Bayesian perception

Daniela Pamplona

U2IS - ENSTA - IPParis

ecampus moodle: MI210 - Modèles neuro-computationnels de la vision (P4 - 2020-21)

daniela.pamplona@ensta.fr

Summary

1. Bayes Theorem and Bayesian modeling
2. Bayesian Brain
   1. Formulation
   2. Examples
3. Bayesian life long learning
4. Critics to the “ideal observer”

Summary

1. Bayes Theorem and Bayesian modeling
2. Bayesian Brain
   1. Formulation
   2. Examples
3. Bayesian life long learning
4. Critics to the “ideal observer”

Frequentist Approach

Event's probability: is the limit of its relative frequency in a large number of trials.

It supports the statistical needs of experimental scientists;

Probabilities can be found by a repeatable objective process.

Number of times it was realized

Prob. Event

$$P(E_1) = \frac{\#E_1}{\sum_i \#E_i}$$

Total number of repetitions

Bayes Theorem

Event's probability: measures a "degree of belief". Bayes' theorem then links the degree of belief in a proposition before and after accounting for an observation

- **Prior:** initial degree of belief in E_1
- **Likelihood:** the degree of belief having accounted for the observation O given that E_1 happend
- **Marginalization:** sum of likelihood over all possible events
- **Posterior:** final degree of believe in E_1

Posterior

Likelihood

$$P(E_1|O) = \frac{P(O|E_1)}{\sum_i P(O|E_i)} P(E_1)$$

E Marginal dist.

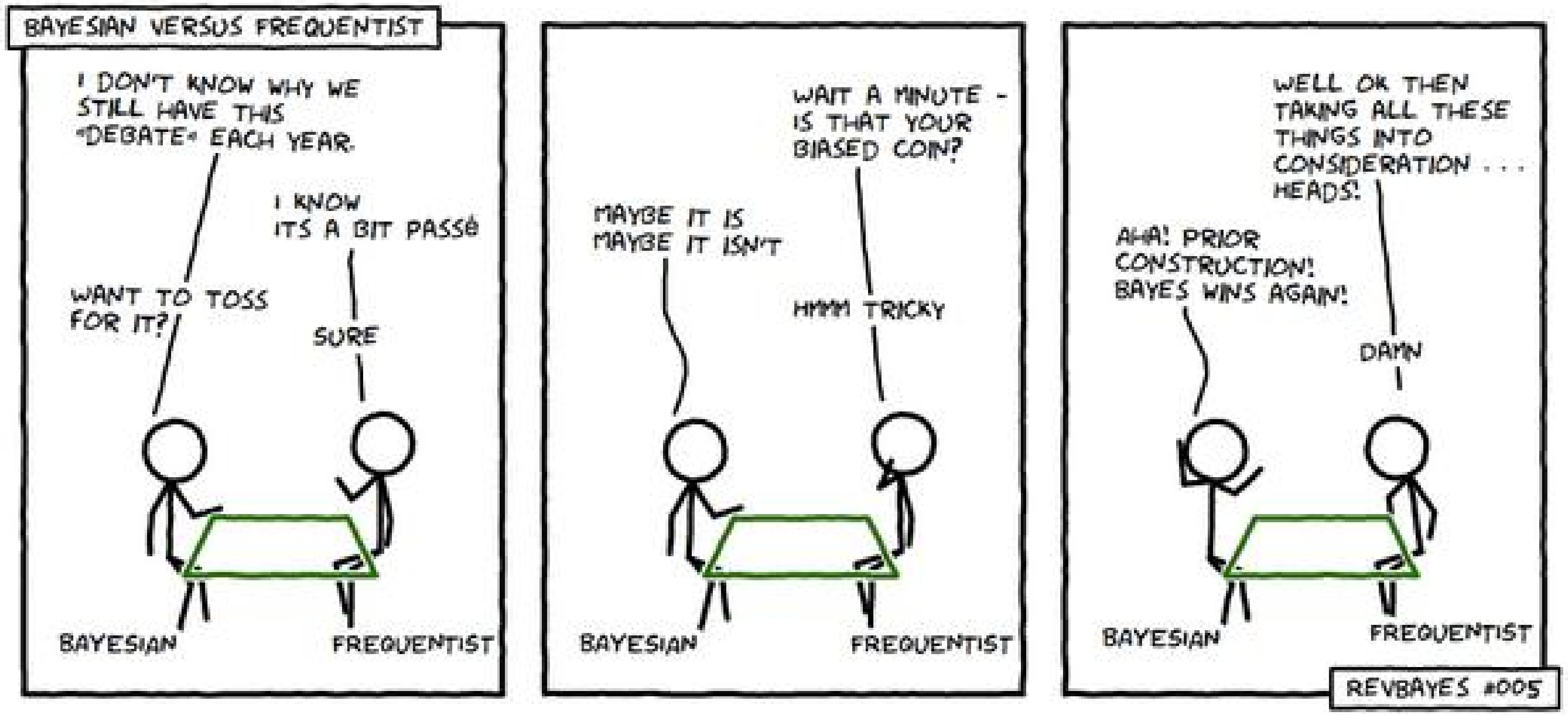
Prior

Frequentist vs Bayesian approach:

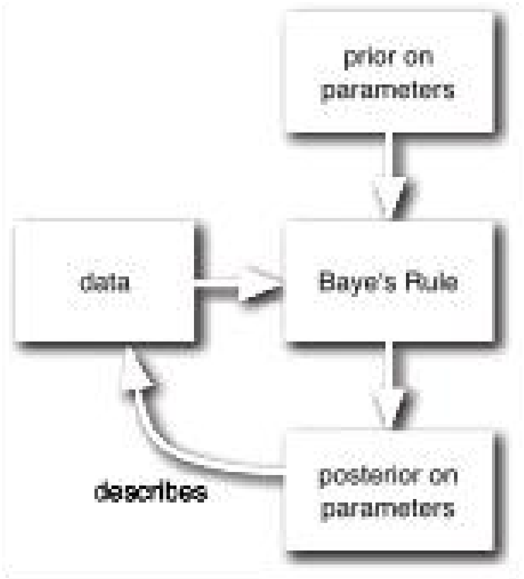
Category	Frequentist	Bayesian
Probability, P:	long-run frequency	degree of belief
Parameters:	fixed but unknown	random variables
Focuses on:	variability of data	uncertainty of knowledge
Mathematical machinery:	sampling distribution, repeated hypothetical experiments	Bayes' theorem, fixed data
Answers/calculates:	$P\,(\text{data}\|\text{hypothesis})$	$P\,(\text{hypothesis}\|\text{data})$
Source of information:	data (observations)	data (observations) + prior belief

Snow extremes and structural reliability,Árpád Rózsás 2016

Frequentist vs Bayesian approach:



Bayesian applications

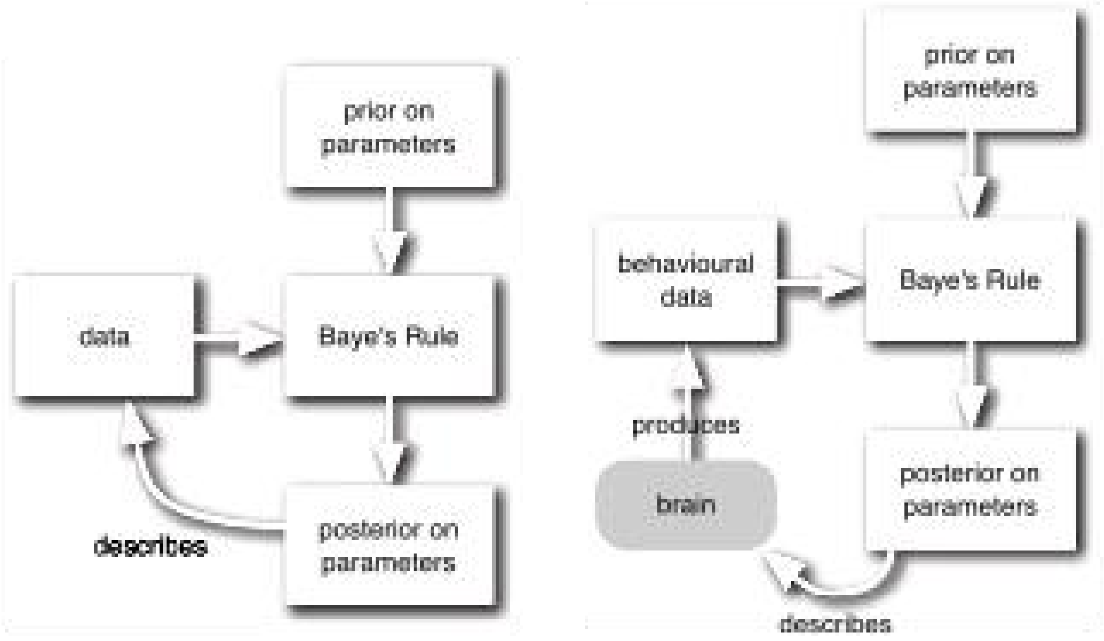


(a) Data analysis models.

Fig. 1.
Source:

A tutorial on Bayesian models of perception, Benjamin T. Vincent

Bayesian applications

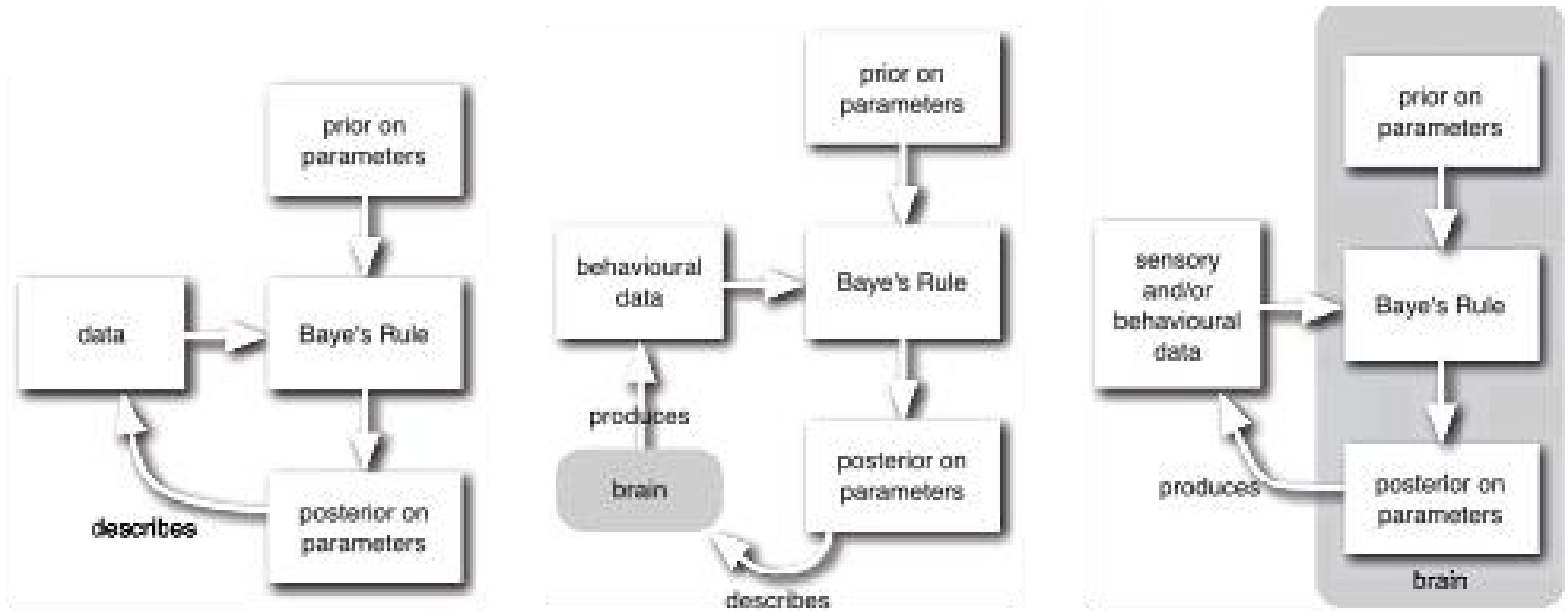


(a) Data analysis models. (b) Psychometric models.

Fig. 1. Three contexts in which Bayesian modelling
Source: Adapted with permission from Kruschke (20

A tutorial on Bayesian models of perception,
Benjamin T. Vincent

Bayesian applications



(a) Data analysis models. (b) Psychometric models. (c) Bayesian Brain models.

Fig. 1. Three contexts in which Bayesian modelling are used.
Source: Adapted with permission from Kruschke (2011).

A tutorial on Bayesian models of perception,
Benjamin T. Vincent

Building a Bayesian model (optimal observer model)

1. **The generative model:** the probabilities of world states and of sensory observations given world states
2. **The inference process:** the probability of the world state given the sensory observations
3. **The distribution of the MAP estimate:** The observer is optimal, it estimates using the maximum a posteriori distribution. The predicted distribution of the observer's estimates.

Building a Bayesian model (optimal observer model)

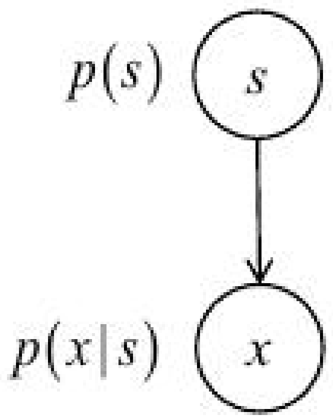
1. **The generative model:** the probabilities of world states and of sensory observations given world states
2. **The inference process:** the probability of the world state given the sensory observations
3. ~~**The distribution of the MAP estimate:** The observer is optimal, it estimates using the maximum a posteriori distribution. The predicted distribution of the observer's estimates.~~

Building a Bayesian model: The generative model

Description of two variables:

1. **Stimulus** or particular relevant feature of it

2. Internal **measurement** of that stimulus



Building a Bayesian model: Inference

1. Derive the likelihood function

$$L(s;x) = p(x|s)$$

2. Write the posterior

$$p(s|x) = \frac{p(x|s)\,p(s)}{p(x)}$$

$$p(s|x) \propto p(x|s)\,p(s)$$

3. Derive the Maximum a Posteriori (MAP) estimator

$$\hat{s}_{\mathrm{MAP}} = \underset{s}{\operatorname{argmax}}\; p(s|x)$$

Building a Bayesian model: The MAP distribution

- Derive the MAP distribution $p\left(\hat{s}_{\mathrm{MAP}} \mid s\right)$

- Estimate the mean and variance $\hat{s}_{\mathrm{MAP}}$ σ^2_{MAP}

Building a Bayesian decision model

1. Bayesian model of all options
2. Build the inner believes
3. Define a risk/cost function
4. Decide for lower risk/cost

Building a Bayesian decision model

1. Bayesian model of all options
2. Build the inner believes
3. ~~Define a risk/cost function~~
4. ~~Decide for lower risk/cost~~ Decide for higher believe

The typical error...

Building a Bayesian decision model

1. Bayesian model of all options
2. Build the inner believes
3. ~~Define a risk/cost function~~
4. ~~Decide for lower risk/cost~~ Decide for higher believe

Bayesian decision modeling

Is the probability of ... find a mosquito smaller or larger ... probability of dying if I find a shark?

Solve on board?

Bayesian decision modeling

Is the probability of dying if I find a mosquito smaller or larger than the probability of dying if I find a shark?

Bayesian decision modeling

Is the probability of dying if I find a mosquito smaller or larger than the probability of dying if I find a shark?

$$\frac{P(D|M)}{P(D|S)} = ???$$

$$\frac{P(D|M)}{P(D|S)} = \frac{\frac{P(M|D)P(D)}{P(M)}}{\frac{P(S|D)P(D)}{P(S)}} = \frac{P(M|D)P(S)}{P(S|D)P(M)}$$

$$= \frac{72500}{7} x \frac{1x10^3}{8x10^{11}} = 1.3x10^{-4}$$

The typical error…

$$P(M|D) \neq P(D|M)$$

$$P(S|D) \neq P(D|S)$$

(Probabilistic) Graphical models

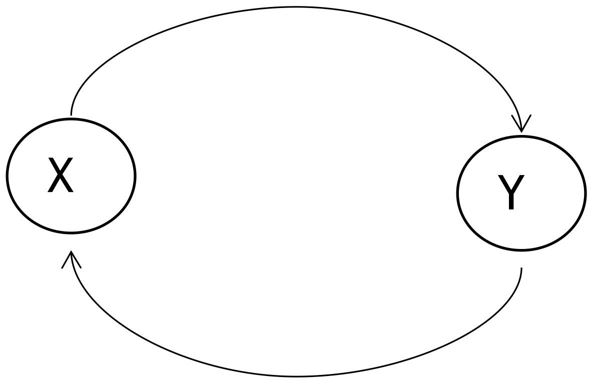
How to represent discrete probabilities?

joint probability table (K-possible states for n variables)

k^n states: easily explode!

Probabilistic graphical model: graph expresses the dependence structure between random variables.

X	Y	P(X,Y)
0	0	0.25
1	0	0.45
0	1	0.15
1	1	0.15



http://frnsys.com/notes/ai/foundations/probabilistic_graphical_models.html

(Probabilistic) Graphical models

Convention:

Every node is dependent on its parent and nothing else that is not a descendant. To put it another way: given its parent, a node is independent of all its non-descendants.

http://frnsys.com/notes/ai/foundations/probabilistic_graphical_models.html

(Probabilistic) Graphical models

Convention:

Every node is dependent on its parent and nothing else that is not a descendant. To put it another way: given its parent, a node is independent of all its non-descendants.

Example

Say we are looking at five events:

- a dog barking (D)
- a raccoon being present (R)
- a burglar being present (B)
- a trash can is heard knocked over (T)
- the police is called (P)

http://frnsys.com/notes/ai/foundations/probabilistic_graphical_models.html

(Probabilistic) Graphical models

Convention:

Every node is dependent on its paren... else that is not a descendant. To put it anoth... its parent, a node is independent of all its non-de...

Example

Say we are looking at f...
- a dog barking (D)
- a raccoon being...
- a burglar being p... B)
- a trash can is hea... nocked over (T)
- the police is called (P)

Can you make the model?

http://frnsys.com/notes/ai/foundations/probabilistic_graphical_models.html

(Probabilistic) Graphical models

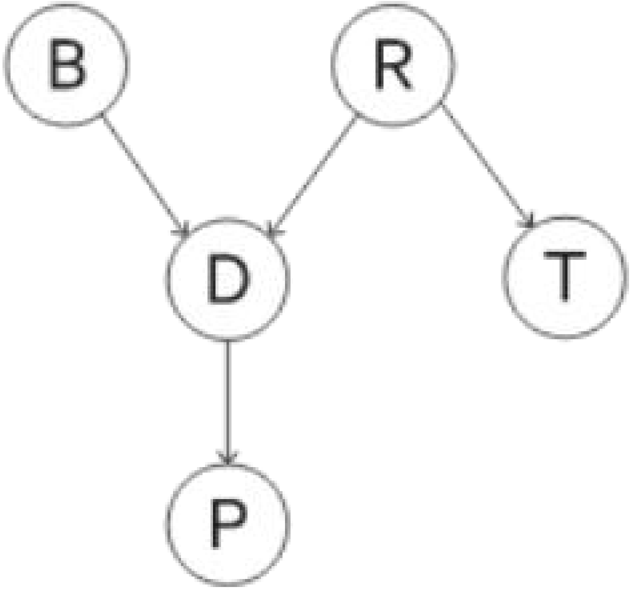
Convention:

Every node is dependent on its parent and nothing else that is not a descendant. To put it another way: given its parent, a node is independent of all its non-descendants.

Example

Say we are looking at five events:

- a dog barking (D)
- a raccoon being present (R)
- a burglar being present (B)
- a trash can is heard knocked over (T)
- the police is called (P)

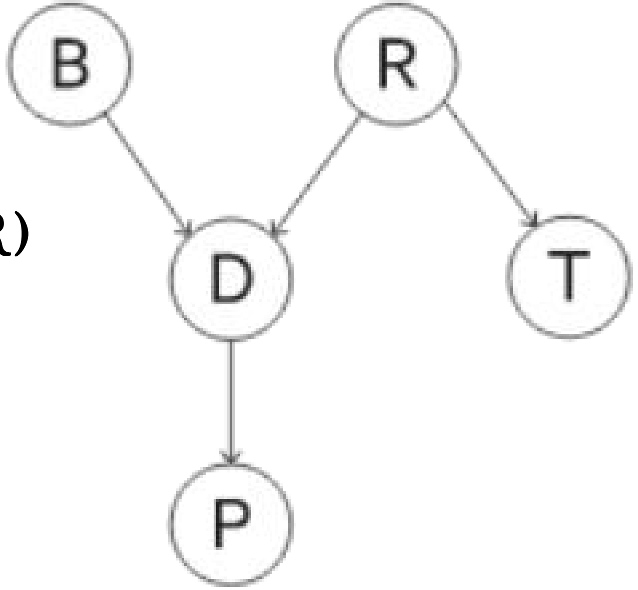


http://frnsys.com/notes/ai/foundations/probabilistic_graphical_models.html

(Probabilistic) Graphical models

Convention:

Every node is dependent on its parent and nothing else that is not a descendant. To put it another way: given its parent, a node is independent of all its non-descendants.



$$P(P,D,B,T,R) = P(P|D,B,T,R)\ P(D|B,T,R)\ P(B|T,R)\ P(T|R)\ P(R)$$

$$P(P,D,B,T,R) = P(P|D)\ P(D|B,R)\ P(B)\ P(T|R)\ P(R)$$

http://frnsys.com/notes/ai/foundations/probabilistic_graphical_models.html

(Probabilistic) Graphical models

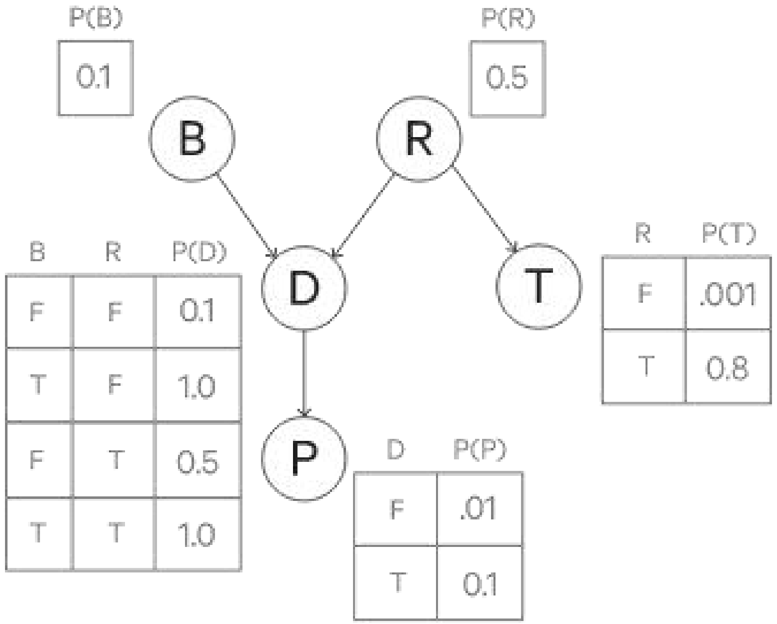
Convention:

Every node is dependent on its parent and nothing else that is not a descendant. To put it another way: given its parent, a node is independent of all its non-descendants.

Example

Say we are looking at five events:

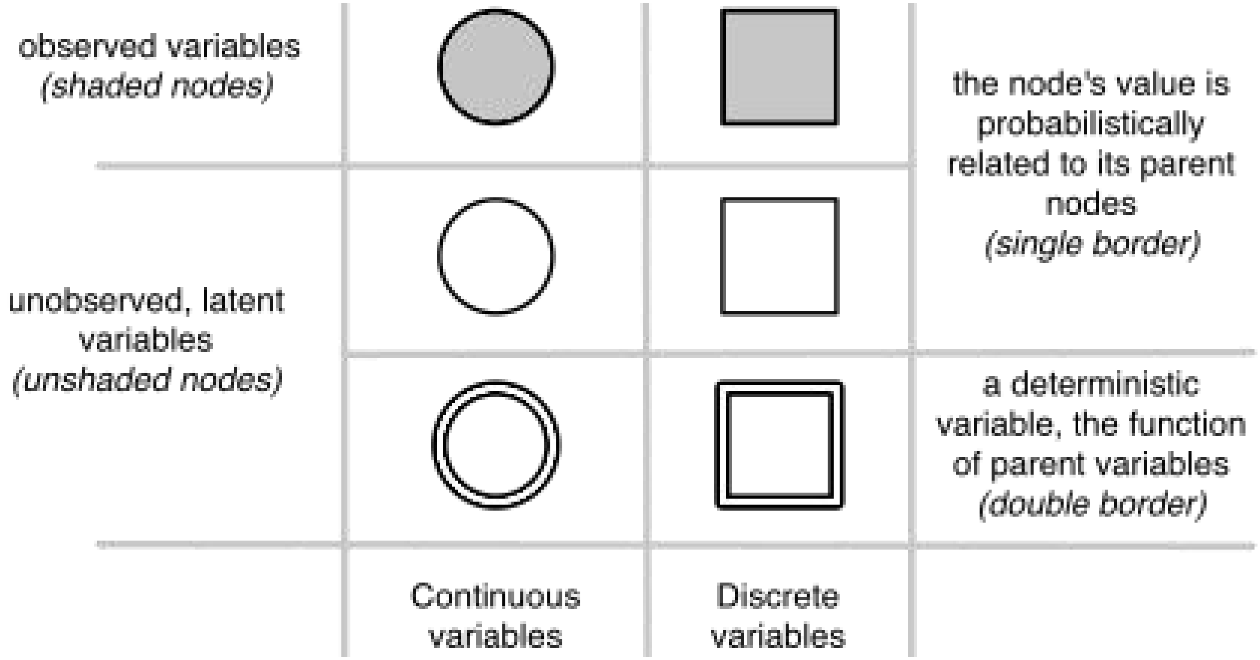
- a dog barking (D)
- a raccoon being present (R)
- a burglar being present (B)
- a trash can is heard knocked over (T)
- the police is called (P)



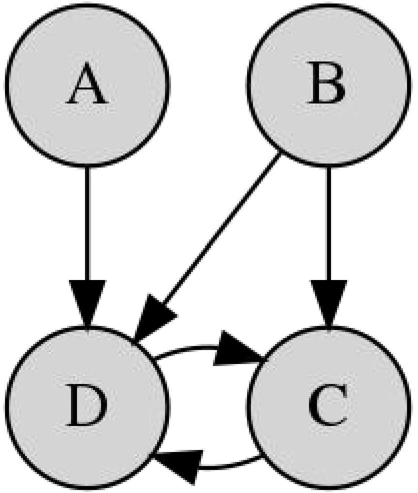
http://frnsys.com/notes/ai/foundations/probabilistic_graphical_models.html

(Probabilistic) Graphical models

Conventions:



(Probabilistic) Graphical models



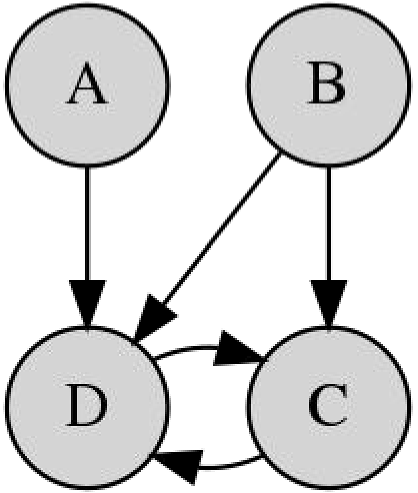
Question:

P(A,B,C,D)=?

http://frnsys.com/notes/ai/foundations/probabilistic_graphical_models.html

https://en.wikipedia.org/wiki/Graphical_model

(Probabilistic) Graphical models



Question:

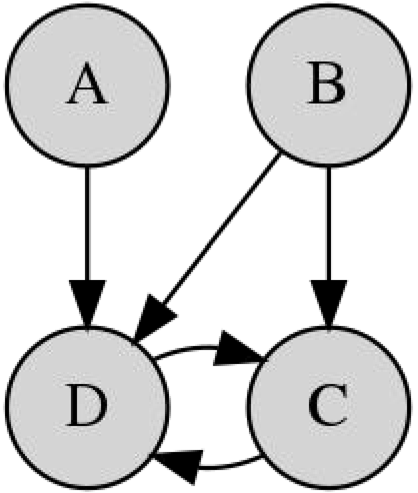
P(A,B,C

Can you make the model?

http://frnsys.com/notes/ai/foundations/probabilistic_graphical_models.html

https://en.wikipedia.org/wiki/Graphical_model

(Probabilistic) Graphical models



Question:
P(A,B,C,D)=?

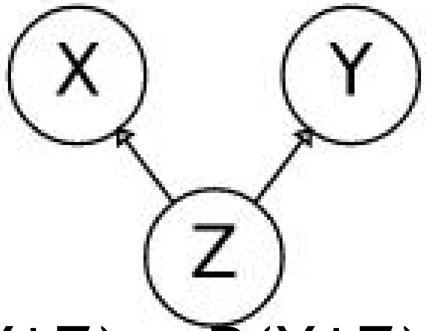
- D depends on A, B, and C;
- C depends on B and D;
- A and B are independent.

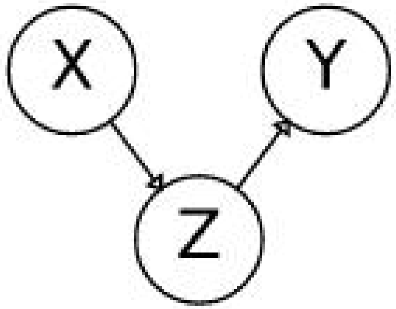
$$P(A,B,C,D) = P(A)xP(B)xP(C,D|A,B)$$

http://frnsys.com/notes/ai/foundations/probabilistic_graphical_models.html
https://en.wikipedia.org/wiki/Graphical_model

(Probabilistic) Graphical models

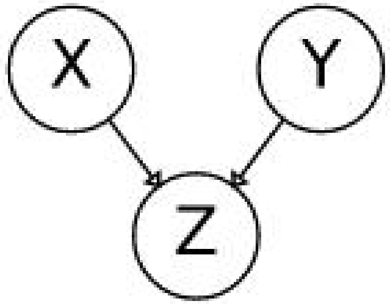
- Conditional independence:





$P(X,Y|Z) = P(X|Z)P(Y|Z)$ $\qquad$ $P(X,Y|Z) = P(Z|X)P(X)P(Y|Z)$

- Conditional dependence:



$P(X,Y|Z) = P(Z|X,Y)P(X)P(Y)$

http://frnsys.com/notes/ai/foundations/probabilistic_graphical_models.html

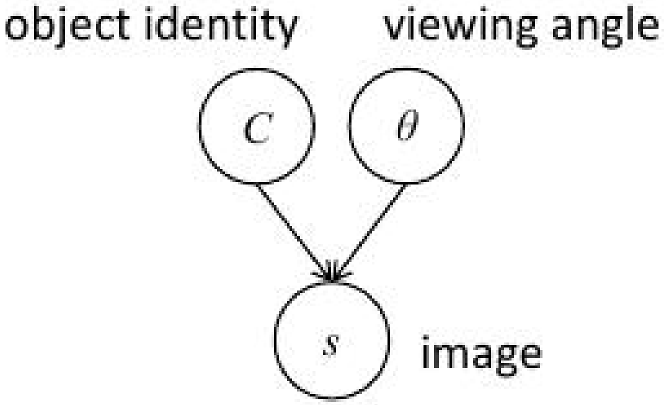
Nuisance variables

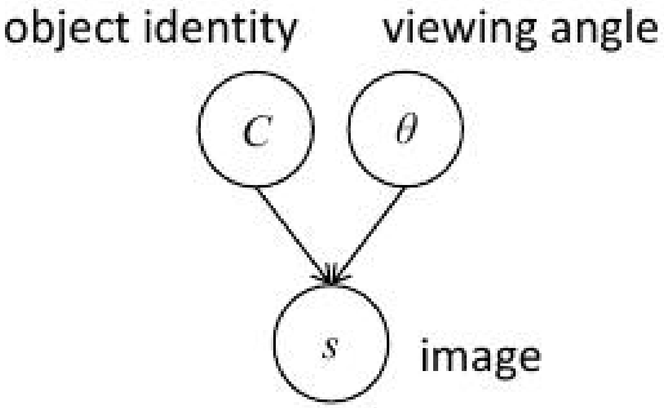
GOAL: To detect the object in the image.
Invariant: Independently of the view point we can recognize these images as bicycle
Nuisance variables: variables that we are not interested in estimating, but are necessary to deal with. They link the variables of interest with measurements

Nuisance variables

Goal: $$p(C \mid s) = \frac{p(s \mid C)\,p(C)}{p(s)}$$



Nuisance variables



Goal: $$p(C \mid s) = \frac{\boxed{p(s \mid C)}\, p(C)}{p(s)}$$

$$p(s \mid C) = \int p(s \mid C, \theta)\, p(\theta \mid C)\, d\theta = \int p(s \mid C, \theta)\, p(\theta)\, d\theta$$

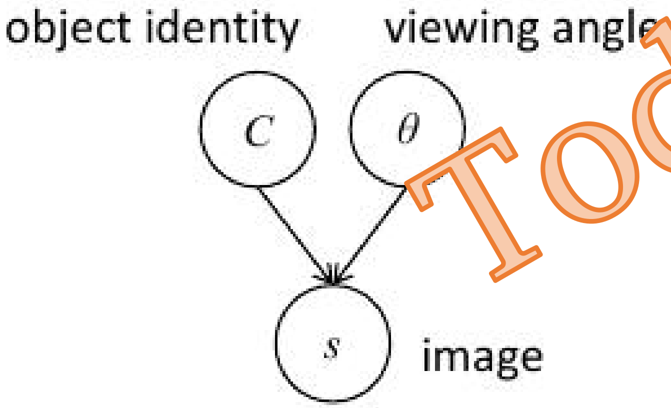
Marginalization

Independence

Nuisance variables

Goal: $p(C \mid s) = \dfrac{\boxed{p(s \mid C)}\, p(C)}{p(s)}$

Today's exercise



$$p(s \mid C) = \int p(s \mid C,\theta)\, p(\theta \mid C)\, d\theta = \int p(s \mid C,\theta)\, p(\theta)\, d\theta$$

Marginalization

Independence

Exercise 2

1. If A is the event "a person is old," and B is the event "a person suffers from Alzheimer's disease," is $p(A|B)$ less than, equal to, or greater than $p(B|A)$? Why?

Summary

1. Bayes Theorem and Bayesian modeling
2. Bayesian Brain
   1. Formulation
   2. Examples
3. Bayesian life long learning
4. Critics to the “ideal observer”

The Bayesian brain

- The environment is predictable and structured
- The sensory inputs are limited and often corrupted

The Bayesian brain

- The environment is predictable and structured
- The sensory inputs are limited and often corrupted
- Decisions should be rapid and effective to search for food, escape predators, and find mates.
- Decisions must reflect the actual nature of the environment, as it is that which determines the effect of an animal's action.

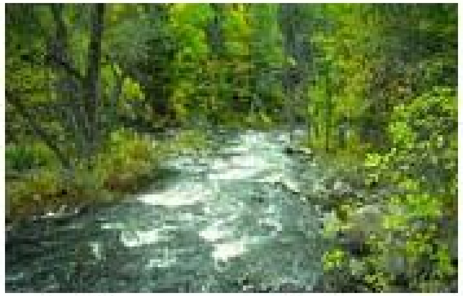
The Bayesian brain

p(my teammate is open to receive my pass | peripheral visual information)

p(a predator is lurking | visual image)

p(this book is worth reading | what I've read so far)

p(this is the one | personality, behavior, appearance)



p(I will successfully jump this stream | its width, my ability)

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

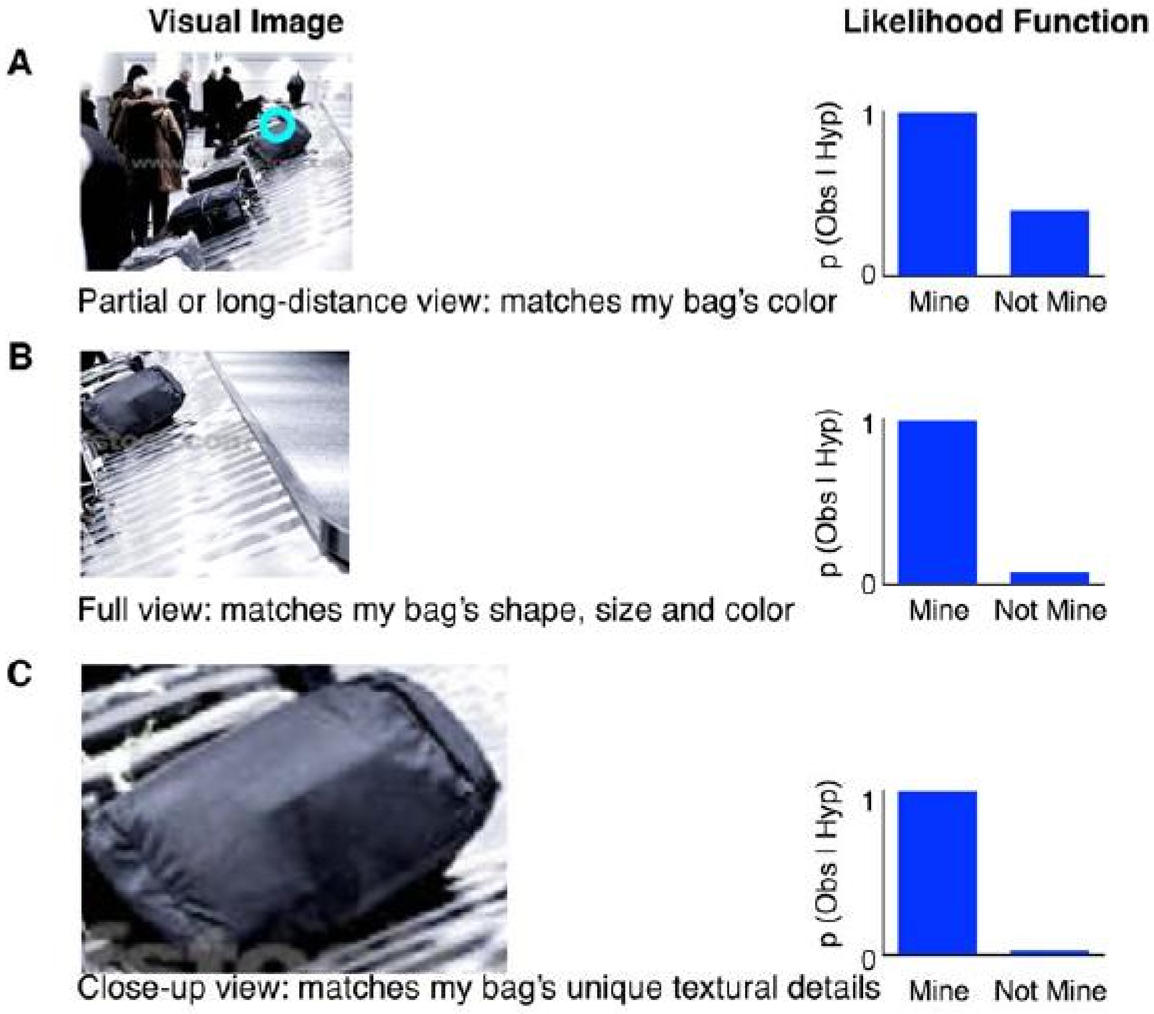
p(my father will laugh when he reads this birthday card | his sense of hum

p(diagnosis | symptoms)

Bayesian perception: an introduction; Ma, Kording and Goldreich

Factors affecting the likelihood

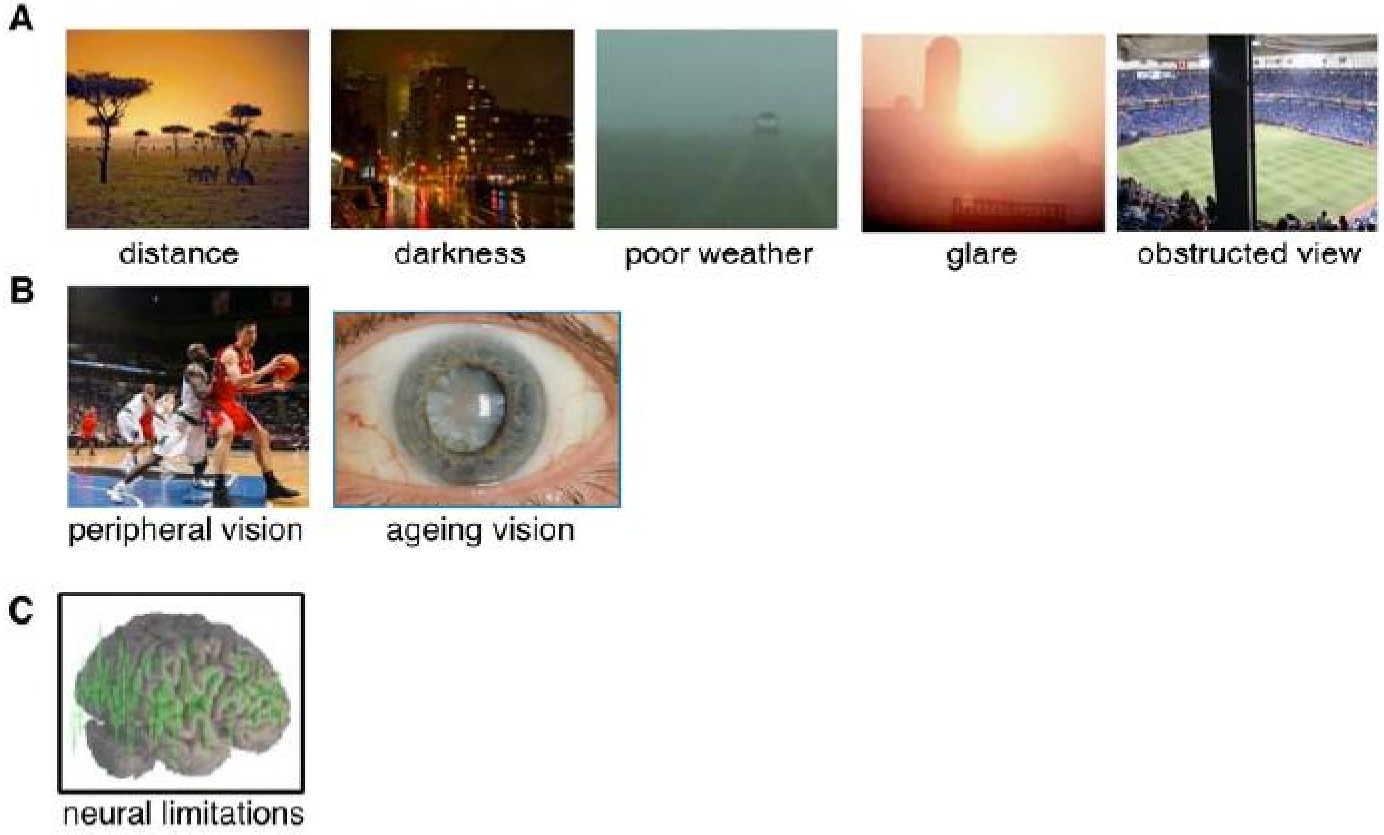
The quality of the observation:
- Environment
- Sensors
- Background knowledge



Factors affecting the likelihood

The quality of the observation:

A. Environment
B. Sensors
C. Background knowledge

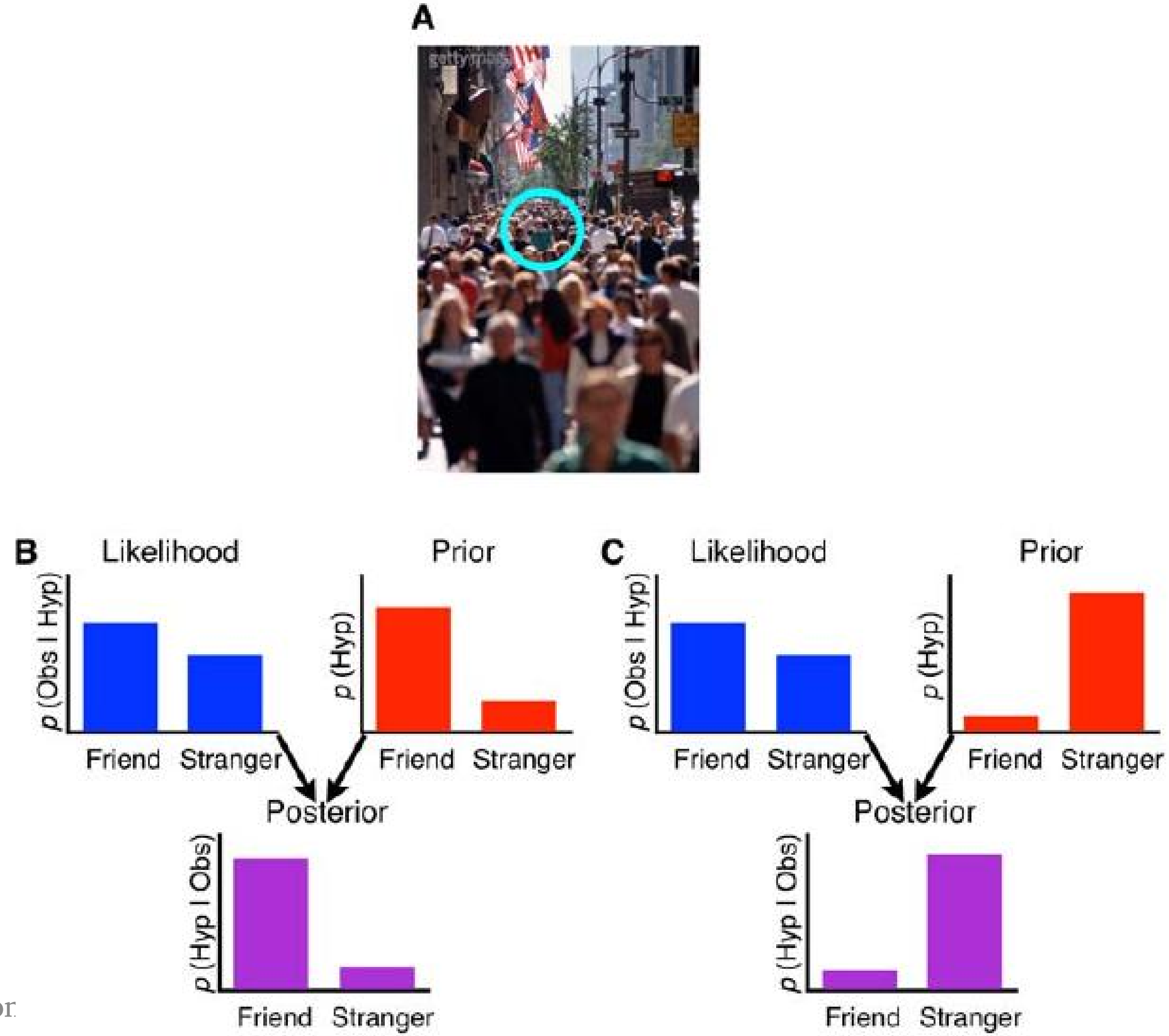


Bayesian perception: an introduction; Ma, Kording and Goldreich

Factors affecting the prior

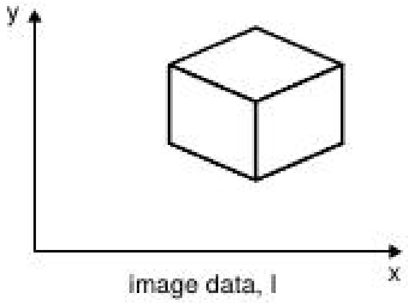
- Evolve over time at several time scales (from instants to life time)

- Depend on the observer



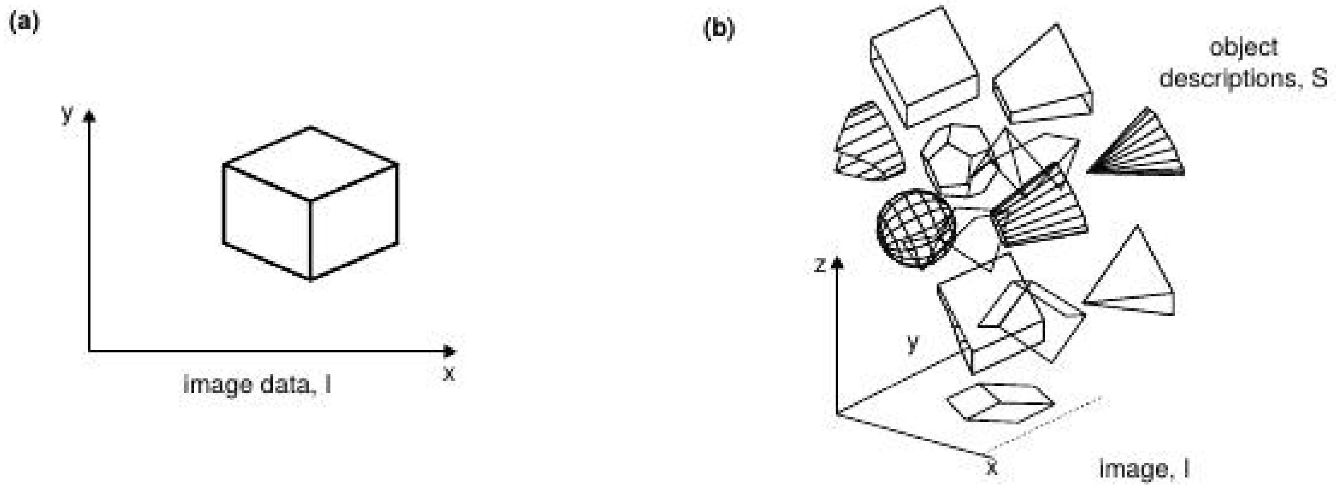
Example 1.
Object perception with priors

(a)



Kersten and Yuille, Bayesian Models of Object Perception, 2011

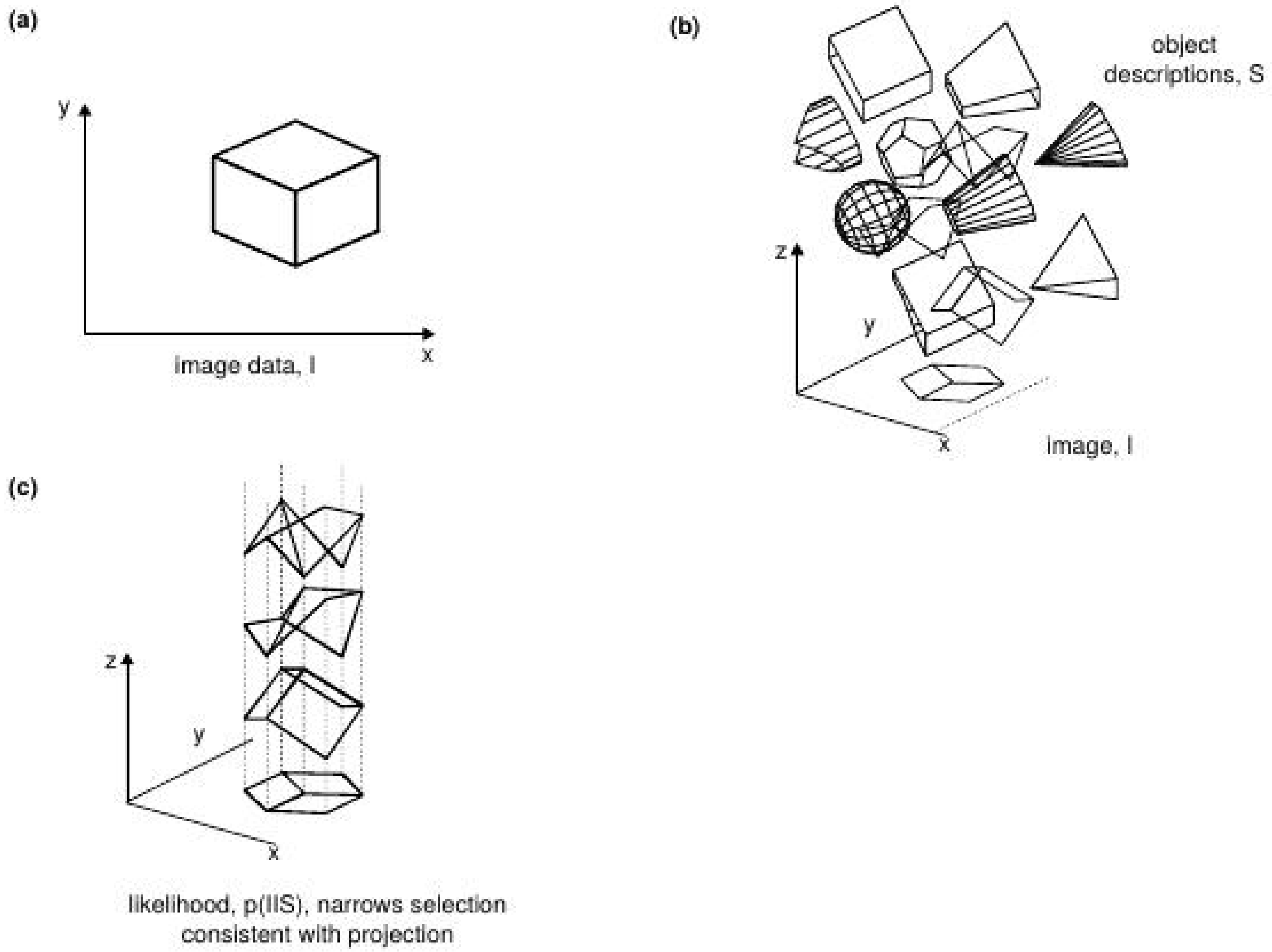
Example 1.
Object perception with priors



Kersten and Yuille, Bayesian Models of Object Perception, 2011

Example 1.
Object perception with priors

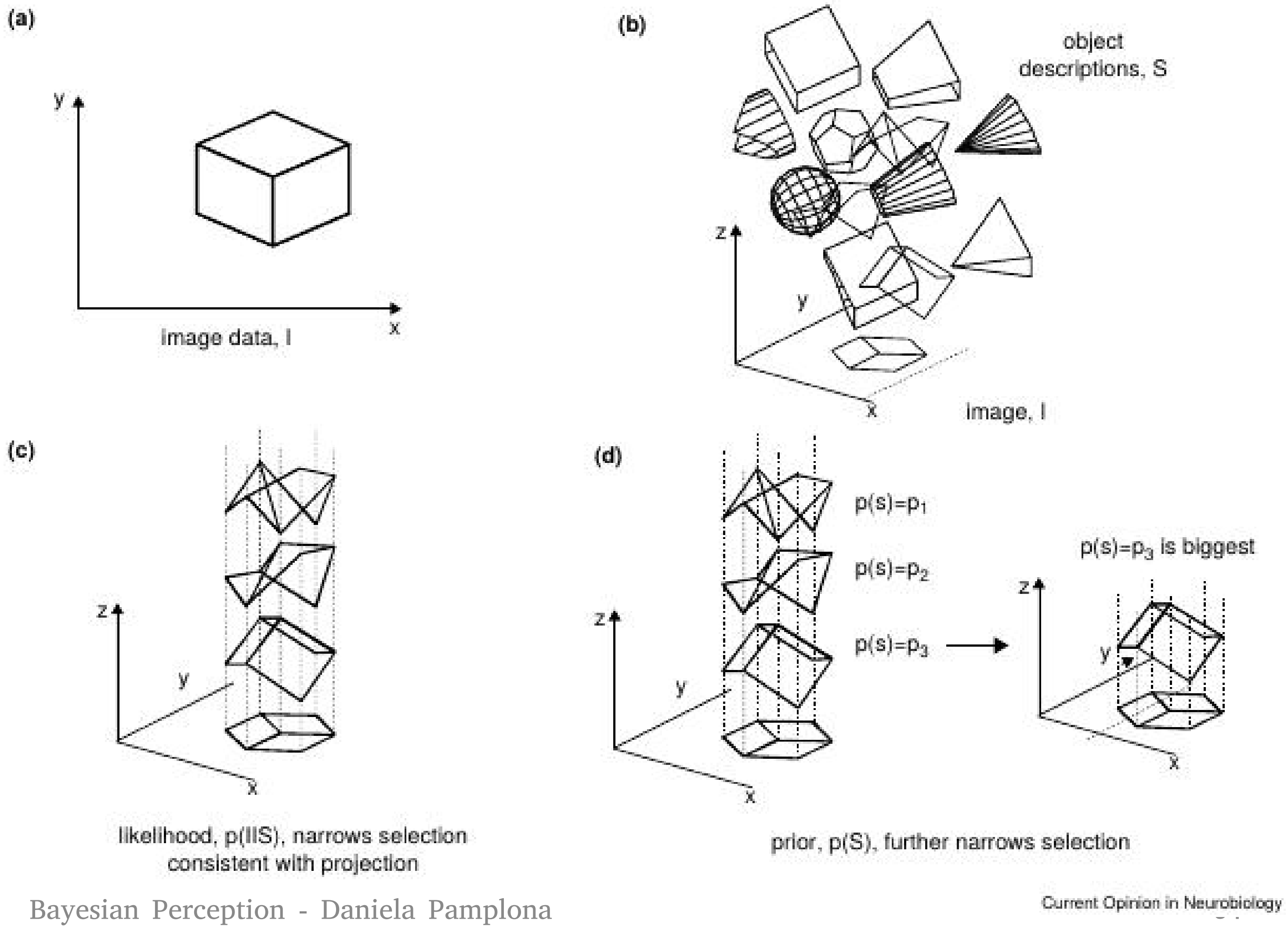
$$p(S|I) = \frac{p(I|S)p(S)}{p(I)}$$



Kersten and Yuille, Bayesian Models of Object Perception,

Example 1.
Object perception with priors

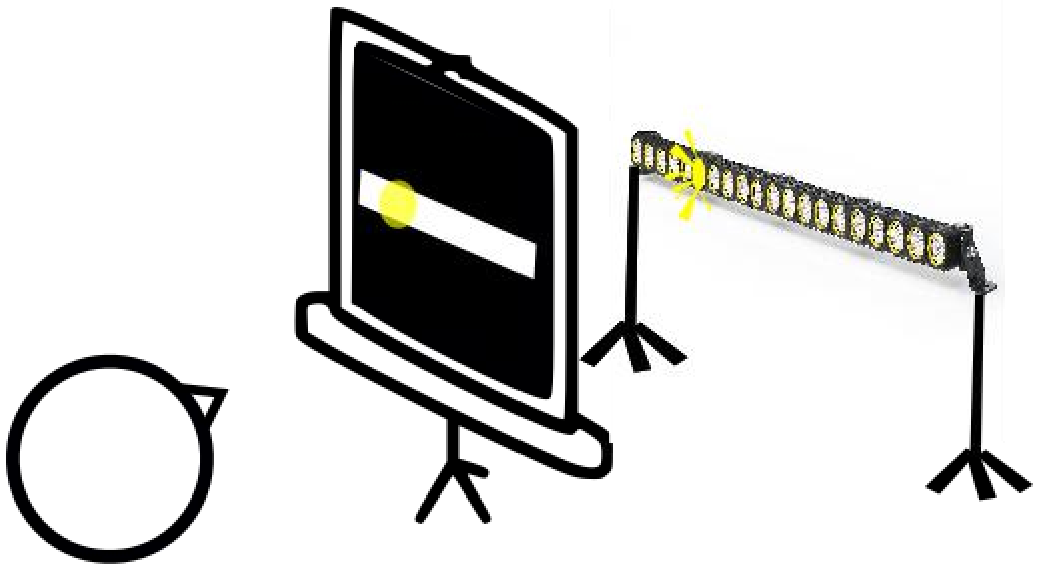
$$p(S|I) = \frac{p(I|S)p(S)}{p(I)}$$



Kersten and Yuille, Bayesian Models of Object Perception,

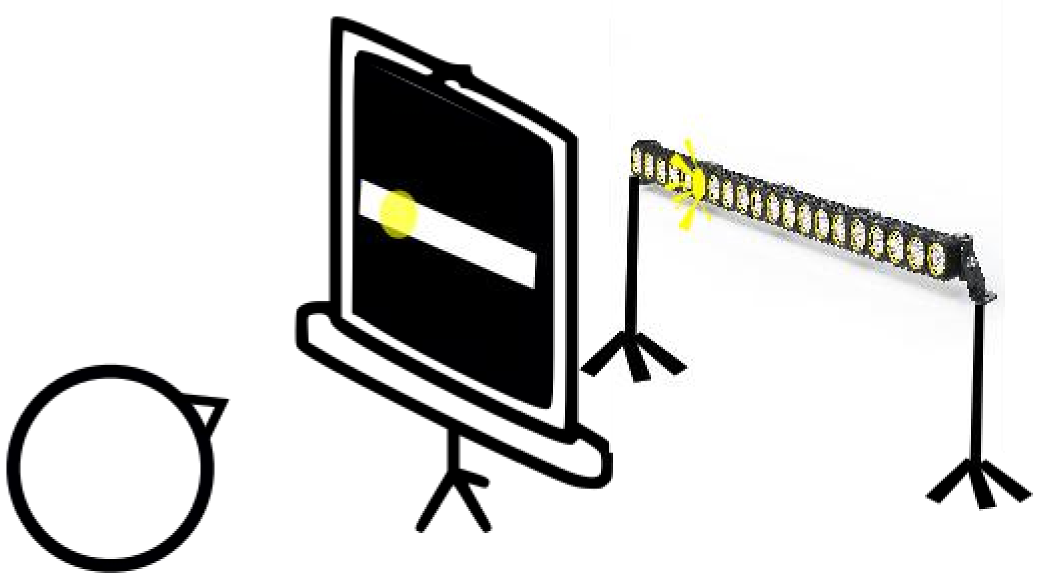
Example 2.
Visual localization

- Subjects are facing a projection screen that displays a horizontal line stretching across the width of the screen.



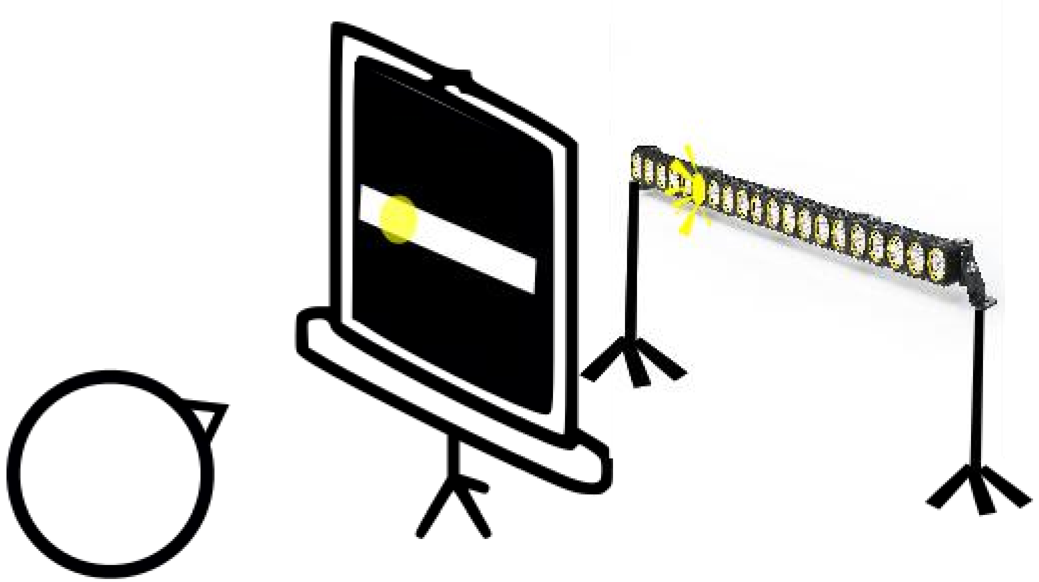
Example 2.
Visual localization

- Subjects are facing a projection screen that displays a horizontal line stretching across the width of the screen.
- Behind the screen, at the same elevation as the line, is a densely spaced array of very many tiny flashlights. A flash will originate from one of these flashlights.

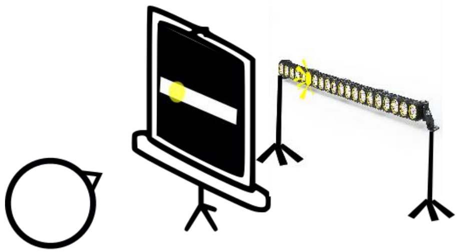


Example 2. Visual localization

- Subjects are facing a projection screen that displays a horizontal line stretching across the width of the screen.
- Behind the screen, at the same elevation as the line, is a densely spaced array of very many tiny flashlights. A flash will originate from one of these flashlights.
- The subject task is to report with a cursor the location from which you perceived the flash to emanate.



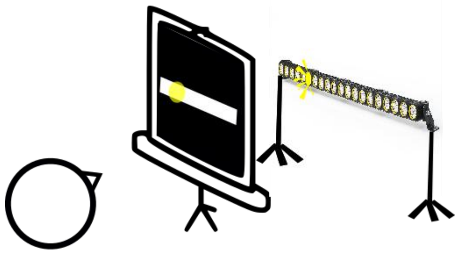
Building a Bayesian model



1. **The generative model:**

2. **The inference process:**

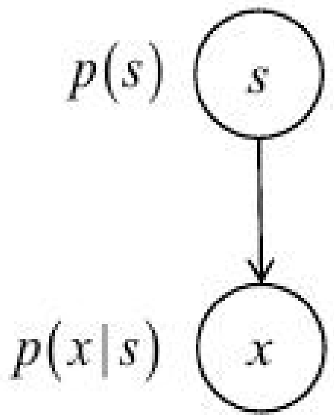
Building a Bayesian model



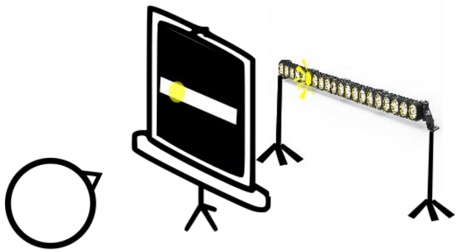
1. **The generative model:**
   - s: stimulus: flash light
   - x: observation: the interpretation that our visual system does of the stimulus
   - l: length of the white bar
   - Objective: MAP: $\text{Max}_s\ P(s|x) = \text{Max}_s\ P(x|s)\ P(s)$

   $$P(s) = G(l/2, \sigma_s); \qquad P(x|s) = G(s, \sigma_v)$$

2. **The inference process:**

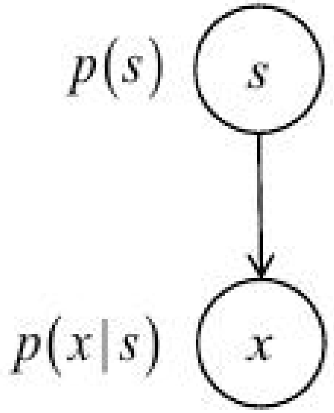


Building a Bayesian model



1. **The generative model:**
   - s: stimulus: flash light
   - x: observation: the interpretation that our visual system does of the stimulus
   - l: length of the white bar
   - Objective: MAP: $\mathrm{Max}_s\ P(s|x) = \mathrm{Max}_s\ P(x|s)\ P(s)$

$$P(s) = G(l/2, \sigma_s); \qquad P(x|s) = G(s, \sigma_v)$$

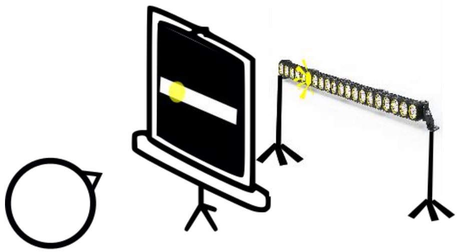


2. **The inference process:**
   - $P(x|s)\ P(s) = K\ G(\mu_F, \sigma_F)$

$$K > 0;\ \mu_F = \frac{l/2\sigma_v^2 + s\sigma_S^2}{\sigma_v^2 + \sigma_S^2};\ \sigma_F = \sqrt{\frac{\sigma_v^2 \sigma_S^2}{\sigma_v^2 + \sigma_S^2}}$$

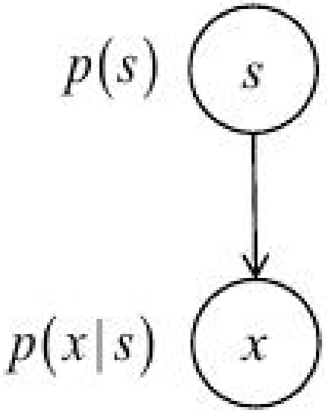
   - $\mathrm{Max}_s\ P(x|s)\ P(s) = \dfrac{l/2\sigma_v^2 + x\sigma_S^2}{\sigma_v^2 + \sigma_S^2}$

Building a Bayesian model



1. **The generative model:**
   - s: stimulus: flash light
   - x: observation: the interpretation that our visual system does of the stimulus
   - l: length of the white bar
   - Objective: MAP: $\mathrm{Max}_s\ P(s|x) = \mathrm{Max}_s\ P(x|s)\ P(s)$

$$P(s) = G(l/2, \sigma_s); \qquad P(x|s) = G(s, \sigma_v)$$



Today's exercise

2. **The inference process:**
   - $P(x|s)\ P(s) = K\ G(\mu_P, \sigma_P)$

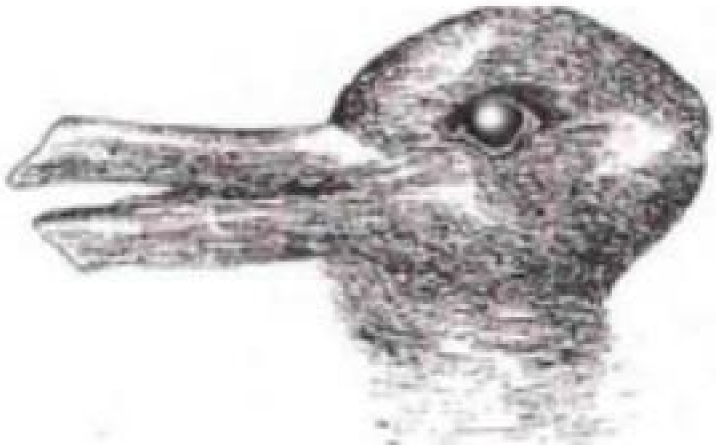
$$K > 0;\ \mu_P = \frac{l/2\sigma_v^2 + s\sigma_S^2}{\sigma_v^2 + \sigma_S^2};\ \sigma_P = \sqrt{\frac{\sigma_v^2\sigma_S^2}{\sigma_v^2 + \sigma_S^2}}$$

   - $\mathrm{Max}_s\ P(x|s)\ P(s) = \dfrac{l/2\sigma_v^2 + x\sigma_S^2}{\sigma_v^2 + \sigma_S^2}$

Exercise 2

- **2.** Can you construct a prior that would give rise to a posterior with two local maxima**?** Show mathematically that your claim is true.

Exercise 2

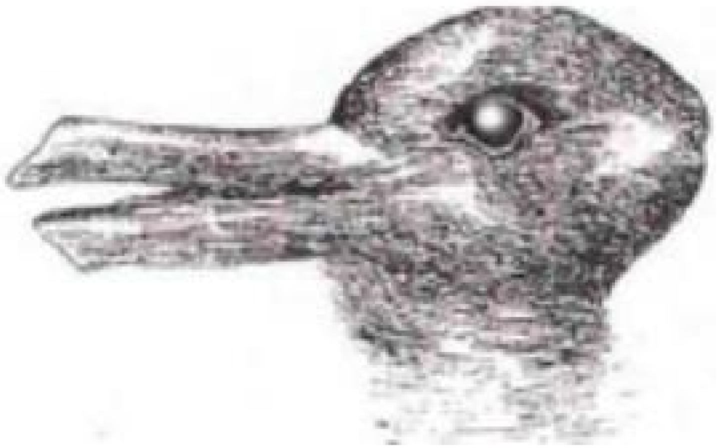


3. Easter Sunday vs a Sunday in October:

> Easter: people recognized as rabbit

> October: people recognized as duck

Exercise 2



3. Easter Sunday vs a Sunday in October:

> Easter: people recognized as rabbit
> October: people recognized as duck

Provide a Bayesian perceptual explanation for the authors' results by first defining all the relevant variables, then the generative process, and finally the graphical model of the problem.

Example 3: Cue integration: Multi-sensory example

Clue: A fact or idea that serves to reveal something or solve a problem (e.g. a crime or a puzzle).

Cue: A signal for action (like an actor entering the stage). Also, a piece of information which aids the memory in retrieving details, or indicates a desired course of action.

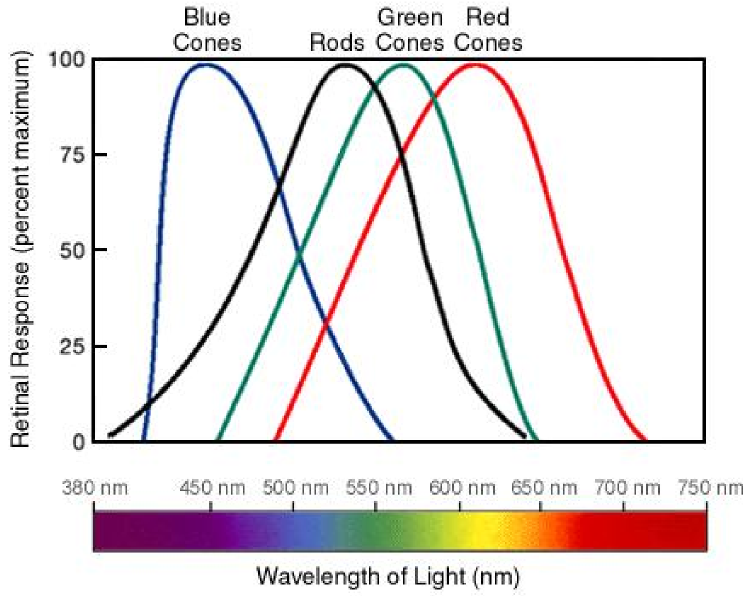
Example 3:
Cue integration: Multi-sensory example

Who is on stage?

1. auditory
2. gestures
3. speaker's facial movements

Example 3:
Cue integration: Uni-sensory example

- Different photoreceptors are tuned for different wave lengths which are combined to give the notion of color
- Two retinal images give the depth perception
- …
- All the pathways are combined to give the environment representation

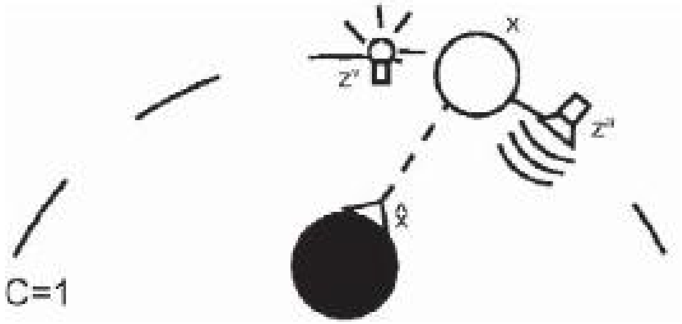


Example 3:
Cue integration: Multi-sensory example

2 cues for the same source:
visual (x_A) and auditory (x_V)

Goal:
to guess where the source is.

$$\hat{s} = w_A \, x_A + w_V \, x_V$$



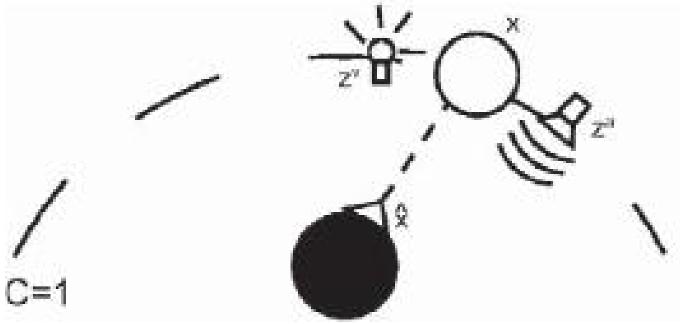
Variations on w:
laud/dark environments
deaf/blind subjects

Example 3:
Cue integration: Multi-sensory example

2 cues for the same source:
visual (x_A) and auditory (x_V)

Goal:
to guess where the source is.

$$p(x|x_{A,}\ x_V) = ??$$



Variations:
Noisy likelihoods: laud/dark environments
Flat likelihoods: deaf/blind subjects

Bayesian Cue Integration as a Developmental Outcome
of Reward Mediated Learning, Weisswange et al.

Recall: Building a Bayesian model

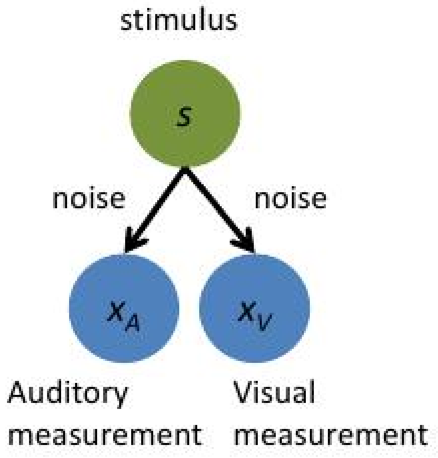
1. The generative model
2. The inference process
3. ~~The distribution of the MAP estimate~~

Example 3: Cue integration: Multi-sensory example

1. **The generative model**
2. The inference process
3. ~~The distribution of the MAP estimate~~

$$p(x_A \mid s) = \frac{1}{\sqrt{2\pi\sigma_A^2}} e^{-\frac{(x_A - s)^2}{2\sigma_A^2}}$$

$$p(x_V \mid s) = \frac{1}{\sqrt{2\pi\sigma_V^2}} e^{-\frac{(x_V - s)^2}{2\sigma_V^2}}$$



Conditional independence

$$p(x_A, x_V \mid s) = p(x_A \mid s)\, p(x_V \mid s)$$

flat prior p(s)

Example 3:
Cue integration: Multi-sensory example

1. The generative model
2. **The inference process**
3. ~~The distribution of the MAP estimate~~

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

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

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

$$p(s \mid x_A, x_V) = \frac{1}{\sqrt{2\pi\sigma^2_{\text{combined}}}} e^{-\frac{\left(s-\mu_{\text{combined}}\right)^2}{2\sigma^2_{\text{combined}}}} \quad \mu_{\text{combined}} = \frac{\frac{x_A}{\sigma_A^2} + \frac{x_V}{\sigma_V^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2}} \quad \sigma^2_{\text{combined}} = \frac{1}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2}}$$

Example 3:
Cue integration: Multi-sensory example

1. The generative model
2. **The inference process**
3. ~~The distribution of the MAP estimate~~

$$\hat{s}_{\text{MAP}} = \frac{\dfrac{x_A}{\sigma_A^2} + \dfrac{x_V}{\sigma_V^2}}{\dfrac{1}{\sigma_A^2} + \dfrac{1}{\sigma_V^2}}$$

$$\hat{s}_{\text{MAP}} = w_A x_A + w_V x_V$$

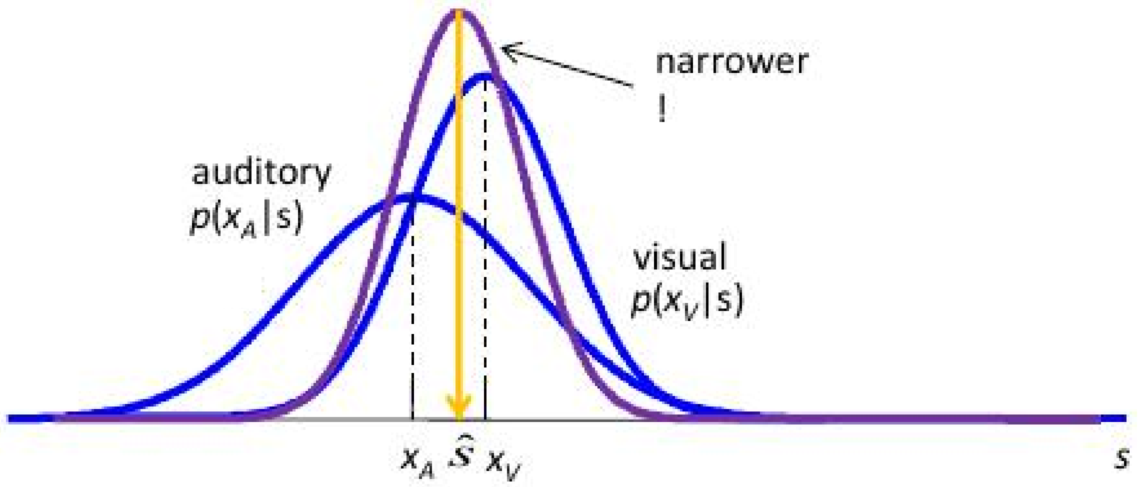
$$w_A = \frac{\dfrac{1}{\sigma_A^2}}{\dfrac{1}{\sigma_A^2} + \dfrac{1}{\sigma_V^2}}, \quad w_V = \frac{\dfrac{1}{\sigma_V^2}}{\dfrac{1}{\sigma_A^2} + \dfrac{1}{\sigma_V^2}}$$

Example 3:
Cue integration: Multi-sensory example

1. The generative model
2. **The inference process**
3. ~~The distribution of the MAP estimate~~

$$p(s \mid x_A, x_V) = \frac{1}{\sqrt{2\pi\sigma^2_{\text{combined}}}} e^{-\frac{(s-\mu_{\text{combined}})^2}{2\sigma^2_{\text{combined}}}}$$

$$\mu_{\text{combined}} = \frac{\frac{x_A}{\sigma_A^2} + \frac{x_V}{\sigma_V^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2}} \qquad \sigma^2_{\text{combined}} = \frac{1}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2}}$$



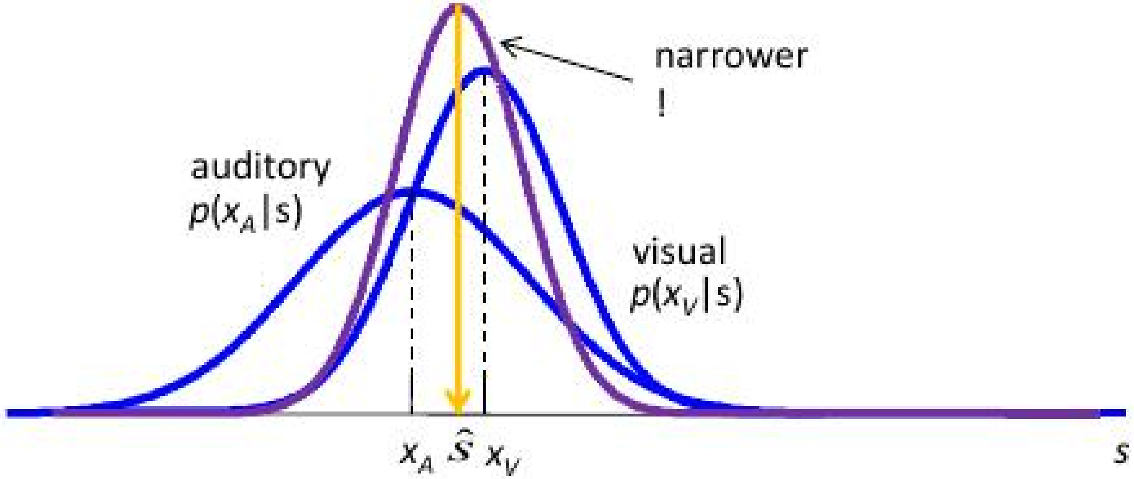
Example 3: Cue integration: Multi-sensory example

1. The generative model
2. **The inference process**
3. ~~The distribution of the MAP estimate~~

Today's exercise

$$p(s \mid x_A, x_V) = \frac{1}{\sqrt{2\pi\sigma^2_{\text{combined}}}} e^{-\frac{(s-\mu_{\text{combined}})^2}{2\sigma^2_{\text{combined}}}}$$

$$\mu_{\text{combined}} = \frac{\frac{x_A}{\sigma_A^2} + \frac{x_V}{\sigma_V^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2}} \qquad \sigma_{\text{combined}} = \frac{1}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2}}$$

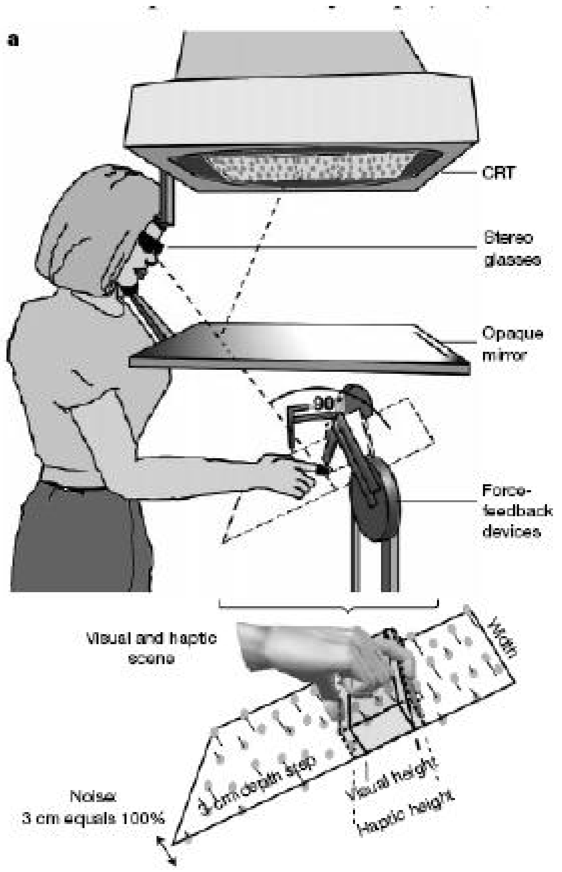


Exercise 2

4.Recall the problem of integration of two Gaussian cues (auditory and visual) described in the class (section 2, slide 42). Starting from the generative model described in the class, derive mathematically the μcombined and σcombined estimators for the posterior distribution of the integrated cues.

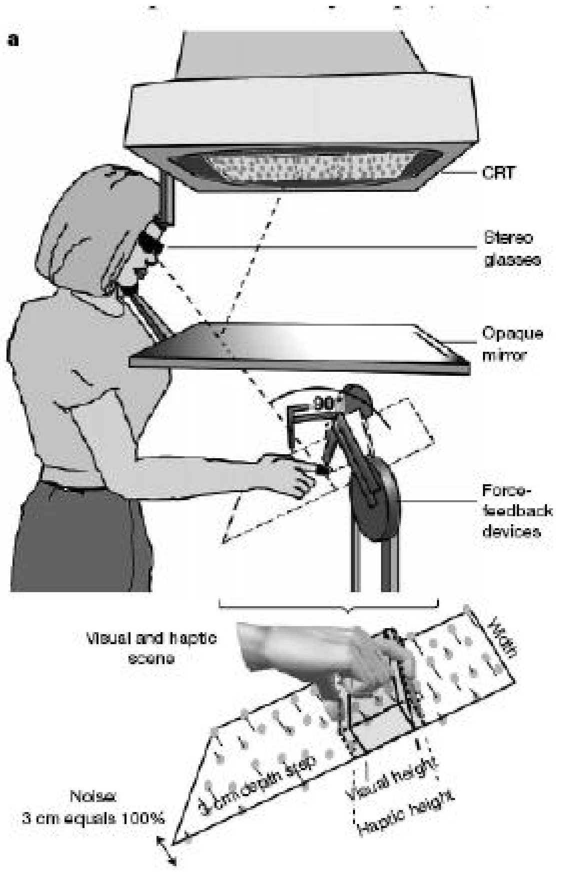
$$p(s \mid x_A, x_V) = \frac{1}{\sqrt{2\pi\sigma^2_{\text{combined}}}} e^{-\frac{(s-\mu_{\text{combined}})^2}{2\sigma^2_{\text{combined}}}} \qquad \mu_{\text{combined}} = \frac{\frac{x_A}{\sigma_A^2} + \frac{x_V}{\sigma_V^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2}} \qquad \sigma^2_{\text{combined}} = \frac{1}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2}}$$

Example 2:
Cue integration: Multisensory example



Ernst M, Banks. M. Humans integrate visual and hapticinformation in a statisticallyoptimal fashion, 2002

Example 2:
Cue integration: Multisensory example



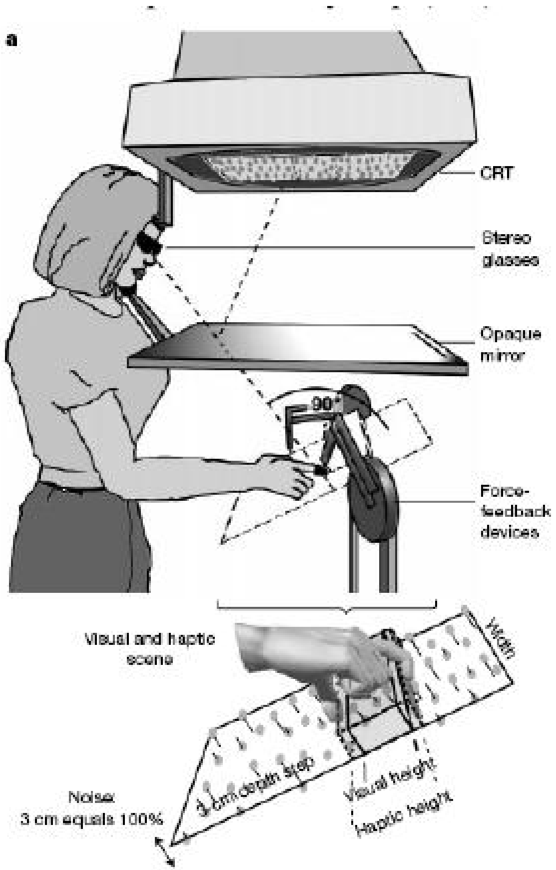
Gaussian prior on visual and haptics observation

Parameters fitted independently

$$\hat{S} = \sum_i w_i \hat{S}_i \qquad w_i = \frac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2}$$

Ernst M, Banks. M. Humans integrate visual and hapticinformation in a statisticallyoptimal fashion, 2002

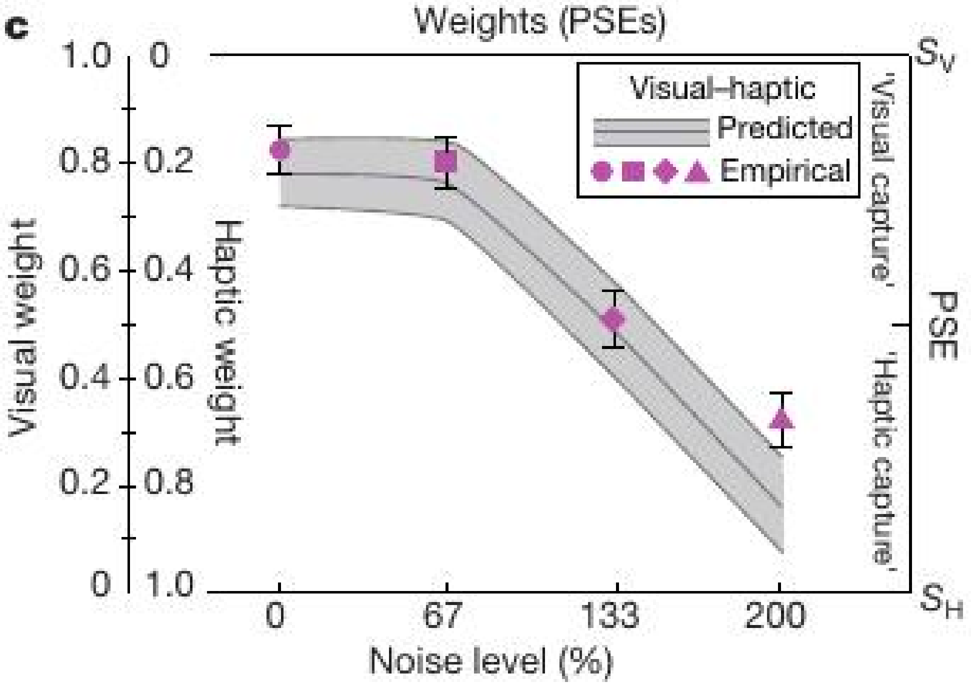
Example 2:
Cue integration: Multisensory example



Gaussian prior on visual and haptics observation

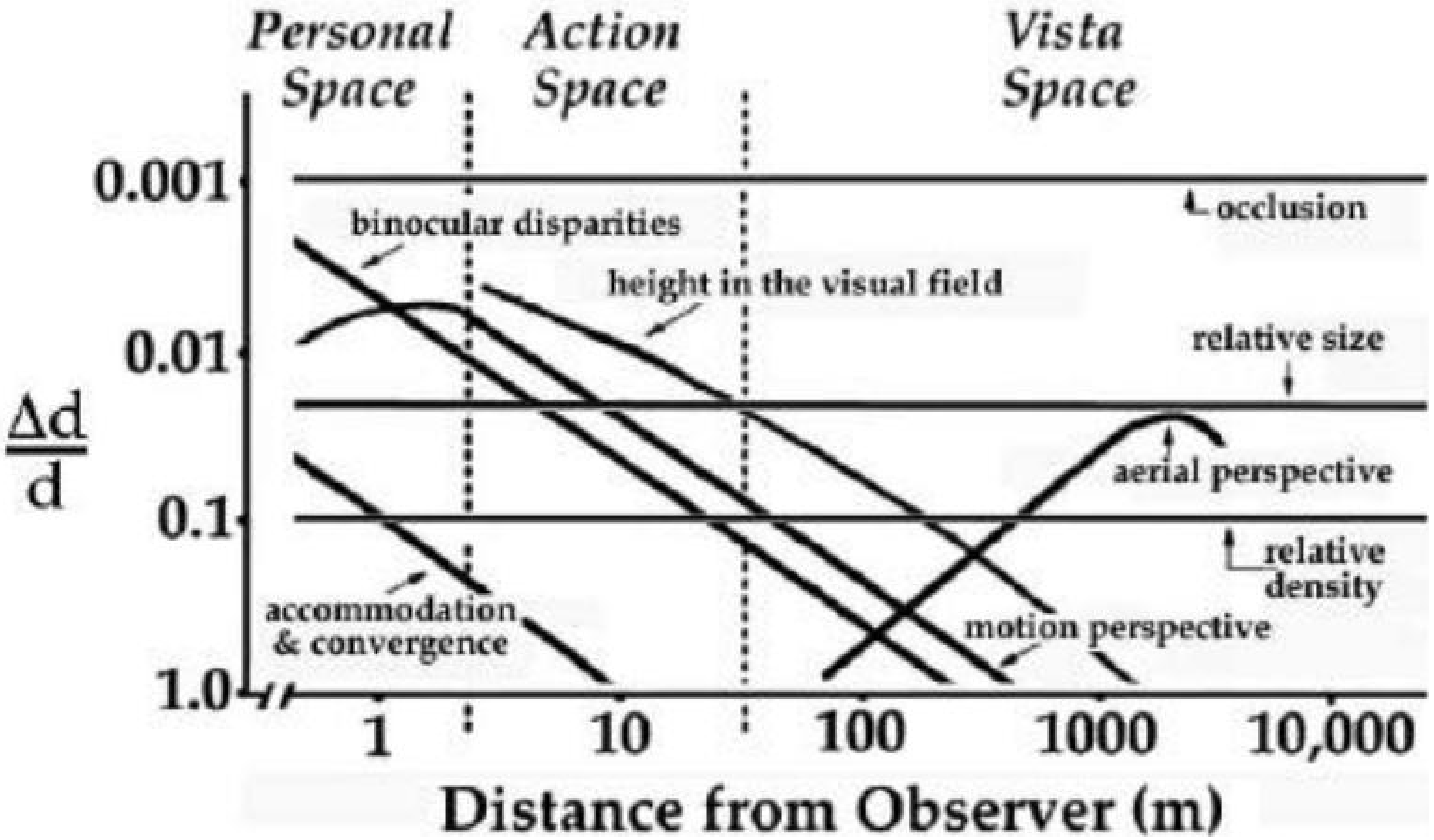
Parameters fitted independently

$$\hat{S} = \sum_i w_i \hat{S}_i \qquad w_i = \frac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2}$$



Ernst M, Banks. M. Humans integrate visual and hapticinformation in a statisticallyoptimal fashion, 2002

Example 4:
Cue integration: Visual cues for depth estimation



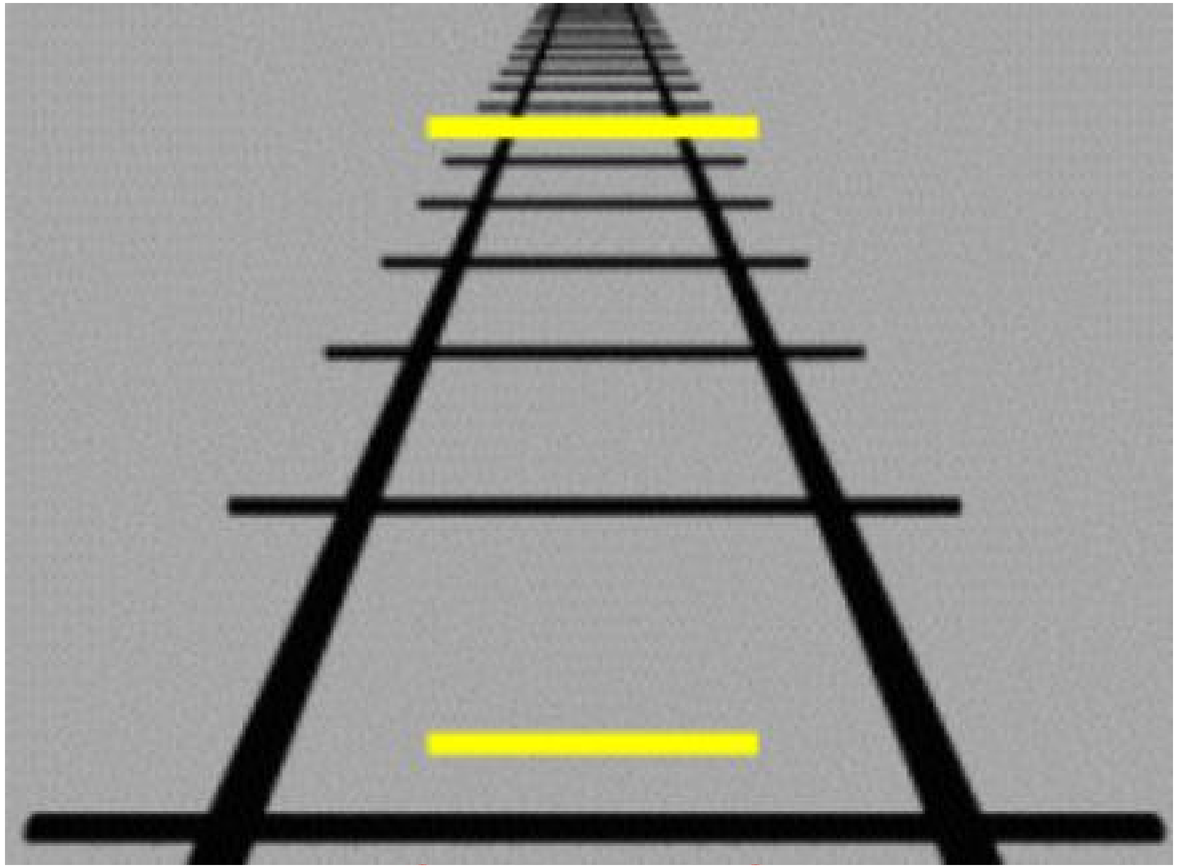
Cutting, J.E. and Vishton, P.M. (1995)

Example 5: Illusions

1. Misleading

2. Bi-stable

Example 5.1.1: Illusions: misleading

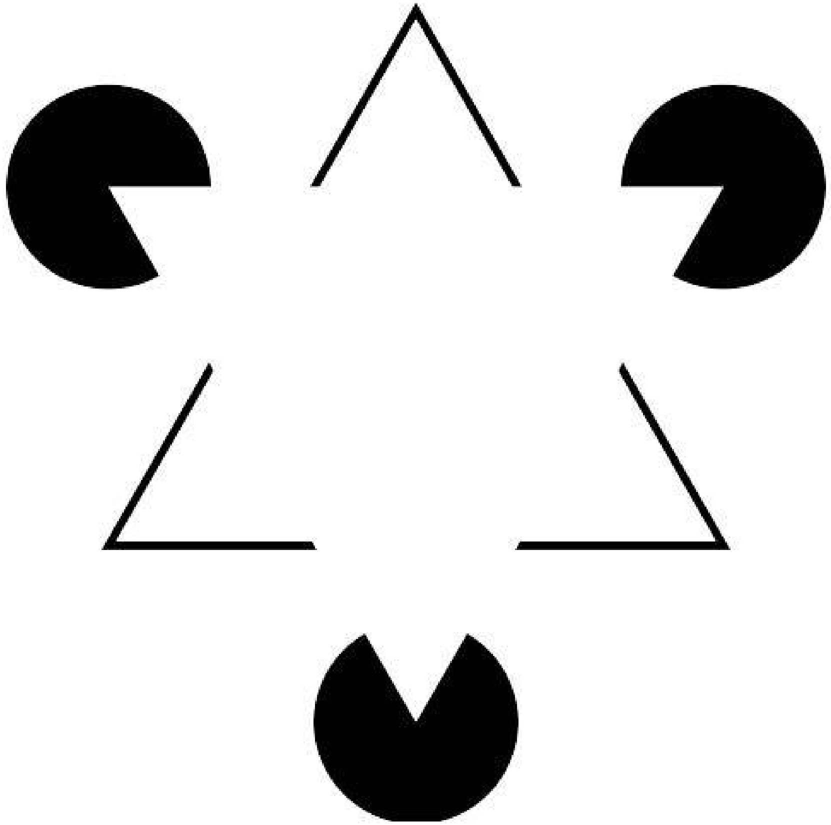


https://www.verywellmind.com/cool-optical-illusions-2795841

Example 5.1
Illusions: misleading

https://www.verywellmind.com/cool-optical-illusions-2795841

Example 5.1.2: Illusions: misleading



https://www.verywellmind.com/cool-optical-illusions-2795841

Example 5.1.2: Illusions: misleading

More on Gestalt laws?
https://www.verywellmind.com/gestalt-laws-of-perceptual-organization-2795835
http://helios.mi.parisdescartes.fr/~moisan/papers/2006-9.pdf

https://www.verywellmind.com/cool-optical-illusions-2795841

Example 5.1.3:
(visual-motor) Illusions: misleading

- https://www.youtube.com/watch?v=pjdVLRyTrZw

Example 5.1.4: Illusions: Rhombus moving

- see on flash content debugger the rhombus illusion

- "Fat" Rhombus moves horizontally, independent of contrast
- "Thin" Rhombus moves horizontally on high contrast and obliquely on low contrast

Weiss et al, 2002

Recall: Building a Bayesian model

1. The generative model
2. The inference process
3. ~~The distribution of the MAP estimate~~

Example 5.1.4: Illusions: Rhombus moving

The generative model: [Assumptions]

- Pixels move but do not change their intensity over time.

$$I(x,y,t) = I(x + v_x\Delta t,\, y + v_y\Delta t,\, t + \Delta t)$$

- Observations are noisy, Gaussian white noise, with variance **σ**

$$I(x,y,t) = I(x + v_x\Delta t,\, y + v_y\Delta t,\, t + \Delta t) + \eta$$

Gaussian noise

Weiss et al, 2002

Example 5.1.4: Illusions: Rhombus moving

The generative model: [Likelihood]

- Velocity is constant in a small window around each pixel
- Intensity surface $I(x_i, y_i, t)$ is sufficiently smooth so it can be approximated by a linear function for small temporal duration (substitution by Taylor series).

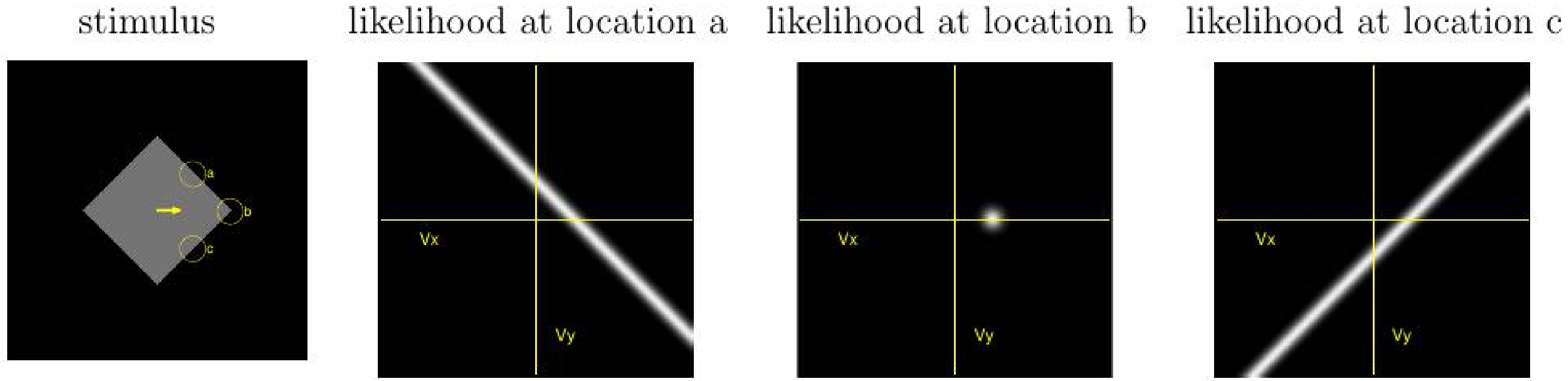
$$P(I(x_i, y_i, t)|v_i) \propto$$

$$\exp\left(-\frac{1}{2\sigma^2}\int_{x,y} \underbrace{w_i(x,y)}_{\text{window}}(\underbrace{I_x(x,y,t)}_{\text{Derivative}}v_x + I_y(x,y,t)\underbrace{v_y}_{\text{Velocity}} + I_t(x,y,t))^2\, dx\, dy\right)$$

Weiss et al, 2002

Example 5.1.4: Illusions: Rhombus moving

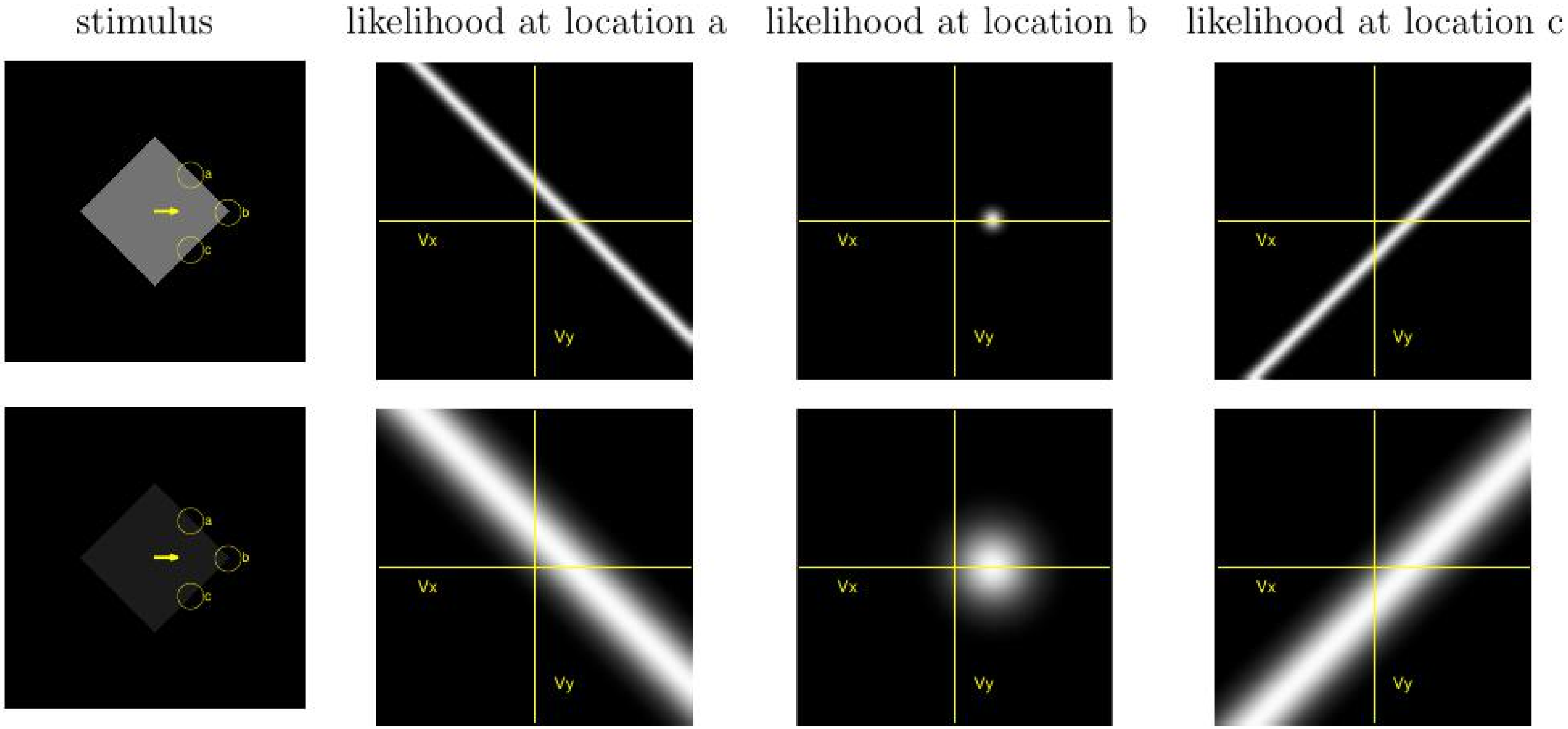
The generative model: [Likelihood] $P(I(x_i, y_i, t) | v_i)$



Likelihood orientation depends on point features

Example 5.1.4: Illusions: Rhombus moving

The generative model: [Likelihood] $P(I(x_i, y_i, t) | v_i)$



Likelihood peakness depends on contrast

Weiss et al, 2002

Example 5.1.4: Illusions: Rhombus moving

The generative model: [Prior distribution]

•prior favoring slow speeds

$$P(v) \propto \exp\left(-\|v\|^2/2\sigma_p^2\right)$$

Example 5.1.4: Illusions: Rhombus moving

The inference process : [Posterior distribution]

- noise is independent over spatial location

$$P(v|I) \propto \exp\left(-\|v\|^2/2\sigma_p^2 - \frac{1}{2\sigma^2}\int_{x,y}\sum_i w_i(x,y)(I_x(x,y)v_x + I_y(x,y)v_y + I_t)^2 dxdy\right)$$

- the entire image moves according to a single translational velocity, and so sum over all spatial positions so $w_i(x, y)$ is a constant

$$P(v|I) \propto \exp\left(-\|v\|^2/2\sigma_p^2 - \frac{1}{2\sigma^2}\int_{x,y}(I_x(x,y)v_x + I_y(x,y)v_y + I_t)^2 dxdy\right)$$

Weiss et al, 2002

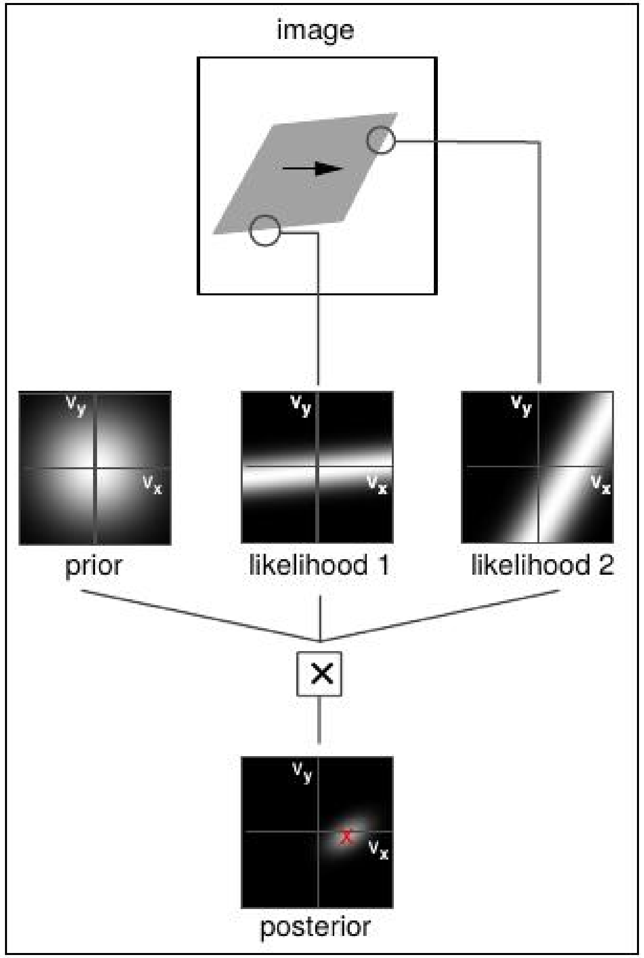
Example 5.1.4: Illusions: Rhombus moving

The inference process: [MAP estimator]

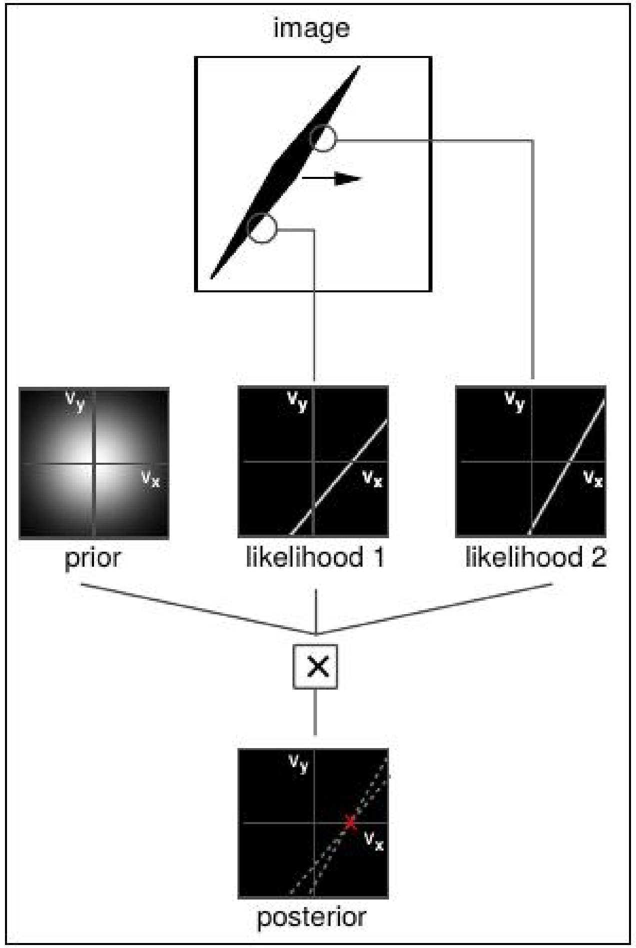
$$v^* = -\begin{pmatrix} \sum I_x^2 + \frac{\sigma^2}{\sigma_p^2} & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 + \frac{\sigma^2}{\sigma_p^2} \end{pmatrix}^{-1} \begin{pmatrix} \sum I_x I_t \\ \sum I_y I_t \end{pmatrix}$$

Weiss et al, 2002

Example 5.1.4: Illusions: Rhombus moving

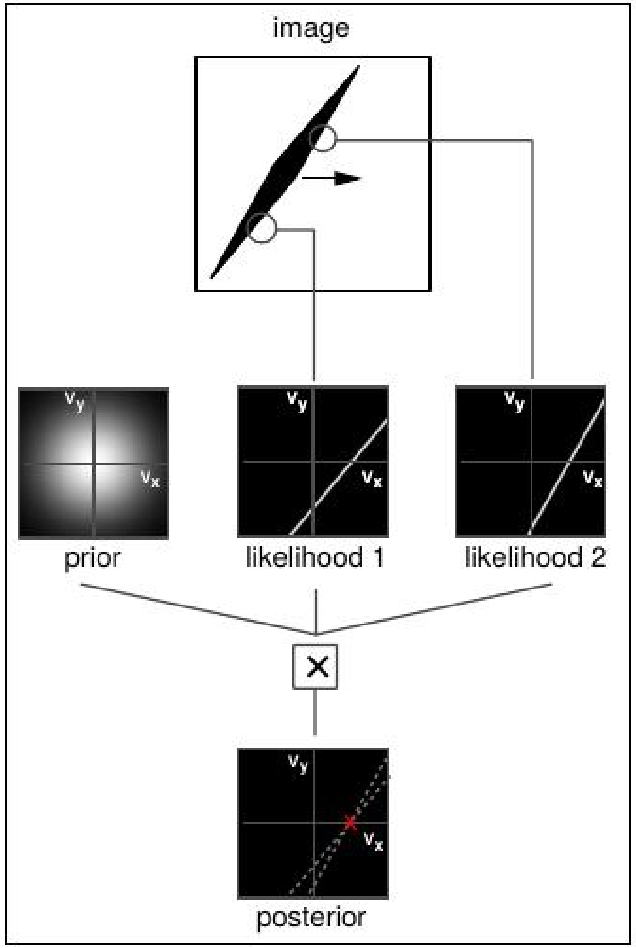


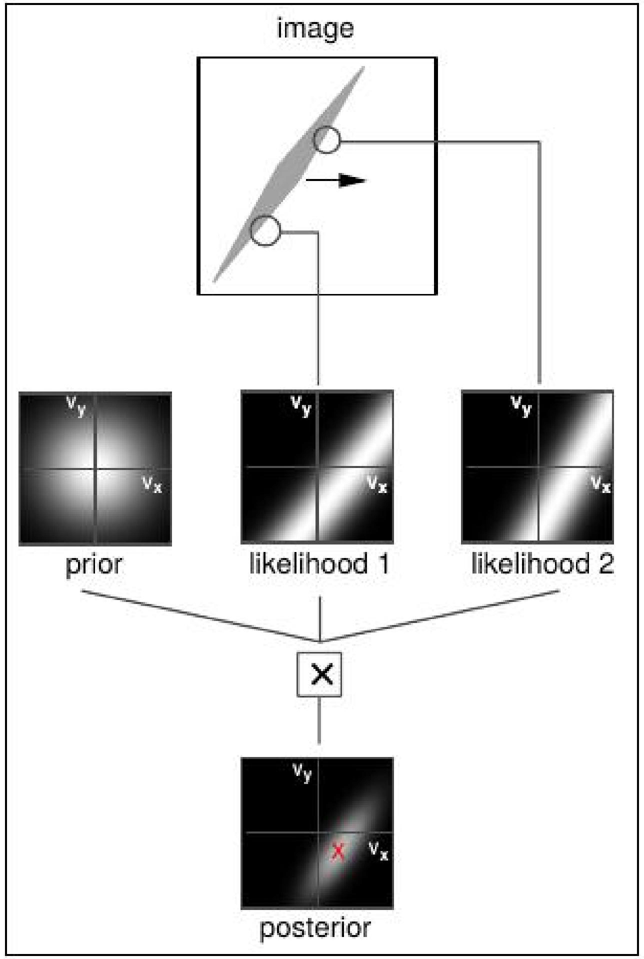
Example 5.1.4: Illusions: Rhombus moving



Weiss et al, 2002

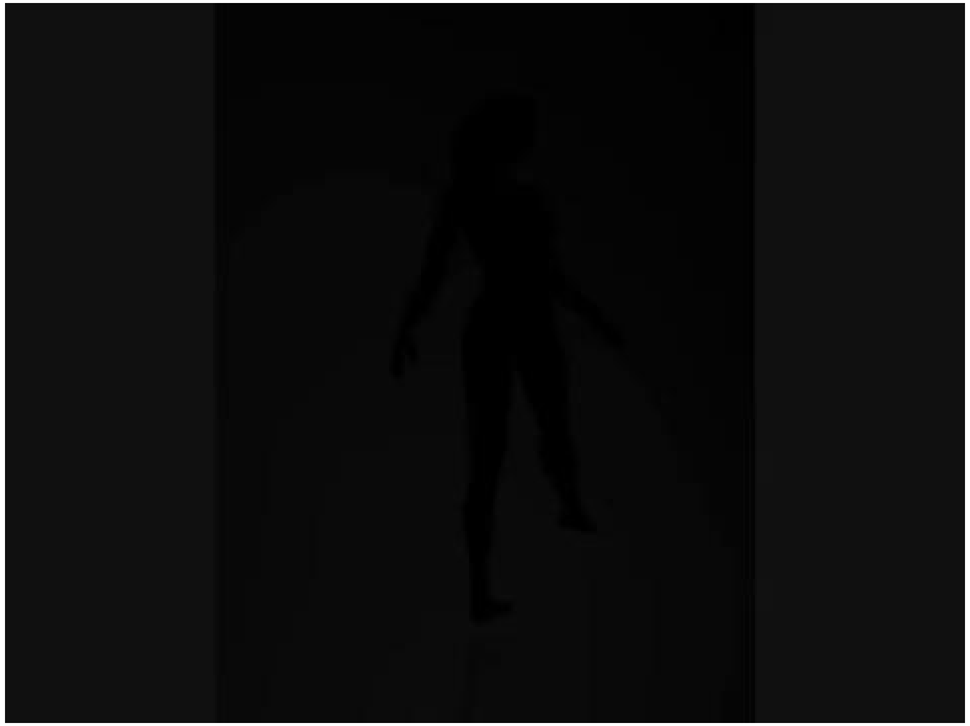
Example 5.1.4: Illusions: Rhombus moving





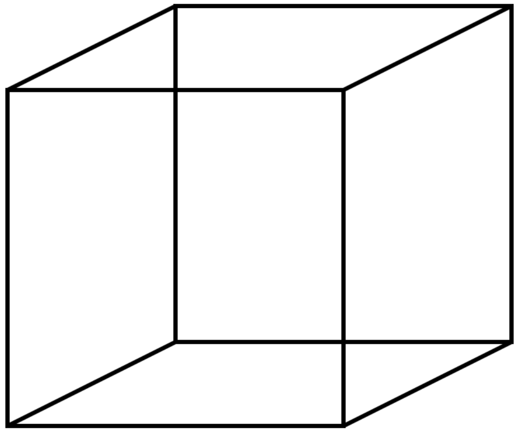
Weiss et al, 2002

Example 5.2.1: Illusions: Spinning ballerina



https://www.verywellmind.com/cool-optical-illusions-2795841

Example 5.2.2: Illusions: Necker cube

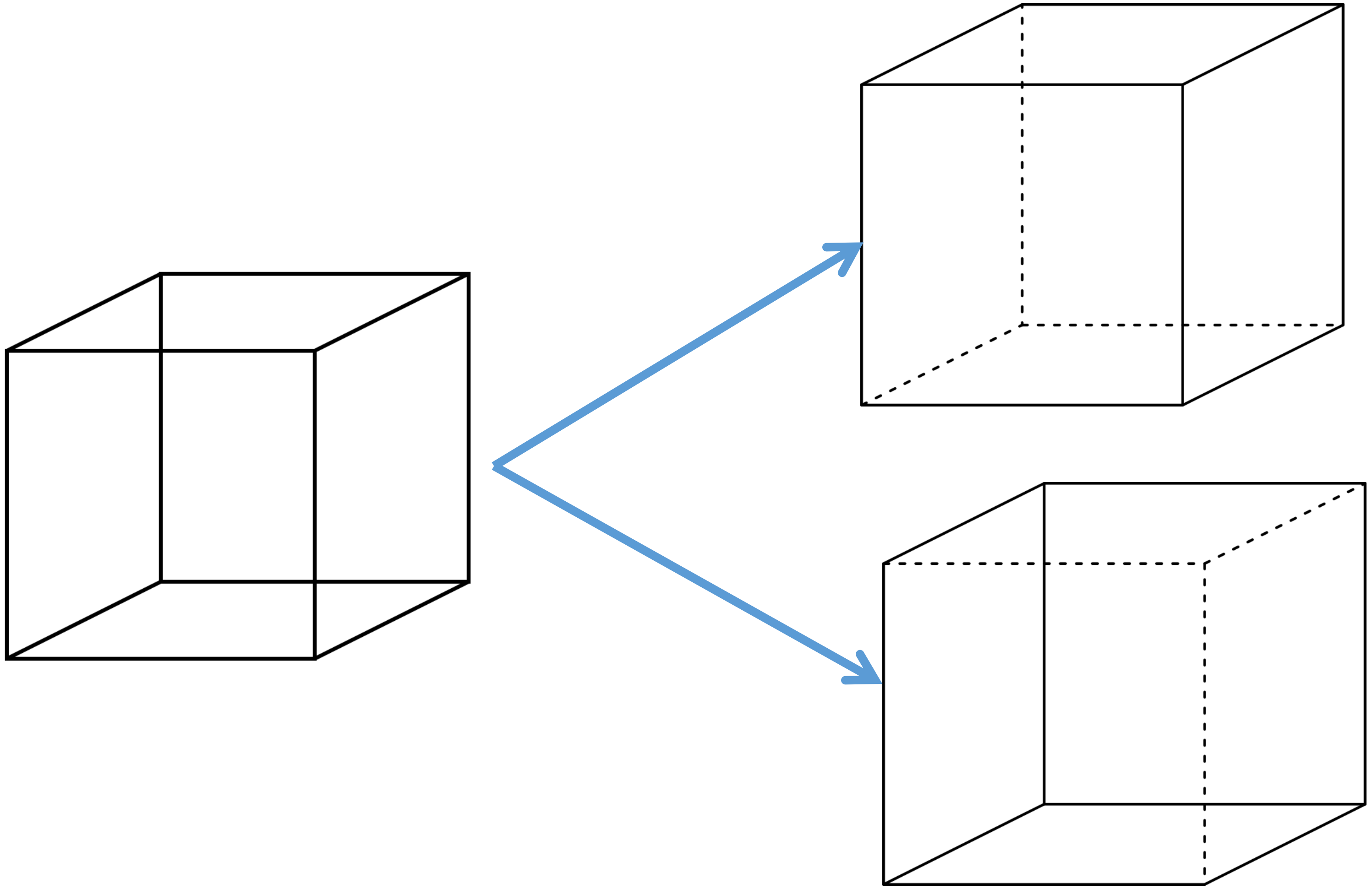


VA (the subject is Above the cube)

VB (the subject is Below the cube)

https://en.wikipedia.org/wiki/Necker_cube

Example 5.2.2: Illusions: Necker cube

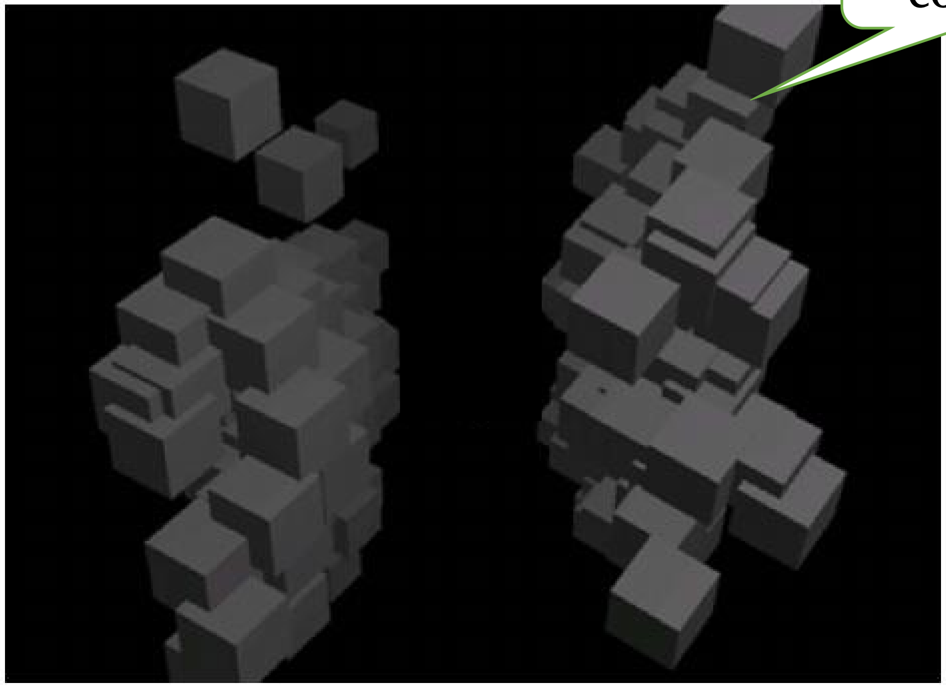


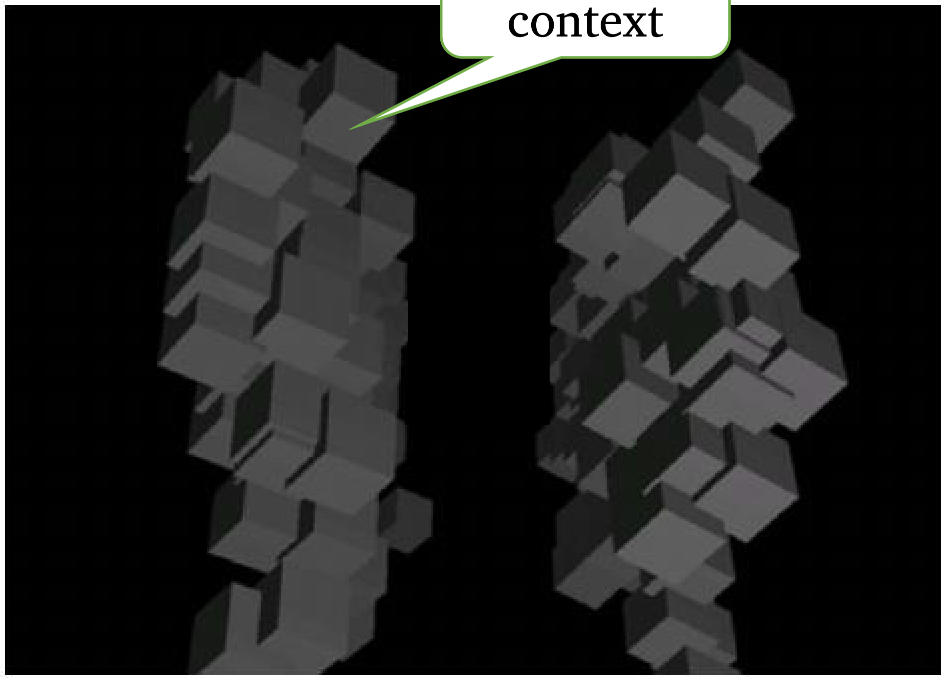
VA (the subject is Above the cube)

VB (the subject is Below the cube)

https://en.wikipedia.org/wiki/Necker_cube

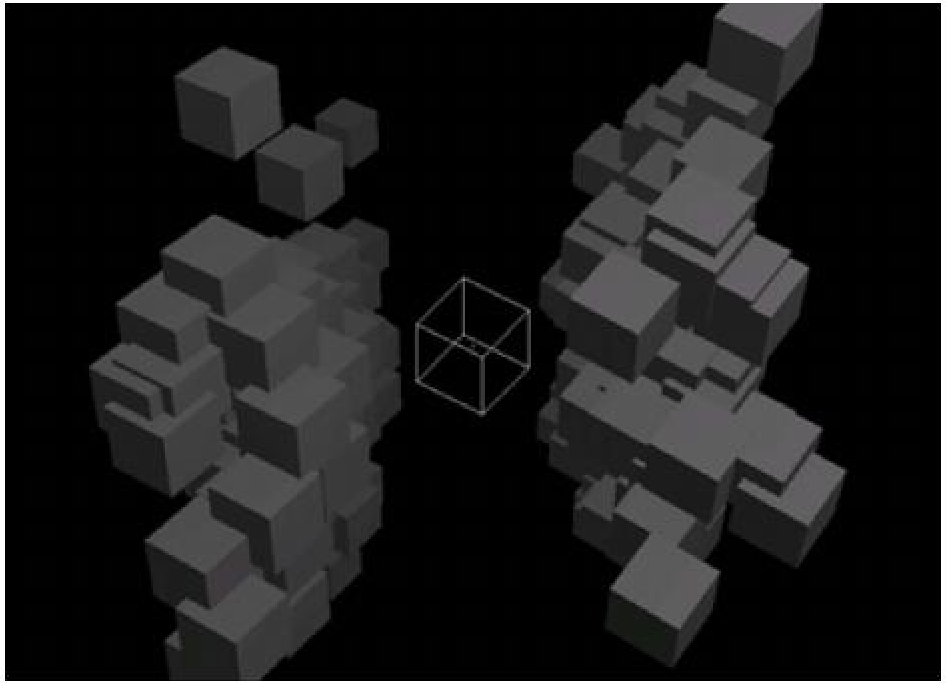
Example 5.2.2:
Illusions: Contextual necker cube

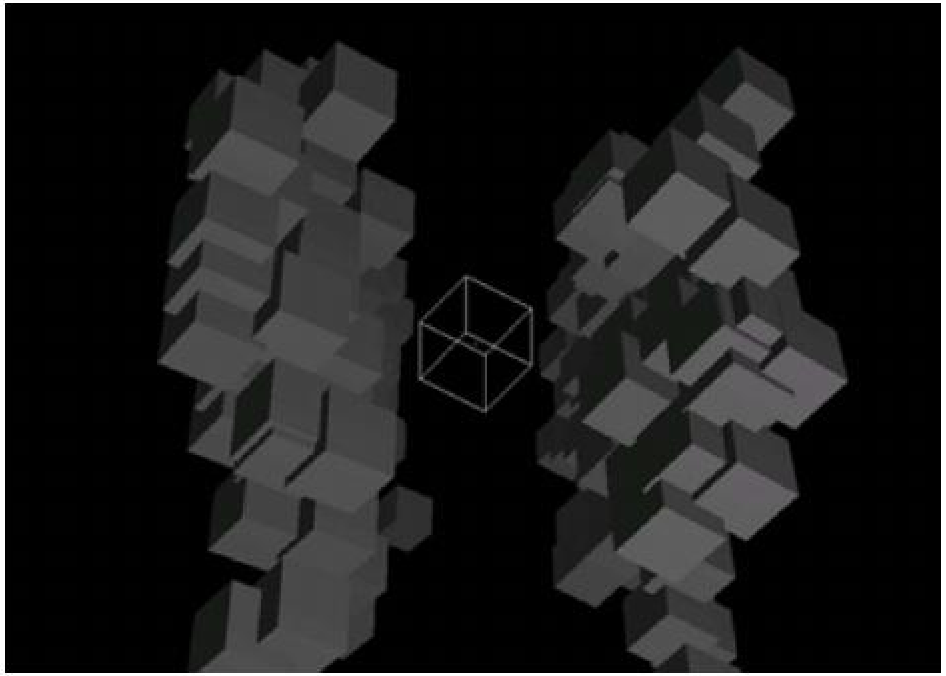




Perceptual multi-stability predicted by search model
for Bayesian decisions, Sundareswara, Schrater,2008

Example 5.2.2: Illusions: Contextual necker cube

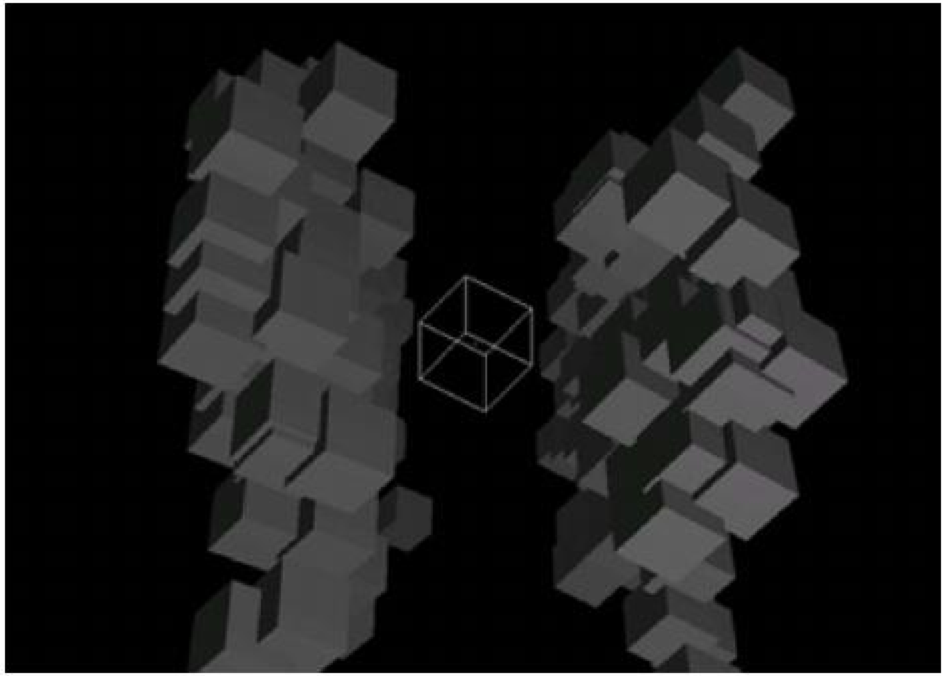




Perceptual multi-stability predicted by search model for Bayesian decisions, Sundareswara, Schrater,2008

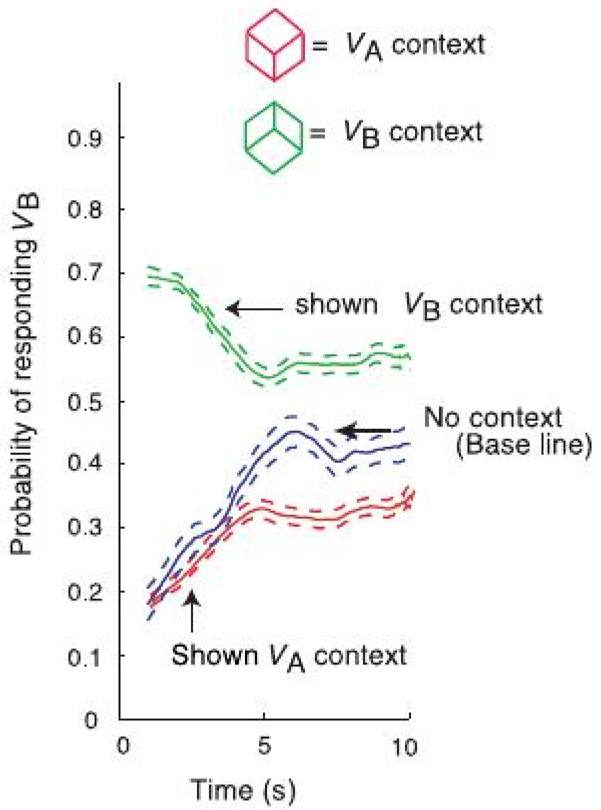
Example 5.2.2: Illusions: Contextual necker cube

Subjects seeing a contextual necker cube in a screen are asked to say which of the view is stable right now: VA or VB



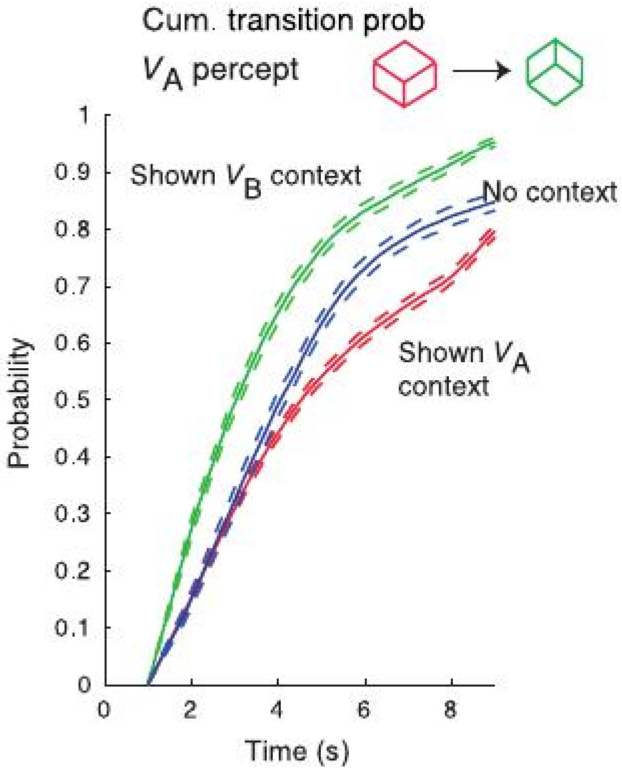
Perceptual multi-stability predicted by search model for Bayesian decisions, Sundareswara, Schrater,2008

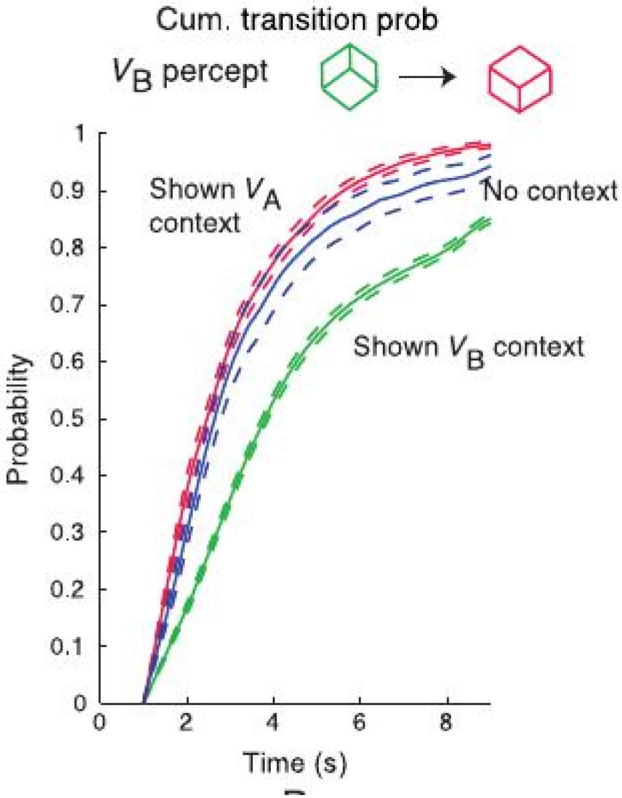
Example 5.2.2: Illusions: Contextual necker cube



Perceptual multi-stability predicted by search model for Bayesian decisions, Sundareswara, Schrater,2008

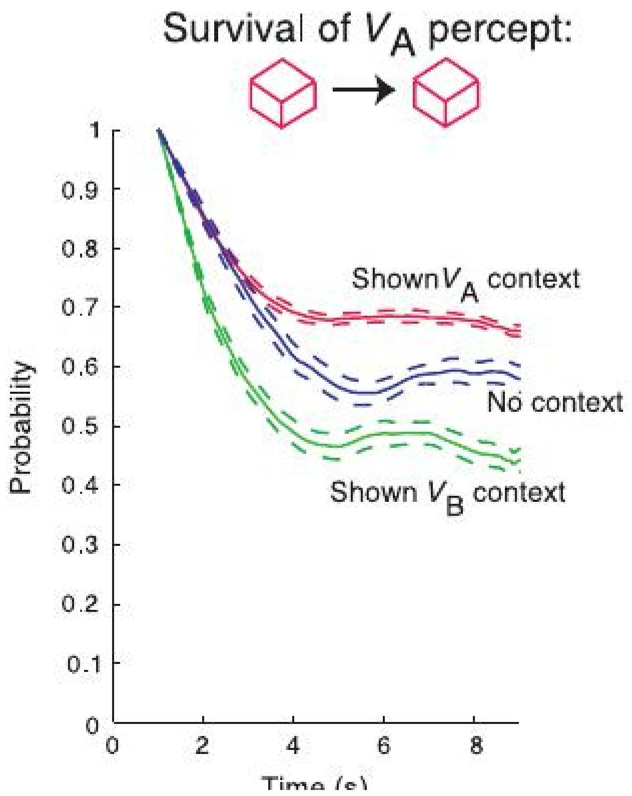
Example 5.2.2: Illusions: Contextual necker cube

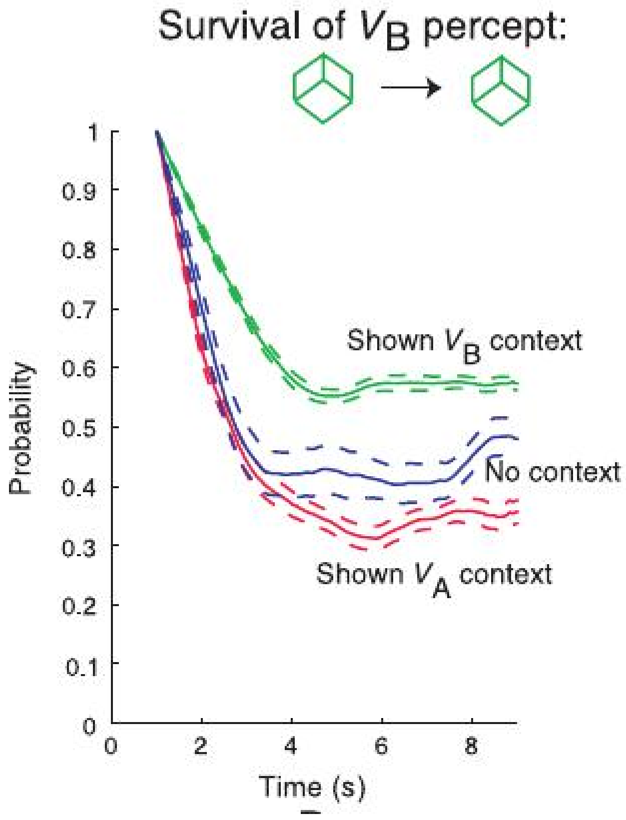




Perceptual multi-stability predicted by search model for Bayesian decisions, Sundareswara, Schrater,2008

Example 5.2.2: Illusions: Contextual necker cube



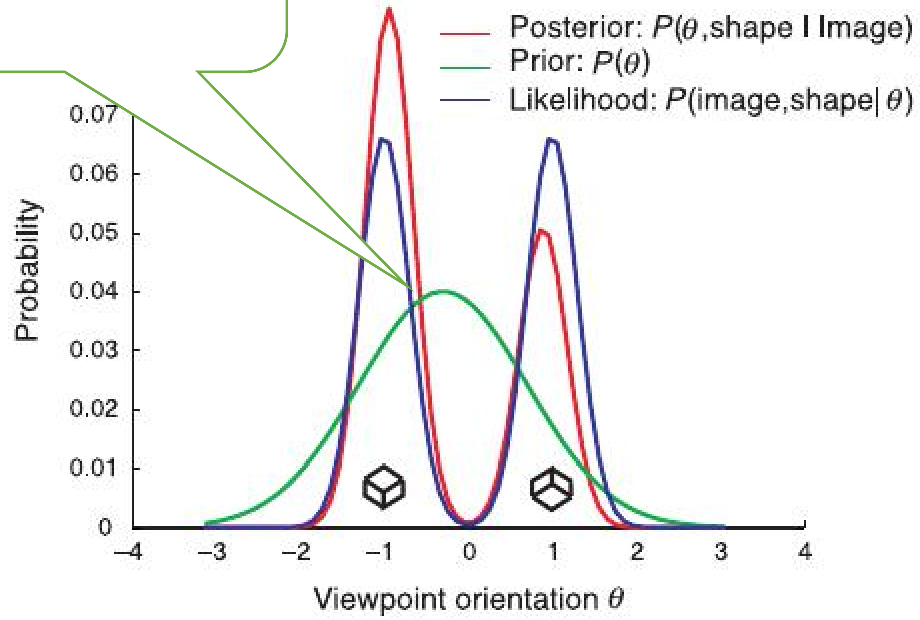


Perceptual multi-stability predicted by search model for Bayesian decisions, Sundareswara, Schrater,2008

Example 5.2.2: Illusions: Contextual ube

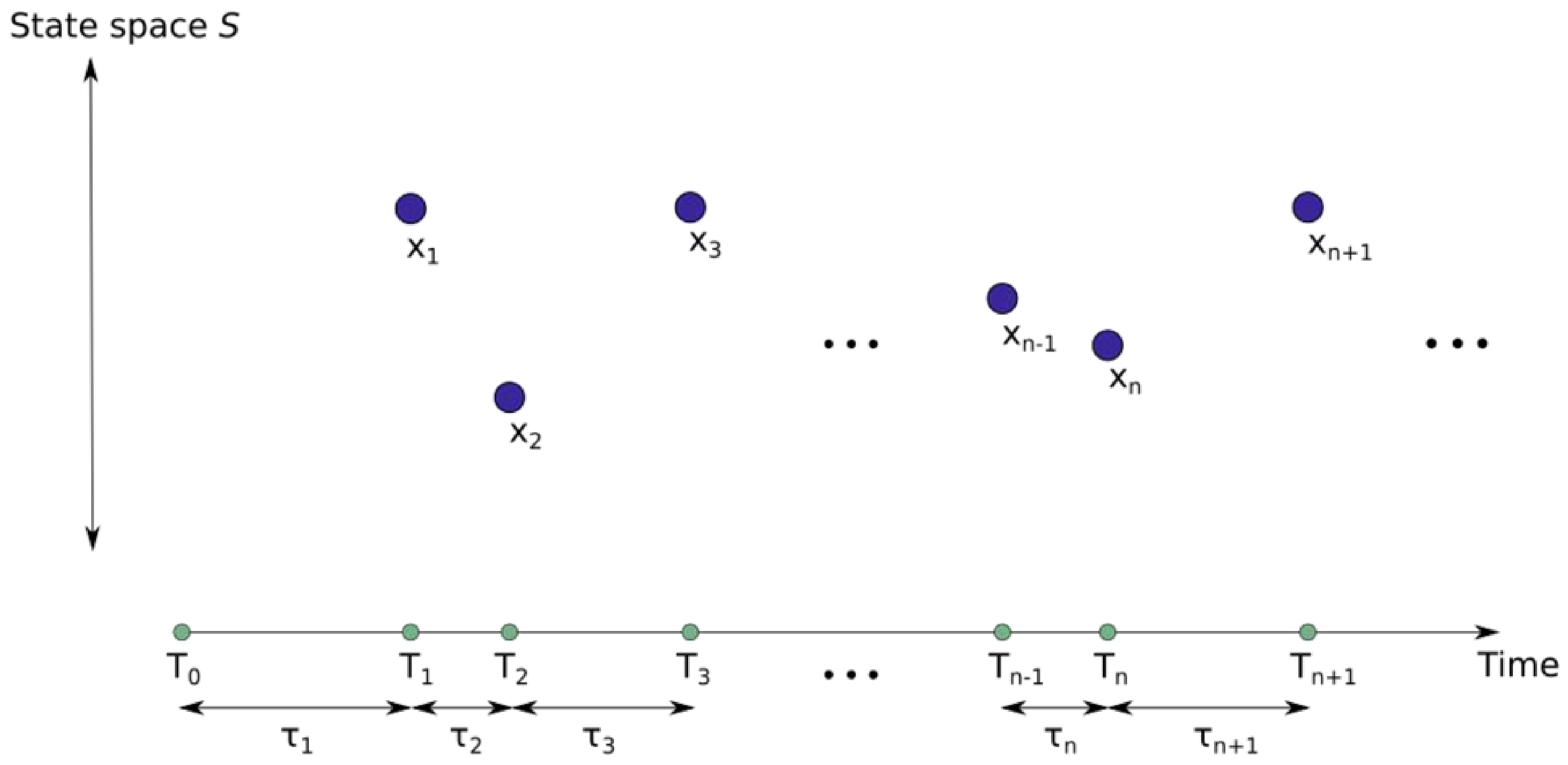
Prior biased VA

General idea: Subjects are sampling from a bi-modal posterior
How to do that? **Markov Renewal Process** (Random process that generalizes the notion of Markov jump processes)



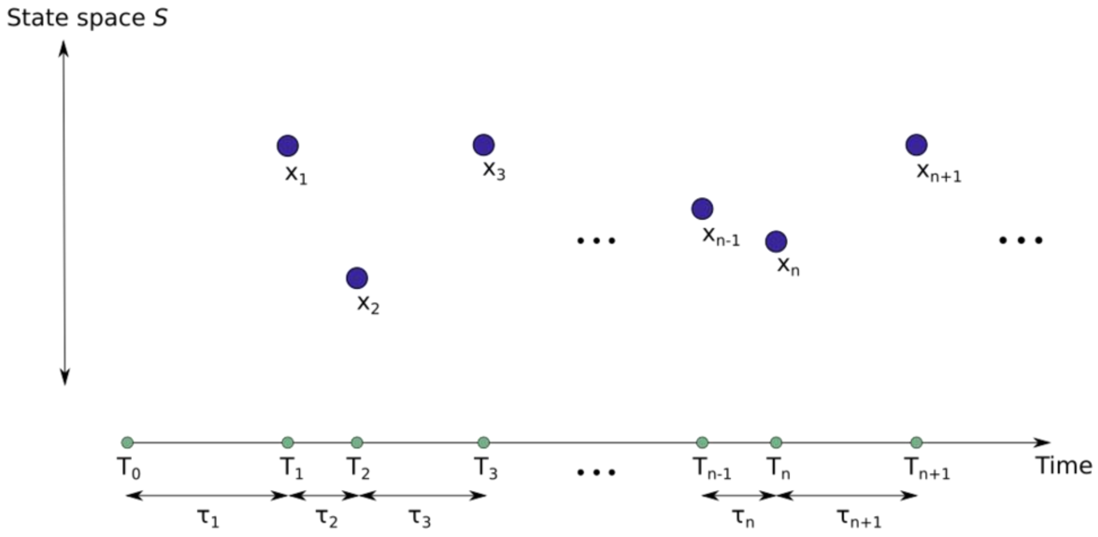
Perceptual multi-stability predicted by search model for Bayesian decisions, Sundareswara, Schrater,2008

Example 5.2.2:
Illusions: Contextual necker cube



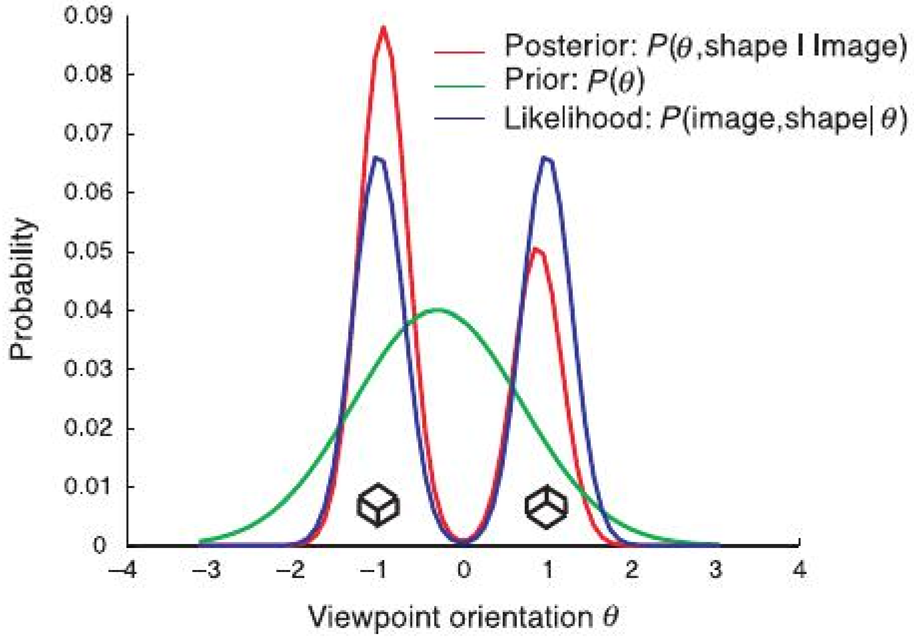
Example 5.2.2: Illusions: Bi-stable

- set of states : S
- (X_n, T_n) Tn jump times and X_n the associated X_n in the Markov Chain
- Inter arrival time: $\tau_n = T_n - T_{n-1}$
- (X_n, T_n) Markov Renewal process if



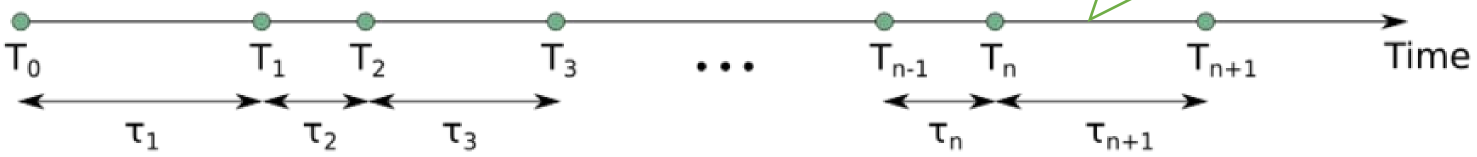
$$\Pr(\tau_{n+1} \le t, X_{n+1} = j \mid (X_0, T_0), (X_1, T_1), \ldots, (X_n = i, T_n))$$
$$= \Pr(\tau_{n+1} \le t, X_{n+1} = j \mid X_n = i) \, \forall n \ge 1, t \ge 0, i, j \in \mathrm{S}$$

Example 5.2.2:
Illusions: Contextual necker cube

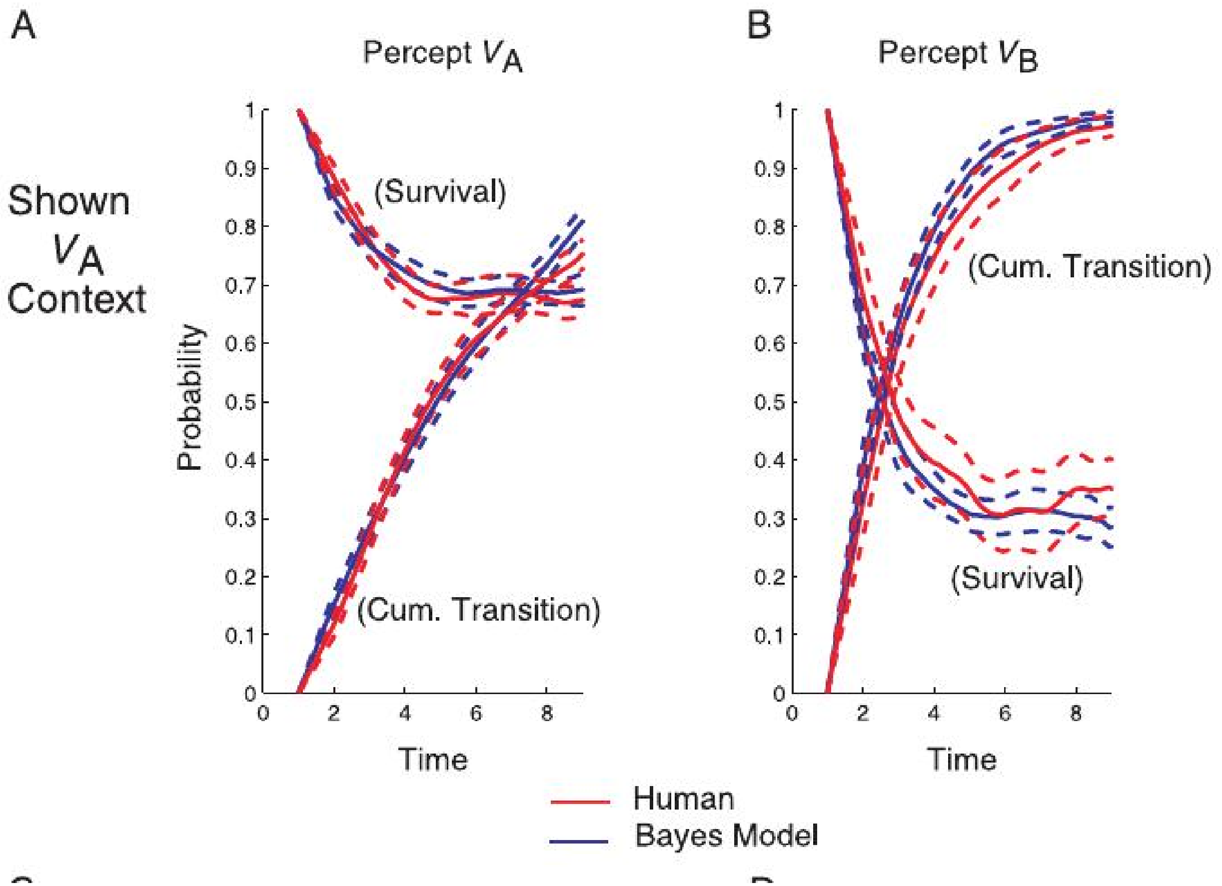


$$\omega = \frac{P(\theta_{V_A})}{P(\theta_{V_B})} = \exp\left(\frac{(\theta_{V_B} - \theta_M)^2 - (\theta_{V_A} - \theta_M)^2}{2\sigma^2}\right)$$

The ratio of the mass under the posterior peaks and the memory decay determine the amount of time spent in each precept

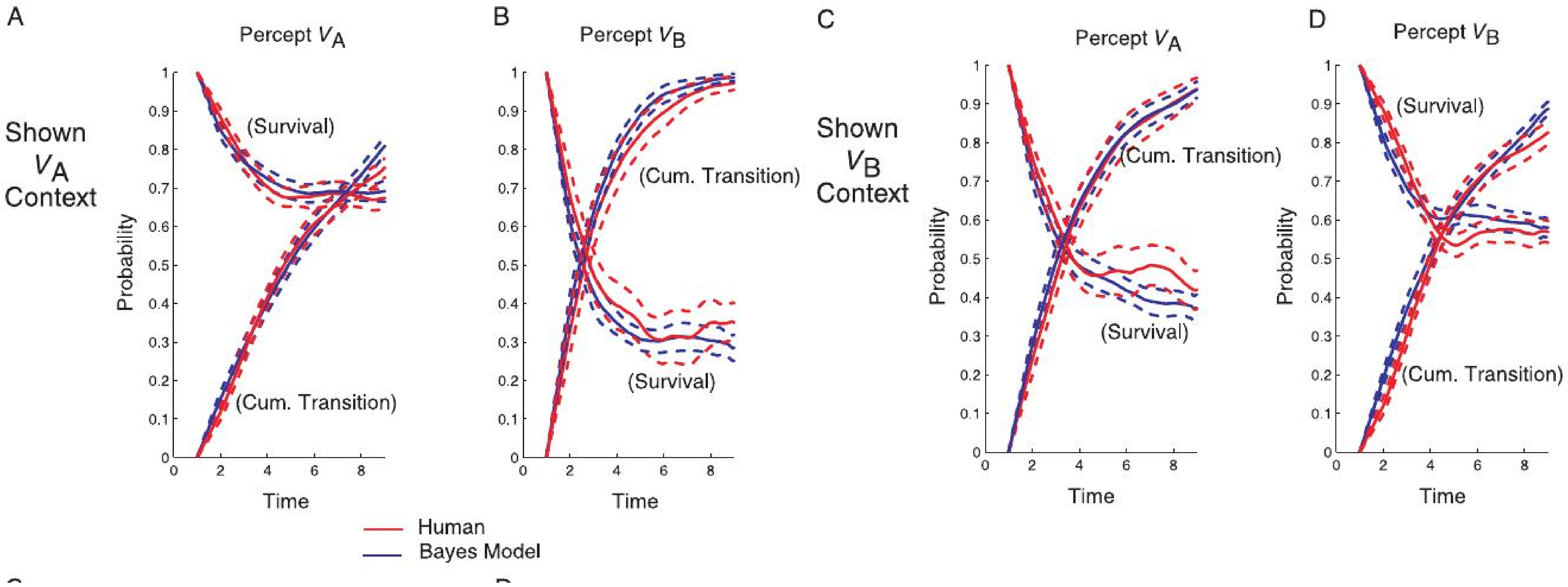


Example 5.2.2: Illusions: Bi-stable



Perceptual multi-stability predicted by search model for Bayesian decisions, Sundareswara, Schrater,2008

Example 5.2.2: Illusions: Bi-stable



Perceptual multi-stability predicted by search model for Bayesian decisions, Sundareswara, Schrater,2008

Example 6: Invariances

Invariants in perception:

- View point
- Color
- Contrast
- Retinal position

...

However, be careful... The perceptual invariance is actually limited...

It is usually modeled as neuronal adaptation

Example 6: Invariances

Invariants in perception:

- View point
- Color
- Contrast
- Retinal position

...

However, be careful... The perceptual invariance is actually limited...

It is usually modeled as neuronal adaptation

Summary

1. Bayes Theorem and Bayesian modeling
2. Bayesian Brain
    1. Formulation
    2. Examples
3. Bayesian life long learning
4. Critics to the “ideal observer”

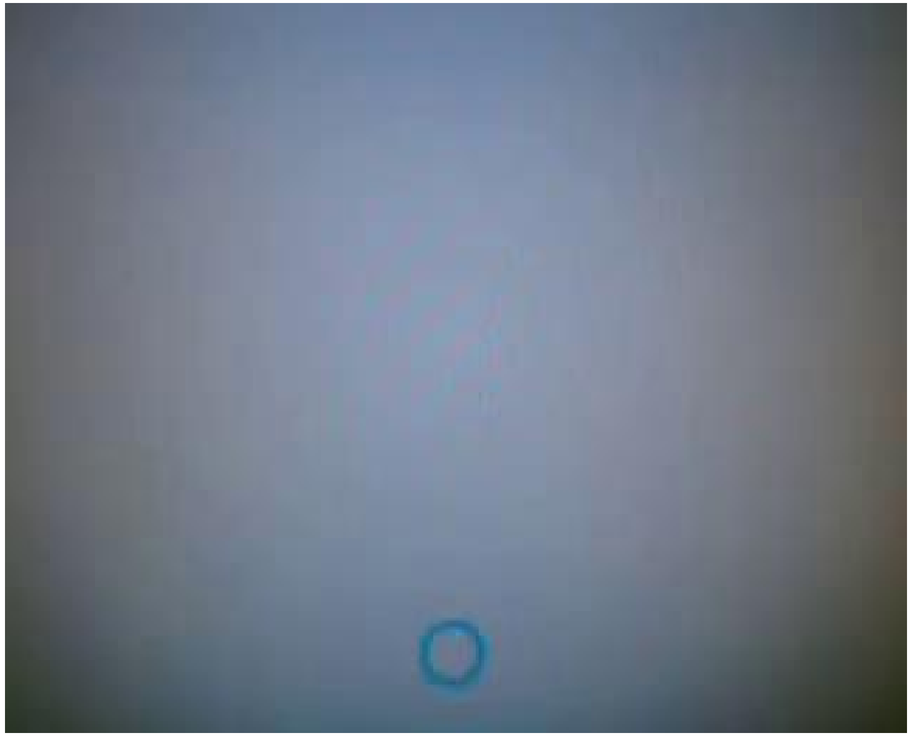
Bayesian life long learning

Learning is a process by which we update what we **believe** about the world around us.

Whenever we make an **observation** we will update these beliefs, which will usually get more precise when we make observations.

In a constantly changing world, the **update** of our beliefs is highly important and relevant

Random dot motion



https://www.youtube.com/watch?v=Cx5Ax68Slvk

Random dot motion

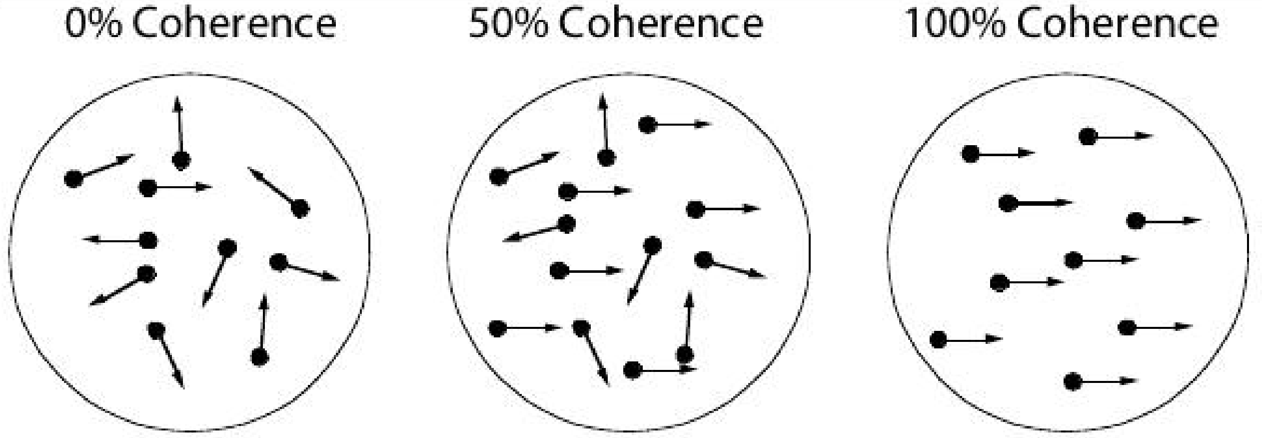


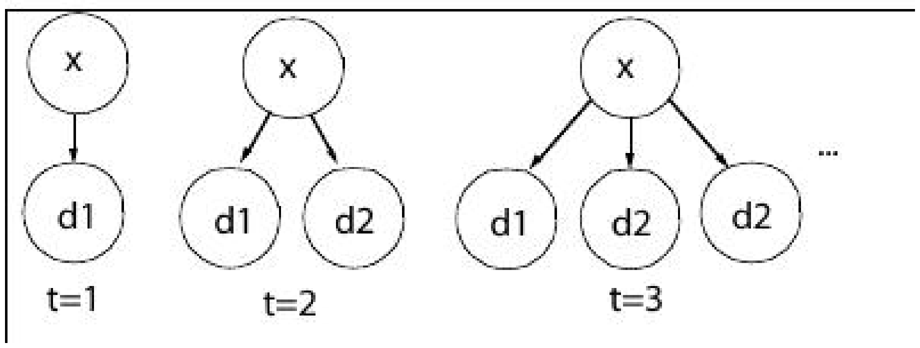
Figure 8.2: Motion stimulus. The subject has to estimate the overall direction of point movement (left or right). The difficulty of the task – i.e., uncertainty in the likelihood function (per time unit) – can be adjusted by changing the proportion of dots that move in the same direction (coherence).

Building a Bayesian decision model

1. Bayesian model of all options
2. Build the inner believes
3. ~~Define a risk/cost function~~
4. ~~Decide for lower risk/cost~~ Decide for higher believe

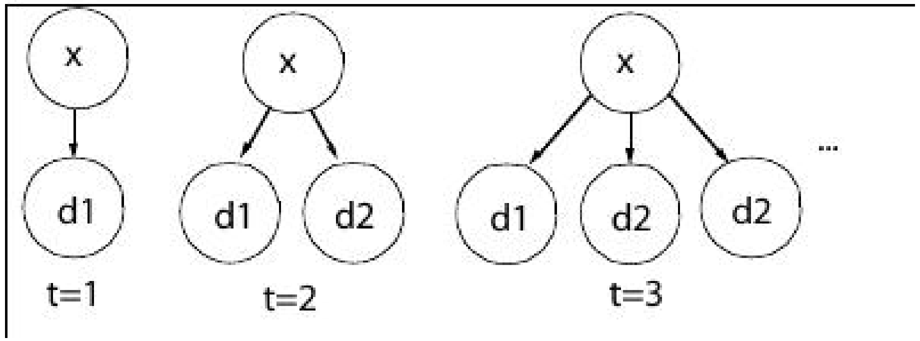
Bayesian life long learning: random dot motion

1. **The generative model**
2. The inference process
3. ~~The distribution of the MAP estimate~~



Bayesian life long learning: random dot motion

1. The generative model
2. **The inference process**
3. ~~The distribution of the MAP estimate~~

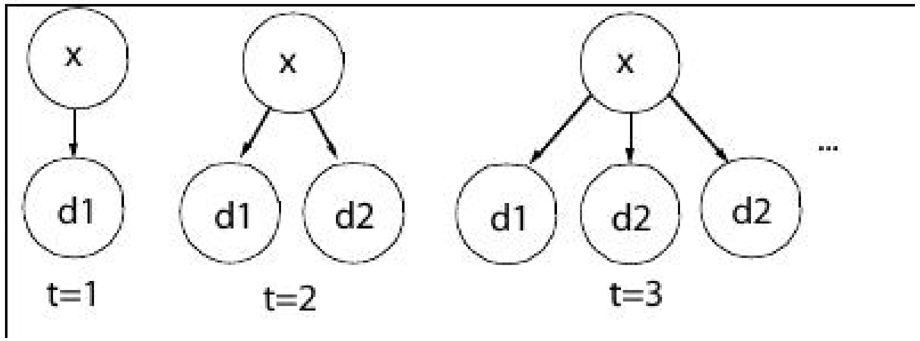


$$p(x \mid d_1 \cdots d_N) \propto p(x) \prod_{i=1}^{N} p(d_i \mid x)$$

$$= p(x) p(d_N \mid x) \prod_{i=1}^{N-1} p(d_i \mid x) \propto p(x \mid d_1 \cdots d_{N-1}) p(d_N \mid x)$$

Bayesian life long learning: random dot motion

1. The generative model
2. **The inference process**
3. ~~The distribution of the MAP estimate~~



The posterior at time N-1 is the prior at time N

$$p(x \mid d_1 \cdots d_N) \propto p(x) \prod_{i=1}^{N} p(d_i \mid x)$$

$$= p(x) p(d_N \mid x) \prod_{i=1}^{N-1} p(d_i \mid x) \propto \boxed{p(x \mid d_1 \cdots d_{N-1})} p(d_N \mid x)$$

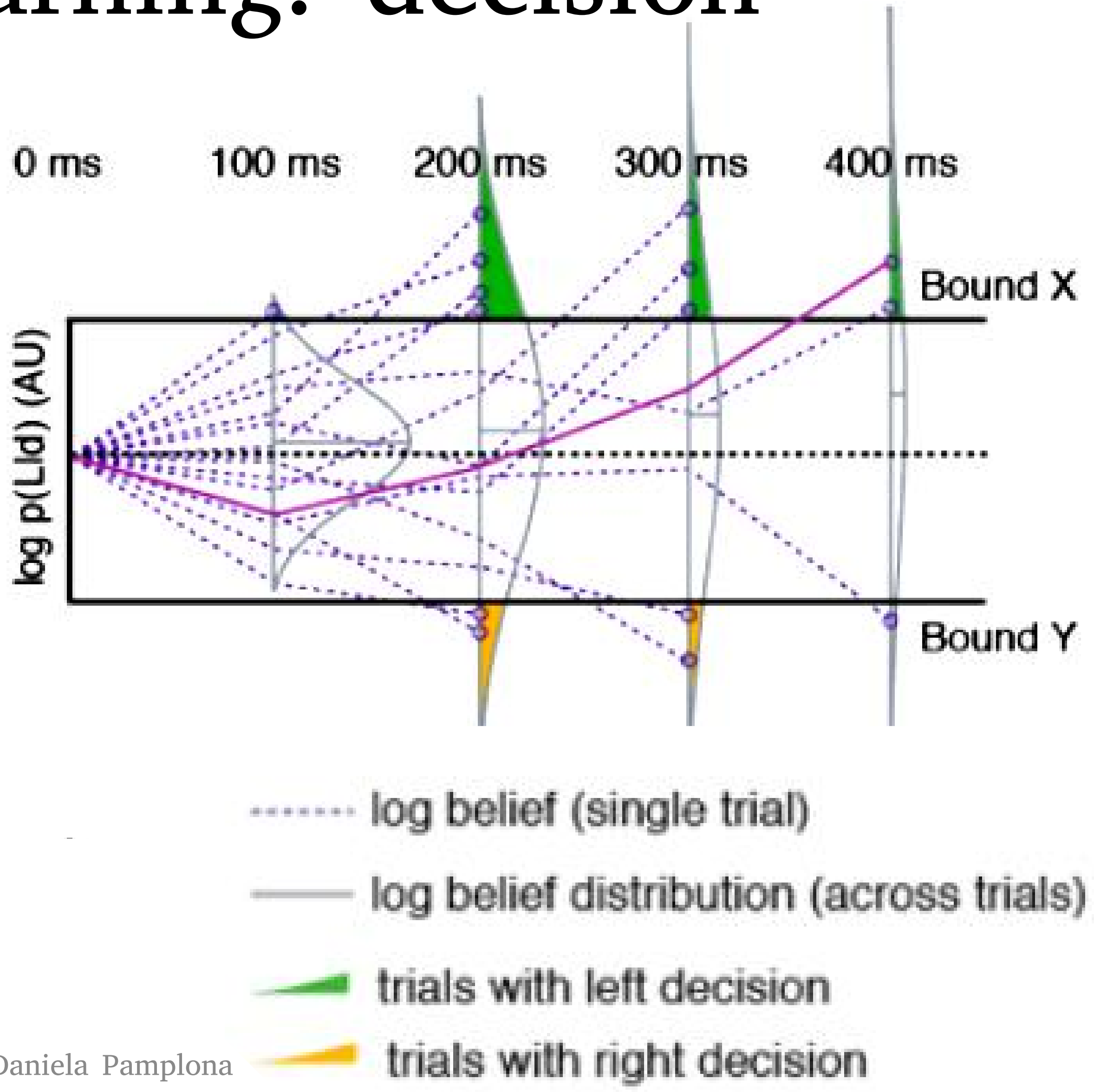
Bayesian life long learning: decision

Decision making process:
(left is the correct decision)

1. calculate d

$$d = Log\left(\frac{p(x = left|d_1,\ldots,d_N))}{p(x = right|d_1,\ldots,d_N))}\right)$$

1. if d> Bound X return left
2. if d< Bound Y return right



Bayesian life long learning: decision

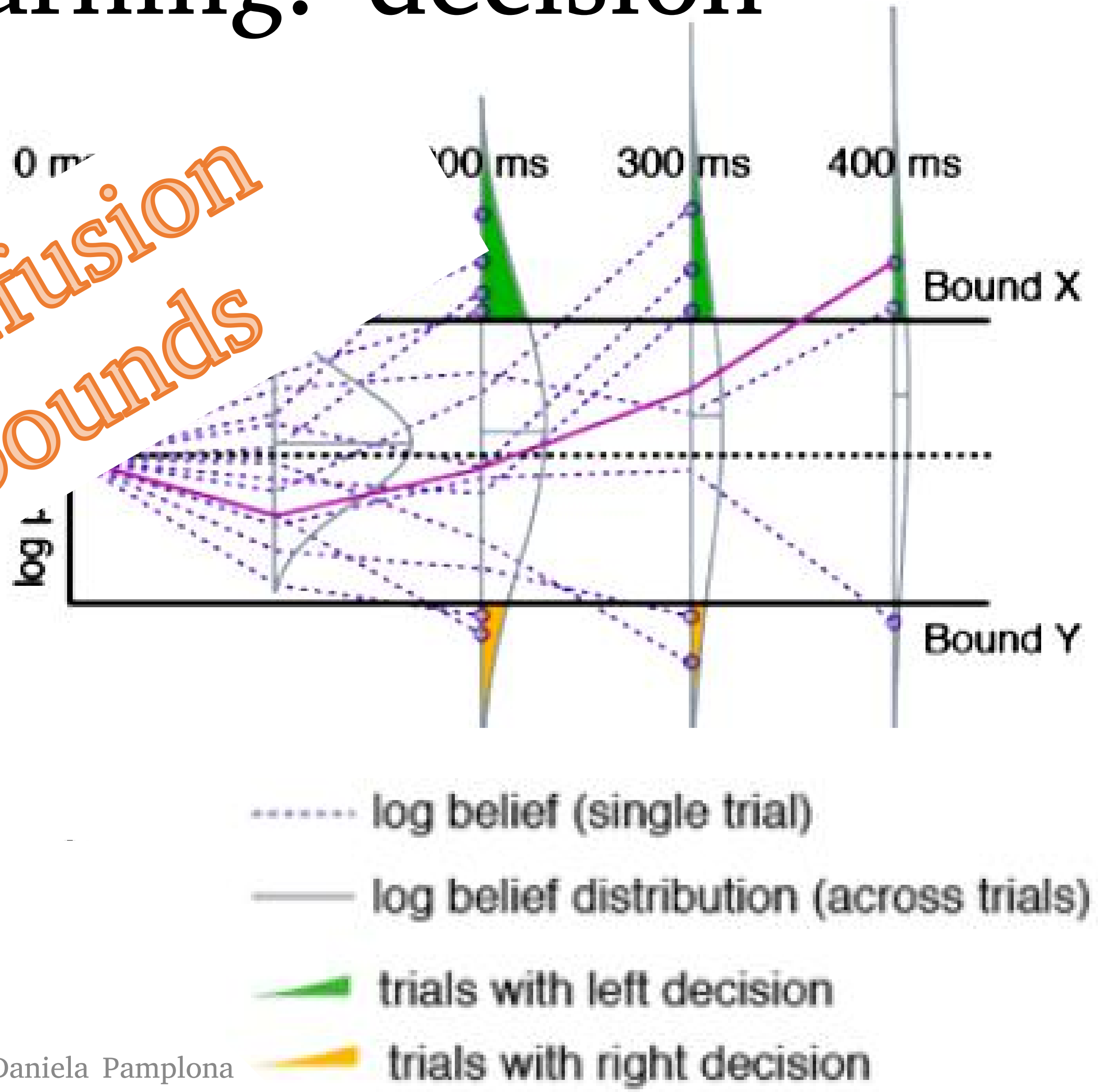
Decision making process:
(left is the correct decision)

1. calculate d

$$d = Log\left(\frac{p(\ldots}{\ldots_N))} \right)$$

1. if d> Bou… …urn left
2. if d< Boun… Y return right

Equivalent to diffusion process with bounds



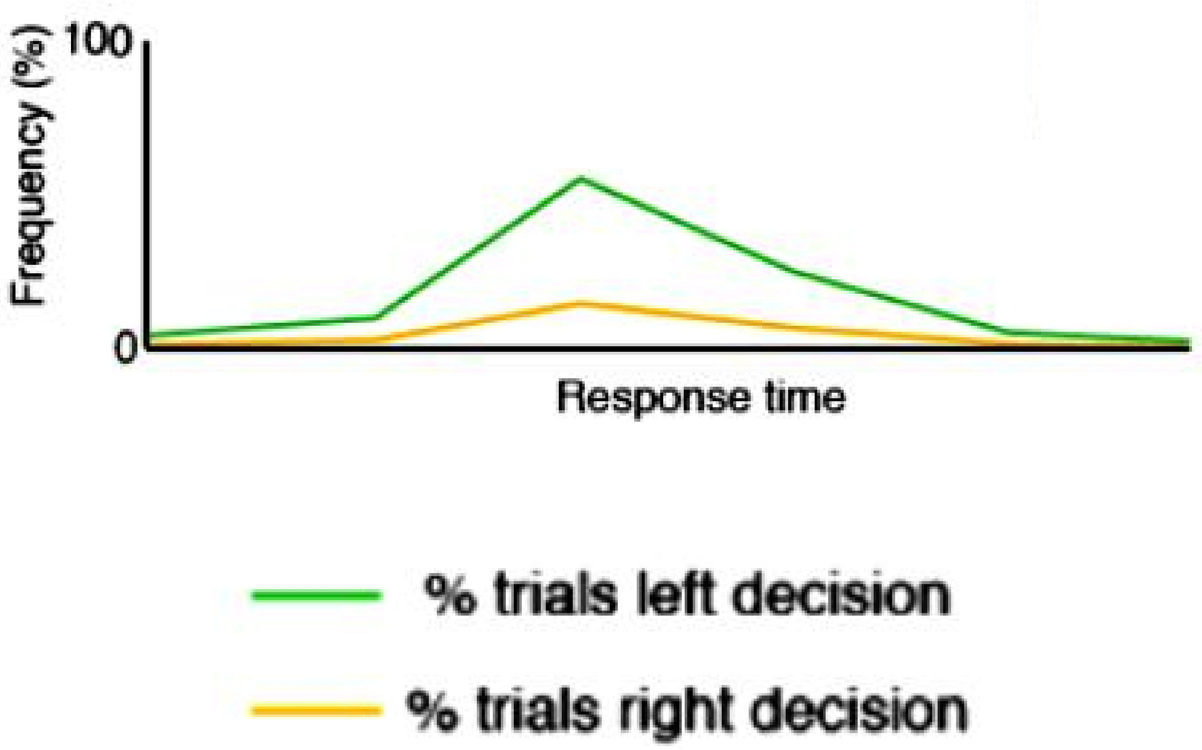
Bayesian life long learning: decision

Decision making process:
(left is the correct decision)

1. calculate d

$$d = Log\left(\frac{p(x = left|d_1, \ldots, d_N))}{p(x = right|d_1, \ldots, d_N))}\right)$$

1. if d> Bound X return left
2. if d< Bound Y return right



Bayesian life long learning: decision

Decision making process:
(left is the correct decision)

1. calculate d

$$d = Log\left(\frac{p(x = left|d_1,\ldots,d_N)}{p(x = right|d_1,\ldots,d_N)}\right)$$

1. if d> Bound X return left
2. if d< Bound Y return right

Today's exercise

Exercise 2

- **7.** Show that the decision is also recursive.

$$p(x \mid d_1 \cdots d_N) \propto p(x) \prod_{i=1}^{N} p(d_i \mid x)$$

$$= p(x) p(d_N \mid x) \prod_{i=1}^{N-1} p(d_i \mid x) \propto p(x \mid d_1 \cdots d_{N-1}) p(d_N \mid x)$$

$$d = Log \left(\frac{p(x = left \mid d_1, \ldots, d_N))}{p(x = right \mid d_1, \ldots, d_N))} \right)$$

Summary

1. Bayes Theorem and Bayesian modeling
2. Bayesian Brain
   1. Formulation
   2. Examples
3. Bayesian life long learning
4. **Critics to the “ideal observer”**

Critics to the “ideal observer”

Ideal observer: the one that follows optimally the Bayes rule /probabilistic approach.

Critics to the “ideal observer”

Ideal observer: the one that follows optimally the Bayes rule /probabilistic approach.

Problems:

- No time constrains
- No energetic constrains
- It assumes perfect knowledge of image statistics
- It does not include the task

Critics to the “ideal observer”

Ideal observer: the one that follows optimally the Bayes rule /probabilistic approach.

Problems:

- No time constrains
- No energetic constrains
- It assumes perfect knowledge of world statistics
- It does not include the task

Good news: There are ways of introducing this constrains: by modeling (bayesian networks), bayesian reinforcement learning, etc... but it is hard...

Critics to the “ideal observer”

Examples of failing:

- Pelli, Farell,and Moore.The remarkable ineficiency of word recognition. Nature, 2003
- Legge, Kersten and Burgess. Contrast discrimination in noise. Journal of the Optical Society of America A, 1987.
- Schrater and Kersten, How Optimal Depth Cue Integration Depends on the Task, International Journal of Computer Vision, 2000

Summary

1. Bayes Theorem and Bayesian modeling
2. Bayesian Brain
   1. Formulation
   2. Examples
3. Bayesian life long learning
4. Critics to the “ideal observer”