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

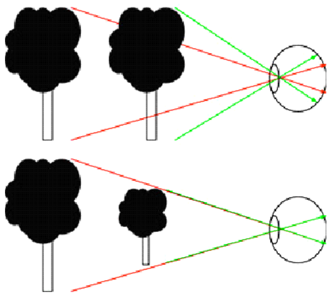
- Pictorial depth cues: familiar size, relative size, [brightness], occlusion, shading and shadows, aerial/atmospheric perspective, linear perspective, height within image, texture gradient, contour
- Other static, monocular cues: accommodation, blur, [astigmatic blur, chromatic aberration]
- Motion 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?

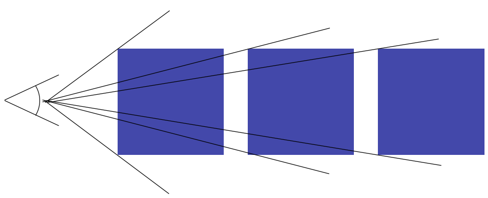
Monocular depth cues

Retinal projection depends on size and distance

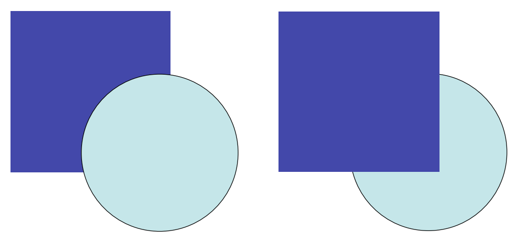


Relative size as a cue to depth

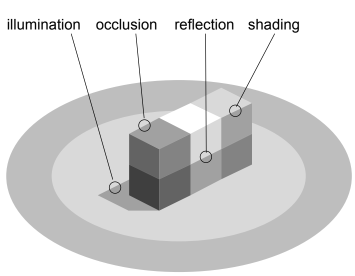
Relative size as a cue to depth



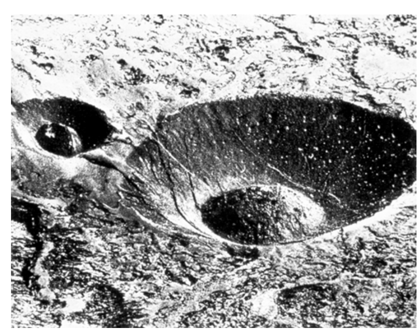
Occlusion as a cue to depth



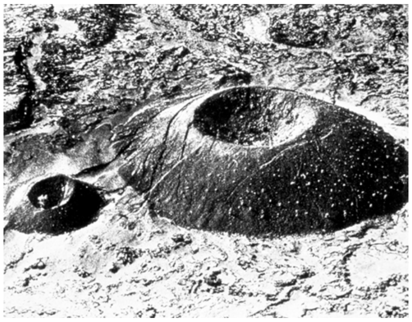
Shading, reflection, and illumination



Shading – prior of light-from-above

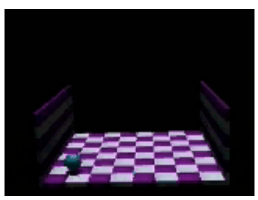


Shading (flip the photo upside-down)

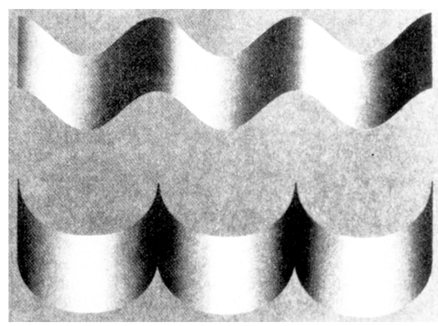


Cast Shadows

Dynamic Cast Shadows



Shading and contour

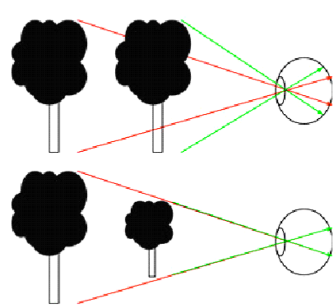


Aerial/Atmospheric Perspective

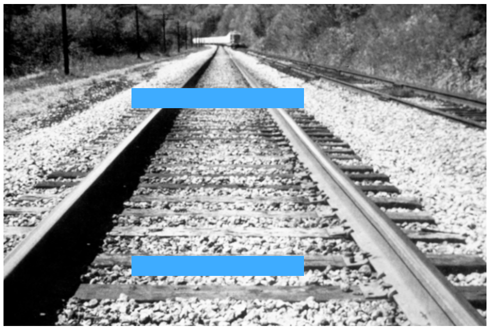
Geometry of Linear Perspective

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



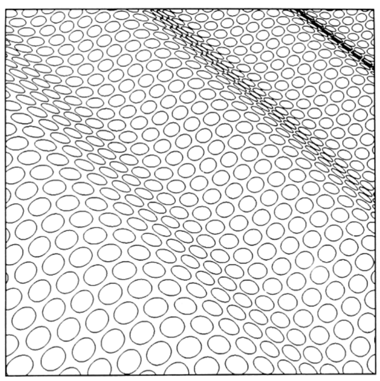
Linear perspective



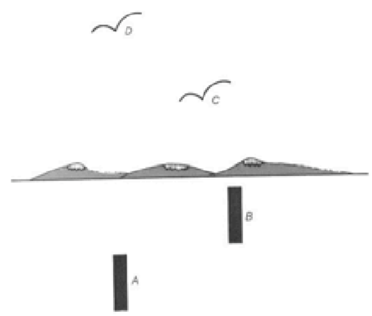
Size constancy

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

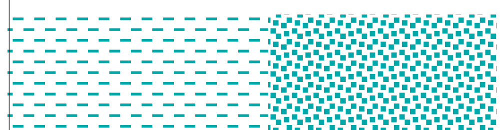
Distance, depth, and 3D shape cues

- 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

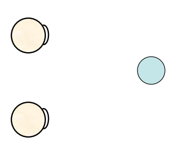
Dynamic (Kinetic) Occlusion



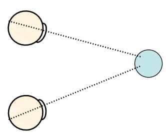
Distance, depth, and 3D shape cues

- 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

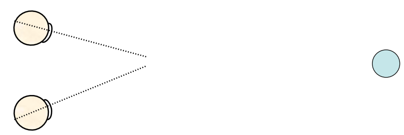
Vergence Angle As One Binocular Source



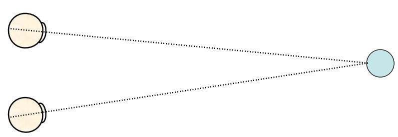
Vergence Angle As One Binocular Source



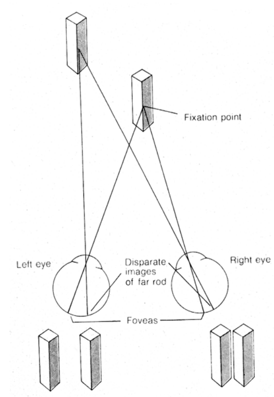
Vergence Angle As One Binocular Source



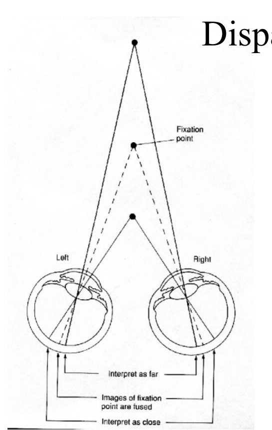
Vergence Angle As One Binocular Source



Binocular disparity



Disparity

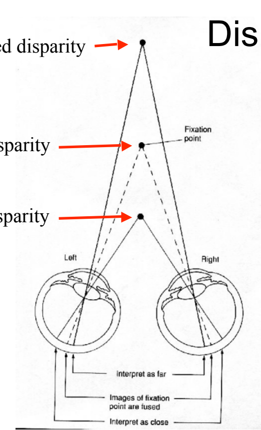


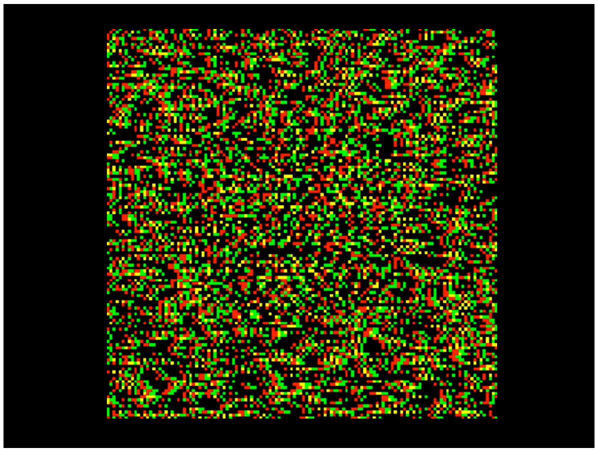
Disparity

Uncrossed disparity

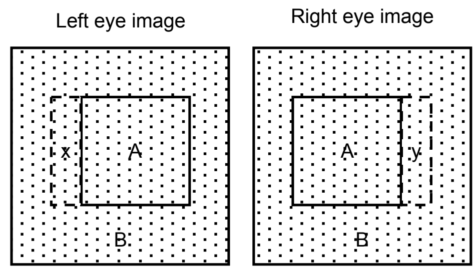
Zero retinal disparity

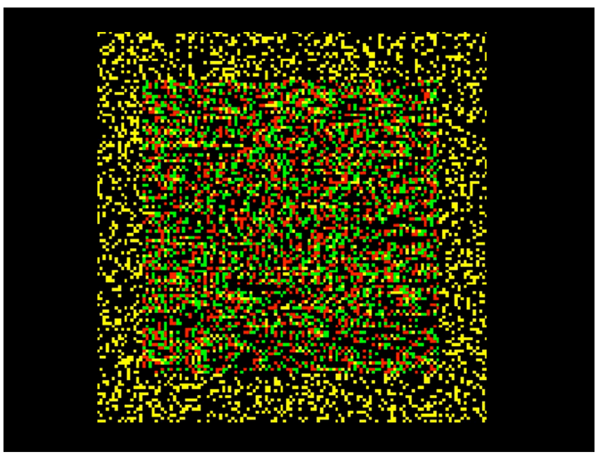
Crossed disparity





How to make a random-dot stereogram





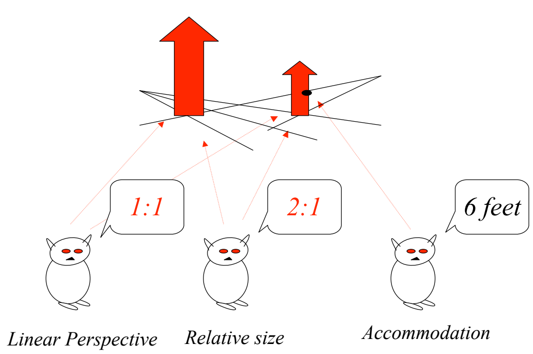
Depth Cue Combination: Issues

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?

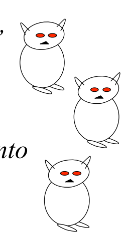
When cues disagree ...



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

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

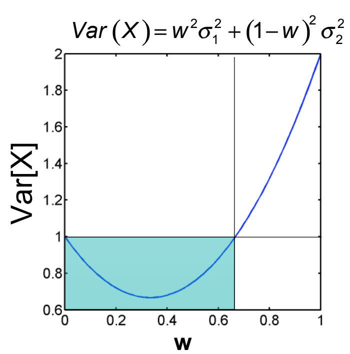
unbiased?

Minimum-Variance Cue Combination

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

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

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



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:

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

Reparamerization

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

$X = w_1 X_1 + w_2 X_2$

weight proportional to reliability

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

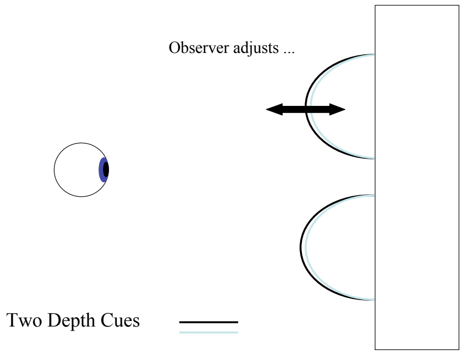
$r = r_1 + r_2$ reliabilities add

Perturbation Methodology and Influence Measures

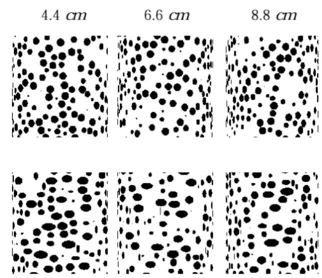
How can we measure the influence of various cues on perceptual judgments in complex scenes?

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

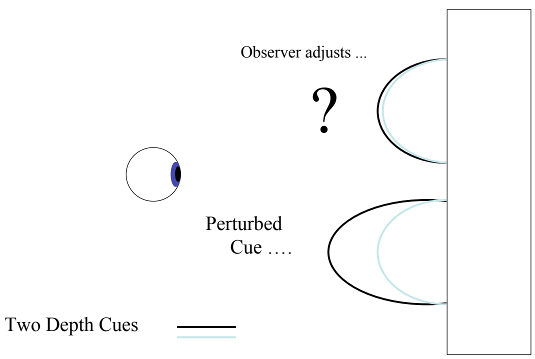
Perturbation Method



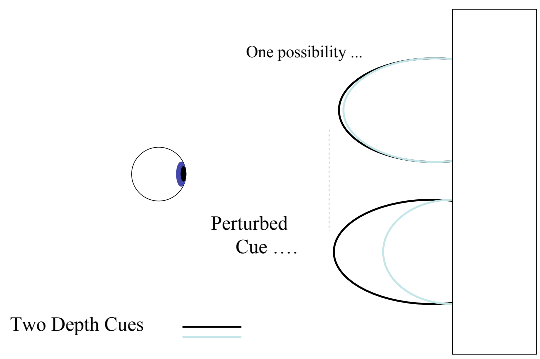
Example: Texture and Motion



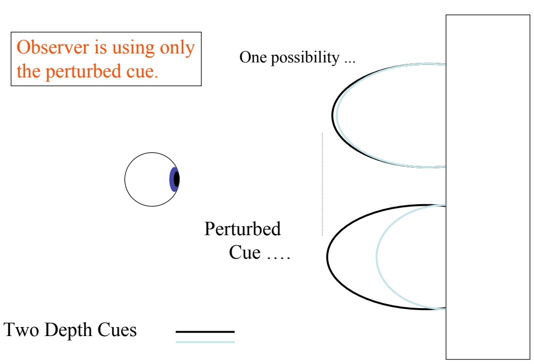
Perturbation Method



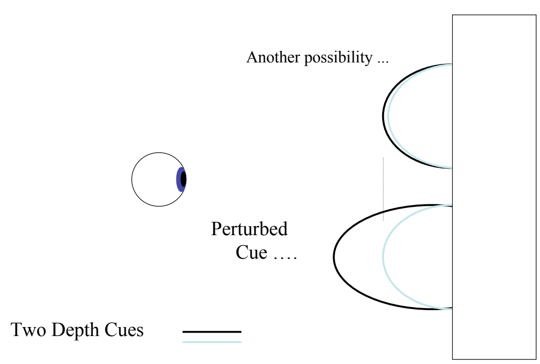
Perturbation Method



Perturbation Method

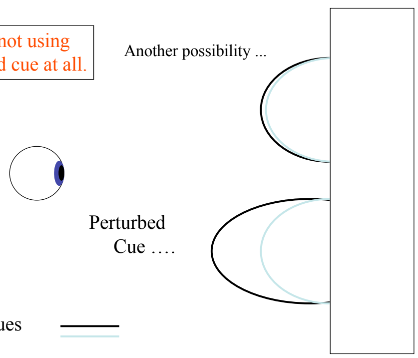


Perturbation Method

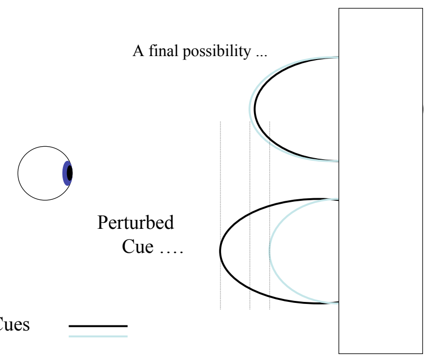


Perturbation Method

Observer is not using the perturbed cue at all.

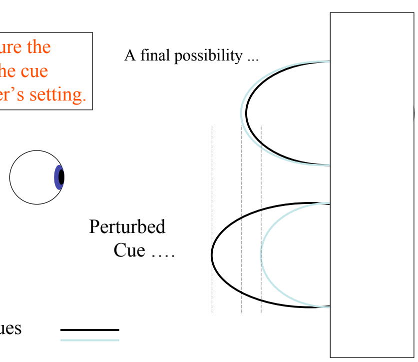


Perturbation Method



Perturbation Method

We will measure the **influence** of the cue on the observer's setting.



Influence Measures

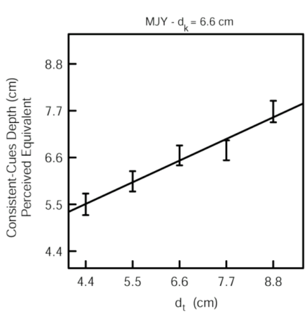
$$I_{cue} = \frac{\Delta_{setting}}{\Delta_{cue}}$$

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:

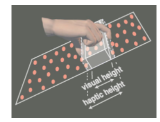
- ML: Maximum-likelihood estimator
- MAP: Maximum a posteriori estimator
- Mean of the posterior
- Median of the posterior
- Etc.

Optimal Cue Combination

Humans integrate visual and haptic information in a statistically optimal fashion

Marc O. Ernst* & Martin S. Banks

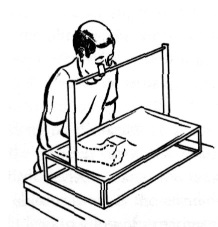
Vision Science Program/School of Optometry, University of California, Berkeley 94720-2020, USA



Rock & Victor (1964)

Irv Rock

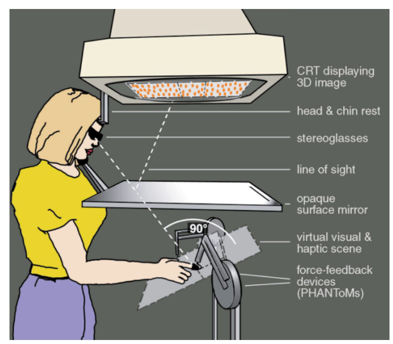
View object through distorting lens while exploring object haptically



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

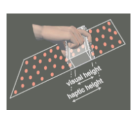
Visual capture

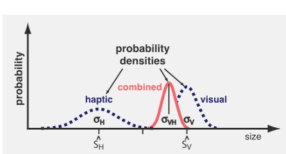
Visual/Haptic Setup



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

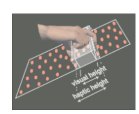
$$w_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2}$$

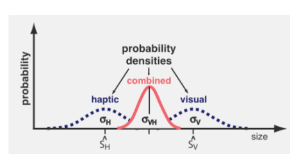
Variance

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

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

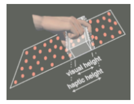
$$w_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2}$$

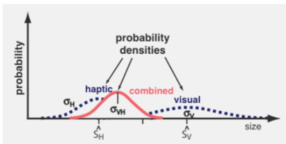
Variance

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

Visual Capture ?

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





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

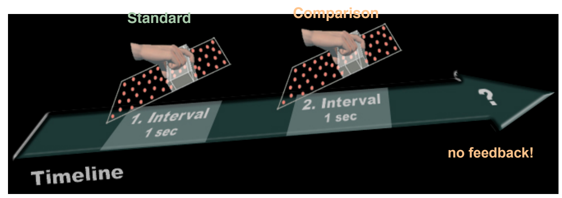
Weights

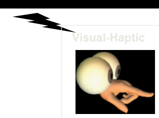
$$w_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2}$$

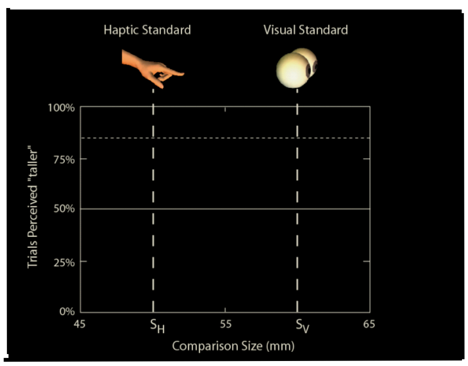
Variance

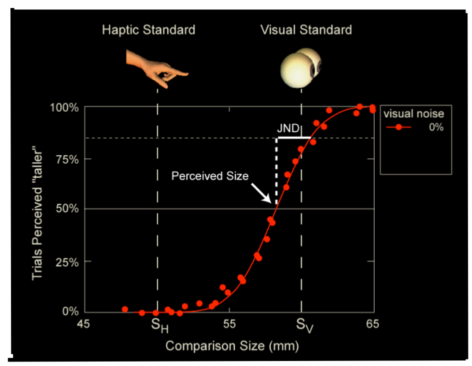
$$\frac{1}{\sigma_{VH}^2} = \frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2}$$

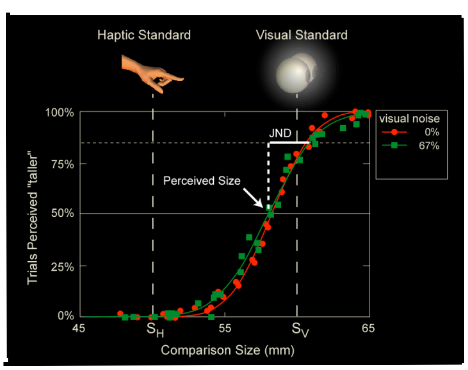
2-IFC Task

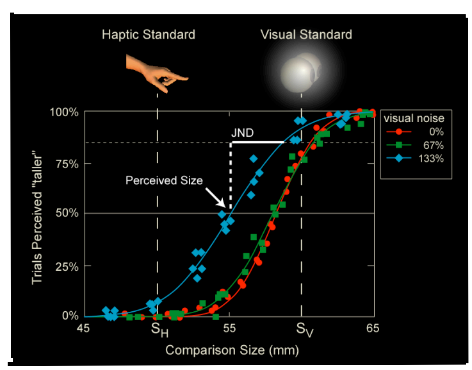


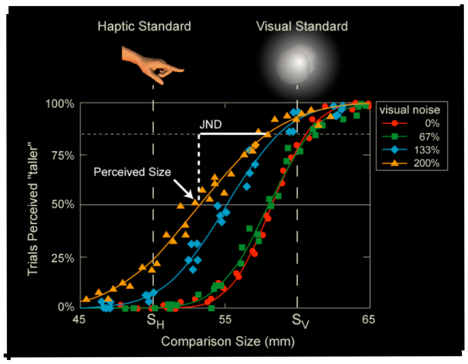




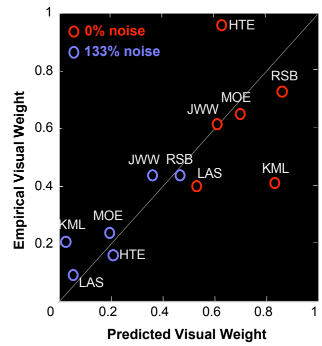








Individual Differences

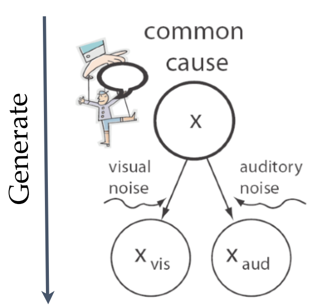


Nonlinear Cue Combination: Causal models

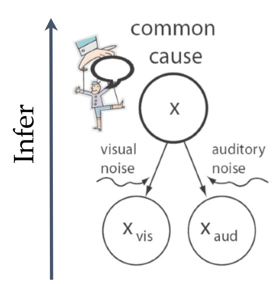
Nonlinear Cue Combination: Causal models

Visual Auditory combination (Ventriloquist effect)

Modeling: Where do cues come from?



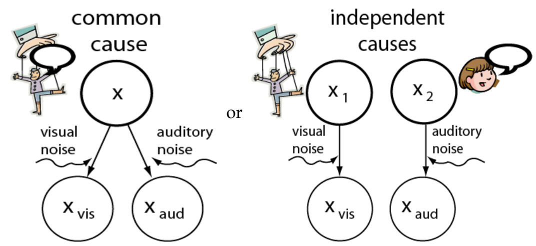
Traditional Bayesian model



What would happen now?

Obviously there may be more than one source.

Mixture model

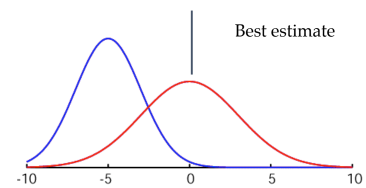


p(causal model)

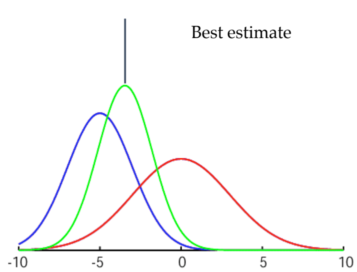
- Using Bayes rule:

$$p(common|x_{vis}, x_{aud}) = \frac{p_{common}p(x_{vis}, x_{aud}|common)}{p_{common}p(x_{vis}, x_{aud}|common) + (1 - p_{common})p(x_{vis}, x_{aud}|\neg common)}$$

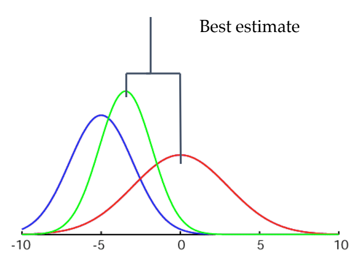
Independent causes: where is the auditory stimulus



Common cause: where is the auditory stimulus

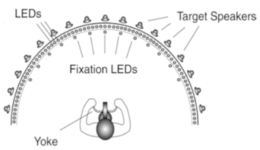


Mean squared error estimate

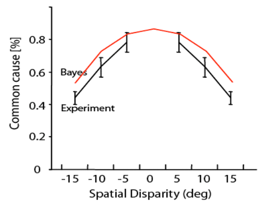


$$\hat{X} = P(C|X_{vis}, X_{aud})\hat{X}_c + P(\neg C|X_{vis}, X_{aud})\hat{X}\neg c$$

Experimental test

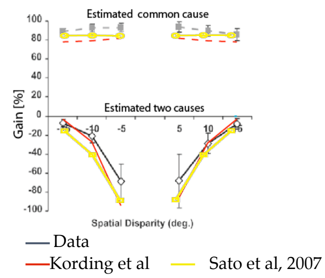


Button: common cause or two

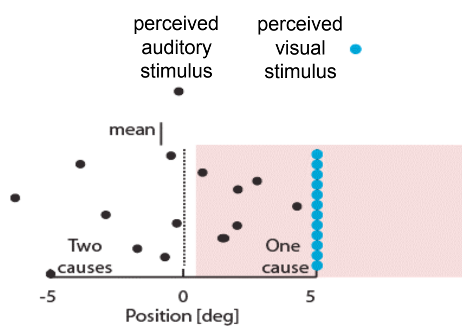


Wallace et al. 2005; Hairston et al. 2004

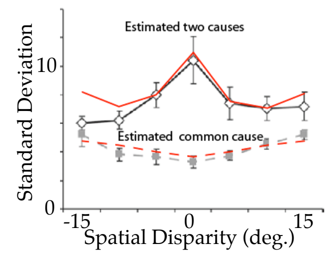
Measured gain



How can the gain be negative?

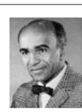


Predicting the variance



Worse prediction if we assume model selection

Statistical Decision Theory

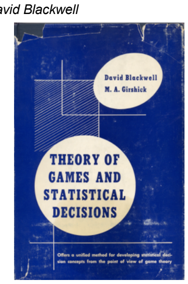


David Blackwell

Abraham Wald

John von Neumann

Oskar Morgenstern



1954

The Three Elements of SDT

$W = \{w_1, w_2, \cdots, w_m\}$ possible states of the world

$A = \{a_1, a_2, \cdots, a_p\}$ possible actions

$X = \{x_1, x_2, \cdots, x_n\}$ possible sensory events

$$X \sim f(x;\theta)$$

The Three Functions of SDT

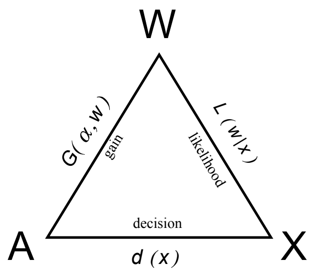


Fig. 1

The Three Functions of SDT

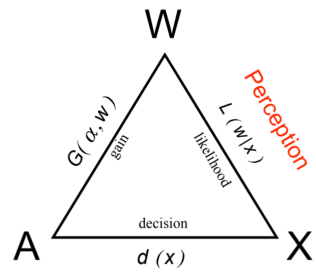


Fig. 1

The Three Functions of SDT

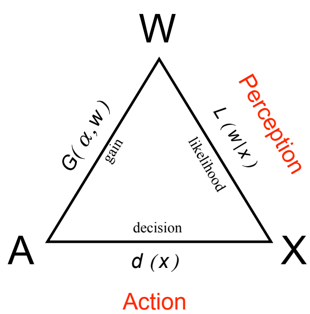


Fig. 1

The Three Functions of SDT

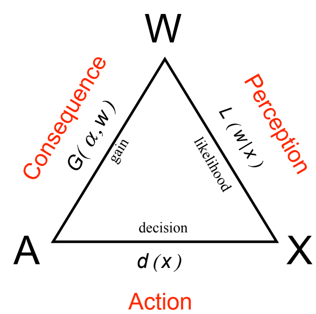
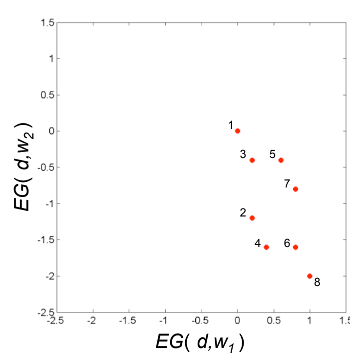


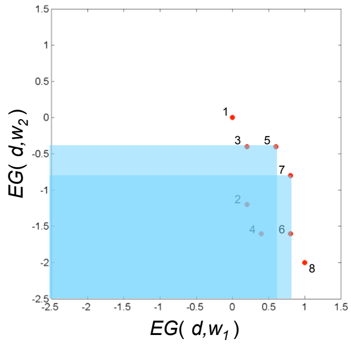
Fig. 1

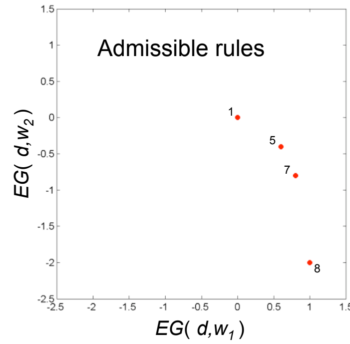
Goal: select

$$d : X \rightarrow A$$

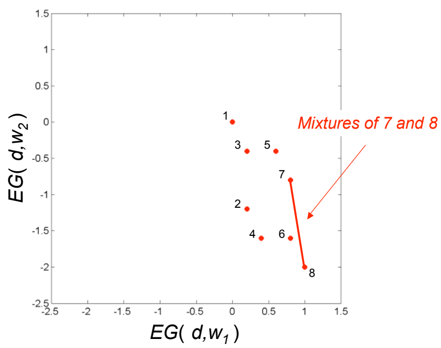


Partial ordering: dominance

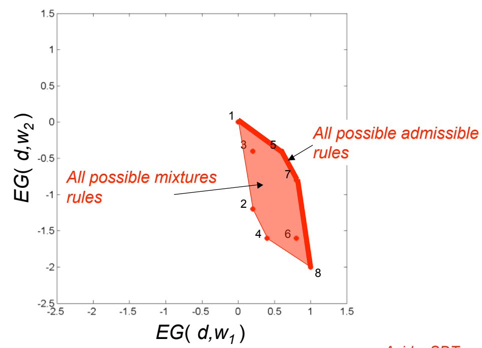




Mixture rules



Mixture rules

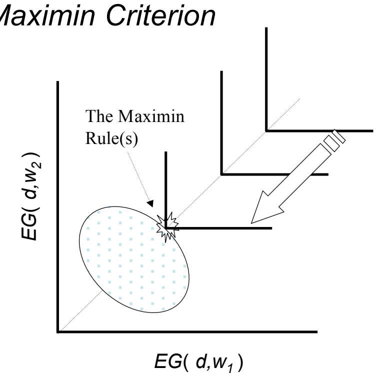


Aside: SDT

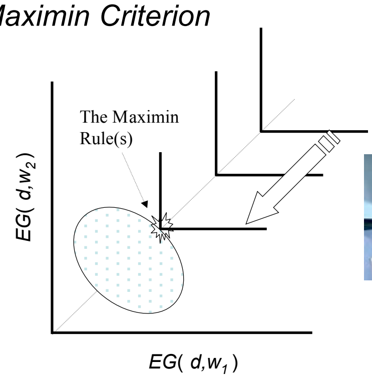
Partial ordering of rules

d: X → A

Maximin Criterion

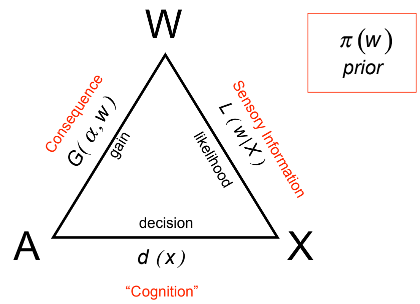


Maximin Criterion



Dr Evil

Bayesian Decision Theory

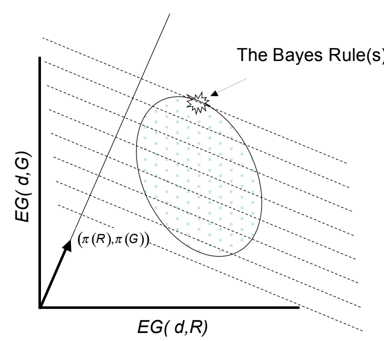


Bayesian Decision Theory

We add a prior probability distribution on
The state of the world and define the
Bayes Risk of each rule.

$$EBR(d) = \sum_{i=1}^{n} EG(d, w_i)\pi(w_i)$$

The rule d* that maximizes Bayes Risk
Is the Bayes Rule (it need not be unique).



Bayesian Decision Theory (Continuous Form)

Maximize expected Bayes gain

$$EBG(d) = \iint G(d(x), w)\, L(w \mid x)\, \pi(w)\, dx\, dw$$

by choice of a decision rule

$$d : X \rightarrow A$$

Bayesian Decision Theory

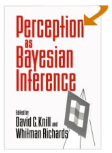
Geisler (1989)

$$\iint G(d(x), w)\, L(w \mid x)\, \pi(w)\, dx\, dw$$

Bayesian Decision Theory

Geisler (1989)

$$\iint G\big(d(x),w\big)\, L(w \mid x)\,\pi(w)\, dx\, dw$$



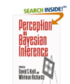
Knill & Richards (1996)

Perception as Bayesian Inference

A choice

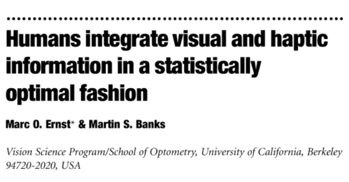
Normative *Geisler (1989)*

Descriptive (process)



How test?

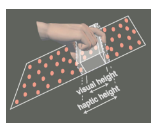
Optimal Cue Combination



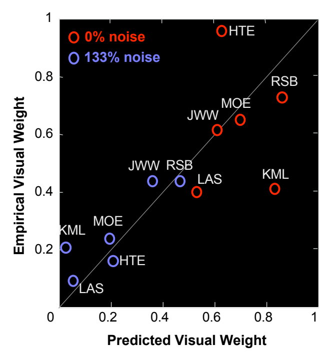
Humans integrate visual and haptic information in a statistically optimal fashion

Marc O. Ernst* & Martin S. Banks

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



Individual Differences



What is the gain/loss function?

Quadratic loss (least squares)

Process model is weighted linear combination minimizing quadratic loss

$$S = w_H S_H + w_V S_V$$

A weak test of BDT …

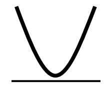
Bayesian Decision Theory

Geisler (1989)

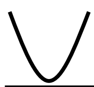
$$\iint G(d(x),w)\, L(w \mid x)\, \pi(w)\, dx\, dw$$

All that really matters …

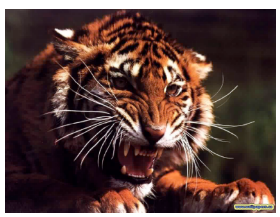
Knill & Richards (1996)



loss function?



loss function?



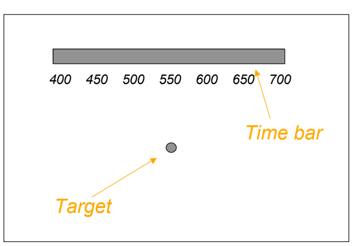
loss function!

Can the visuo-motor system, presented with arbitrary gain functions, select decision rules that maximize expected gain?

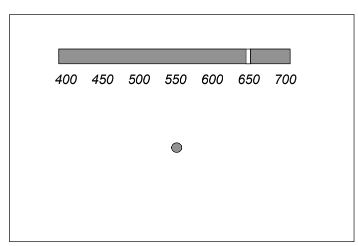
Direct manipulation of gain/loss function

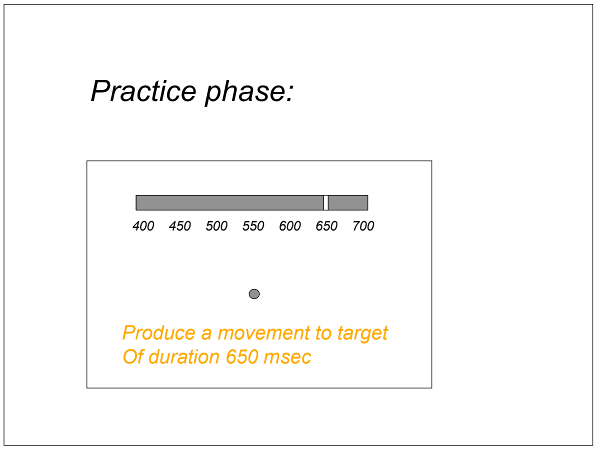
Strong test of BDT

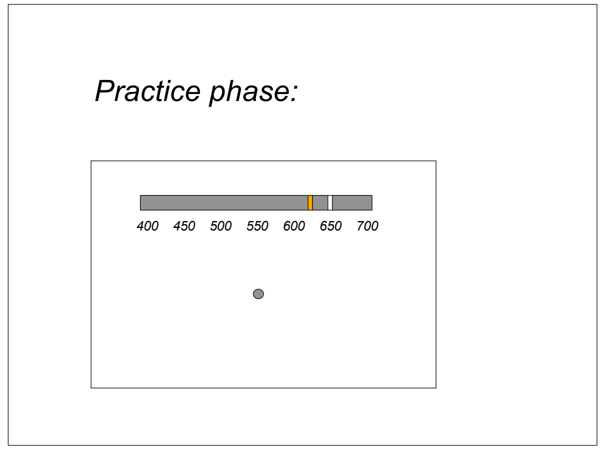
Motor Timing Experiment:

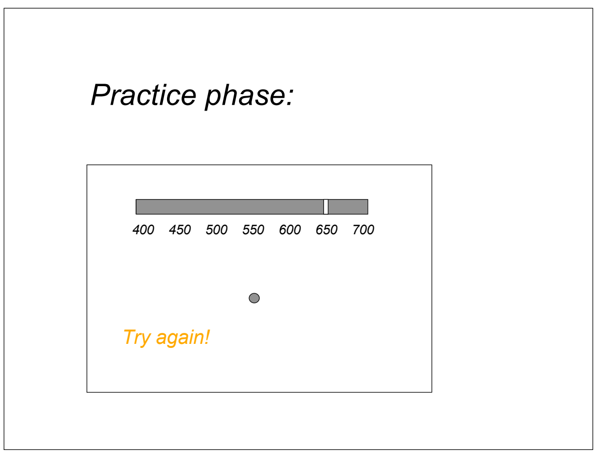


Practice phase:

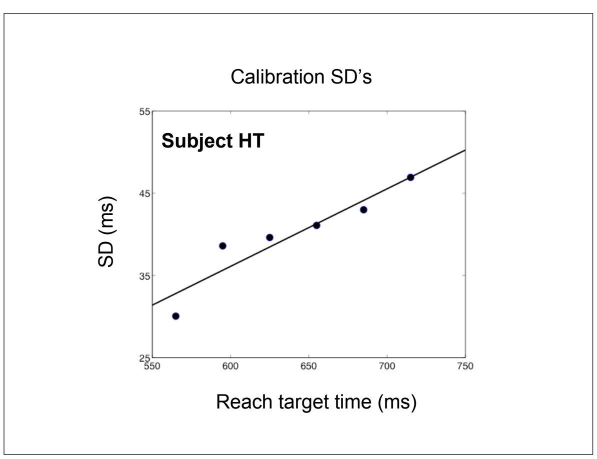


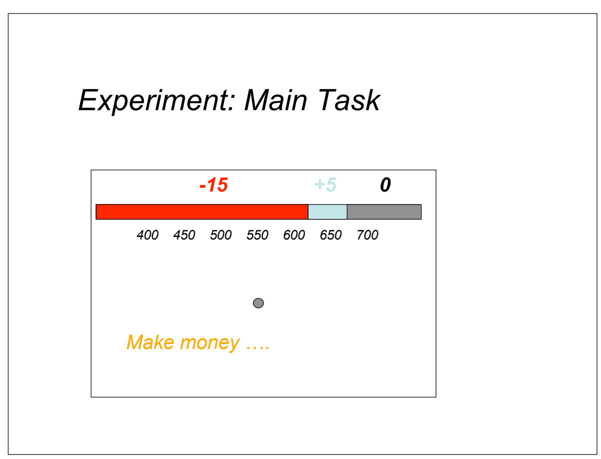

Practice phase:
400 450 500 550 600 650 700
Produce a movement to target
Of duration 650 msec


Practice phase:
400 450 500 550 600 650 700


Practice phase:
400 450 500 550 600 650 700
Try again!

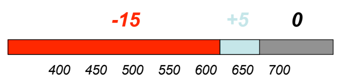

Practice phase:
400 450 500 550 600 650 700
Better!


Calibration SD's
Subject HT
SD (ms)
55
45
35
25
550
600
650
700
750
Reach target time (ms)

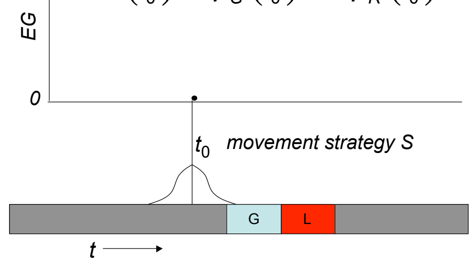

Experiment: Main Task
-15
+5
0
400 450 500 550 600 650 700
Make money ….

Experiment: Main Task

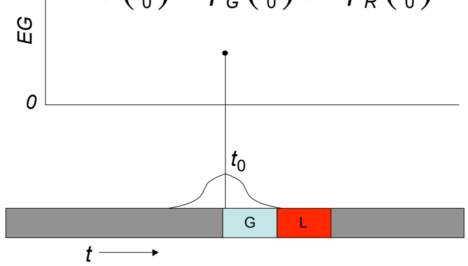
How to maximize expected gain?



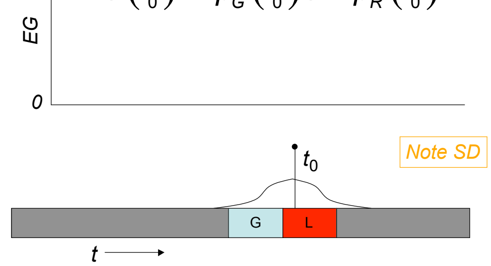
$$EG(t_0) = p_G(t_0)G + p_R(t_0)R$$



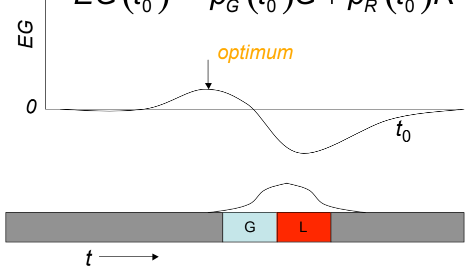
$$EG(t_0) = p_G(t_0)G + p_R(t_0)R$$



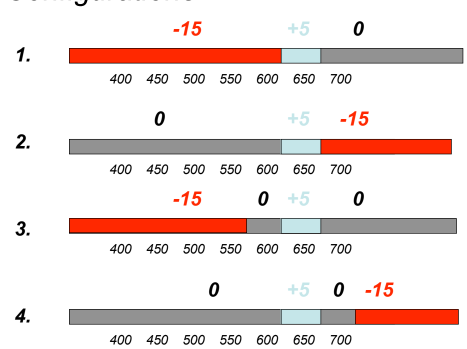
$$EG(t_0) = p_G(t_0)G + p_R(t_0)R$$

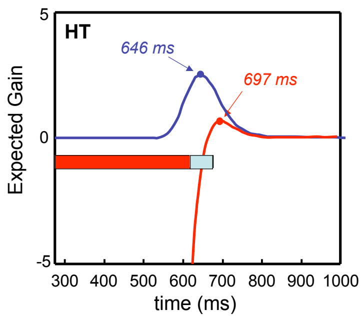


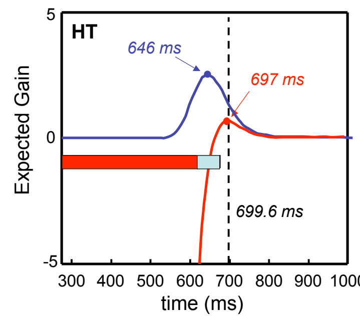
$$EG(t_0) = p_G(t_0)G + p_R(t_0)R$$

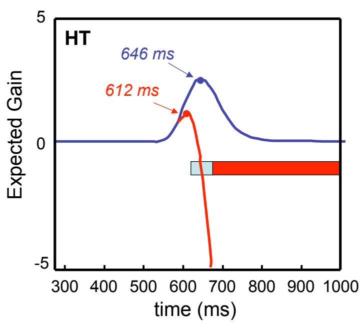


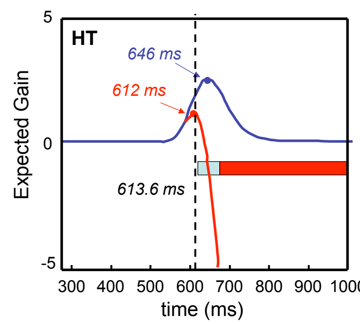
Configurations

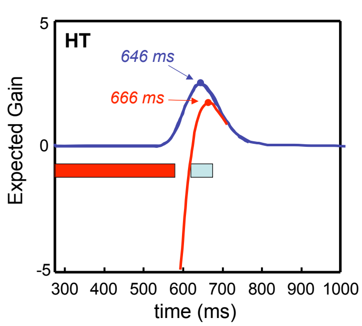


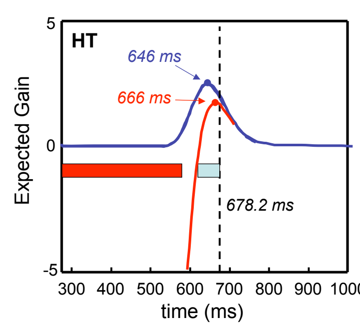

HT
646 ms
697 ms
Expected Gain
time (ms)

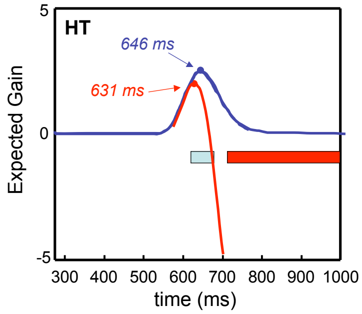

HT
646 ms
697 ms
699.6 ms
Expected Gain
time (ms)

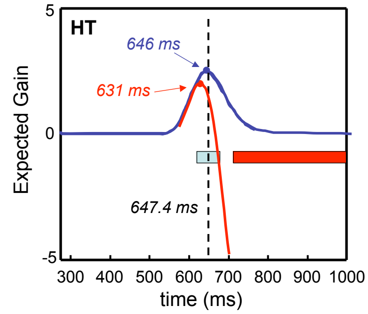

HT
646 ms
612 ms
Expected Gain
time (ms)

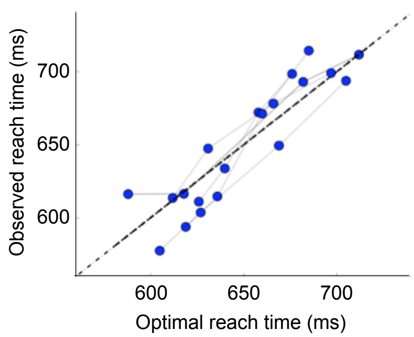

HT
646 ms
612 ms
613.6 ms
Expected Gain
time (ms)


HT
646 ms
666 ms
Expected Gain
time (ms)


HT
646 ms
666 ms
678.2 ms
Expected Gain
time (ms)







Summary

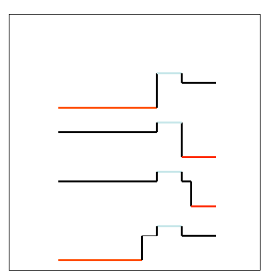
Subjects chose movements whose mean time came close to maximizing expected gain.
No patterned deviations.

Bayesian Decision Theory

$$\iint G(d(x),w)\; L(w \mid x)\,\pi(w)\; dx\,dw$$

Bayesian Decision Theory

$$\iint G(d(x),w)\; L(w \mid x)\,\pi(w)\; dx\,dw$$



Conclusions

Gain/loss functions are problems posed by the environment to the organism

They are an useful as independent variables in exploring visuo-motor function

Their manipulation allows us to test performance against ideal in a wide range of economic games