COMP 546

Lecture 15

Cue combinations, Bayesian models

Thurs. March 1, 2018

Visual Cues:

image properties that can tell us about scene properties

Image	Scene
texture - size, shape, density	depth gradient - slant, tilt
shading	surface curvature
binocular disparities	depth
motion (from moving observer)	
defocus blur	

Last lecture: Likelihood

$$p(\, I = i \mid S = s \,)$$

- Probability of measuring image $I = i,$ when the scene is $S = s$. (called "likelihood" of scene $S = s,$ given the image $I = i$).

- Maximum likelihood method:

  Choose $S = s$ that maximizes $p(\, I = i \mid S = s \,)$

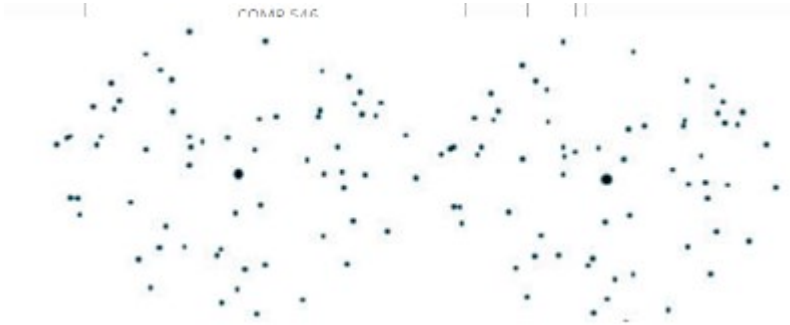
This lecture:

How to combine cues ?

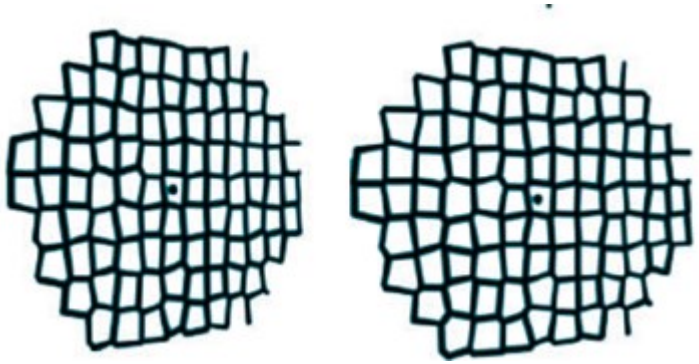
$$p(\, I_1,\, I_2 \mid S\,)$$

Example:

texture only
(monocular)



stereo only



texture and stereo

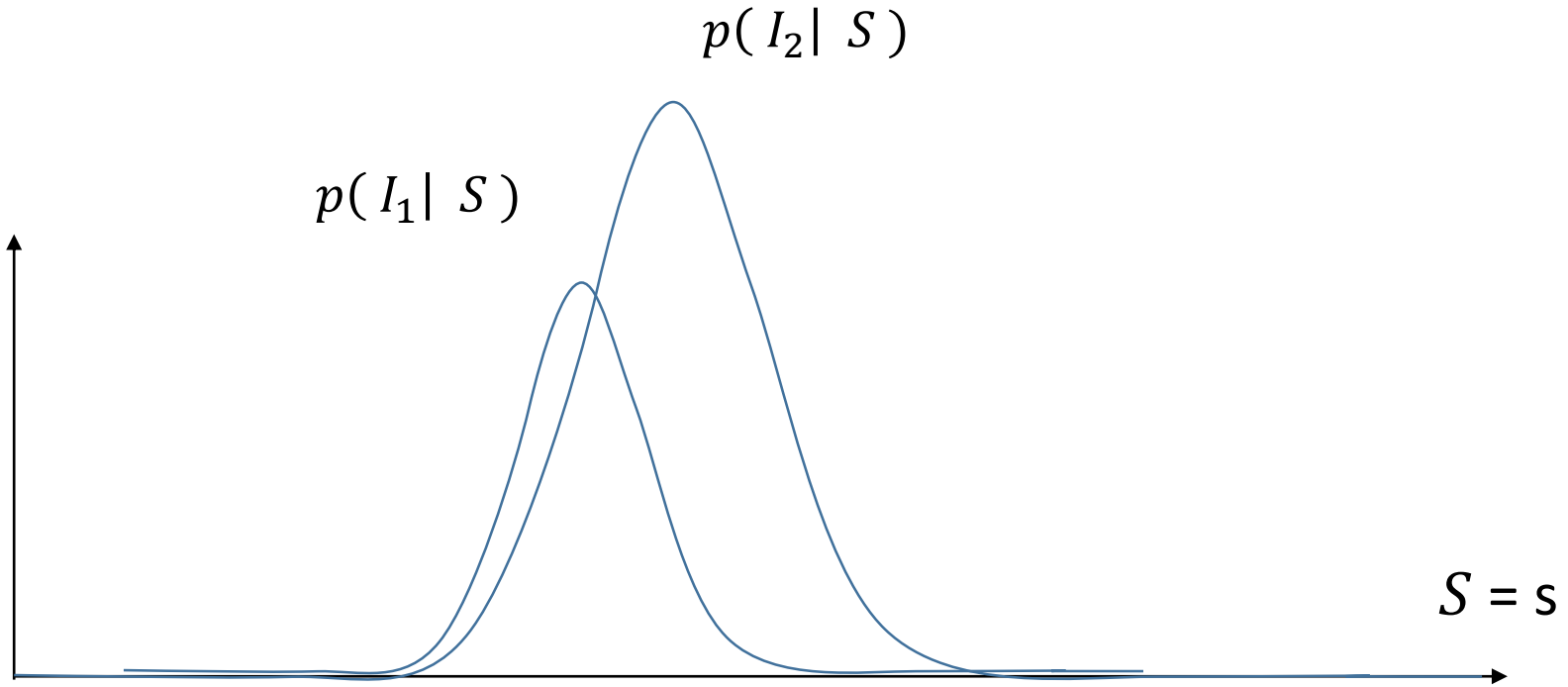
[Hillis 2004]

Assume likelihood function is “conditionally independent”:

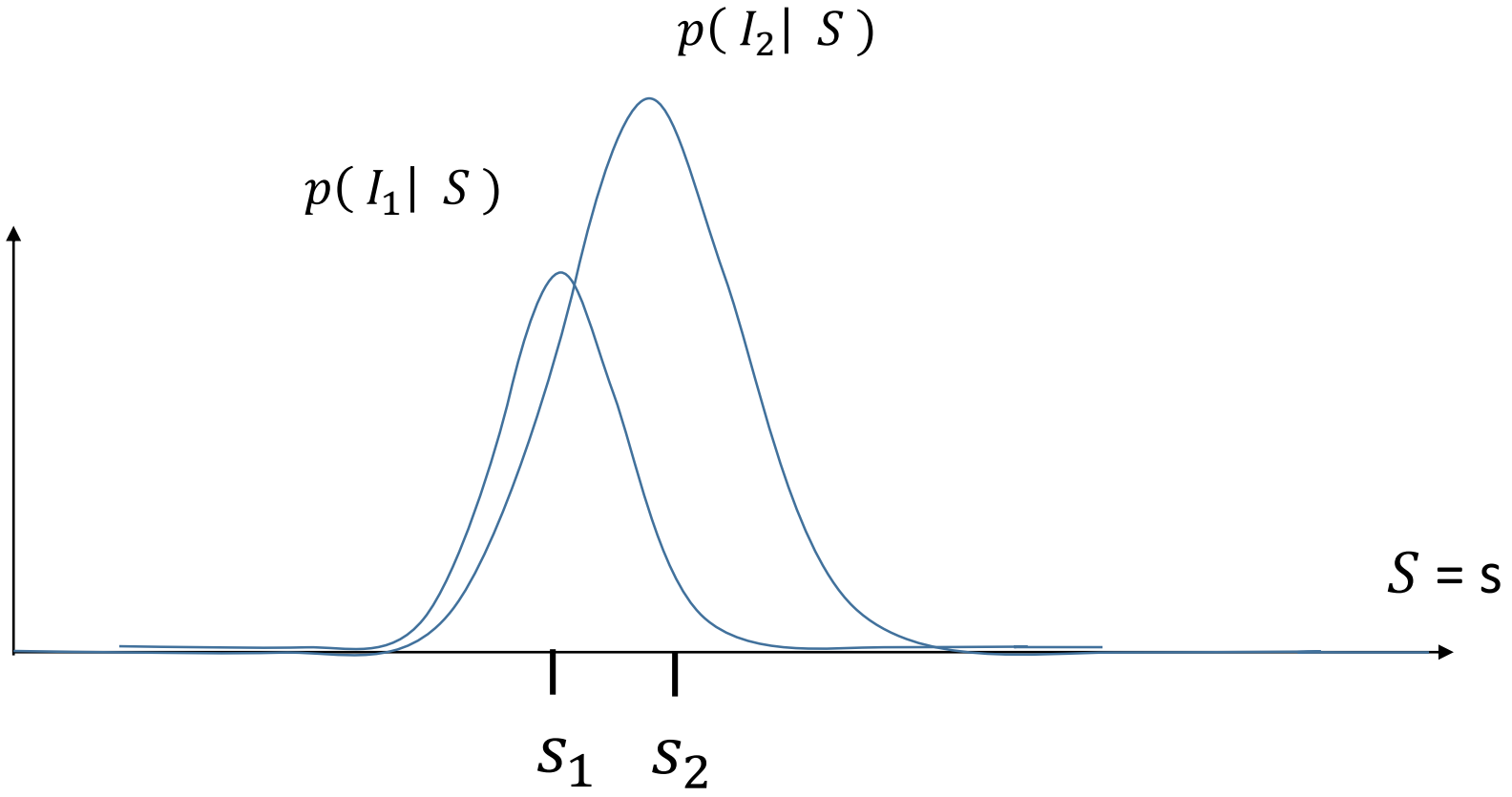
$$p(I_1, I_2 \mid S) = p(I_1 \mid S)\ p(I_2 \mid S)$$

e.g. I_1 is texture.

I_2 is binocular disparity.



Assume $p(I_1 = i_1|\ S = s)$ and $p(I_2 = i_2\ |\ S = s)$ are Gaussian shaped.



Assume $p(I_1 = i_1|\ S = s)$ and $p(I_2 = i_2\ |\ S = s)$ are Gaussian shaped.

Their maxima might occur at different values of s. Why ?

We want to find the s that maximizes:

$$p(I_1 \mid S = s)\ \ p(I_2 \mid S = s) = e^{-\frac{(s - s_1)^2}{2\,{\sigma_1}^2}}\ \ e^{-\frac{(s - s_2)^2}{2\,{\sigma_2}^2}}$$

We want to find the s that maximizes:

$$p(I_1 \mid S = s)\ \ p(I_2 \mid S = s) \ = \ \ e^{-\frac{(s - s_1)^2}{2\,{\sigma_1}^2}} \ \ e^{-\frac{(s - s_2)^2}{2\,{\sigma_2}^2}}$$

So, we want to find the s that minimizes:

$$\frac{(s - s_1)^2}{2\sigma_1^2} + \frac{(s - s_2)^2}{2\sigma_2^2}.$$

The lecture notes show that the solution $S = s$ is

$$s = w_1 s_1 + w_2 s_2$$

where

$$w_1 + w_2 = 1 \qquad 0 < w_i < 1$$

"Linear Cue Combination"

The lecture notes show that the solution $S = s$ is

$$s = w_1 s_1 + w_2 s_2$$

where

$$w_1 + w_2 = 1 \qquad 0 < w_i < 1$$

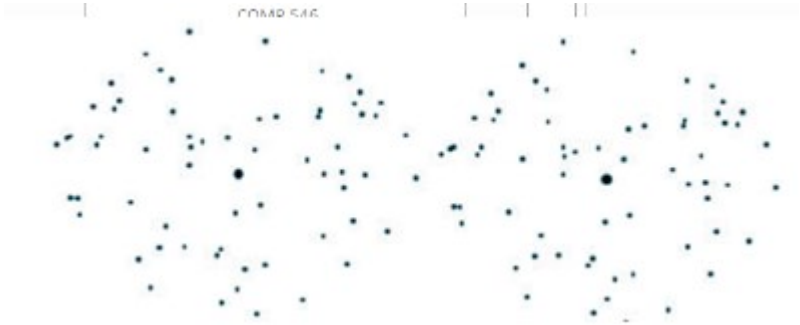
$$w_1 = \frac{{\sigma_2}^2}{{\sigma_1}^2 + {\sigma_2}^2} \qquad w_2 = \frac{{\sigma_1}^2}{{\sigma_1}^2 + {\sigma_2}^2}$$

Thus, less reliable cue (larger σ) get less weight.

Example:

[Hillis 2004]

texture only
(monocular)

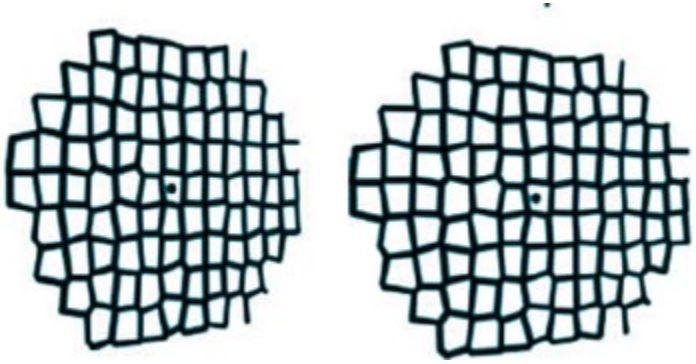


stereo only

Measure slant discrimination thresholds for cues *in isolation*.
Estimate likelihood function parameters ($s_1,\ \sigma_1\ , s_2,\ \sigma_2$).

… then

- present cues together

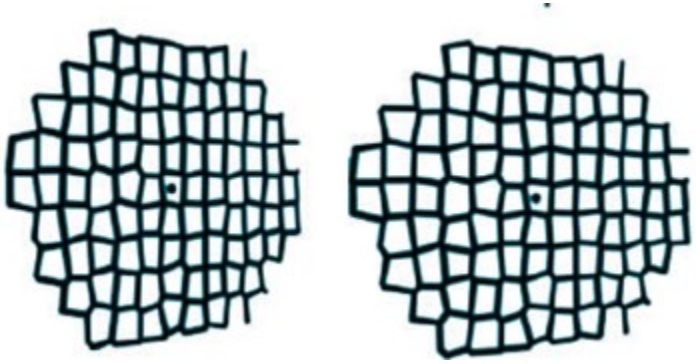


texture and stereo

- measure thresholds for S
- convert thresholds to likelihood parameters (s , σ)

… then

- present cues together

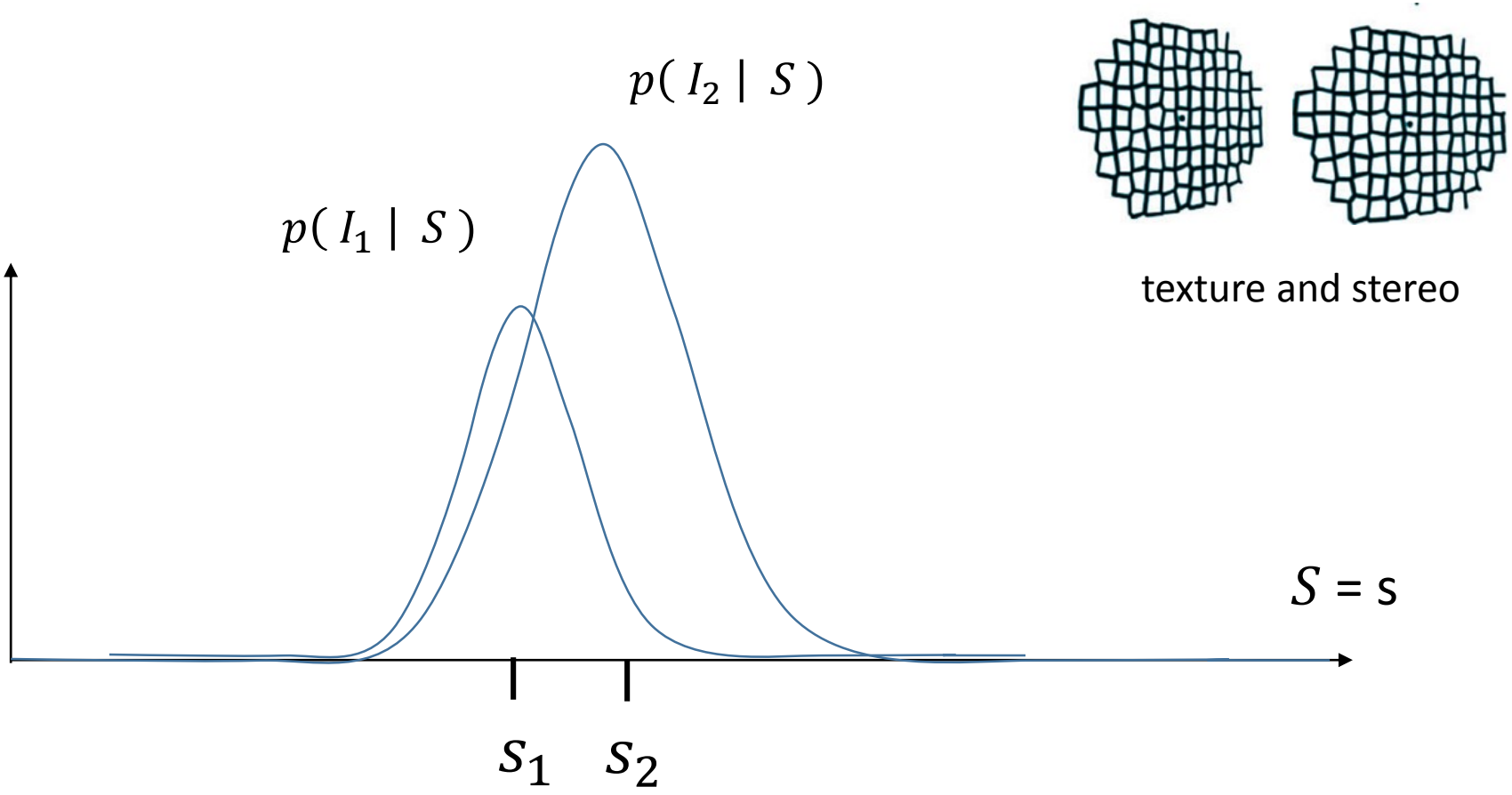


texture and stereo

- measure thresholds for S
- convert thresholds to likelihood parameters $(s\ ,\ \sigma)$
- examine if these values are consistent with the model*

$$s = w_1 s_1 + w_2 s_2$$

*Model also makes prediction about σ in combined case.



texture and stereo

Experimenter can manipulate s_1 , s_2 , σ_1 , σ_2 and predict effect on perception of slant.

COMP 546

Lecture 15

Cue combinations,
Bayesian models

Thurs. March 1, 2018

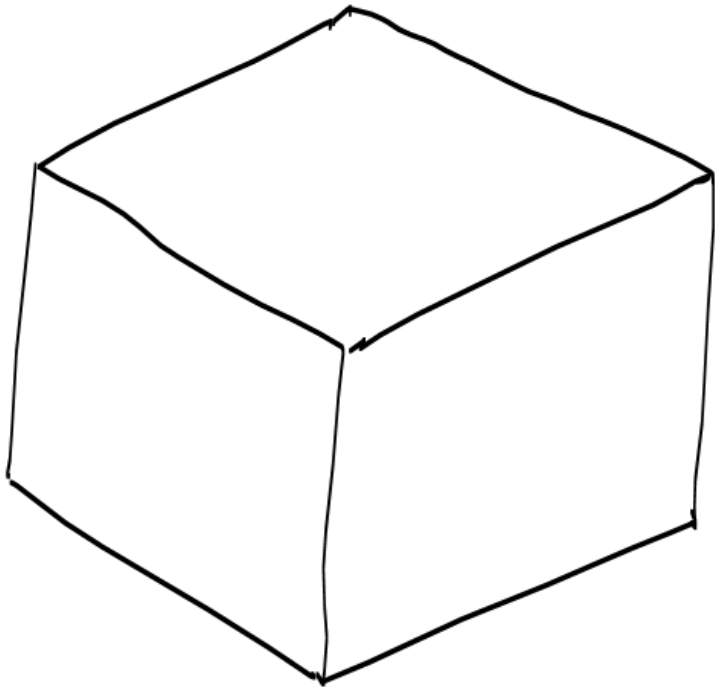
$$p(I = i \mid S = s) \neq p(S = s \mid I = i)$$

Likelihood of scene s, given image i

Probability of scene s, given image i

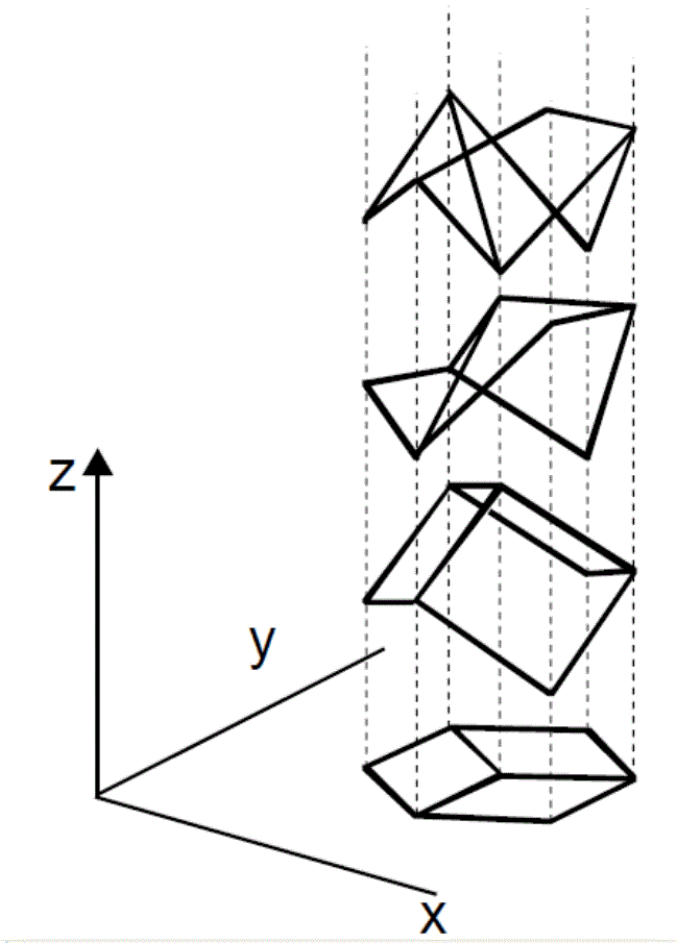
What is the crucial difference ?

Example : interpreting a line drawing



: Are there possible 3D interpretations here other than a box ?

Yes, a flat drawing. What else ?



wire frame with independently chosen depths

regular solid cube

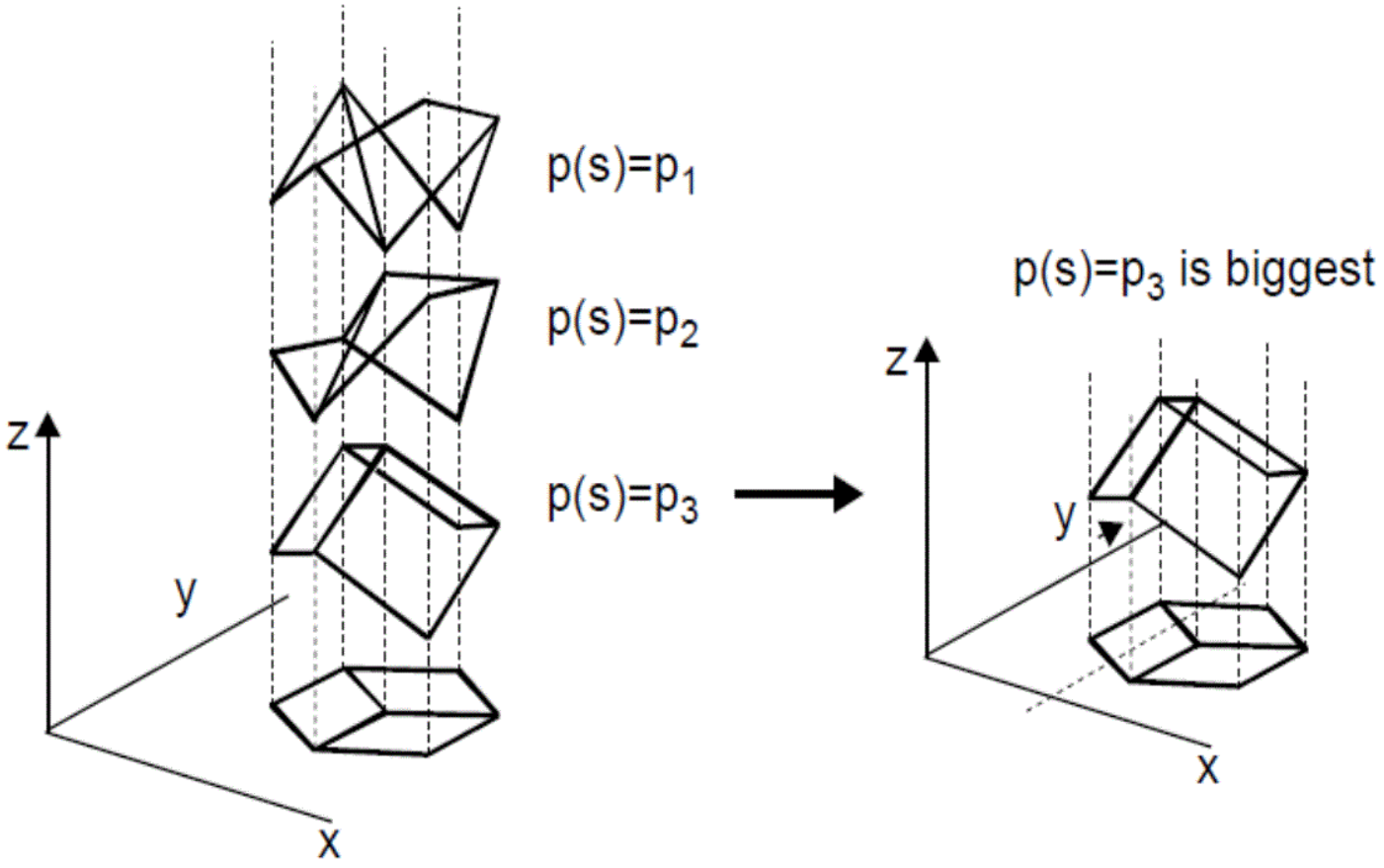
flat drawing

All scenes above have the same likelihood $p(\ I = i \mid\ S = s\ )$.
Why do we prefer the regular solid cube?

[Kersten & Yuille 2003]

Some *scenes* may have a larger probability $p(S = s)$.

The marginal probably $p(S = s)$ is called the "prior".



$$p(I \mid S) \equiv \frac{p(I, S)}{p(S)}$$

$$p(S \mid I) \equiv \frac{p(I, S)}{p(I)}$$

Thus,

$$p(I \mid S)\ p(S) = p(S \mid I)\ p(I)$$

Bayes Theorem

$$p(I \mid S) \equiv \frac{p(I, S)}{p(S)}$$

$$p(S \mid I) \equiv \frac{p(I, S)}{p(I)}$$

Thus,

likelihood scene prior

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

posterior image prior

Maximum '*a Posteriori*' (MAP)

Given an image, $I = i$, find the scene $S = s$ that maximizes $p(\ S = s \mid I = i\ )$.

likelihood scene prior

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

posterior image prior

Maximum '*a Posteriori*' (MAP)

Given an image, $I = i$, find the scene $S = s$ that maximizes $p(\ S = s \mid I = i\ )$.

We don't care about $p(\ I = i\ )$. Why not ?

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

likelihood

scene prior

posterior

image prior

If the prior p(S) is uniform then maximum likelihood gives the same solution as maximum posterior (MAP).

likelihood scene prior

$$p(S \mid I) = \frac{p(I \mid S)\,\cancel{p(S)}}{p(I)}$$

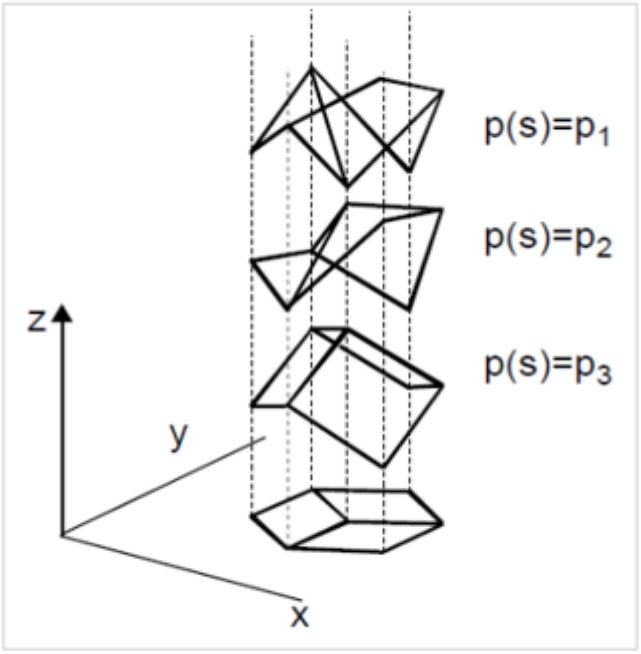
constant

posterior image prior

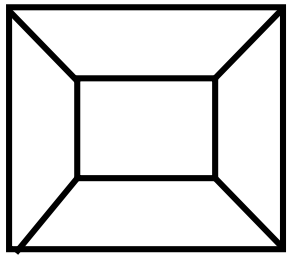
Interesting cases arise when the prior is non-uniform.

likelihood

prior



Ames Room



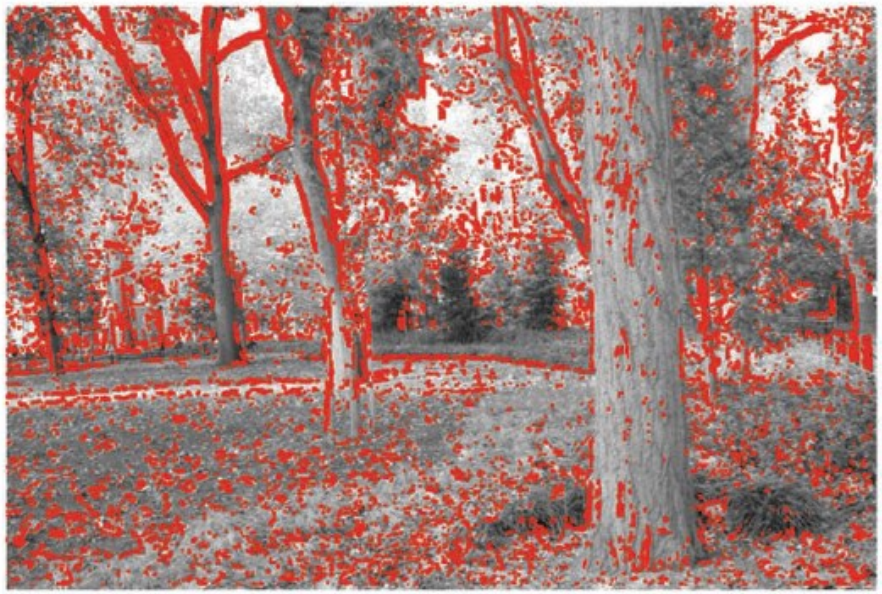
http://www.youtube.com/watch?v=Ttd0YjXF0no

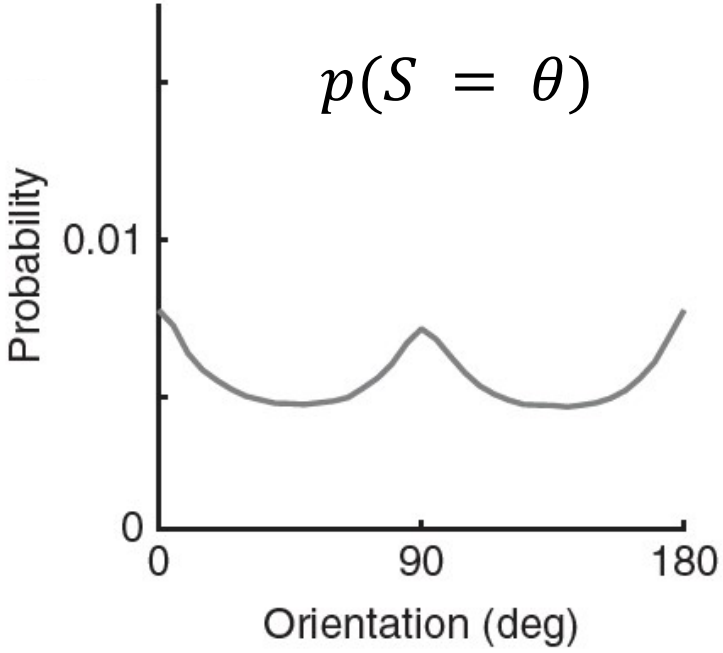
https://www.youtube.com/watch?v=gJhyu6nlGt8

Priors (“Natural Scenes Statistics”)

- intensity
- orientation of image lines, edges
- disparity
- motion
- surface slant, tilt

orientation θ of lines, edges

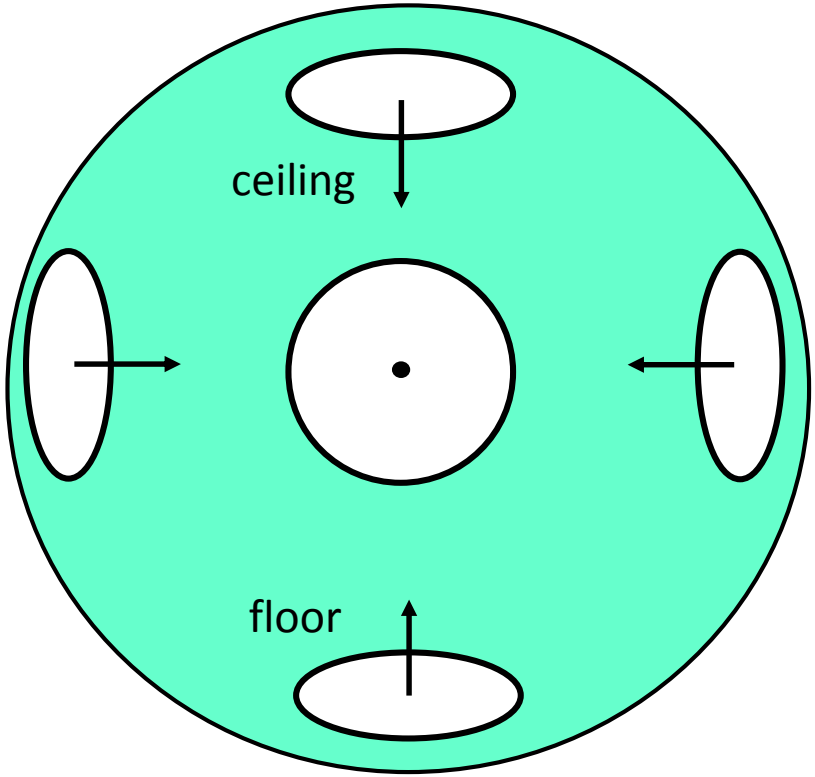




[Girshick 2011]

People are indeed better at discriminating vertical and horizontal orientations than oblique orientations. Why? Because they use a prior ?

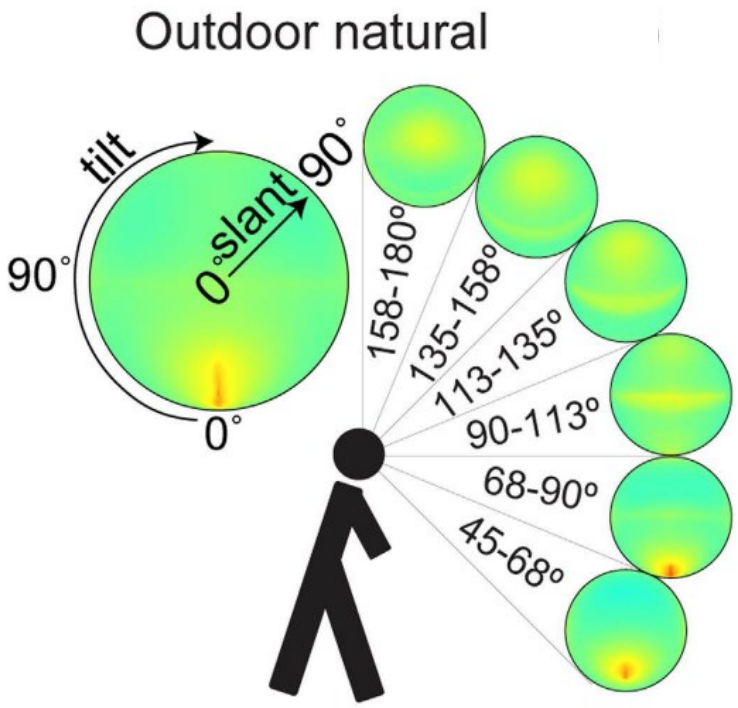
surface slant σ and tilt τ



Here we represent (slant, tilt) using a *concave* hemisphere.

See next slide.

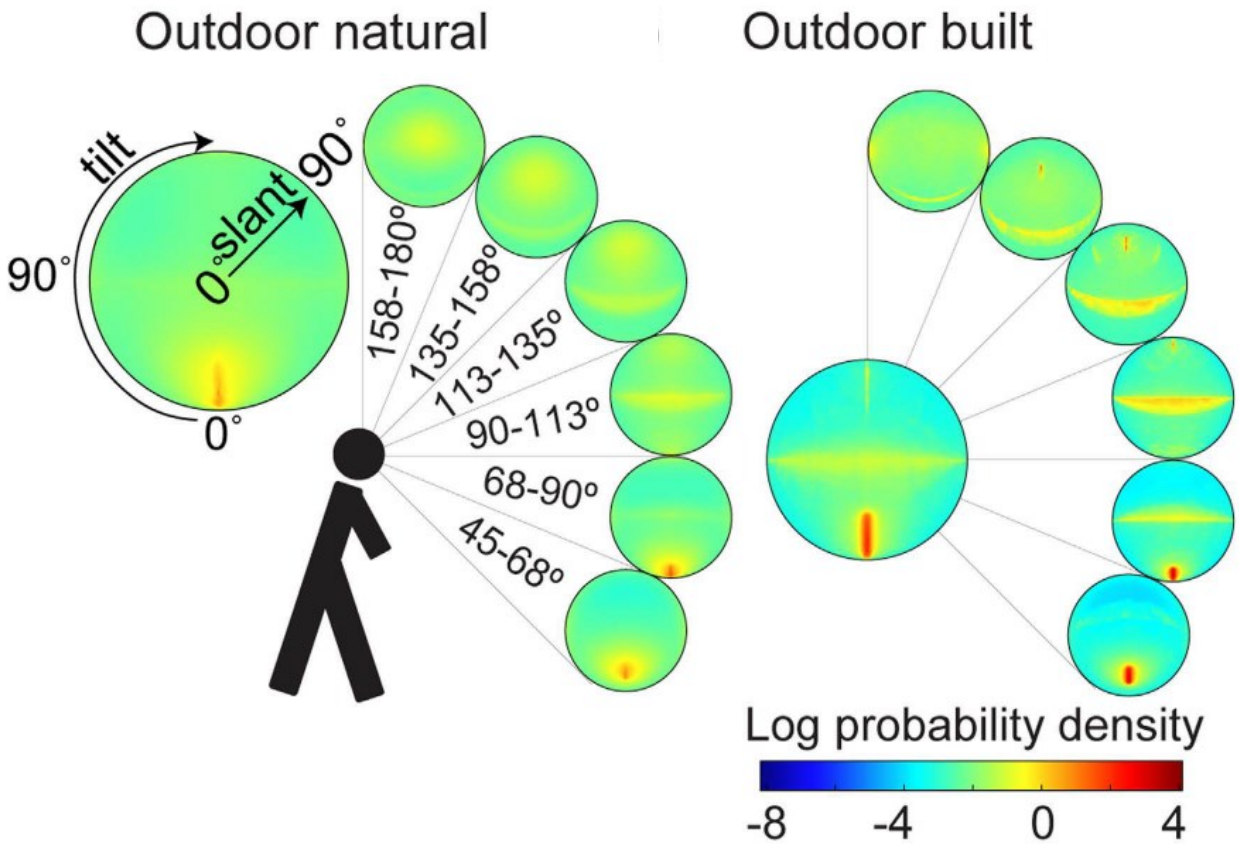
$$p(S = (\sigma, \tau))$$



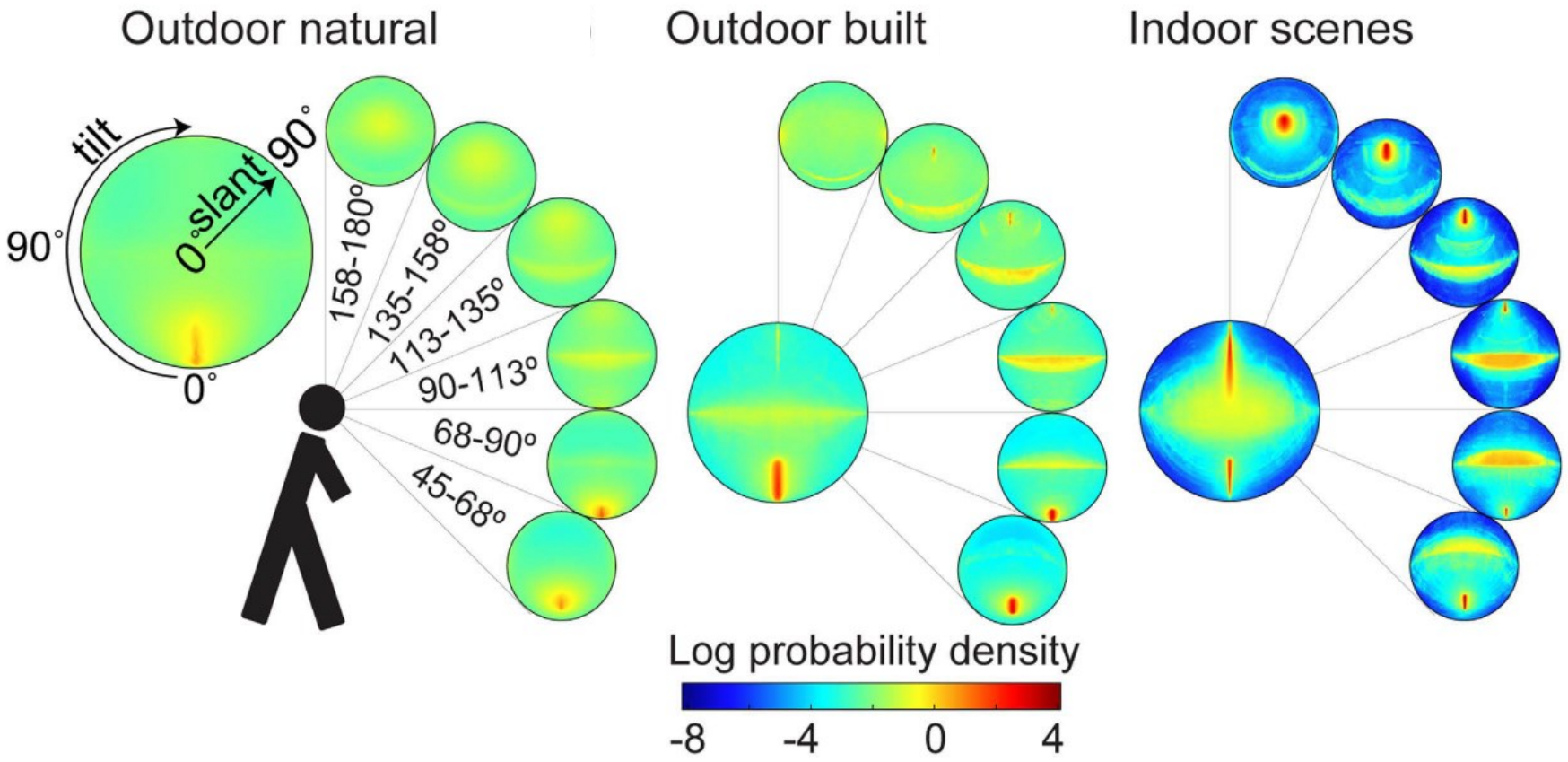
Each disk shows $p(\sigma, \tau)$ for surfaces visible over a range of viewing direction elevations, relative to line of sight.

[Adams & Elder 2016]

$$p(S = (\sigma, \tau))$$

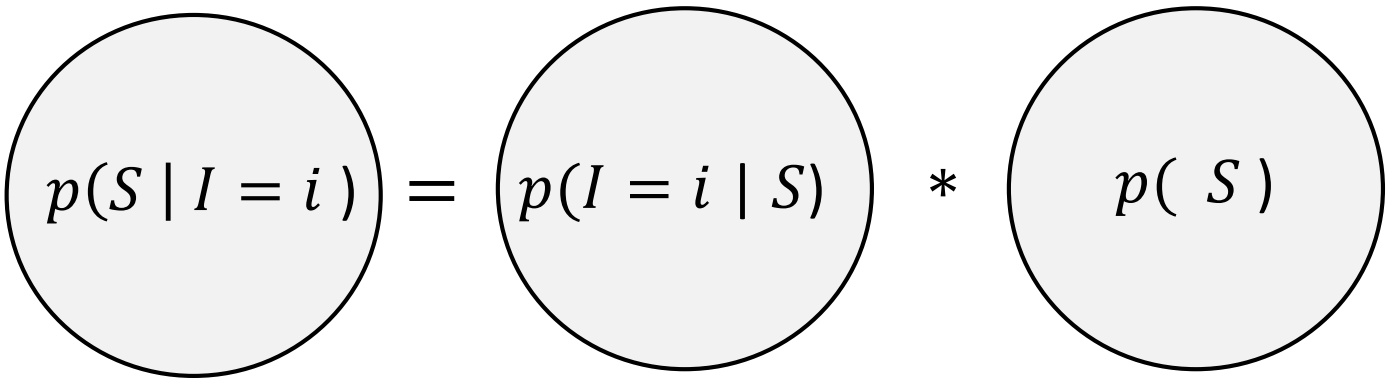

Outdoor natural
Outdoor built
tilt
90°
0°slant
90°
0°
158-180°
135-158°
113-135°
90-113°
68-90°
45-68°
Log probability density
-8
-4
0
4

$$p(S = (\sigma, \tau))$$


Outdoor natural
Outdoor built
Indoor scenes
tilt
90°
0°slant
90°
90°
158-180°
135-158°
113-135°
90-113°
68-90°
45-68°
Log probability density
-8
-4
0
4

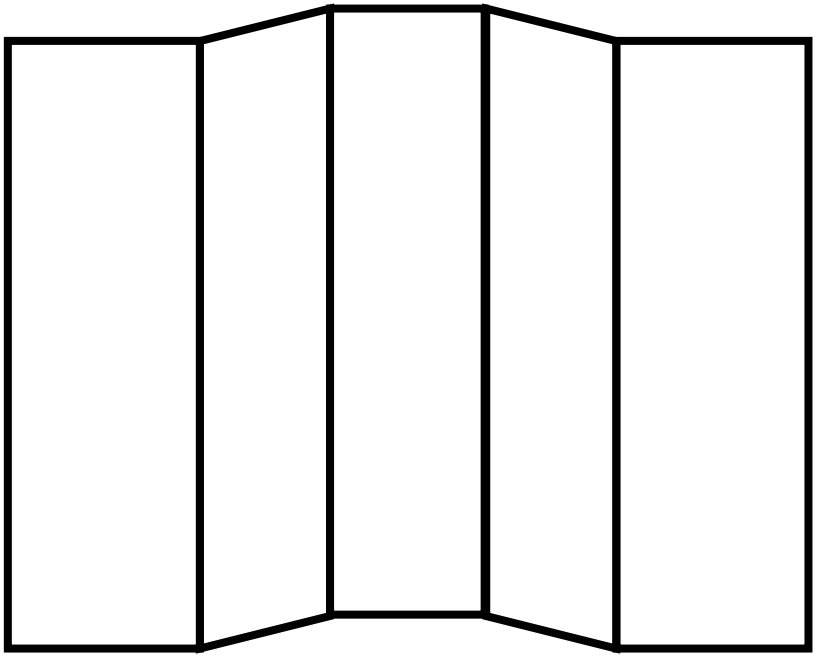
Maximum *a Posteriori* (MAP)

Choose the S = (slant,tilt) that maximizes the posterior.

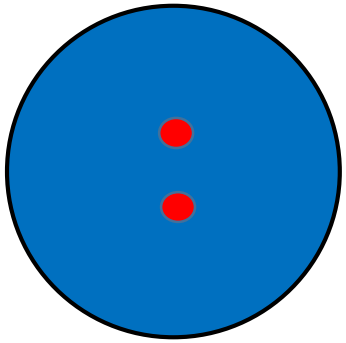


posterior likelihood prior

Likelihood functions can have more than one maximum.



overall
(slant, tilt)

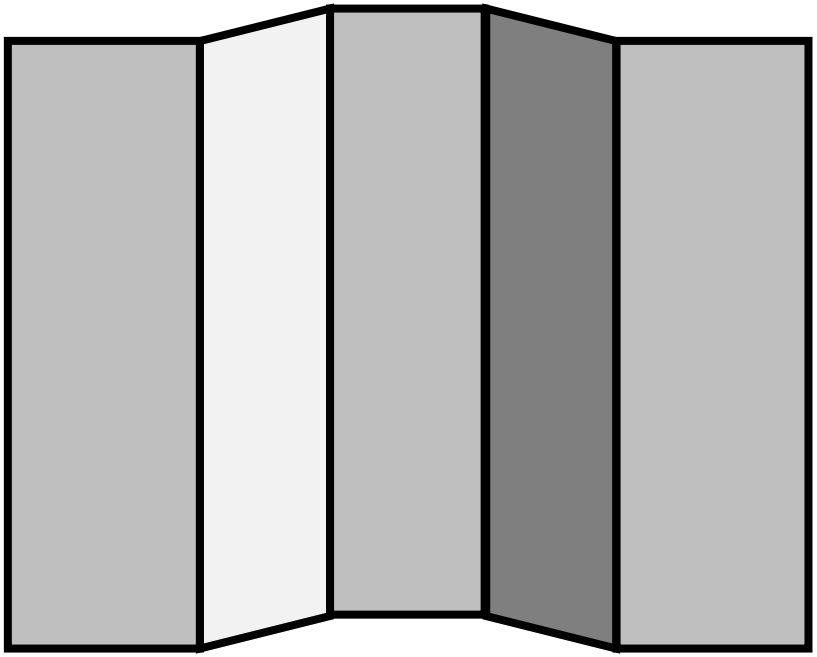


$p(I = i \mid S)$

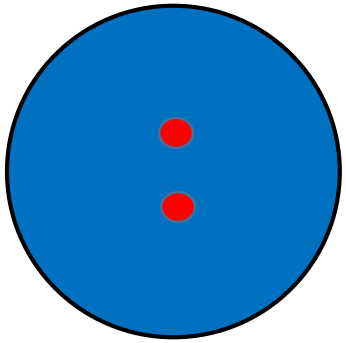
i.e. convex or concave ?

Depth Reversal Ambiguity and Shading

(see Exercise)



Likelihood
(slant, tilt)

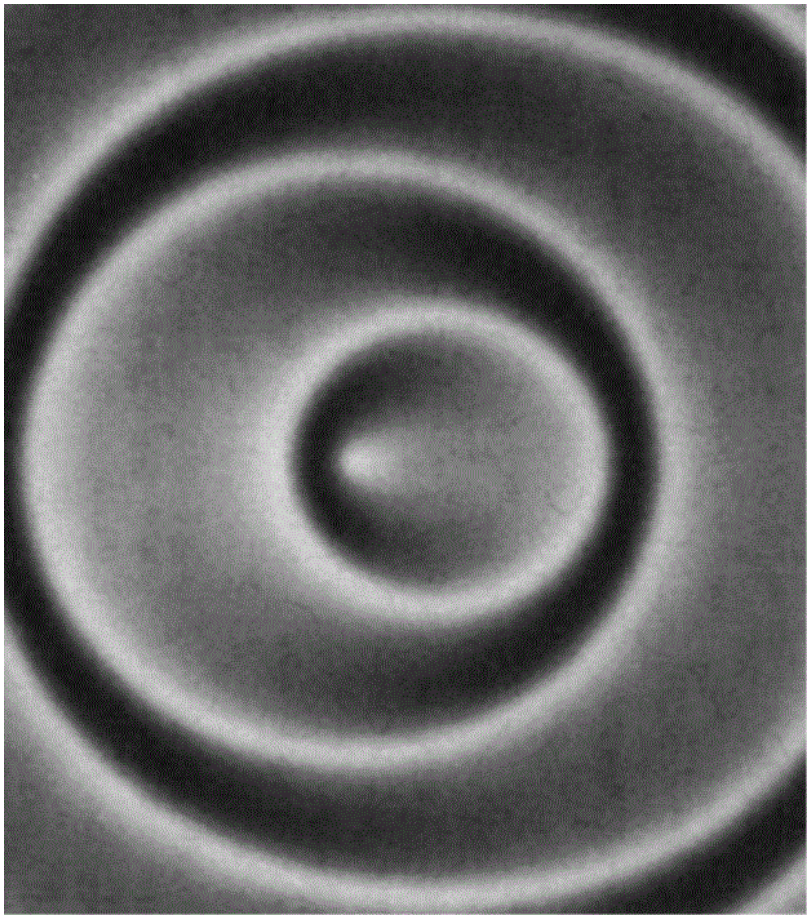


$p(I = i \mid S)$

A valley illuminated from the right produces the same shading as a hill illuminated from the left.

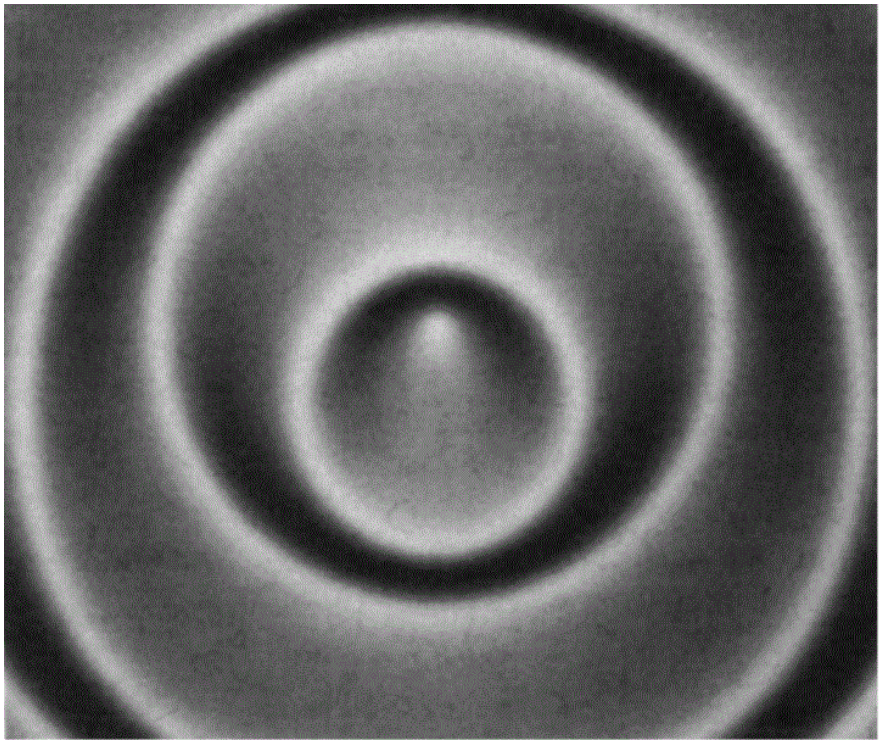
What “priors” does the visual system use to resolve such twofold ambiguities ?

Let’s look at a few related examples.



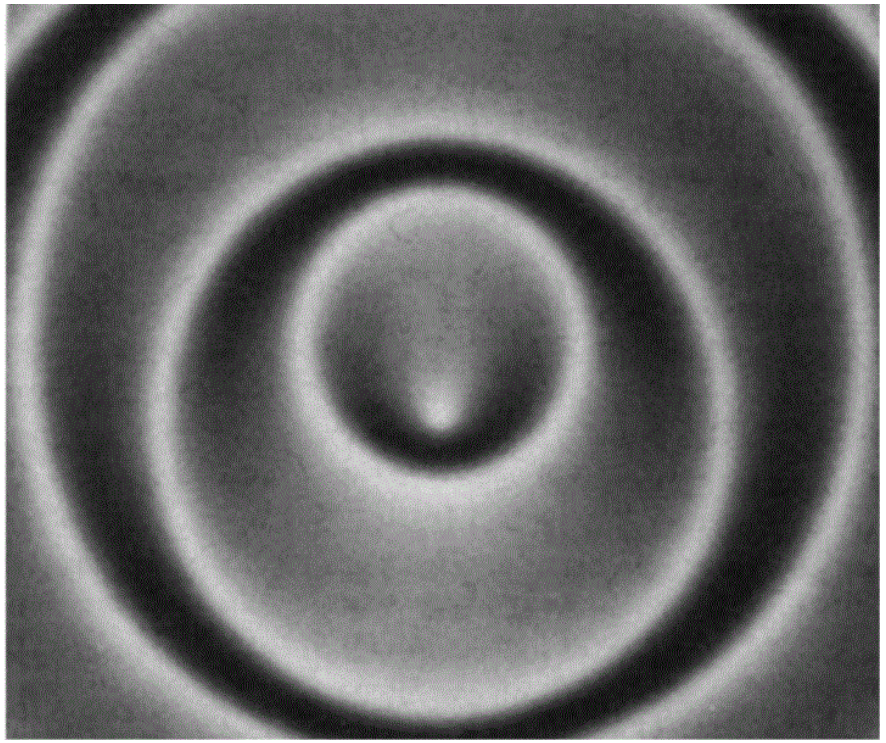
You can perceive the center point as a hill or a valley.

When you see it as a hill, you perceive the tilt as 180 deg (leftward).
But when you see it as a valley, the slant is 0 (rightward).



We tend to see the center as a hill.

Why ?



We tend to see the center as a valley.

Why ?

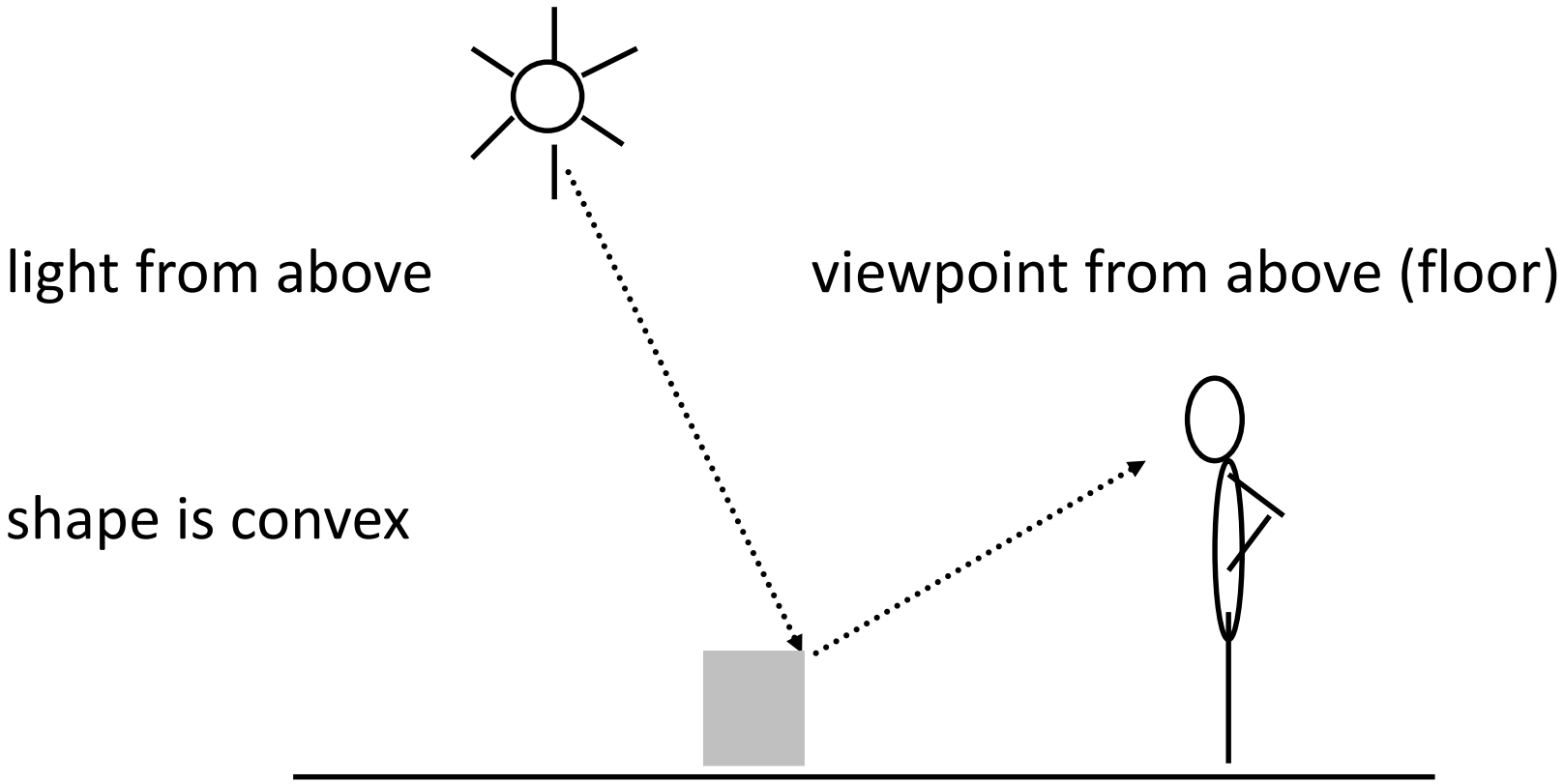
The visual system uses three priors to resolve the depth reversal ambiguity:

- surface orientation: p(floor) > p(ceiling)

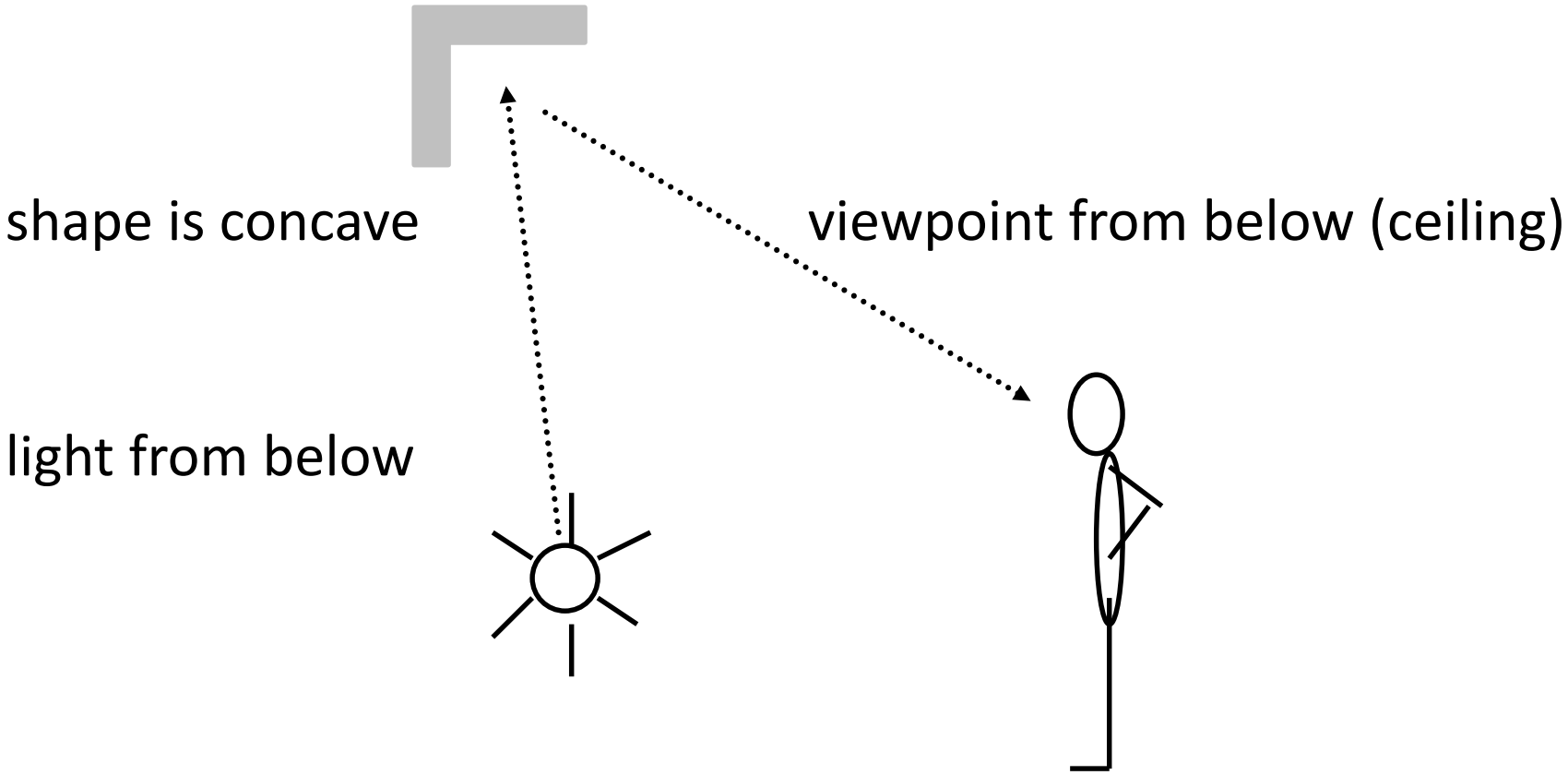
- light source direction: p(above) > p(below)

- ‘global’ surface curvature: p(convex) > p(concave)

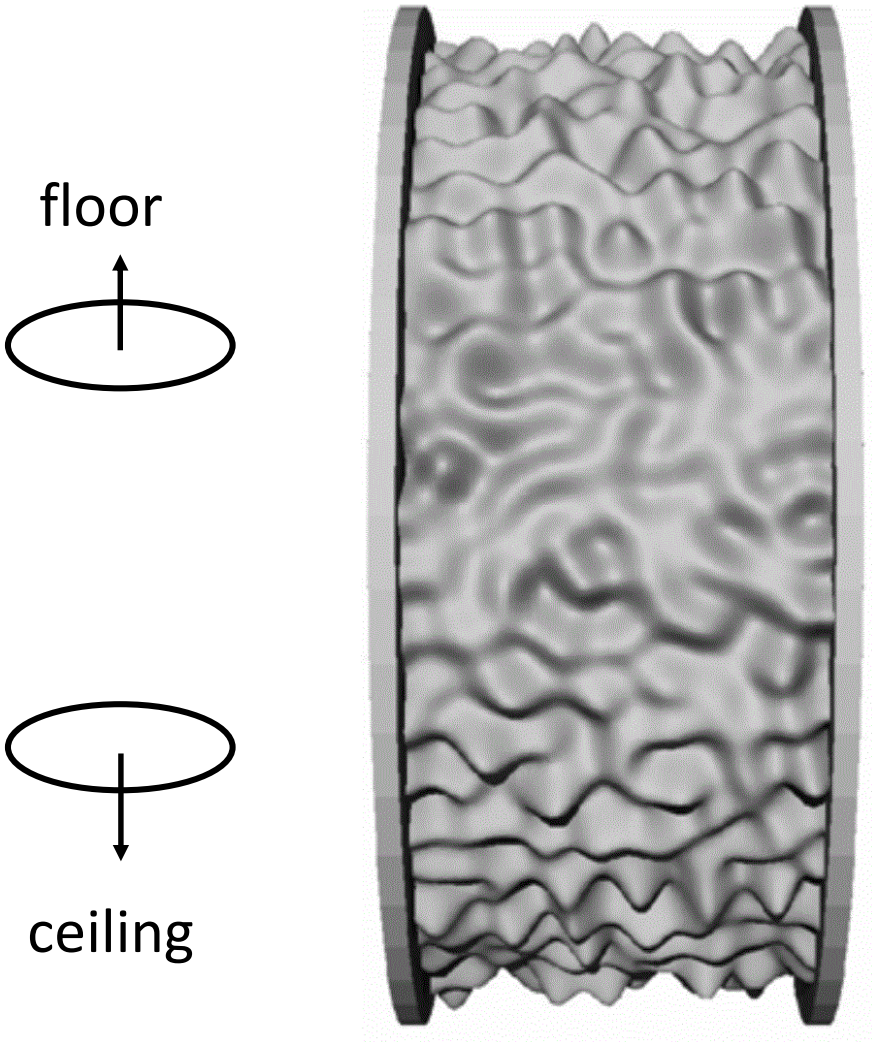
Example in which all three priors assumptions are met



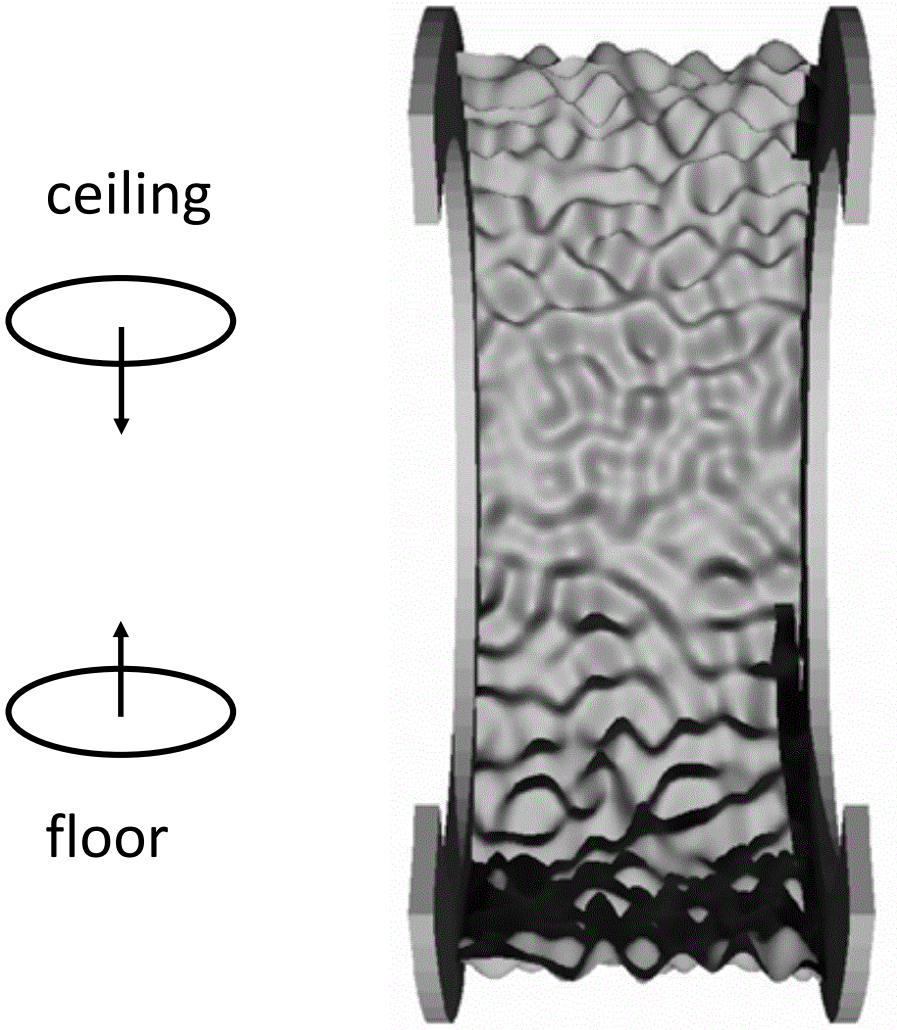
Example in which all three prior assumptions fail



Convex shape, illuminated from above the line of sight



Concave shape, illuminated from below the line of sight



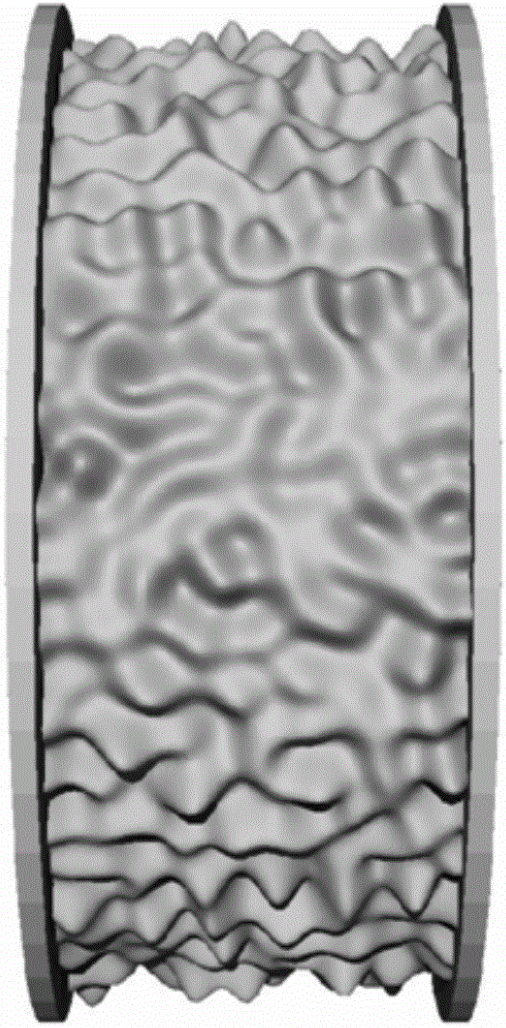
We showed how people combined the three different "priors":

Percent correct in judging local "hill" or "valley":

=	50	+/- 10	floor vs. ceiling
		+/- 10	light from above vs. below
		+/- 10	globally convex/concave

[Langer and Buelthoff, 2001]

Best
(80%)



Worst
(20%)

These look weird, but in different ways. How ?

Reminder

- A2 is due tonight
- Midterm (optional) is first class after Study Break