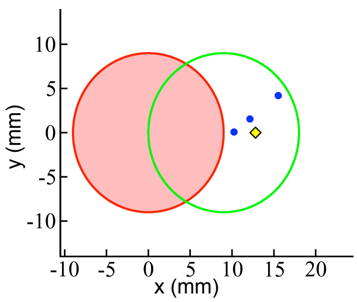
100 points
100 points

200 points

σ = 4.83 mm

Thought Experiment

: -500 : 100 points (2.5 ¢)



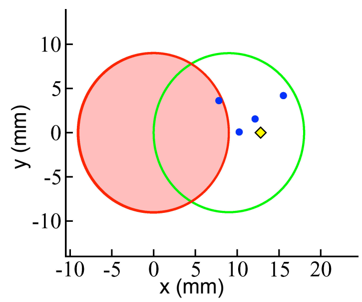
100 points
100 points
100 points

300 points

σ = 4.83 mm

Thought Experiment

: -500 : 100 points (2.5 ¢)



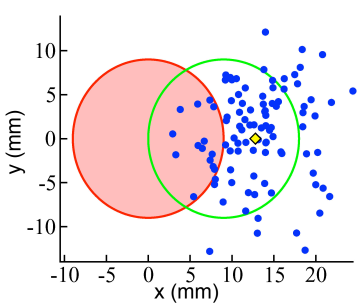
100 points
100 points
100 points
-400 points

-100 points

σ = 4.83 mm

Thought Experiment

: -500 : 100 points (2.5 ¢)



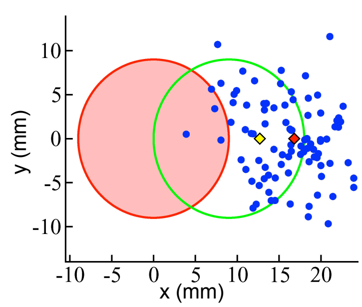
100 points
100 points
100 points
-400 points
⋮

-0.3 pts. per trial

σ = 4.83 mm

Thought Experiment

: -500 : 100 points (2.5 ¢)

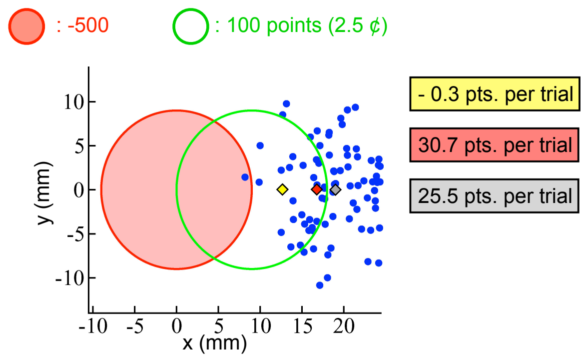


- 0.3 pts. per trial

30.7 pts. per trial

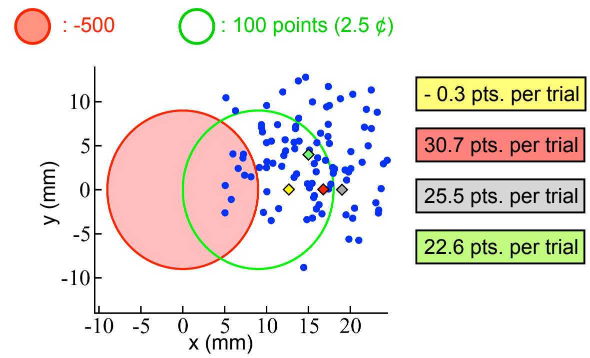
σ = 4.83 mm

Thought Experiment



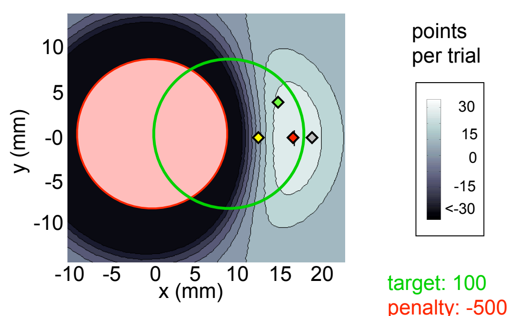
σ = 4.83 mm

Thought Experiment



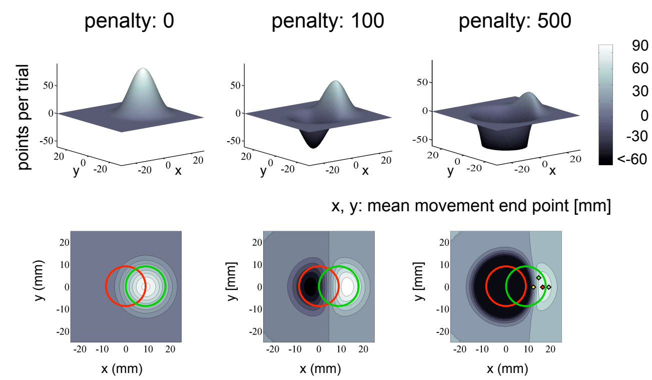
σ = 4.83 mm

Expected gain as function of mean movement end point (x,y):



σ = 4.83 mm

Thought Experiment

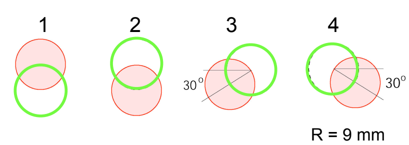


σ = 4.83 mm

Experiment: Movement Under Risk

Movement endpoints in response to novel stimulus configurations.

4 stimulus configurations:
(varied within block)



2 penalty conditions:
0 and -500 points (varied between blocks)

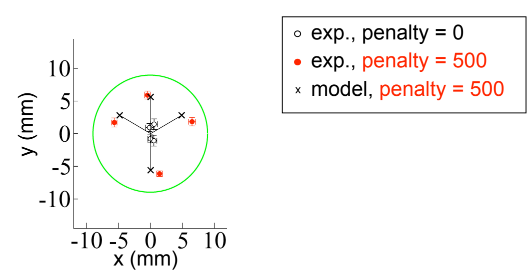
practice session: 300 trials, decreasing time limit

1 session of data collection: 360 trials
24 data points per condition

Trommershäuser, Maloney & Landy (2003). *JOSA A*, *20*,1419.

Results

Comparison with experiment



Subject S5, σ = 2.99 mm

Trommershäuser, Maloney & Landy (2003). *JOSA A*, *20*,1419.

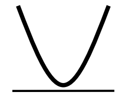
Summary: Movement Under Risk

Subjects' movement endpoints match those that, for their motor variability and the experimenter-imposed task conditions and risk, *do* maximize expected gain.

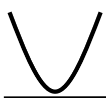
Subjects appear to do this, even when confronting novel configurations, from the first trial, with no apparent learning.

Subjects effectively take into account their own motor variability in planning movements.

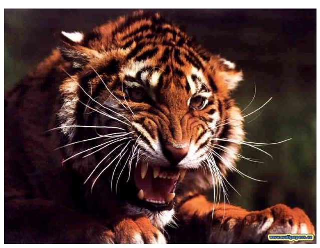
Trommershäuser, Maloney & Landy (2003). *JOSA A*, *20*,1419.



loss function?

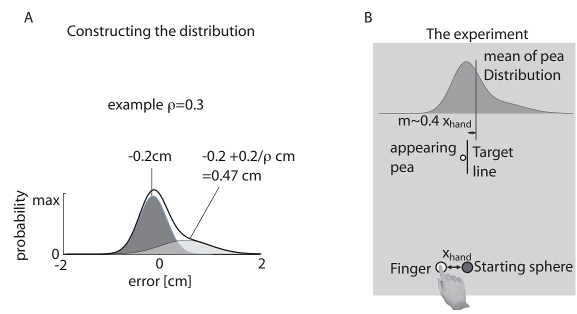


loss function?

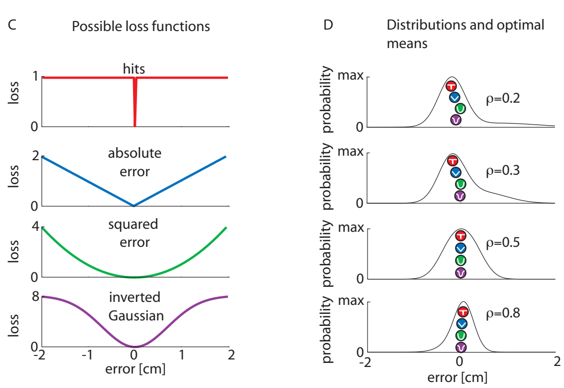


loss function!

Estimating the human cost function: Körding & Wolpert (2004)



Estimating the human cost function: Körding & Wolpert (2004)



Estimating the human cost function: Körding & Wolpert (2004)