Bayesian modeling of behavior

Wei Ji Ma

New York University
Center for Neural Science and Department of Psychology

Teaching assistants

Group 1: Anna Kutschireiter: postdoc at University of Bern
Group 2: Anne-Lene Sax: PhD student at University of Bristol
Group 3: Jenn Laura Lee: PhD student at New York University
Group 4: Jorge Menéndez: PhD student at University College London
Group 5: Julie Lee: PhD student at University College London
Group 6: Lucy Lai: PhD student at Harvard
Group 7: Sashank Pisupati: PhD student at Cold Spring Harbor Laboratory

WiFi:

- Cosyne2019 Password: lisboa2y9, **or**
- Username: epicsana19, Password: epicsana19

Go to www.cns.nyu.edu/malab/courses.html

Download Exercises.pdf (link top right)

If you want these slides, also download Slides.pdf

What is your background?

Why are you here?

Why do we fit models?

“The purpose of models is not to fit the data but to sharpen the questions.”

— Samuel Karlin, R.A. Fisher Memorial Lecture, 1983

“If a principled model with a few parameters fits a rich behavioral data set, I feel that I have really understood something about the world” — Wei Ji Ma, Cosyne Tutorial, 2019

Why do we fit models?

From Ma lab survey by Bas van Opheusden, 201703

Why do we fit models?

Maslow's hammer

Because we can and we are good at it

To get into a higher impact journal

Because Weiji says so

To drown a conceptually uninteresting question in math

From Ma lab survey by Bas van Opheusden, 201703

Why do we fit models?

To make inferences about latent causes of behavior that we cannot observe directly

To get closer to a simplified form of people's cognitive processes

Because we want to infer latent variables/mechanisms

To say something about the potential computations involved when completing a task

Infer what's really happening inside the black box

To create order in the universe

Models let us ask questions that are hard to answer with experiments

To quantify evidence for our theories and hypotheses

To produce good models according to well-considered criteria

From Ma lab survey by Bas van Opheusden, 201703

Schedule for today

		Concept
12:10-13:10	• Why Bayesian modeling	
	• Bayesian explanations for illusions	priors
	• Case 1: Gestalt perception	likelihoods
	• Case 2: Motion sickness	prior/likelihood interplay
13:30-14:40	• Case 3: Color perception	nuisance parameters
	• Case 4: Sound localization	measurement noise
	• Case 5: Change point detection	hierarchical inference
15:00-16:00	• Model fitting and model comparison	
	• Critiques of Bayesian modeling	

Schedule for today

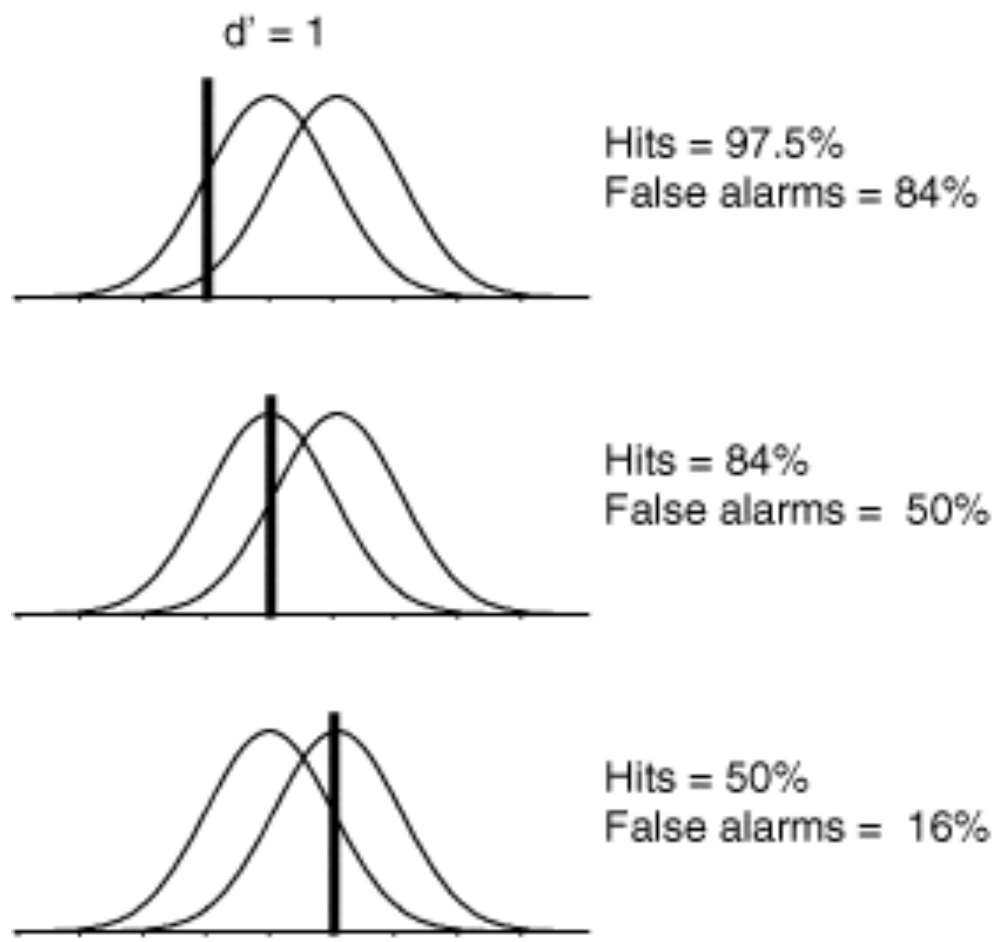
Time	Topic	*Concept*
12:10-13:10	• Why Bayesian modeling	
	• Bayesian explanations for illusions	priors
	• Case 1: Gestalt perception	likelihoods
	• Case 2: Motion sickness	prior/likelihood interplay
13:30-14:40	• Case 3: Color perception	nuisance parameters
	• Case 4: Sound localization	measurement noise
	• Case 5: Change point detection	hierarchical inference
15:00-16:00	• Model fitting and model comparison	
	• Critiques of Bayesian modeling	

Two kinds of models

- **Descriptive model:** describe the data using a function with parameters
  - E.g. neural networks
  - Danger: arbitrarily throwing parameters at it, problems with understanding and generalization
- **Process model:** model based on psychological hypotheses about the process by which a mind makes a decision
  - Usually few parameters
  - Interpretable!
  - Potentially not as powerful

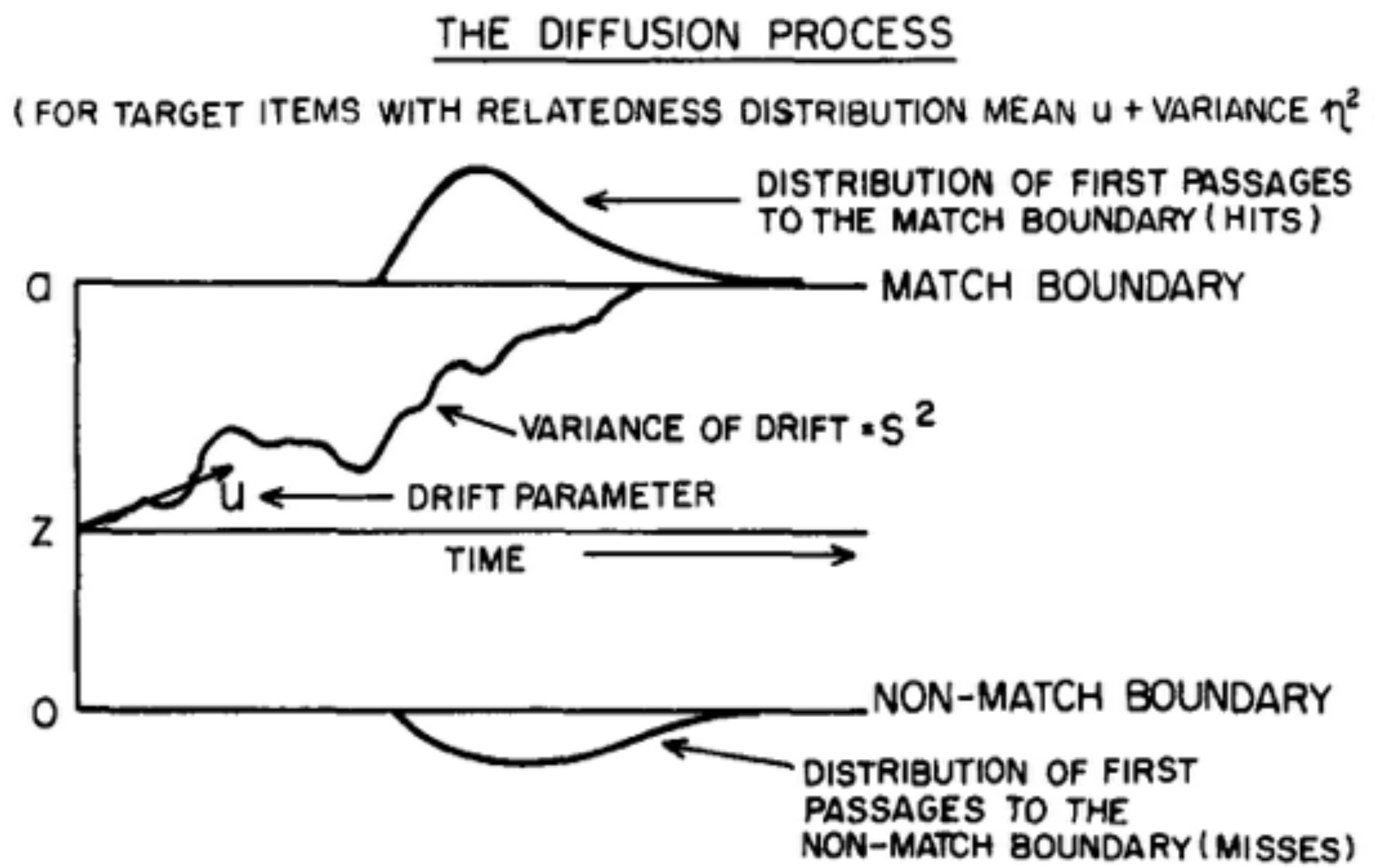
Process models

- Signal detection theory



David Heeger lecture notes

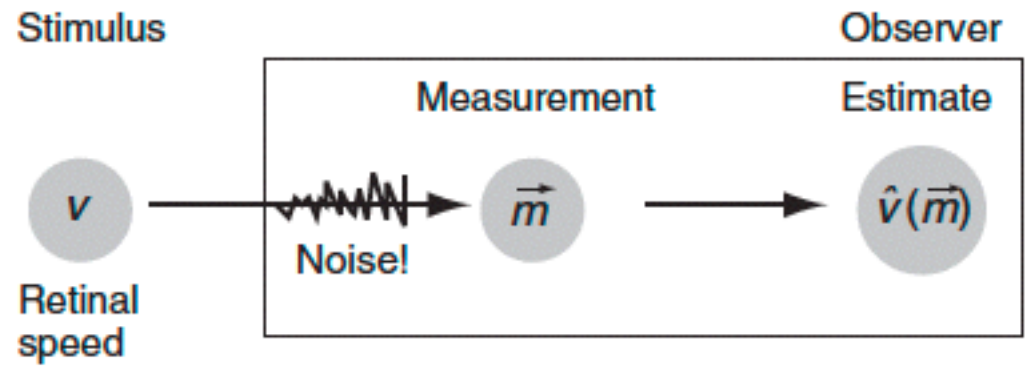
- Drift-diffusion model



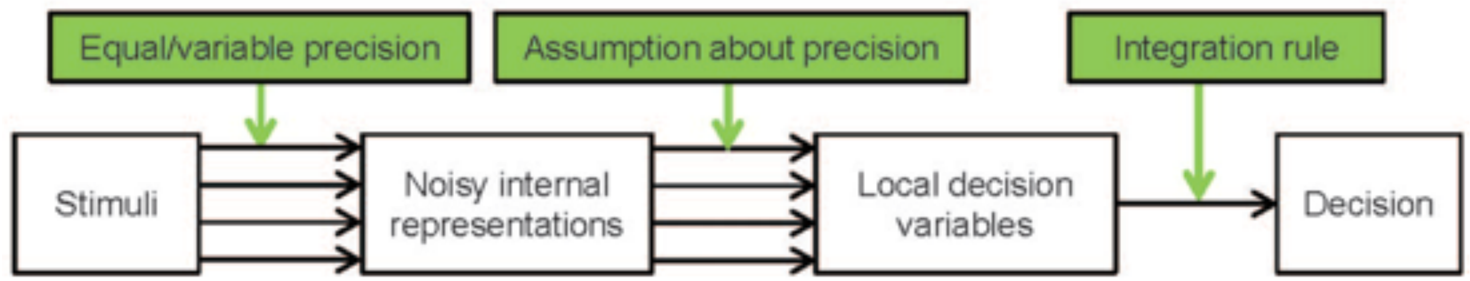
Ratcliff 1978

A special kind of process model: Bayesian

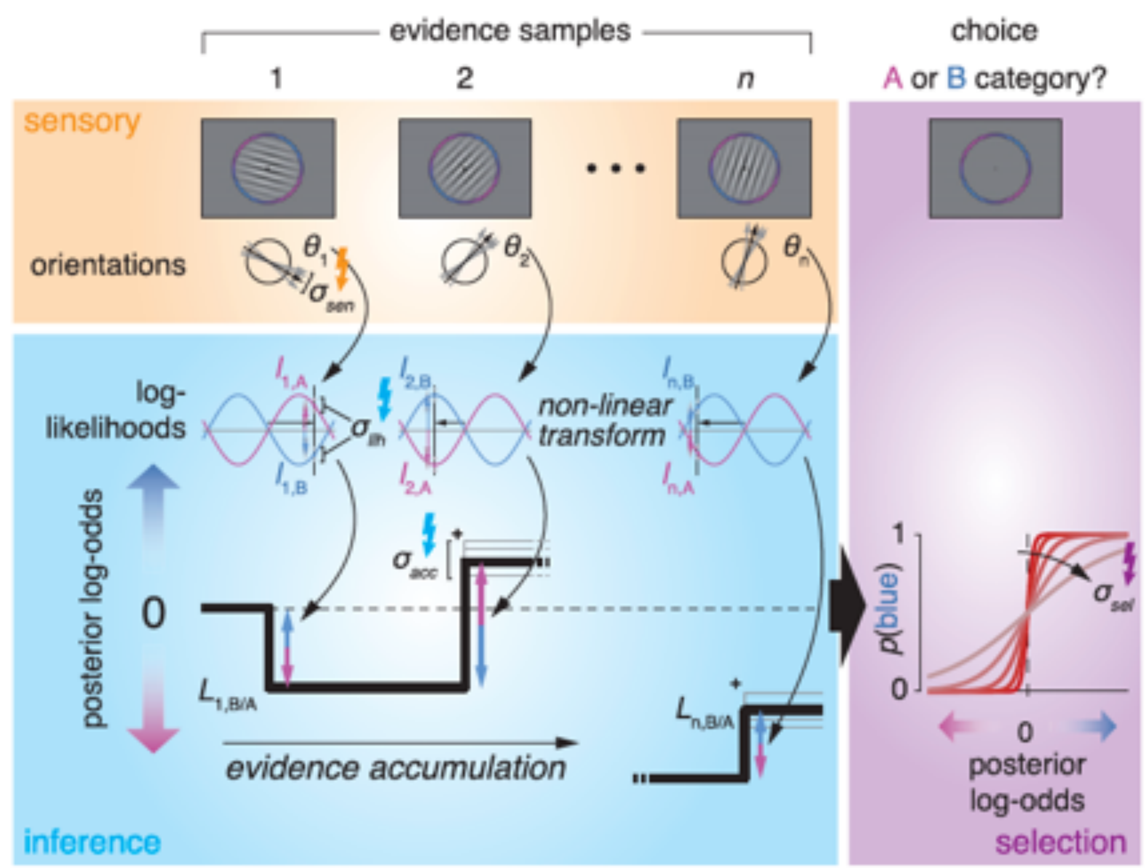
- **State of the world unknown to decision-maker**
  - Uncertainty!
- **Decision-maker maximizes an objective function**
  - In categorical perception: accuracy
  - But could be hitting error, point rewards, survival
- Stronger claim: brain represents probability distributions



Stocker and Simoncelli, 2006



Keshvari et al., 2012



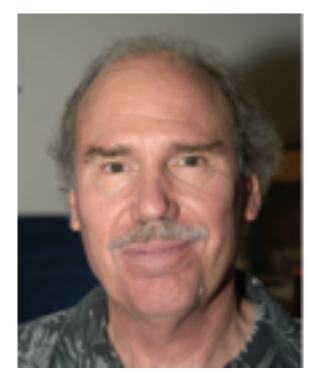
Drugowitsch et al., 2016

Why Bayesian models?

- Evolutionary/normative: Bayesian inference optimizes performance or minimizes cost. The brain might have near-optimized processes crucial for survival.

- Empirical: in many tasks, people are close to Bayesian.

- Bill Geisler's couch argument:



"It is harder to come up with a good model sitting on your couch than to work out the Bayesian model."

- Basis for suboptimal models: Other models can often be constructed by modifying the assumptions in the Bayesian model. Thus, the Bayesian model is a good starting point for model generation.

Where does uncertainty come from?

- Noise
- Ambiguity

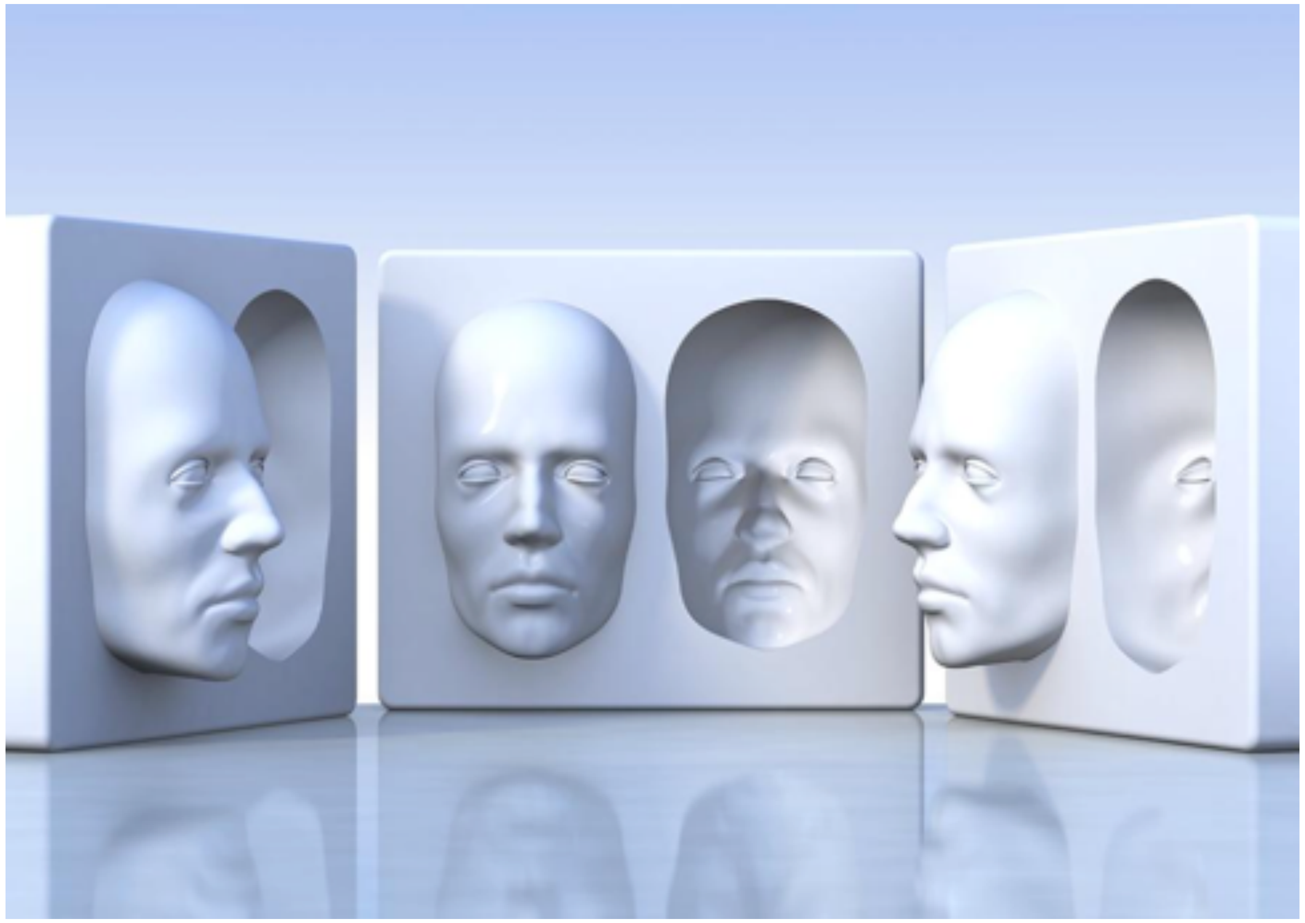
Schedule for today

		Concept
12:10-13:10	• Why Bayesian modeling	
	• Bayesian explanations for illusions	priors
	• Case 1: Gestalt perception	likelihoods
	• Case 2: Motion sickness	prior/likelihood interplay
13:30-14:40	• Case 3: Color perception	nuisance parameters
	• Case 4: Sound localization	measurement noise
	• Case 5: Change point detection	hierarchical inference
15:00-16:00	• Model fitting and model comparison	
	• Critiques of Bayesian modeling	


GATHERING FOR
GARDNER


CUT OUT, FOLD, AND ASSEMBLE AS ILLUSTRATED.
VIEW FROM 2'... CLOSE ONE EYE LOOK INTO THE EYES OF THE CREATURE AND ROCK BACK AND FORTH. WOW!
VALLEY FOLD
MOUNTAIN FOLD
FOLD TO HERE
CUT TO WHITE DOT
GATHERING FOR
GARDNER
BINARY ARTS
Phone: 703-549-4999 Fax: 703-549-6210
MOUNTAIN FOLD TO WHITE DOT
MOUNTAIN FOLD TO ARROW
TAB A
TAB B

Hollow-face illusion

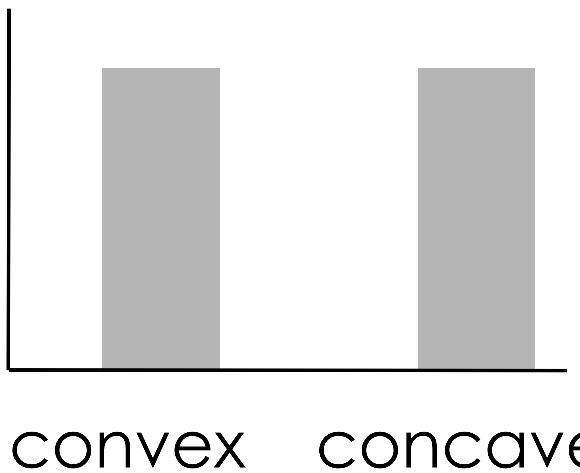


David Mack

Why do we see the dragon/the hollow face as convex?

Likelihood

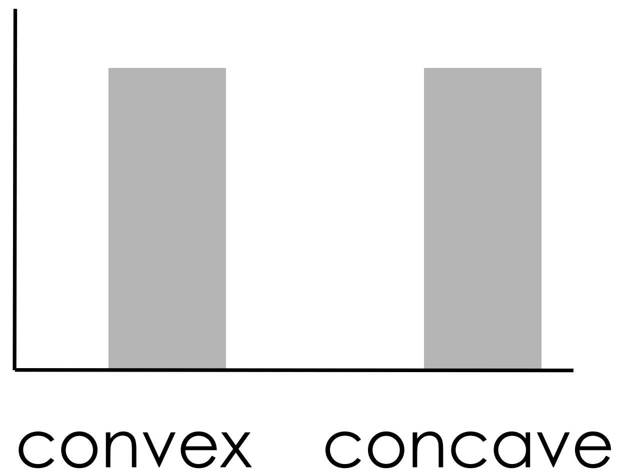
how probable are the retinal image is if the hypothesis were true

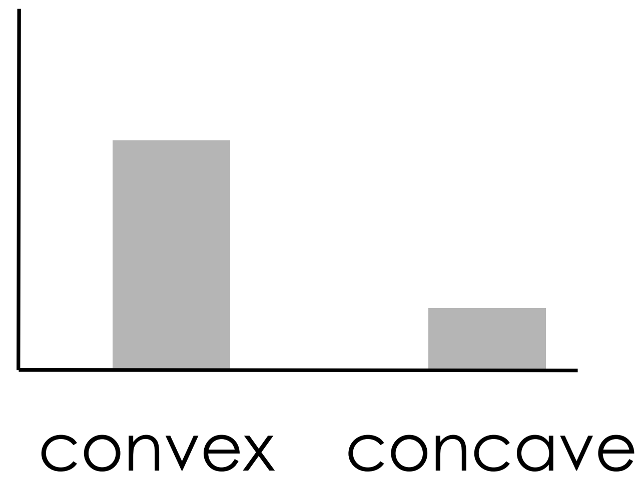


Likelihood x Prior

how probable are the retinal image is if the hypothesis were true

how much do you expect the hypothesis based on your experiences

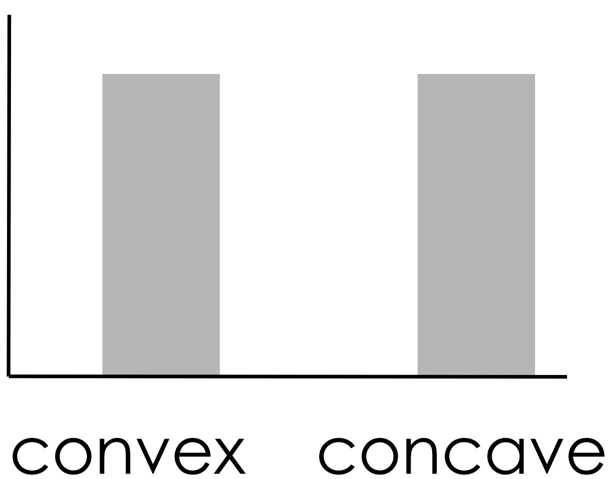


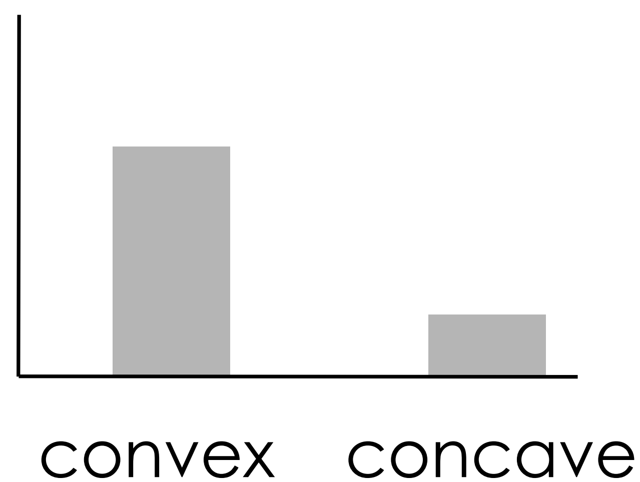


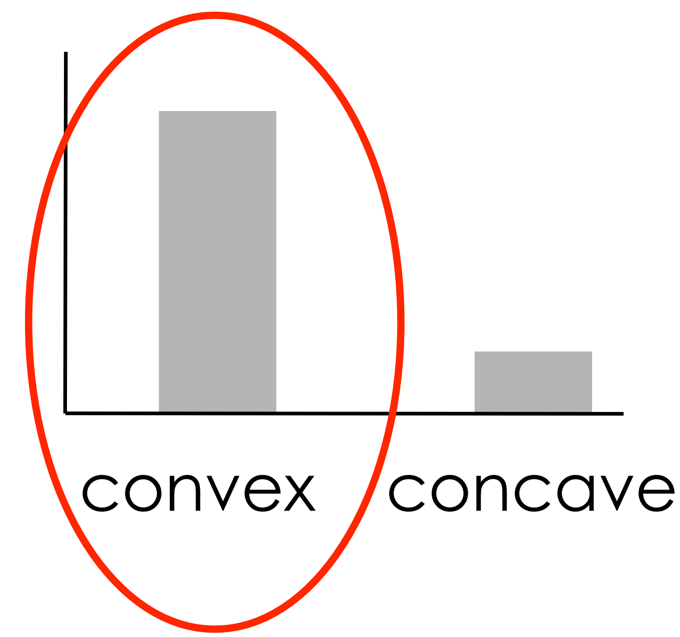
Likelihood x Prior ∝ Posterior probability

how probable are the retinal image is if the hypothesis were true

how much do you expect the hypothesis based on your experiences



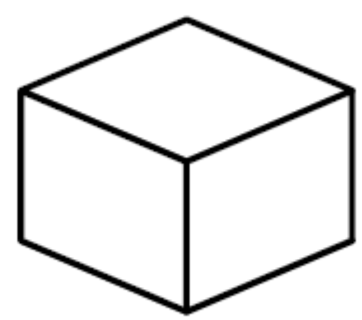




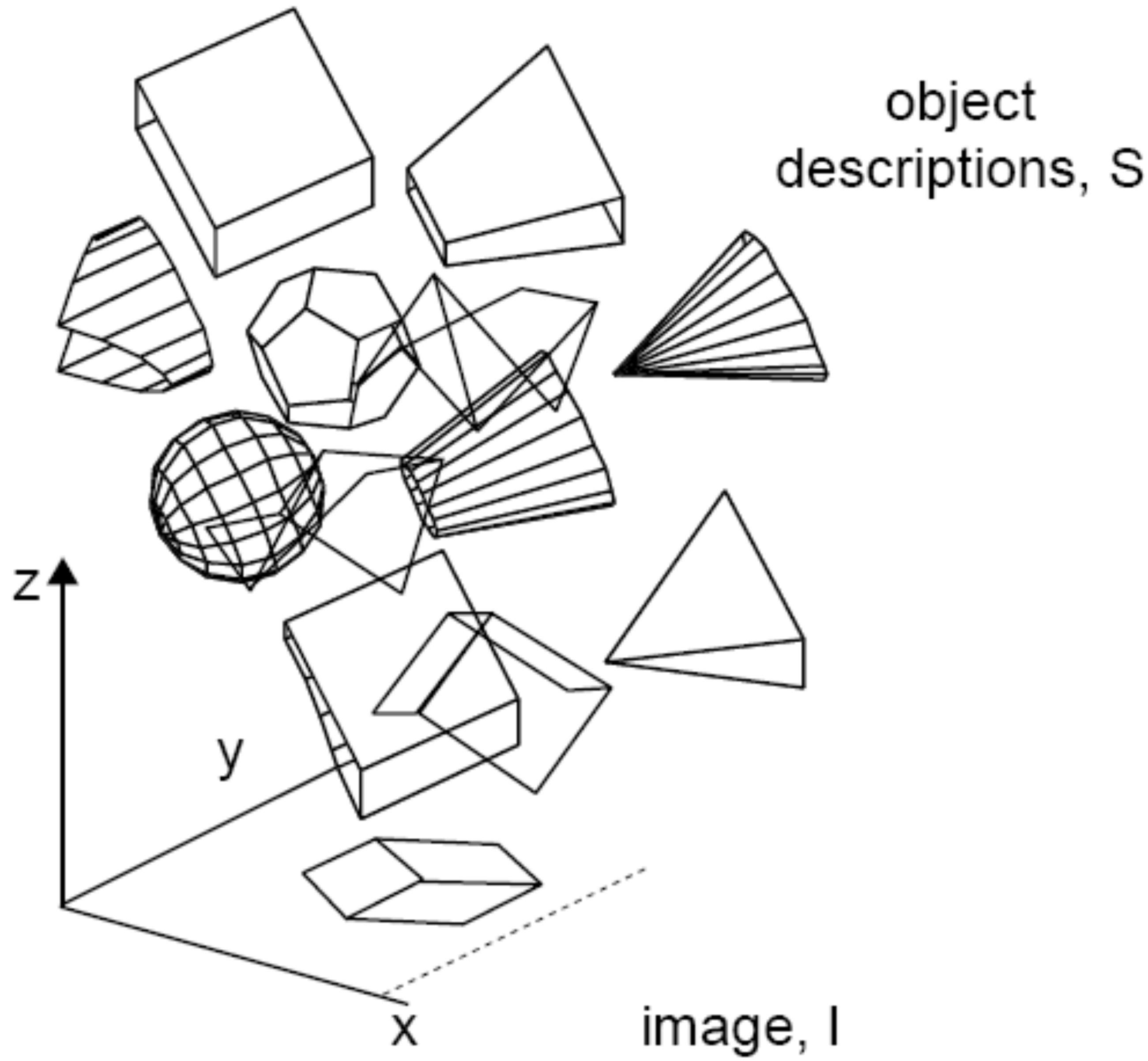
This hypothesis becomes your percept!

Anamorphic illusion by Kurt Wenner

- Where is the ambiguity?
- What role do priors play?
- What happens if you view with two eyes, and why?

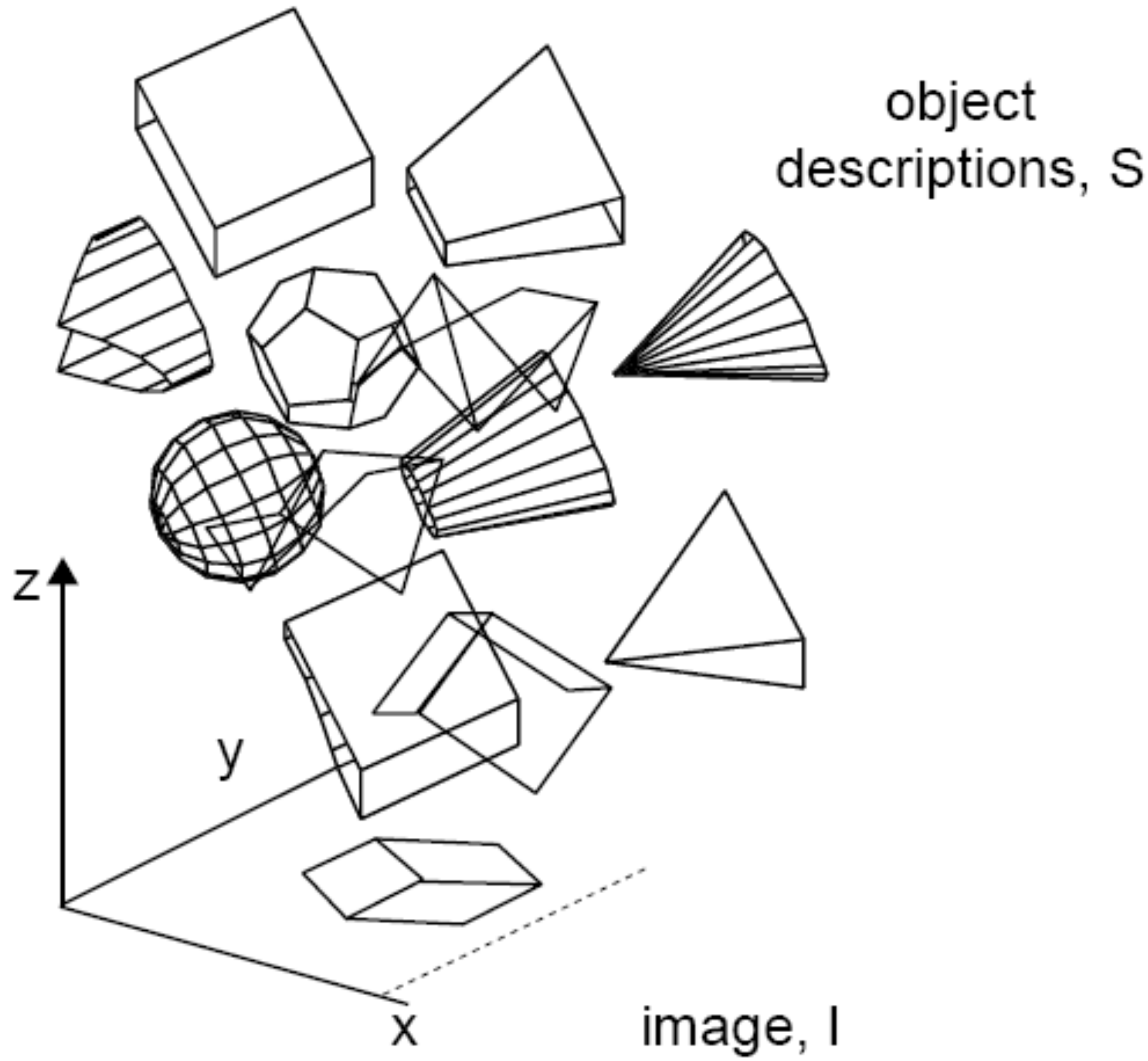


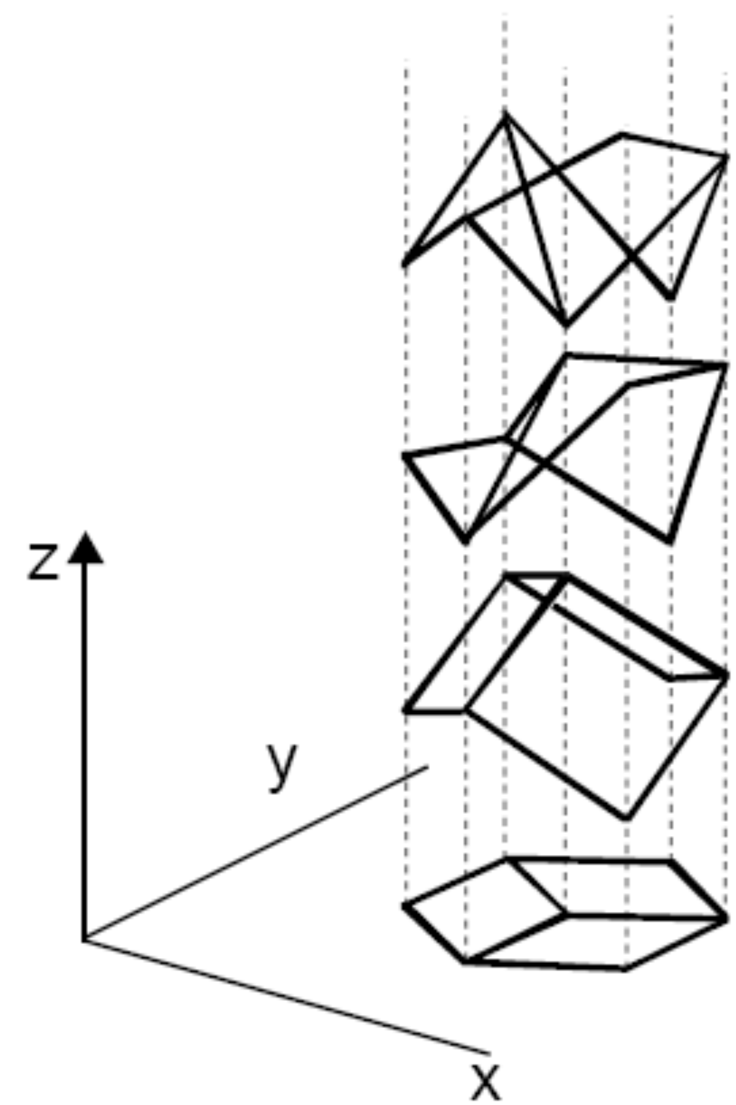
Prior over objects
$p(s)$



Prior over objects
$p(s)$

Likelihood over objects given 2D image
$L(s) = p(I|s)$





Kersten and Yuille, 2003

Examples of priors:

- Convex faces are more common than concave ones
- Priors at the object level (Kersten and Yuille)
- Light usually comes from above (Adams and Ernst)
- Slower speeds are more common (Simoncelli and Stocker)
- Cardinal orientations are more common (Landy and Simoncelli)

Take-home messages from these illusions:

- Illusions are not just “bugs in the system”, they may be the product of an inference machine trying to do the right thing.
- “Wrong” percepts can often be traced back to a prior.

~~Bayesian models are about priors~~

Fake news

Bayesian models are about:

- the decision-maker making the best possible decision (given an objective function)
- the brain representing probability distributions

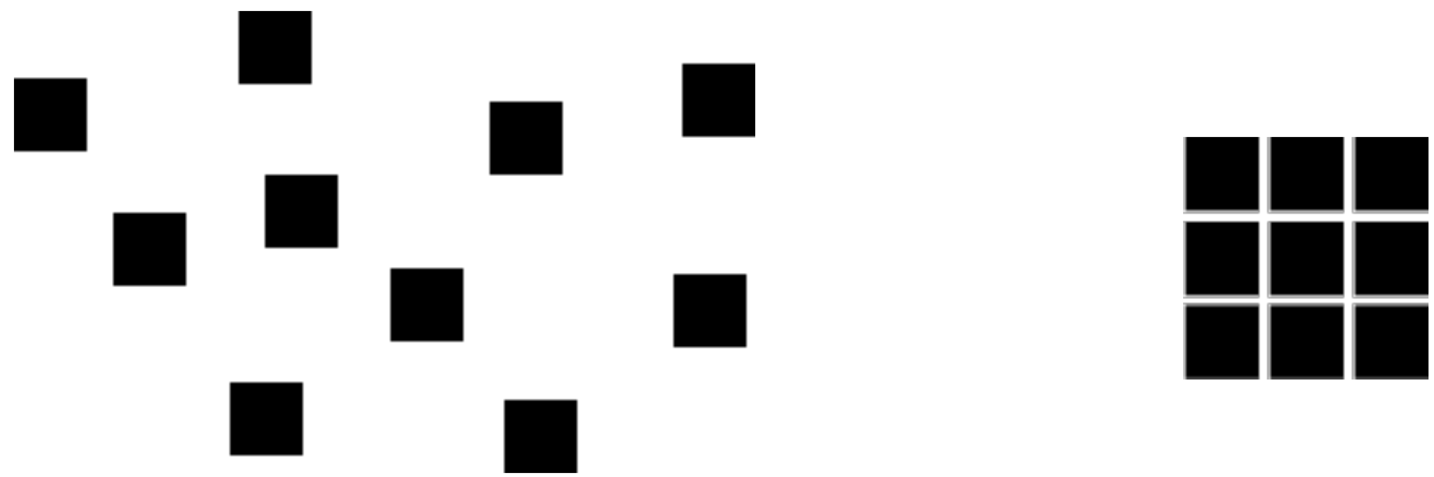
Schedule for today

		Concept
12:10-13:10	• Why Bayesian modeling	
	• Bayesian explanations for illusions	priors
	• Case 1: Gestalt perception	likelihoods
	• Case 2: Motion sickness	prior/likelihood interplay
13:30-14:40	• Case 3: Color perception	nuisance parameters
	• Case 4: Sound localization	measurement noise
	• Case 5: Change point detection	hierarchical inference
15:00-16:00	• Model fitting and model comparison	
	• Critiques of Bayesian modeling	

OKGo, *The writing's on the wall.* Music video by Aaron Duffy and 1stAveMachine

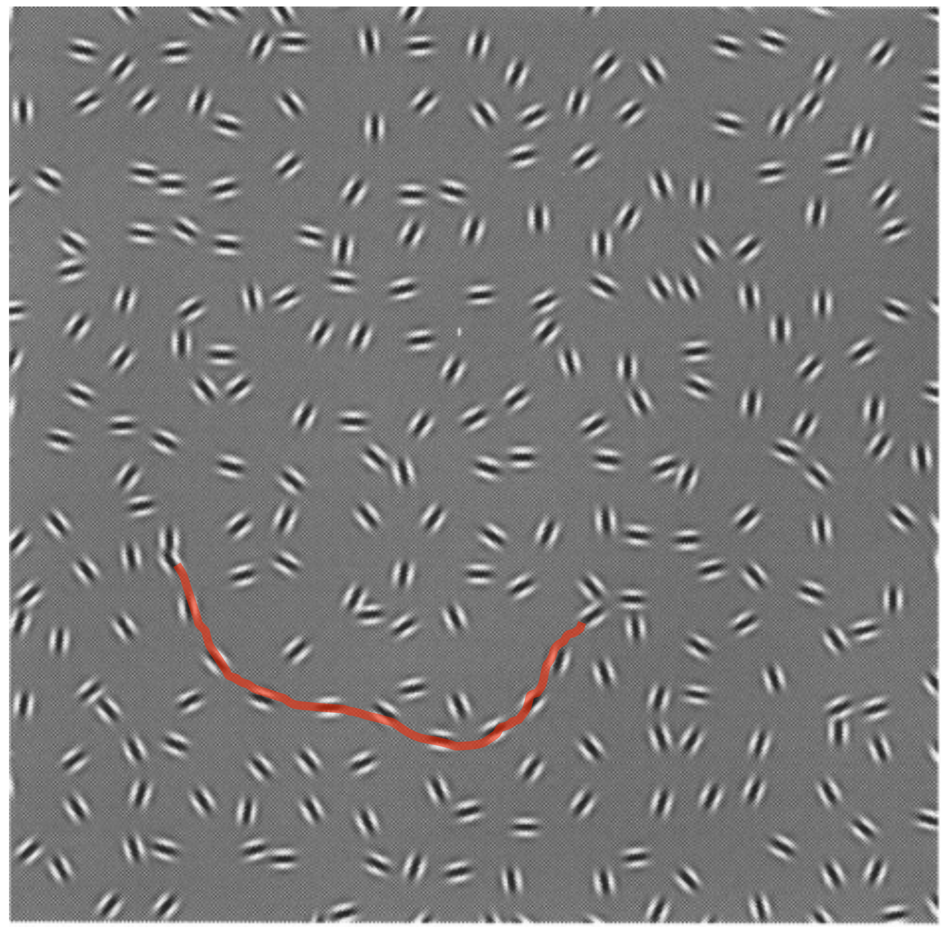
How would a perceptual psychologist describe this kind of percept?

Law of proximity



“Spatial or temporal proximity of elements may induce the mind to perceive a collective entity.”

Law of continuity



Field, Hayes, Hess

"Elements that are aligned tend to be grouped together."

Law of common fate

"When elements move in the same direction, we tend to perceive them as a collective entity."

Bayesian account of Gestalt percepts?

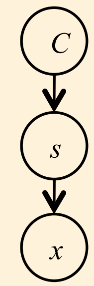
Open Case 1 on page 3

The four steps of Bayesian modeling

Example: categorization task

STEP 1: GENERATIVE MODEL

a) Draw a diagram with each node a variable and each arrow a statistical dependency. Observation is at the bottom.
b) For each variable, write down an equation for its probability distribution. For the observation, assume a noise model. For others, get the distribution from your experimental design. If there are incoming arrows, the distribution is a conditional one.



World state of interest
$$p(C) = 0.5$$

Stimulus
$$p(s \mid C) = \mathrm{N}\left(s; \mu_C, \sigma_C^2\right)$$

Observation
$$p(x \mid s) = \mathrm{N}\left(x; s, \sigma^2\right)$$

STEP 2: BAYESIAN INFERENCE (DECISION RULE)

a) Compute the posterior over the world state of interest given an observation. The optimal observer does this using the distributions in the generative model. Alternatively, the observer might assume different distributions (natural statistics, wrong beliefs). Marginalize (integrate) over variables other than the observation and the world state of interest.

$$p(C \mid s) \propto p(C)\, p(x \mid C) = p(C) \int p(x \mid s)\, p(s \mid C)\, ds = \ldots = \mathrm{N}\left(x; \mu_C, \sigma^2 + \sigma_C^2\right)$$

b) Specify the read-out of the posterior. Assume a utility function, then maximize expected utility under posterior. (Alternative: sample from the posterior.) Result: decision rule (mapping from observation to decision). When utility is accuracy, the read-out is to maximize the posterior (MAP decision rule).

$$\hat{C} = 1 \text{ when } \mathrm{N}\left(x; \mu_1, \sigma^2 + \sigma_1^2\right) > \mathrm{N}\left(x; \mu_2, \sigma^2 + \sigma_2^2\right)$$

STEP 3: RESPONSE PROBABILITIES

For every unique trial in the experiment, compute the probability that the observer will choose each decision option given the stimuli on that trial using the distribution of the observation given those stimuli (from Step 1) and the decision rule (from Step 2).

$$p\left(\hat{C} = 1 \mid x\right) = \Pr_{x \mid s; \sigma}\left(\mathrm{N}\left(x; \mu_1, \sigma^2 + \sigma_1^2\right) > \mathrm{N}\left(x; \mu_2, \sigma^2 + \sigma_2^2\right)\right)$$

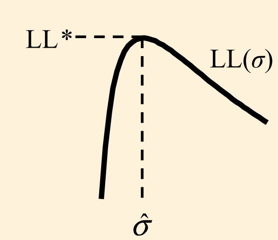
- Good method: sample observation according to Step 1; for each, apply decision rule; tabulate responses. Better: integrate numerically over observation. Best (when possible): integrate analytically.
- Optional: add response noise or lapses.

STEP 4: MODEL FITTING AND MODEL COMPARISON

a) Compute the parameter log likelihood, the log probability of the subject's actual responses across all trials for a hypothesized parameter combination.

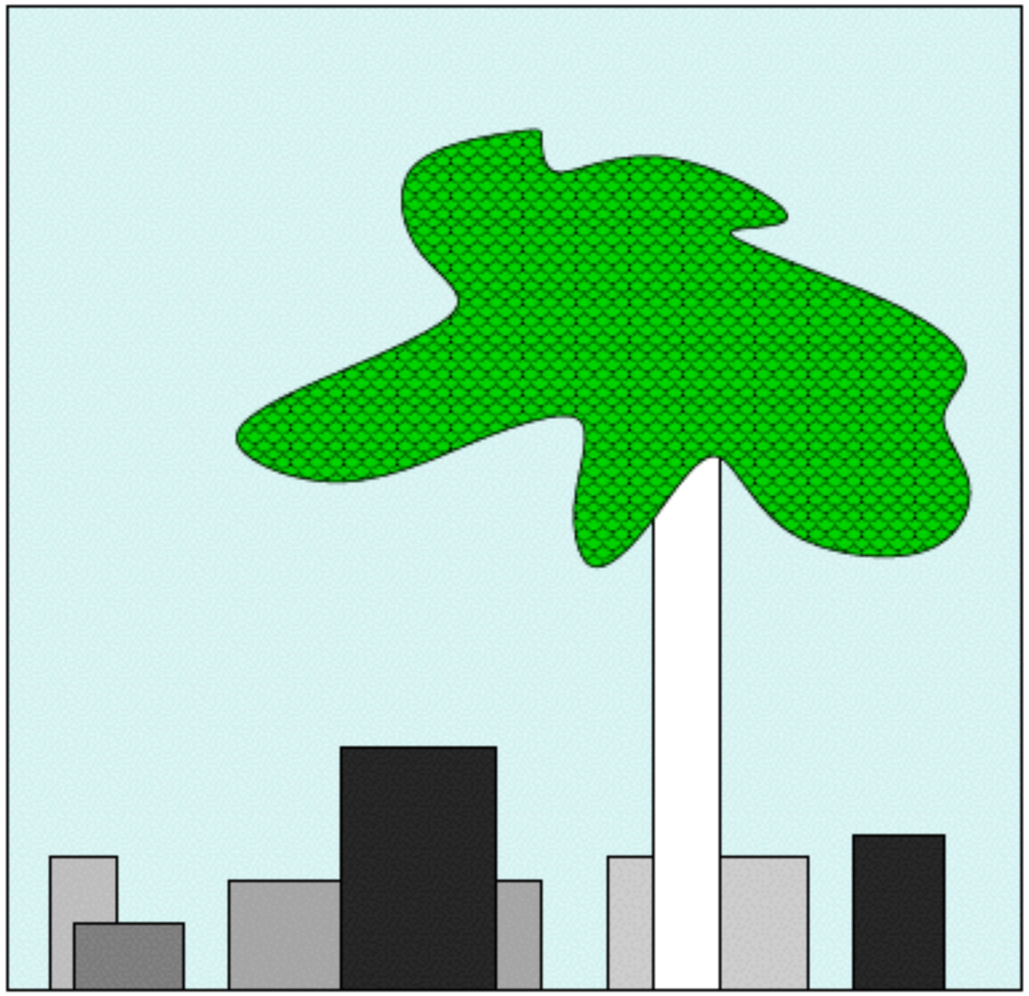
$$\mathrm{LL}(\sigma) = \sum_{i=1}^{\#\text{trials}} \log p\left(\hat{C}_i \mid s_i; \sigma\right)$$

b) Maximize the parameter log likelihood. Result: parameter estimates and maximum log likelihood. Test for parameter recovery and summary statistics recovery using synthetic data.
c) Obtain fits to summary statistics by rerunning the fitted model.
d) Formulate alternative models (e.g. vary Step 2). Compare maximum log likelihood across models. Correct for number of parameters (e.g. AIC). (Advanced: Bayesian model comparison, uses log marginal likelihood of model.) Test for model recovery using synthetic data.
e) Check model comparison results using summary statistics.



Take-home message from Case 1:
With likelihoods like these, who needs priors?

Bayesian models are about the *best possible* decision, not necessarily about priors.



MacKay (2003), *Information theory, inference, and learning algorithms*, Sections 28.1-2

Schedule for today

		Concept
12:10-13:10	• Why Bayesian modeling	
	• Bayesian explanations for illusions	priors
	• Case 1: Gestalt perception	likelihoods
	• Case 2: Motion sickness	prior/likelihood interplay
13:30-14:40	• Case 3: Color perception	nuisance parameters
	• Case 4: Sound localization	measurement noise
	• Case 5: Change point detection	hierarchical inference
15:00-16:00	• Model fitting and model comparison	
	• Critiques of Bayesian modeling	

Motion Sickness: An Evolutionary Hypothesis

Abstract. *Since the occurrence of vomiting as a response to motion is both widespread and apparently disadvantageous, it presents a problem for evolutionary theory. An hypothesis is proposed suggesting that motion sickness is triggered by difficulties which arise in the programming of movements of the eyes or head when the relations between the spatial frameworks defined by the visual, vestibular, or proprioceptive inputs are repeatedly and unpredictably perturbed. Such perturbations may be produced by certain types of motion, or by disturbances in sensory input or motor control produced by ingested toxins. The last would be the important cause in nature, the main function of the emesis being to rid the individual of ingested neurotoxins. Its occurrence in response to motion would be an accidental by-product of this system.*

Michel Treisman, *Science*, 1977

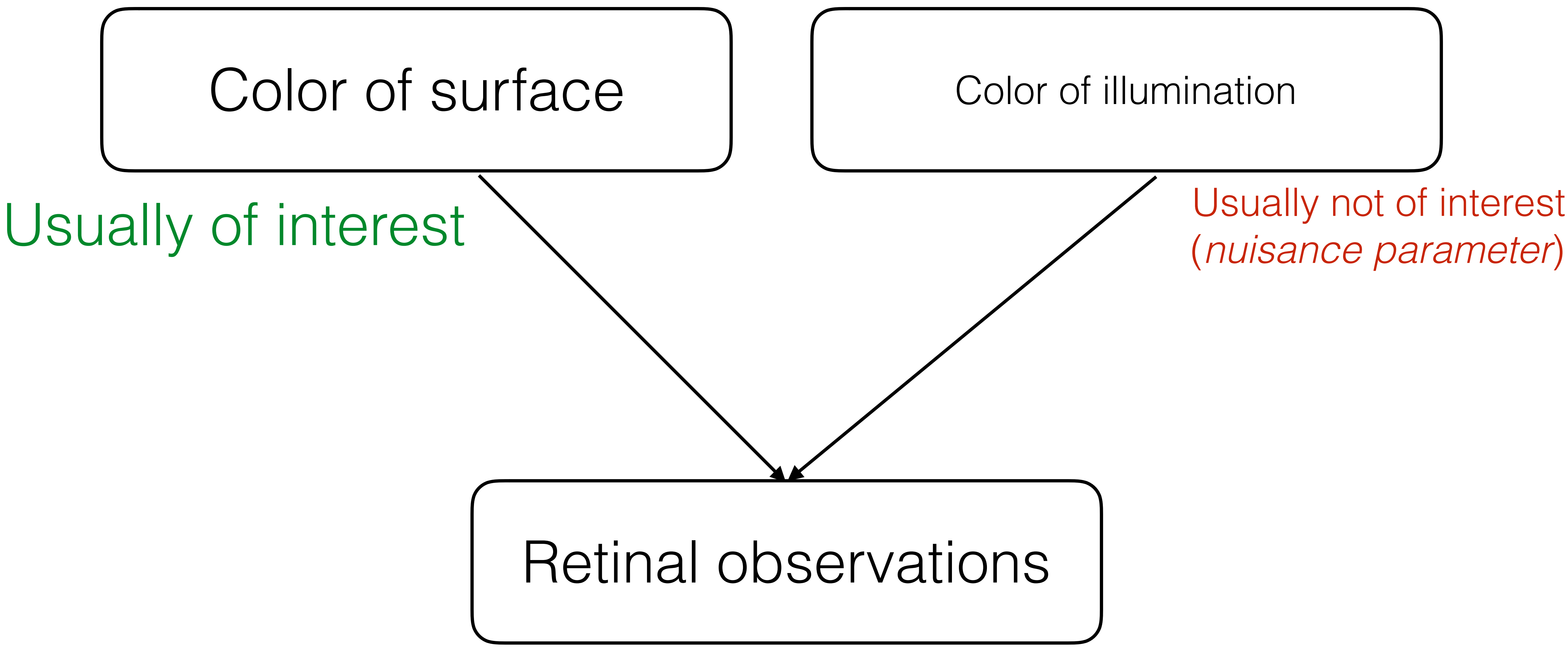
Take-home messages from Case 2:

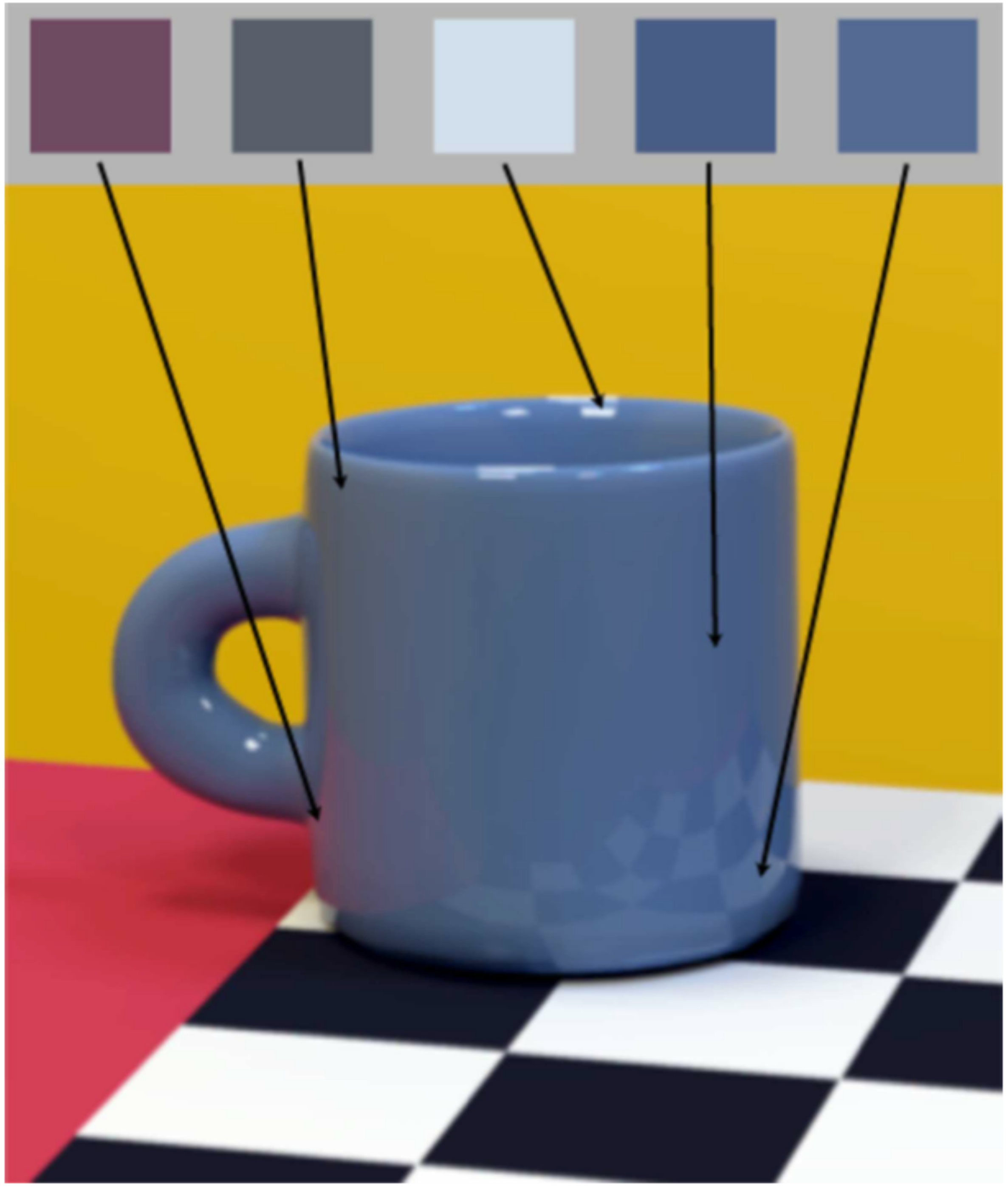
- Likelihoods and priors can compete with each other.
- Where priors come from is an interesting question.

Schedule for today

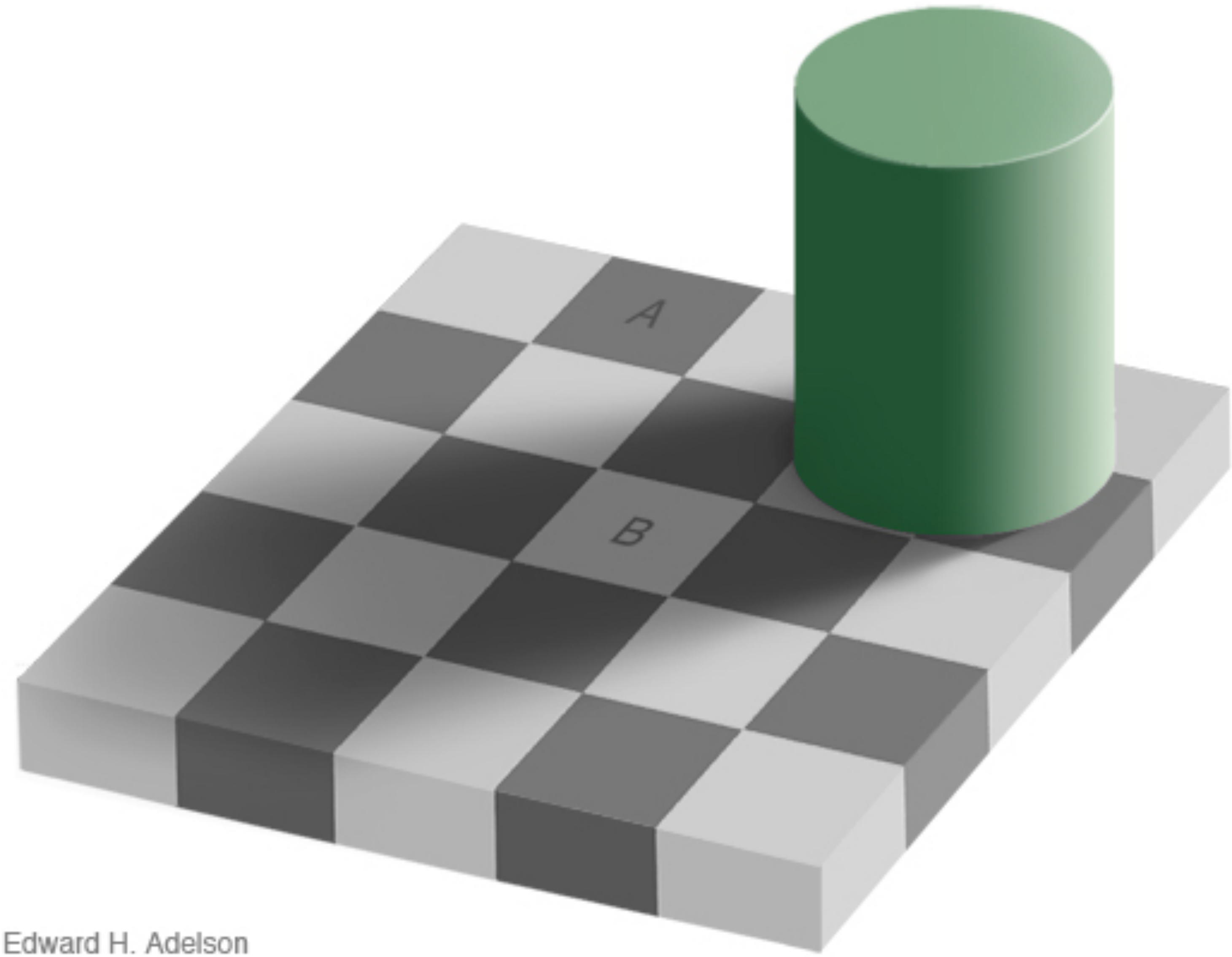
		Concept
12:10-13:10	• Why Bayesian modeling	
	• Bayesian explanations for illusions	priors
	• Case 1: Gestalt perception	likelihoods
	• Case 2: Motion sickness	prior/likelihood interplay
13:30-14:40	• Case 3: Color perception	nuisance parameters
	• Case 4: Sound localization	measurement noise
	• Case 5: Change point detection	hierarchical inference
15:00-16:00	• Model fitting and model comparison	
	• Critiques of Bayesian modeling	

Fundamental problem of color perception

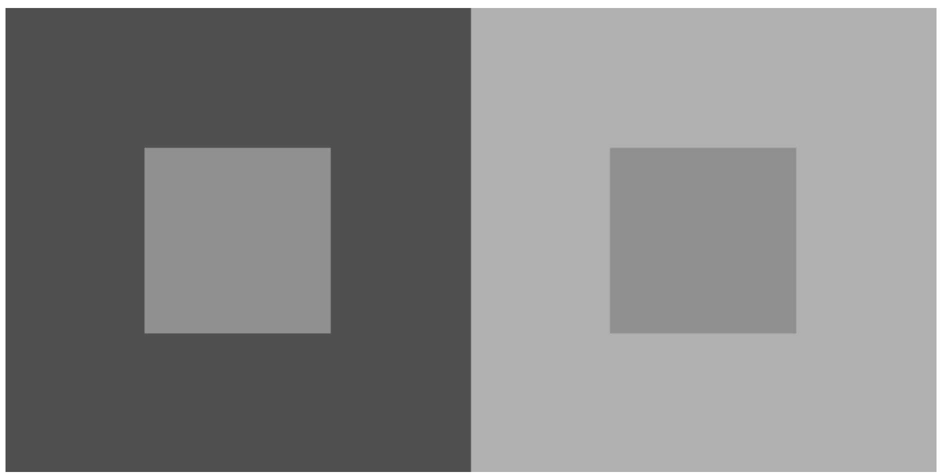




David Brainard

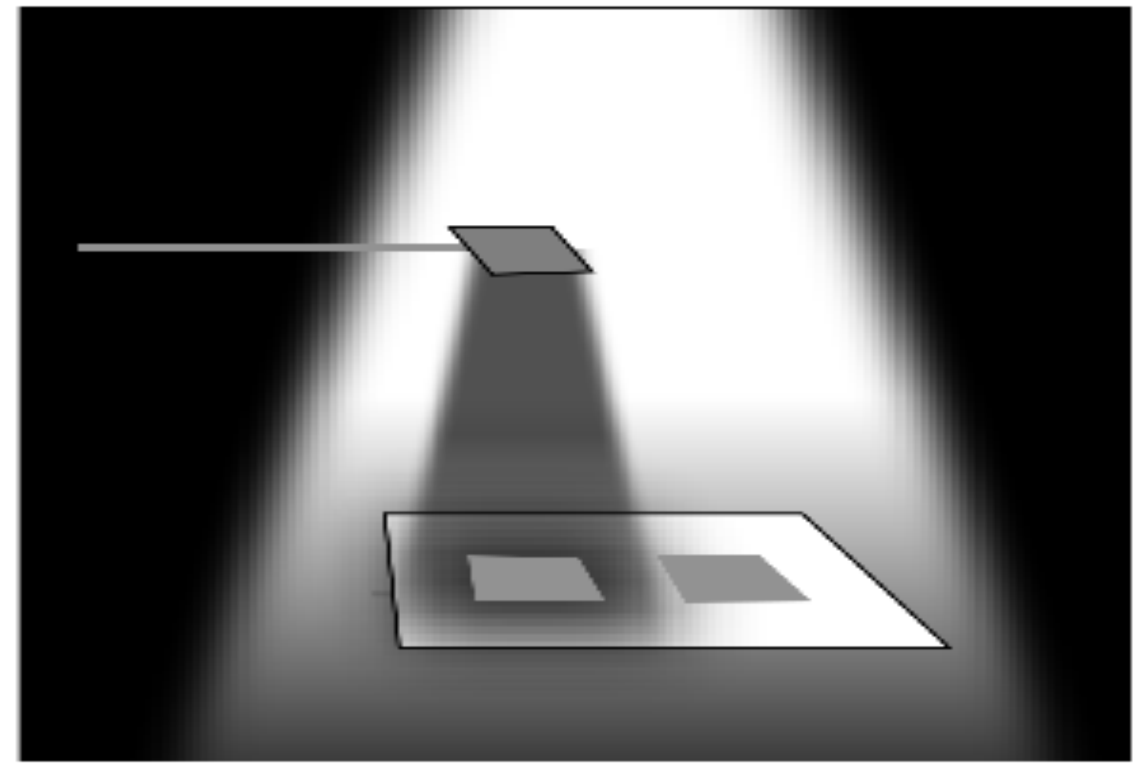


Edward H. Adelson



Light patch in
dim illumination

Dark patch in
bright illumination



Ted Adelson

Take-home messages from Case 3:

- Uncertainty often arises from nuisance parameters.
- A Bayesian observer computes a joint posterior over all variables including nuisance parameters.
- Priors over nuisance parameters matter!

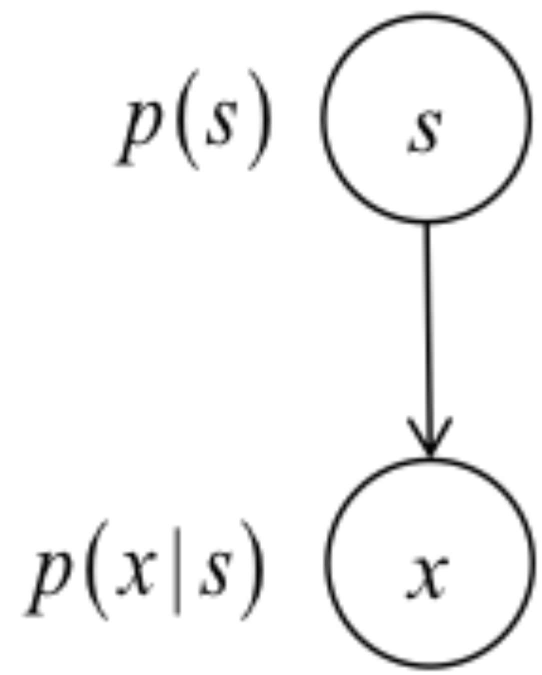
“The Dress”

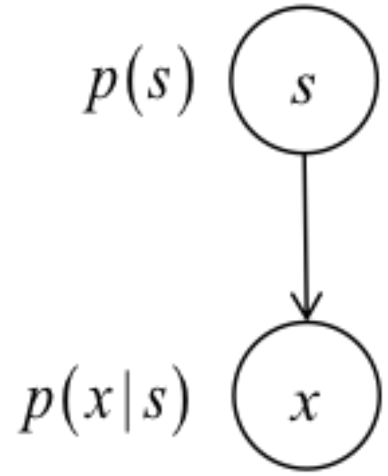
Schedule for today

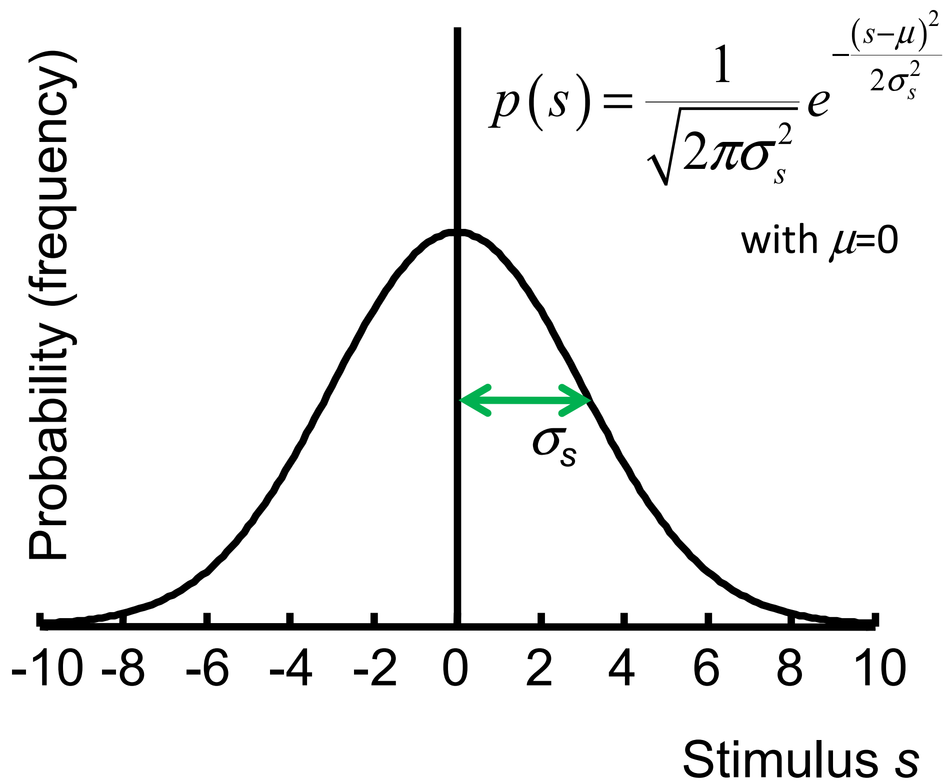
		Concept
12:10-13:10	• Why Bayesian modeling	
	• Bayesian explanations for illusions	priors
	• Case 1: Gestalt perception	likelihoods
	• Case 2: Motion sickness	prior/likelihood interplay
13:30-14:40	• Case 3: Color perception	nuisance parameters
	• Case 4: Sound localization	measurement noise
	• Case 5: Change point detection	hierarchical inference
15:00-16:00	• Model fitting and model comparison	
	• Critiques of Bayesian modeling	

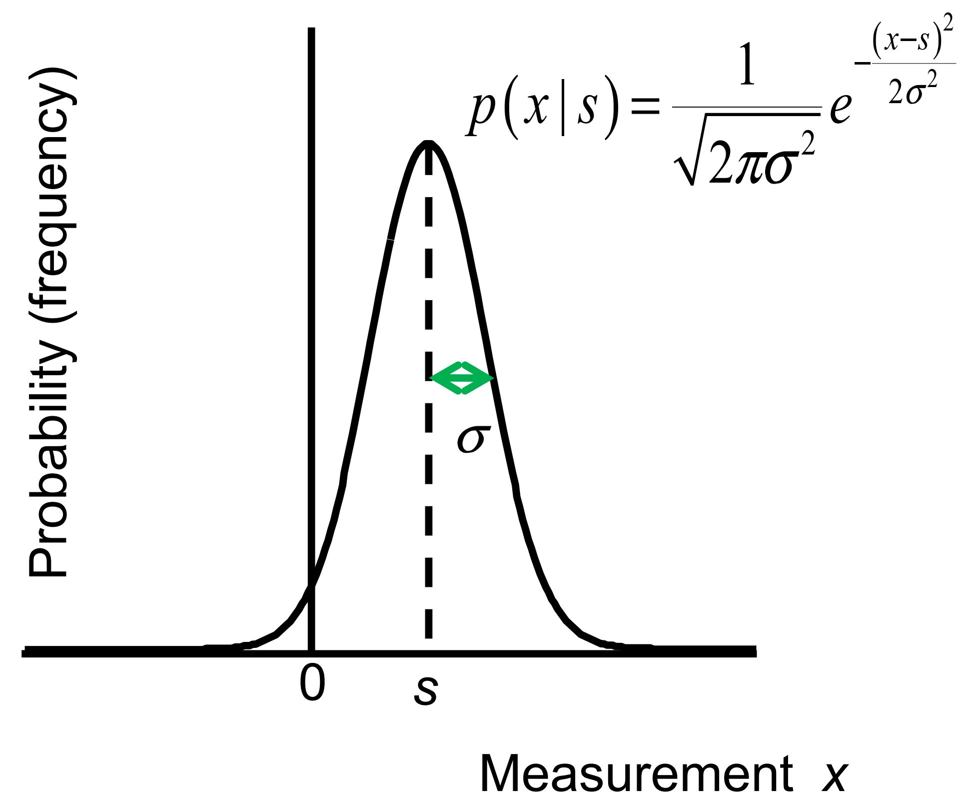
Demo of sound localization

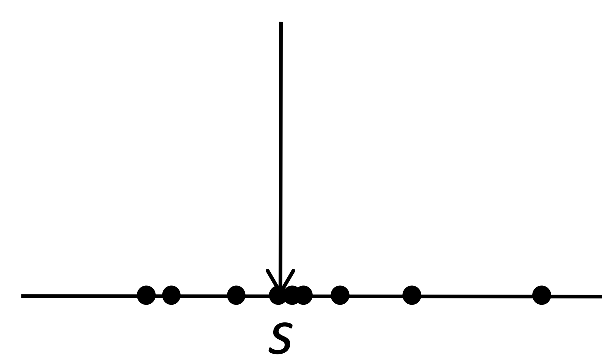
Step 1: Generative model




$p(s)$
s
$p(x|s)$
x

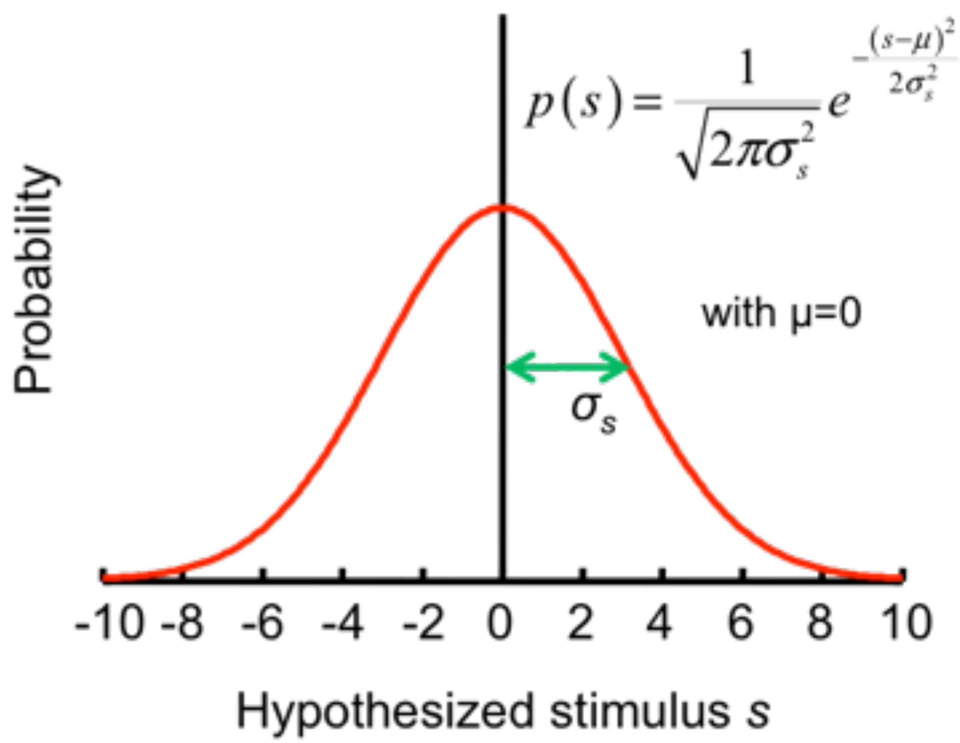

$p(s)=\frac{1}{\sqrt{2\pi\sigma_s^2}}e^{-\frac{(s-\mu)^2}{2\sigma_s^2}}$
with μ=0
Probability (frequency)
σ_s
-10 -8 -6 -4 -2 0 2 4 6 8 10
Stimulus s


$p(x|s)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-s)^2}{2\sigma^2}}$
Probability (frequency)
σ
0
s
Measurement x

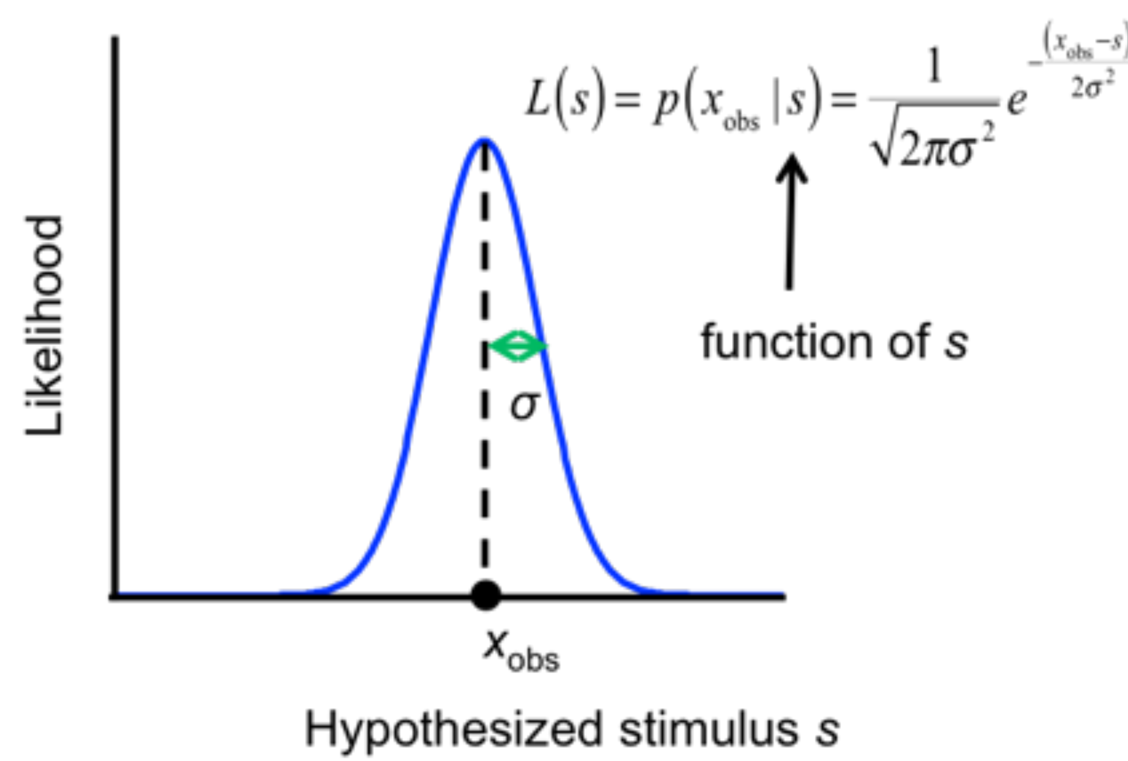

s

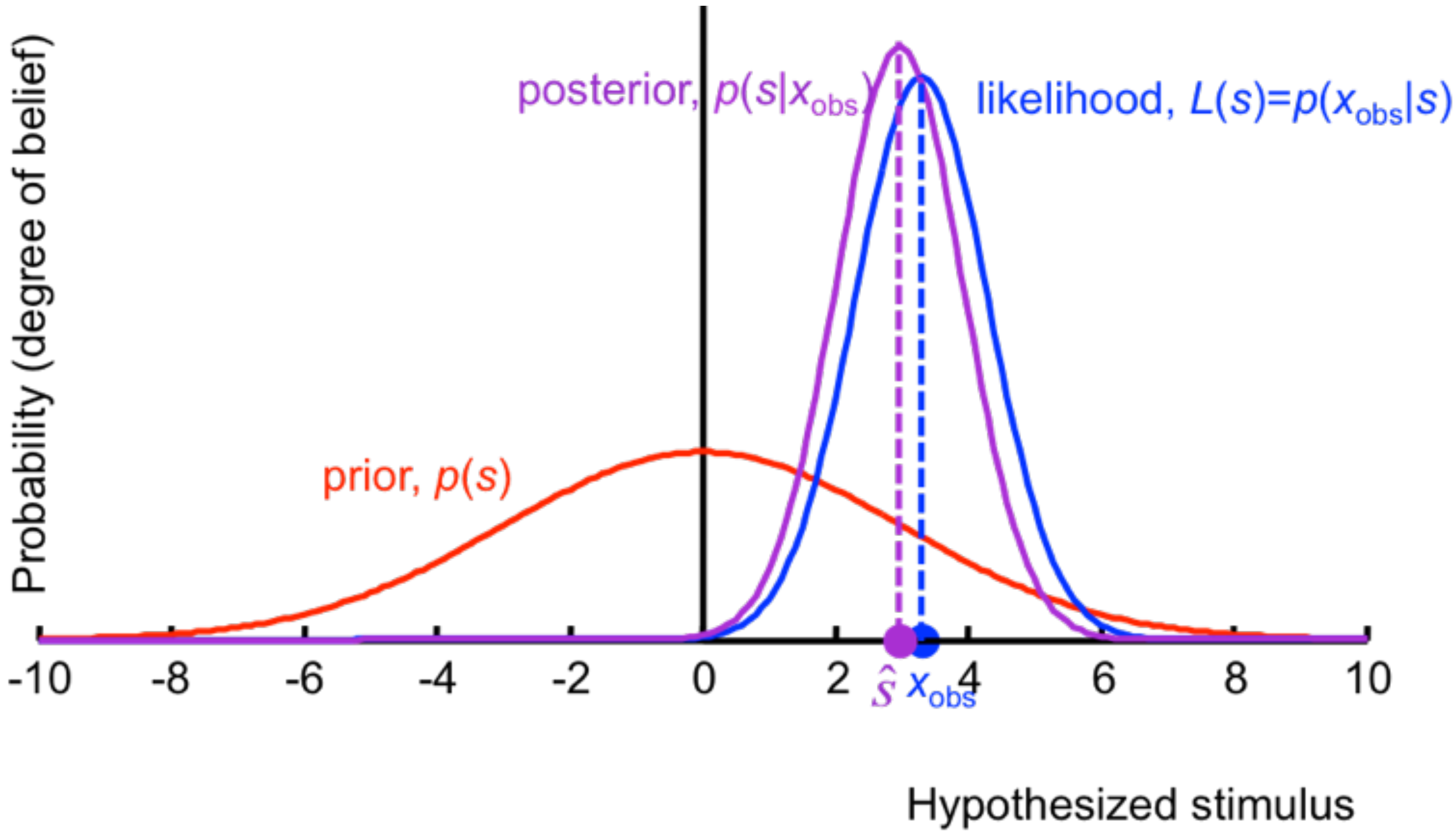
Step 2: Inference, deriving the decision rule

Prior

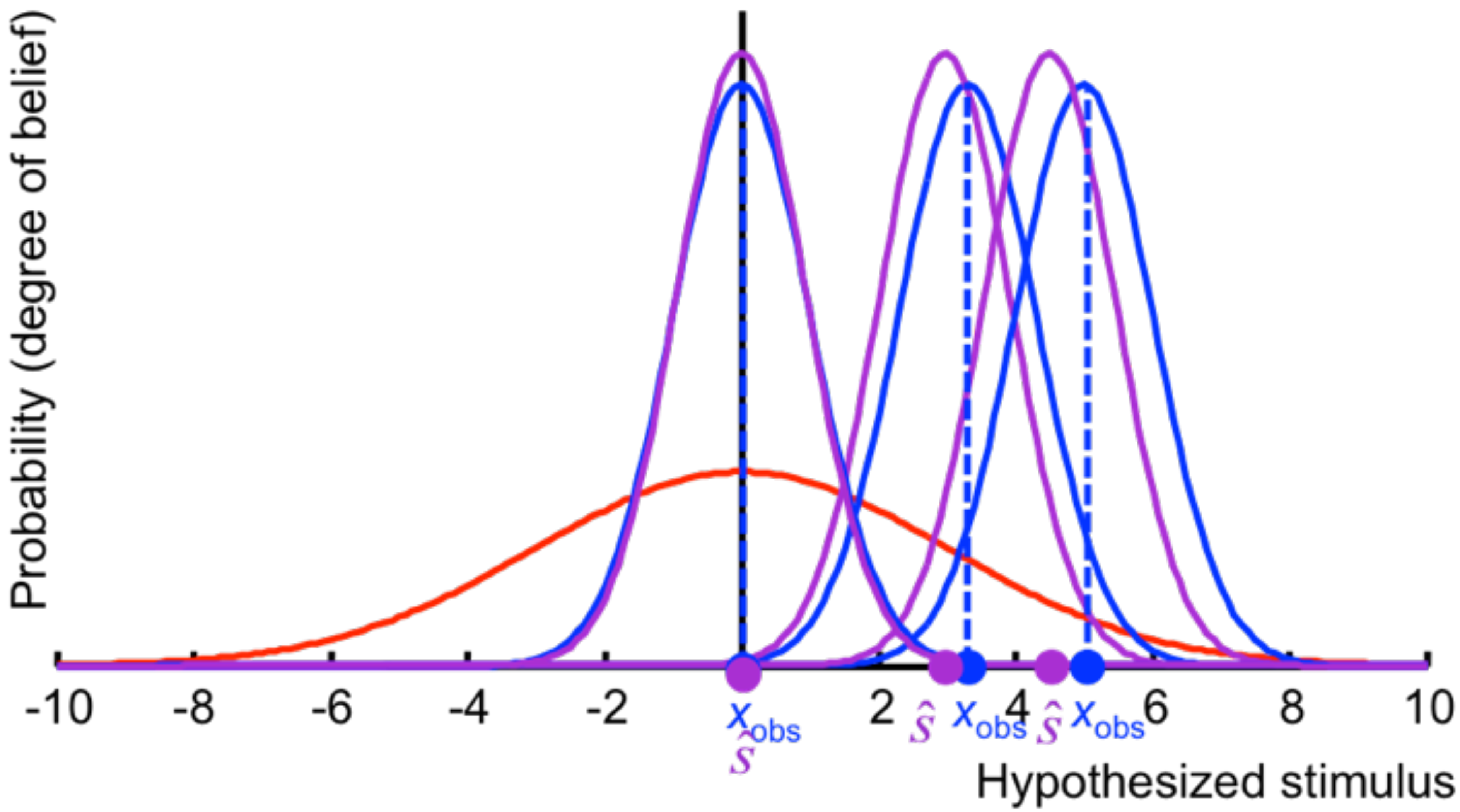


Likelihood





Does the model deterministically predict the posterior for a given stimulus and given parameters?


Probability (degree of belief)
-10
-8
-6
-4
-2
2
4
6
8
10
x_{obs}
$\hat{s}$
$\hat{s}$
x_{obs}
$\hat{s}$
x_{obs}
Hypothesized stimulus

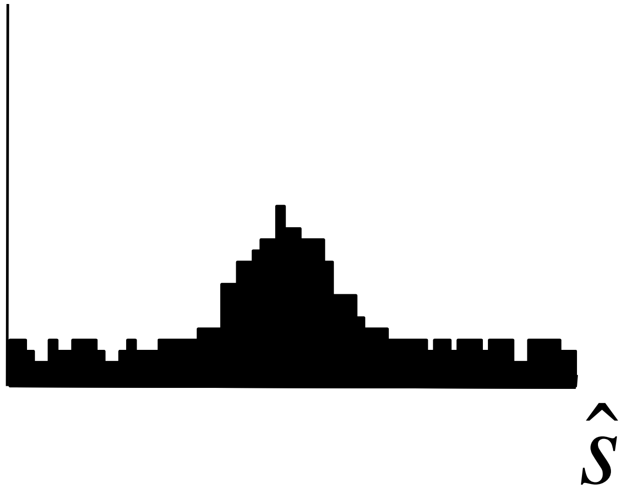
Step 3: Response probabilities (predictions for your behavioral experiment)

Decision rule: mapping $x \rightarrow \hat{s}$

But *x* is itself a random variable for given *s*

Therefore $\hat{s}$ is a random variable for given *s*

$$p\left(\hat{s}|s\right)$$



Can compare this to data!!

Take-home messages from Case 4:

- Uncertainty can also arise from measurement noise
- Such noise is often modeled using a Gaussian
- Bayesian inference proceeds in 3 steps.
- The final result is a predicted response distribution.

Schedule for today

		Concept
12:10-13:10	• Why Bayesian modeling	
	• Bayesian explanations for illusions	priors
	• Case 1: Gestalt perception	likelihoods
	• Case 2: Motion sickness	prior/likelihood interplay
13:30-14:40	• Case 3: Color perception	nuisance parameters
	• Case 4: Sound localization	measurement noise
	• Case 5: Change point detection	hierarchical inference
15:00-16:00	• Model fitting and model comparison	
	• Critiques of Bayesian modeling	

Well known

Bayesian integration (prior x simple likelihood)

Bayesian integration in sensorimotor learning

Konrad P. Körding & Daniel M. Wolpert

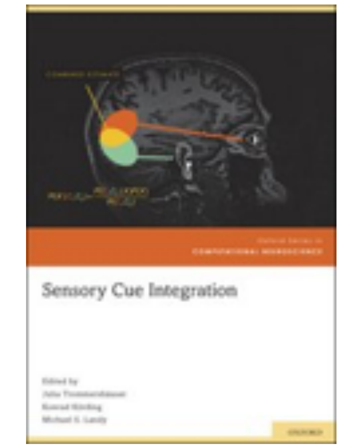
Noise characteristics and prior expectations in human visual speed perception

Alan A Stocker & Eero P Simoncelli

Cardinal rules: visual orientation perception reflects knowledge of environmental statistics

Ahna R Girshick[1,2], Michael S Landy[1,2] & Eero P Simoncelli[1–4]

Cue combination



Less well known but often more interesting

- Complex categorization
- Combining information across multiple items (visual search)
- Combining information across multiple items and across a memory delay (change detection)
- Inferring a changing world state (tracking, sequential effects)
- Evidence accumulation and learning

A simple change point detection task

Take-home messages from Case 5:

- Inference is often hierarchical.
- In such situations, the Bayesian observer marginalizes over the "intermediate" variables (compare this to Case 3)

Topics not addressed

- Lapse rates and response noise
- Utility and reward
- Partially observable Markov decision processes
- Wrong beliefs (model mismatch)
- Learning
- Approximate inference (e.g. sampling, variational approximations)
- How the brain represents probability distributions

Bayesian models are about:

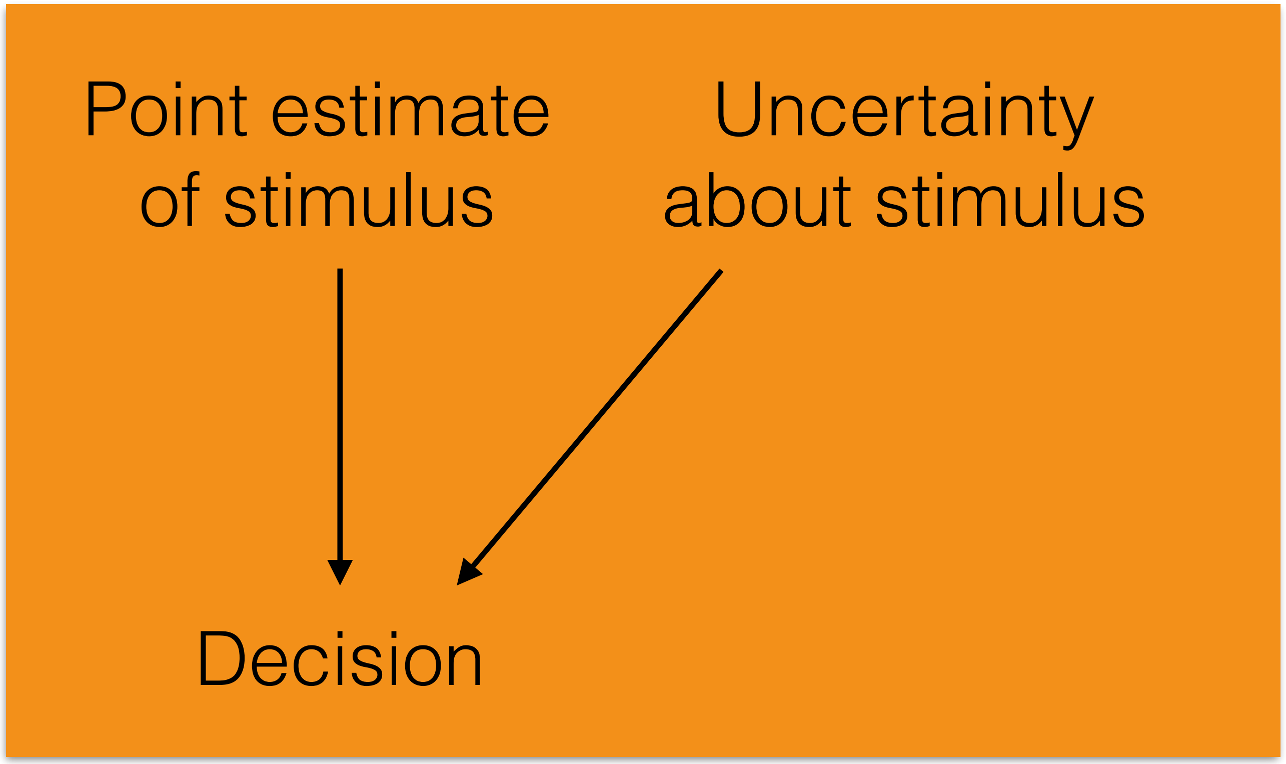
- the decision-maker making the best possible decision (given an objective function)
- **the brain representing probability distributions**

- Lower-contrast patterns appear to move slower than higher-contrast patterns at the same speed (Stone and Thompson 1990)

- This may underlie drivers' tendency to speed up in the fog (Snowden, Stimpson, Ruddle 1998)
- Possible explanation: lower contrast → greater uncertainty → greater effect of prior beliefs (which might favor low speeds) (Weiss, Adelson, Simoncelli 2002)

Probabilistic computation

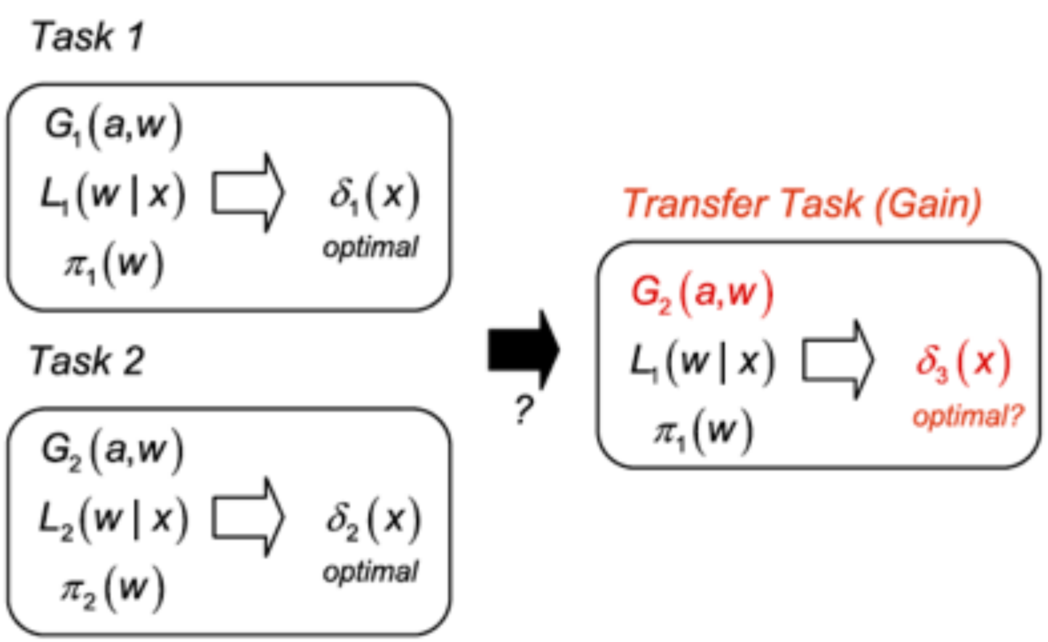
Decisions in which the brain takes into account trial-to-trial knowledge of uncertainty (or even entire probability distributions), instead of only point estimates



What does probabilistic computation "feel like"?

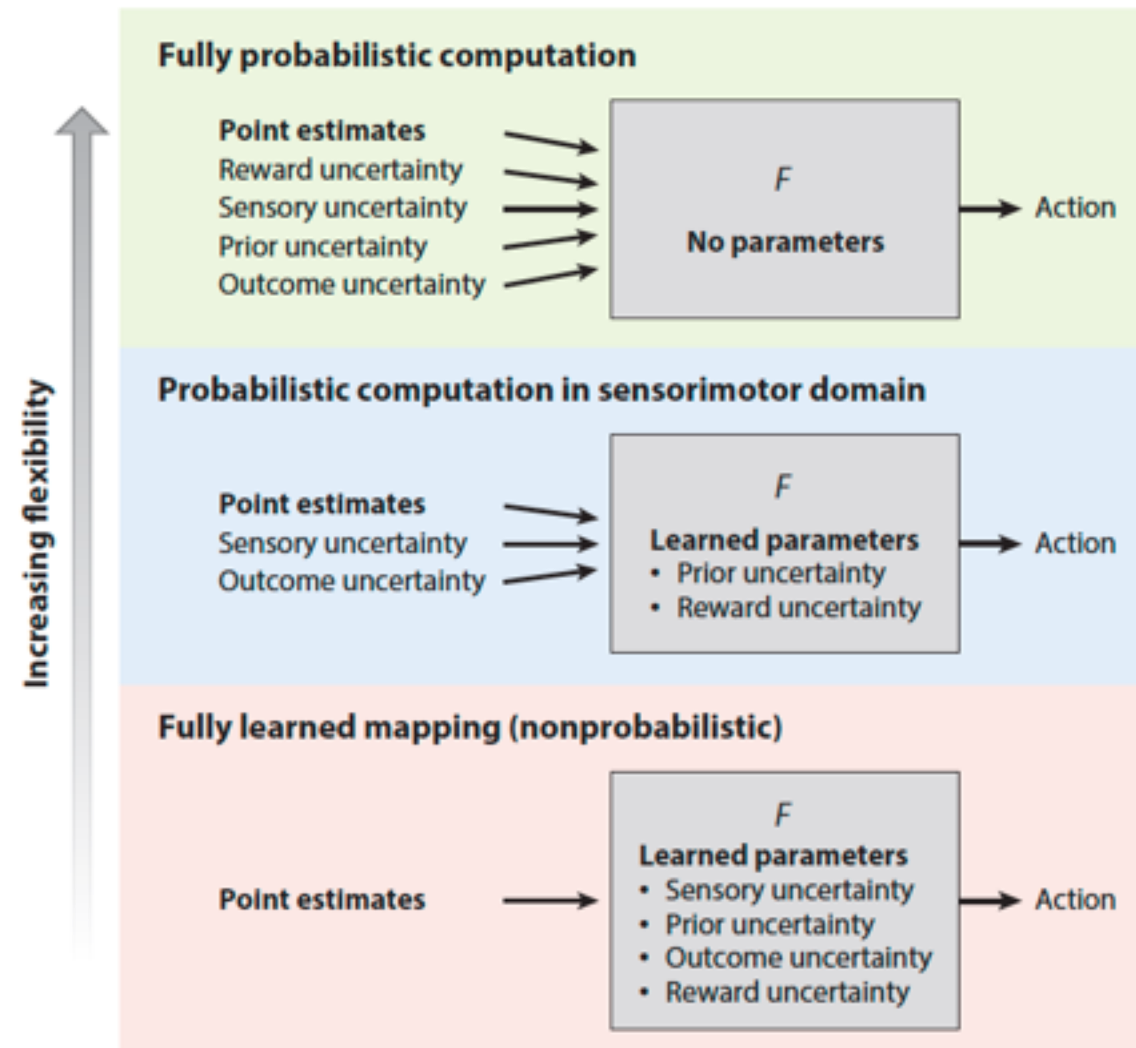
Does the brain represent probability distributions?

Bayesian transfer



Maloney and Mamassian, 2009

Different degrees of probabilistic computation



Ma and Jazayeri, 2014

2006
theory, networks

Bayesian inference with probabilistic population codes

Wei Ji Ma[1,3], Jeffrey M Beck[1,3], Peter E Latham[2] & Alexandre Pouget[1]

2013
behavior, networks

Trial-to-trial, uncertainty-based adjustment of decision boundaries in visual categorization

Ahmad T. Qamar[a,1], R. James Cotton[a,1], Ryan G. George[a,1], Jeffrey M. Beck[b], Eugenia Prezhdo[a], Allison Laudano[a], Andreas S. Tolias[a], and Wei Ji Ma[a,2,3]

2015
behavior,
human fMRI

Sensory uncertainty decoded from visual cortex predicts behavior

Ruben S van Bergen[1], Wei Ji Ma[2], Michael S Pratte[3] & Janneke F M Jehee[1]

2017
trained networks

Efficient probabilistic inference in generic neural networks trained with non-probabilistic feedback

A. Emin Orhan[1] & Wei Ji Ma[1,2]

2018
behavior,
monkey physiology

A neural basis of probabilistic computation in visual cortex

Edgar Y. Walker,[1,2†] R. James Cotton,[1,2,3†] Wei Ji Ma,[4‡] Andreas S. Tolias[1,2,5‡*]

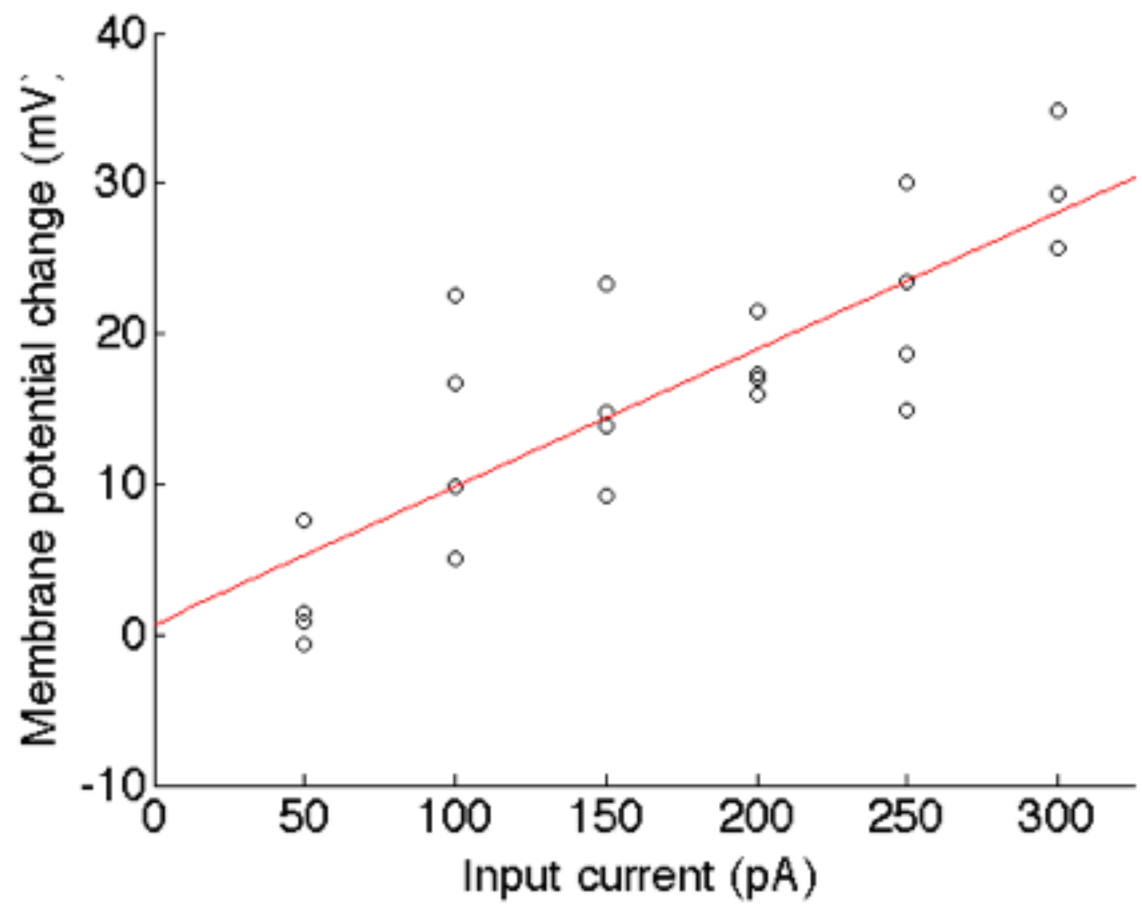
Schedule for today

Time	Topic	*Concept*
12:10-13:10	• Why Bayesian modeling	
	• Bayesian explanations for illusions	priors
	• Case 1: Gestalt perception	likelihoods
	• Case 2: Motion sickness	prior/likelihood interplay
13:30-14:40	• Case 3: Color perception	nuisance parameters
	• Case 4: Sound localization	measurement noise
	• Case 5: Change point detection	hierarchical inference
15:00-16:00	• $\boxed{\text{Model fitting and model comparison}}$	
	• Critiques of Bayesian modeling	

a. What to minimize/maximize when fitting parameters?
b. What fitting algorithm to use?
c. Validating your model fitting method

What to minimize/maximize when fitting a model?

Try #1: Minimize sum squared error



Only principled if your model has independent, fixed-variance Gaussian noise
Otherwise arbitrary and suboptimal

Try #2: Maximize likelihood

Output of Step 3:

***p*(response I stimulus, parameter combination)**

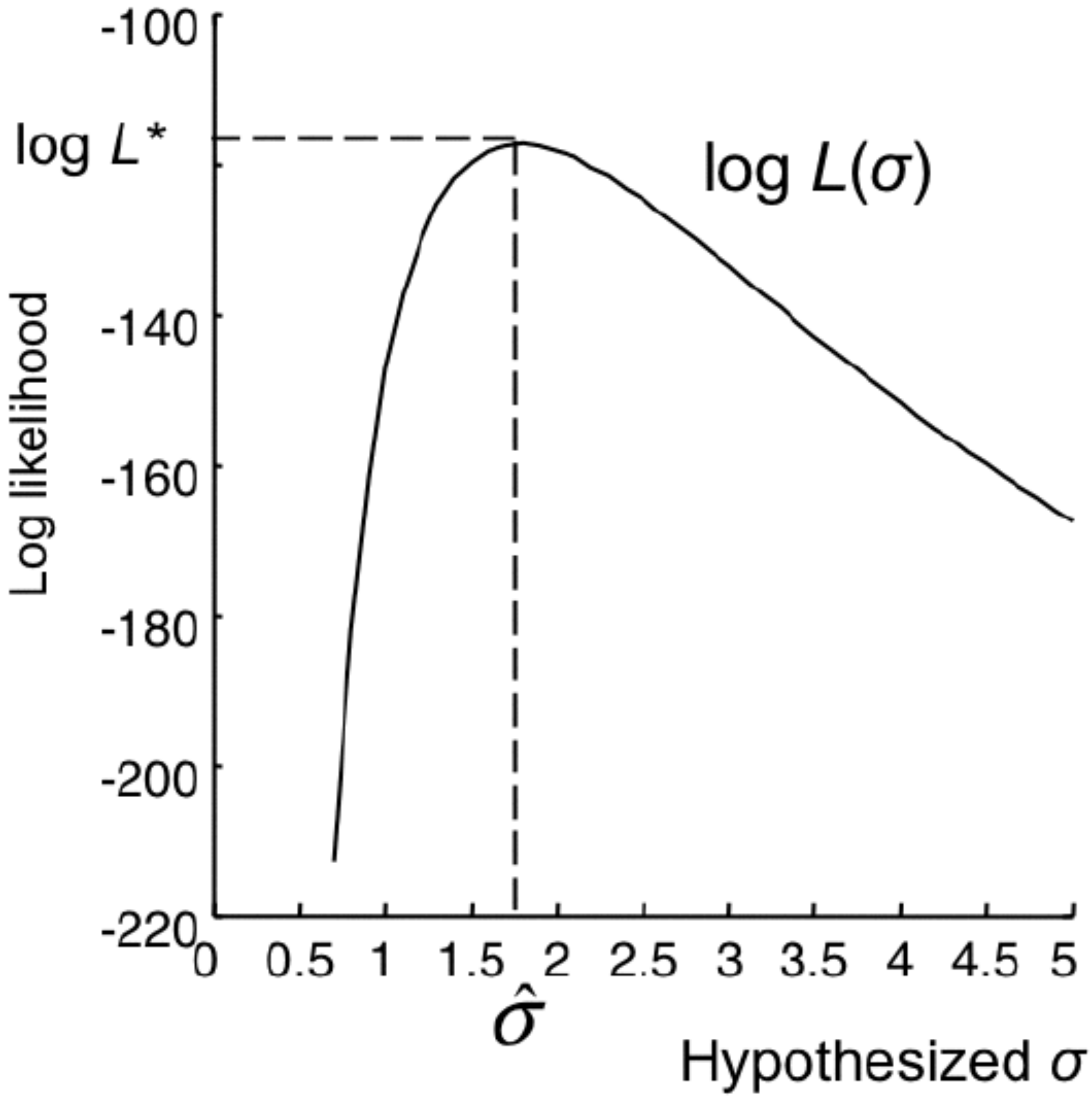
Likelihood of parameter combination

= p(data | parameter combination)

$$= \prod_{\text{trials } i} p\left(\text{response}_i \middle| \text{stimulus}_i,\ \text{parameter combination}\right)$$

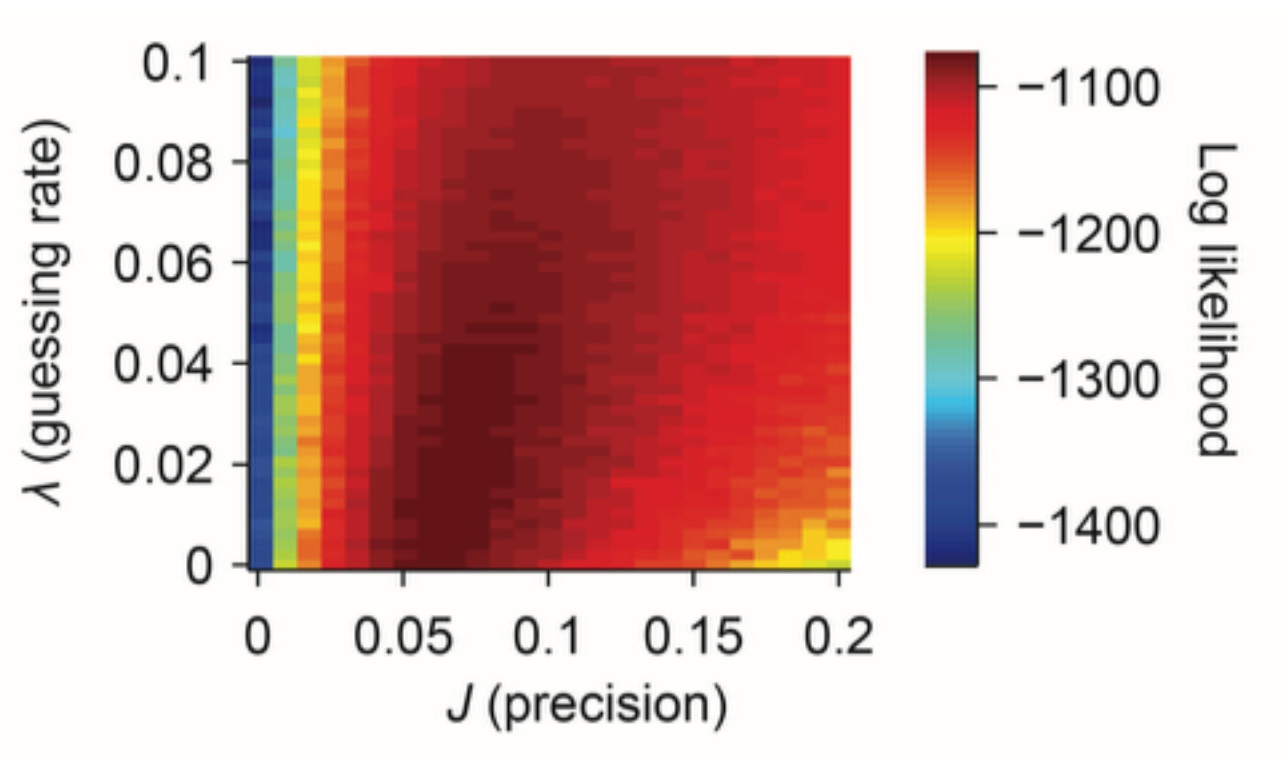
What fitting algorithm to use?

- Search on a fine grid

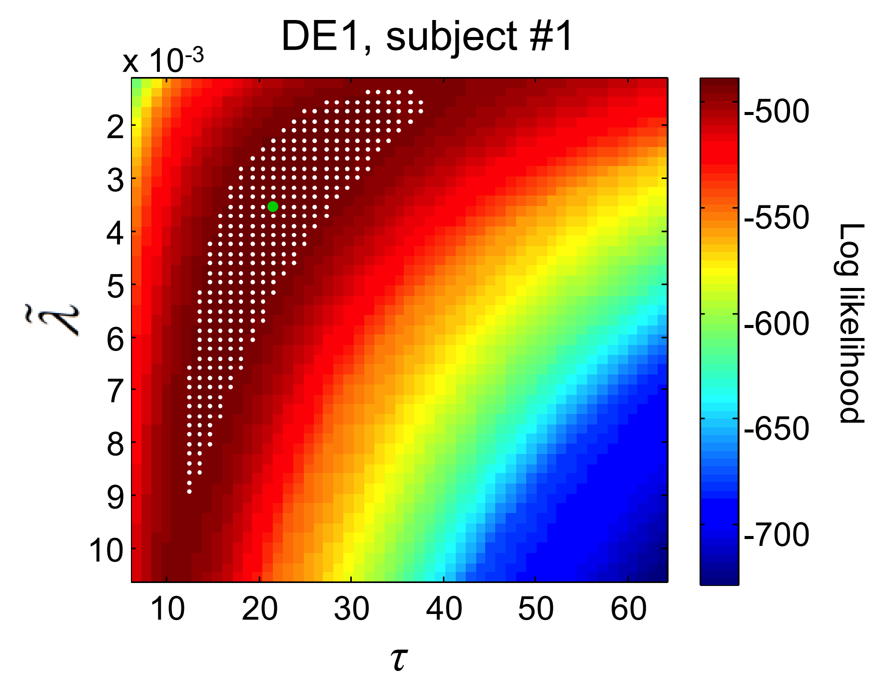


-100
log L^*
log $L(\sigma)$
-140
-160
-180
-200
-220
Log likelihood
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
$\hat{\sigma}$
Hypothesized σ

Parameter trade-offs



Shen and Ma 2017



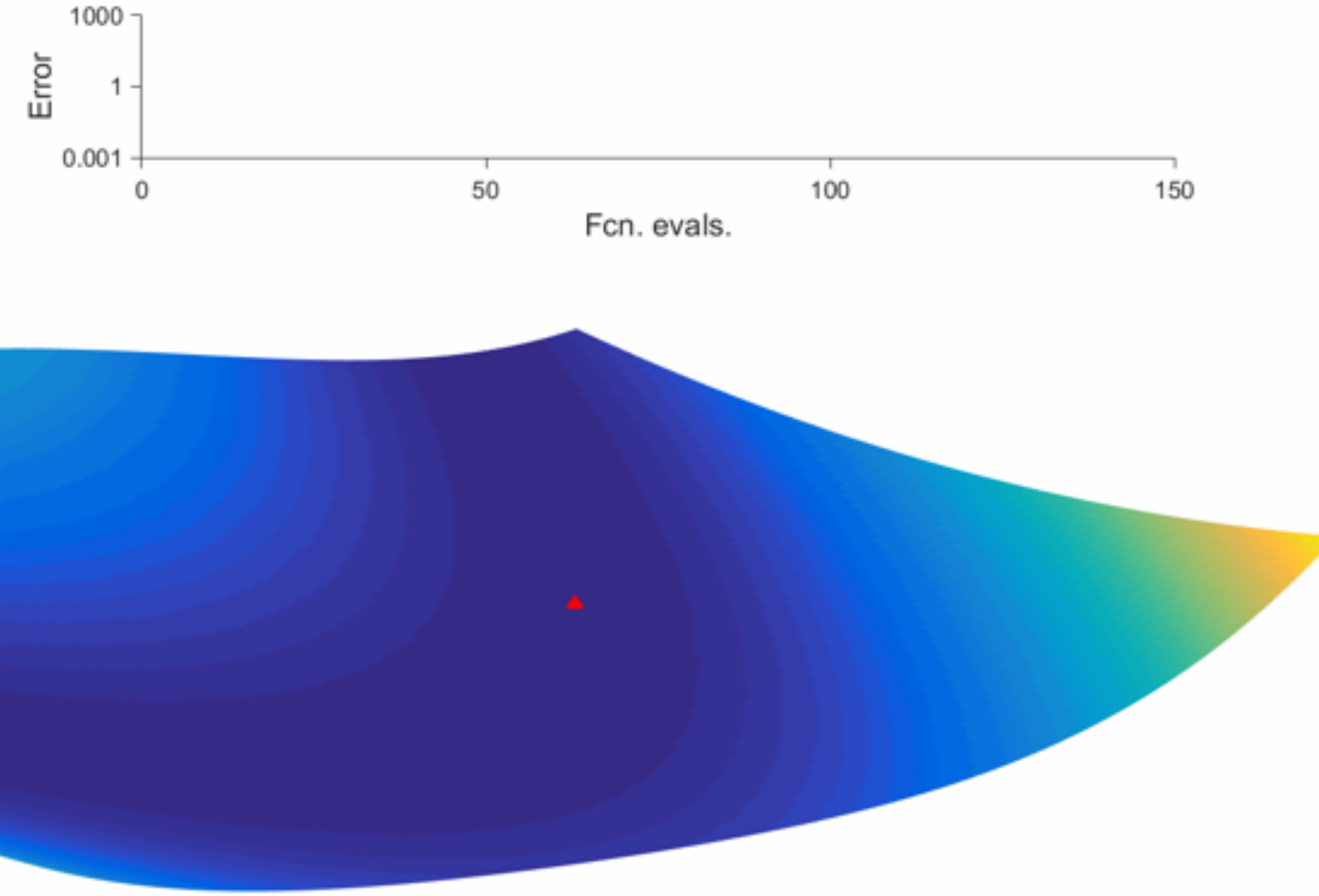
Van den Berg and Ma 2018

What fitting algorithm to use?

- Search on a fine grid
- fmincon or fminsearch in Matlab

OptimViz (Rosenbrock function)

fminsearch

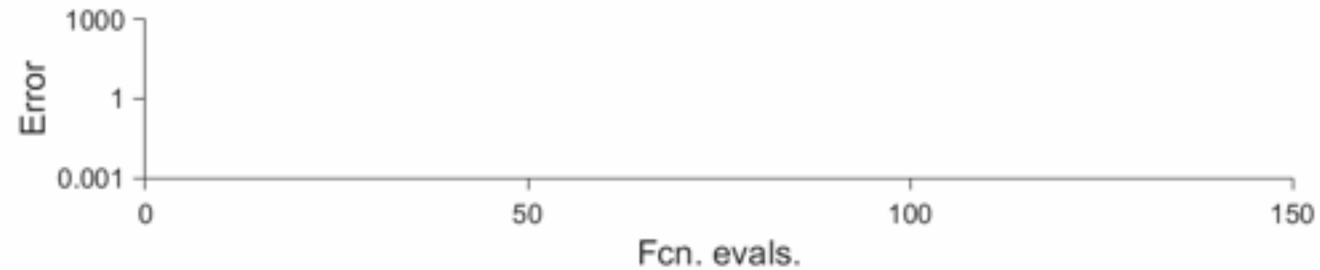


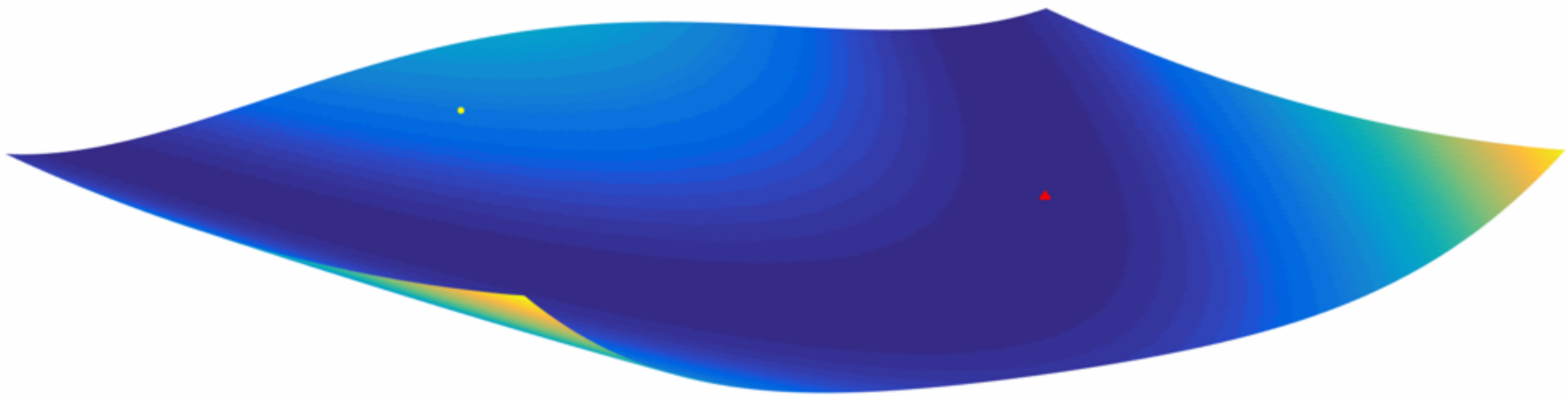
What fitting algorithm to use?

- Search on a fine grid
- fmincon or fminsearch in Matlab
- Bayesian Adaptive Direct Search (Acerbi and Ma 2016)

OptimViz (Rosenbrock function)

BADS





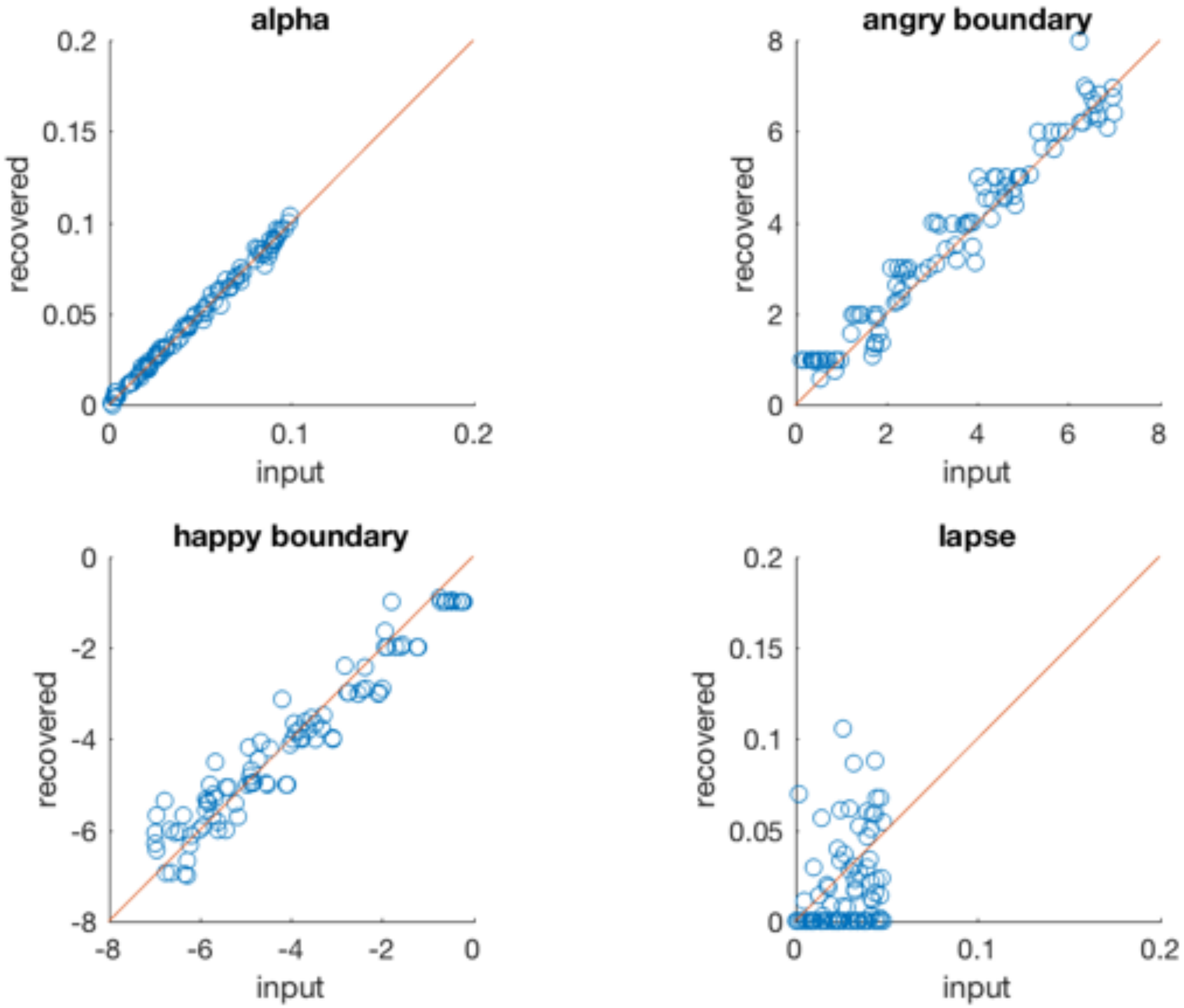
#useBADS

https://github.com/lacerbi/bads

https://github.com/lacerbi/optimviz

Validating your method: Parameter recovery

parameter recovery test



Jenn Laura Lee

Take-home messages model fitting

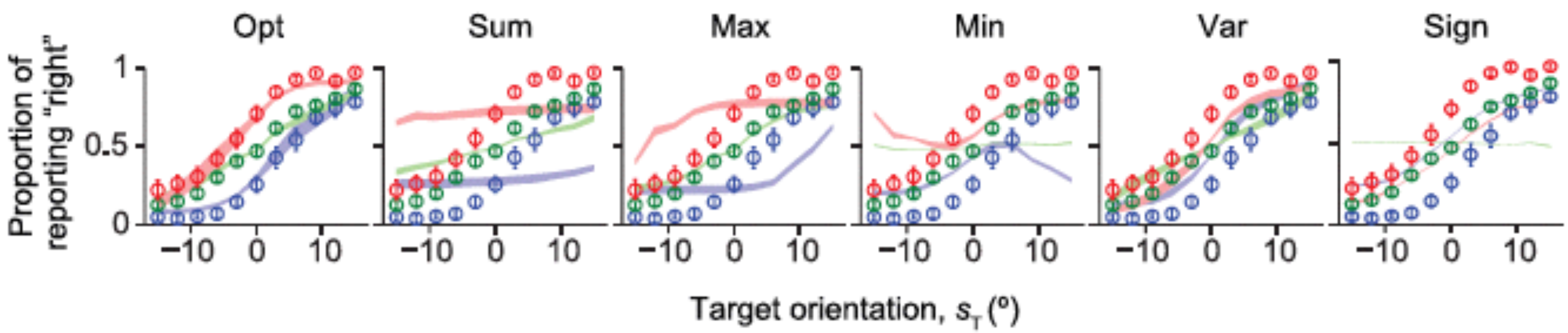
- If you can, maximize the likelihood (probability of individual-trial responses) if you can.
  - *Do not* minimize squared error!
  - *Do not* fit summary statistics (instead fit the raw data)
- Use more than one algorithm
- Consider BADS when you don't trust fmincon/fminsearch
- Multistart
- Do parameter recovery

Model comparison

a. Choosing a model comparison metric
b. Validating your model comparison method
c. Factorial model comparison
d. Absolute goodness of fit
e. Heterogeneous populations

a. Choosing a model comparison metric

Try #1: Visual similarity to the data



Shen and Ma, 2016

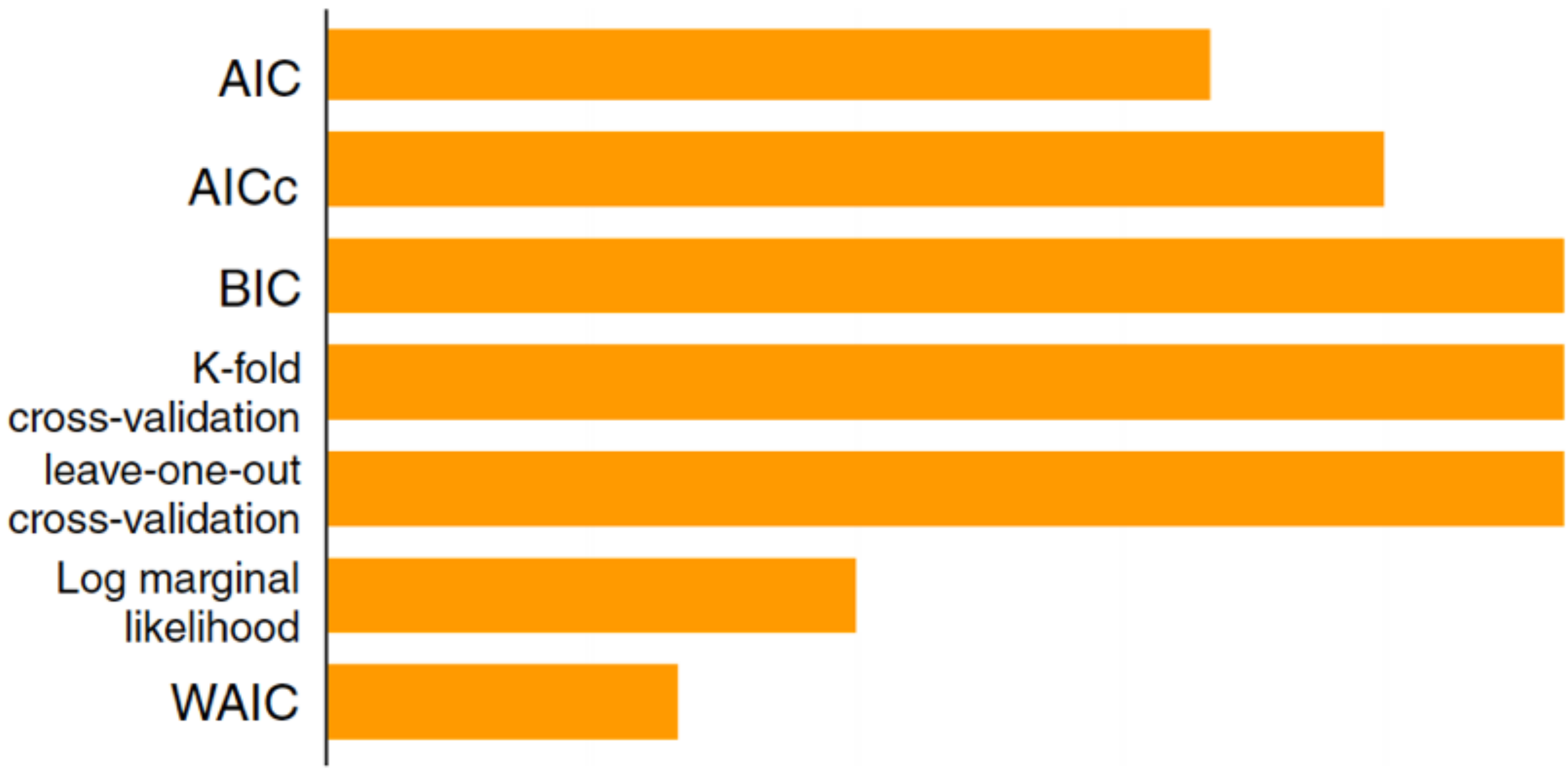
Fine, but not very quantitative

Try #2: R^2

- Just don't do it
  - Unless you have only linear models
    - Which almost never happens

Try #3: Likelihood-based metrics

Good!
Problem: there are many!



From Ma lab survey by Bas van Opheusden, 201703

Metrics based on *maximum likelihood:*

- Akaike Information Criterion (AIC or AICc)
- Bayesian Information Criterion (BIC)

Metrics based on *the full likelihood function* (often sampled using Markov Chain Monte Carlo)*:*

- Marginal likelihood (model evidence, Bayes' factor)
- Watanabe-Akaike Information criterion

Cross-validation can be either

Metrics based on *explanation:*

- Bayesian Information Criterion (BIC)
- Marginal likelihoods (model evidence, Bayes' factors)

Metrics based on *prediction:*

- Akaike Information Criterion (AIC or AICc)
- Watanabe-Akaike Information criterion
- Most forms of cross-validation

Practical considerations:

- No metric is always unbiased for finite data.
- AIC tends to underpenalize free parameters, BIC tends to overpenalize.
- **Do not trust conclusions that are metric-dependent**. Report multiple metrics if you can.

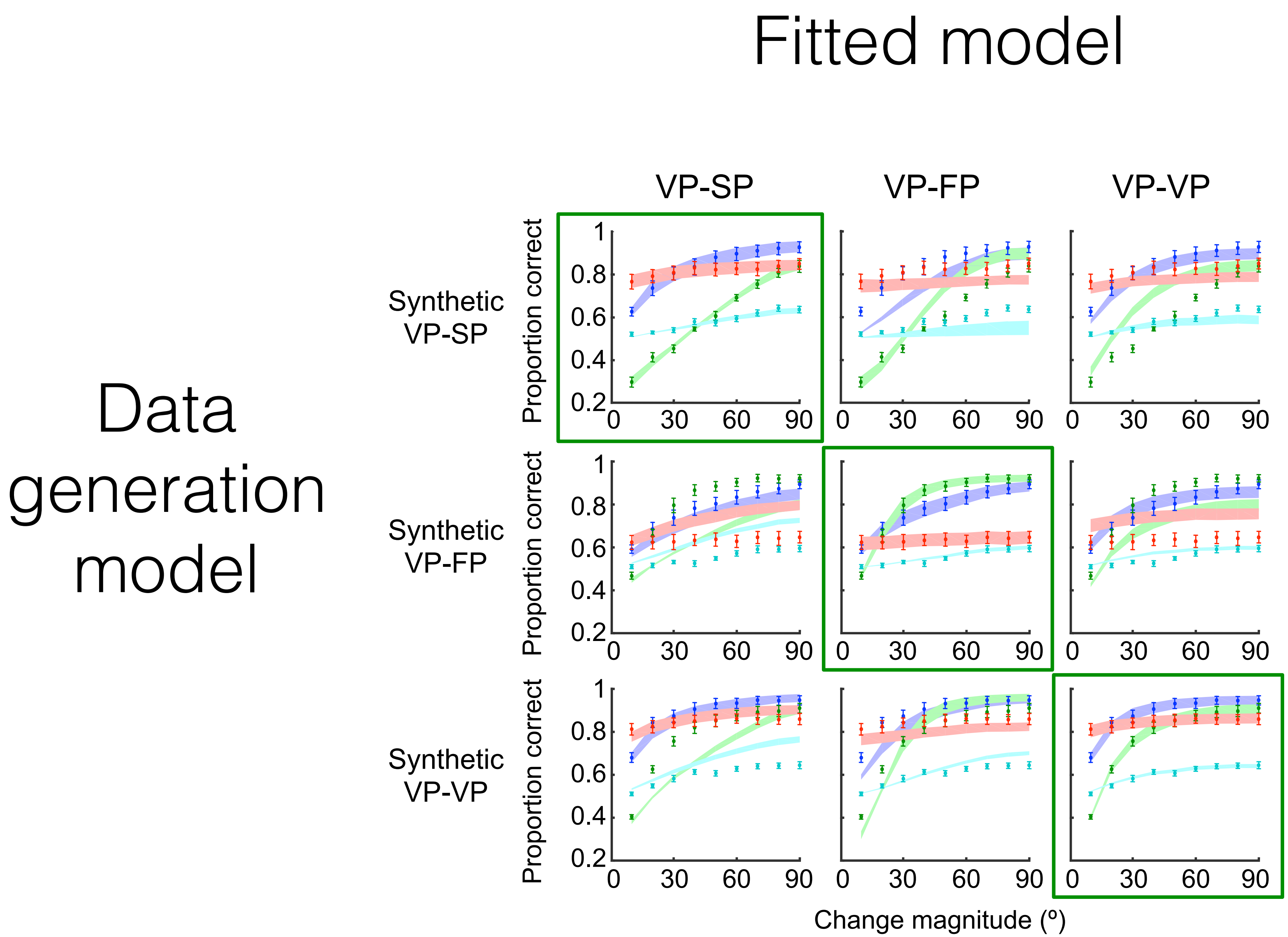
	AICc*(model) − AICc*(VP)	BIC*(model) − BIC*(VP)		LML(model) − LML(VP)	
Model	Mean	Mean	Standard error of the mean	Mean	Standard error of the mean
IL					
M1	−125	−122	15	−121	15
M2	−183	−180	18	−180	18
M3	−167	−164	18	−163	18
Humans	−47.2	−45.7	6.8	−47.1	6.6
EP					
M1	−47.5	−44.8	9.2	−48.9	9.1
M2	−12.8	−10.1	4.6	−12.7	4.8
M3	−30.3	−27.6	7.8	−31.3	8.1
Humans	−12.9	−11.4	1.5	−14.4	1.7
EPF					
M1	−40.2	−40.2	7.9	−39.0	7.8
M2	−9.3	−9.3	4.4	−6.7	4.6
M3	−24.0	−24.0	6.7	−22.6	6.9
Humans	−7.6	−7.6	1.5	−6.2	1.6
VPF					
M1	−1.3	−4.18	0.83	1.5	1.5
M2	−2.2	−4.00	0.91	1.20	0.81
M3	−0.56	−3.2	1.5	2.0	1.1
Humans	−1.46	−3.00	0.32	−0.57	0.31

Devkar, Wright, Ma 2015

Challenge: your model comparison metric and how you compute it might have issues. How to validate it?

b. Model recovery

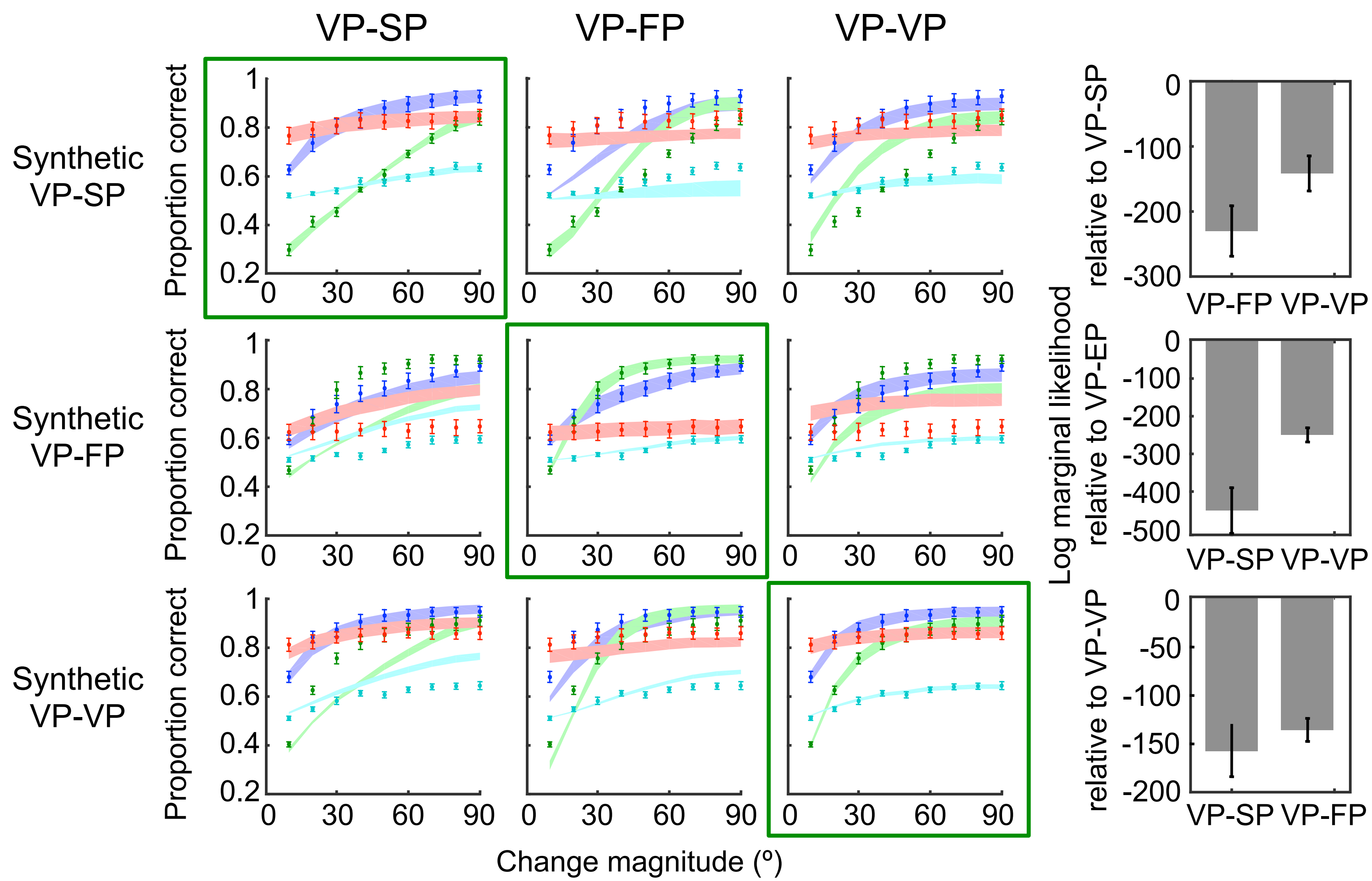
Model recovery example



Devkar, Wright, Ma, Journal of Vision, in press

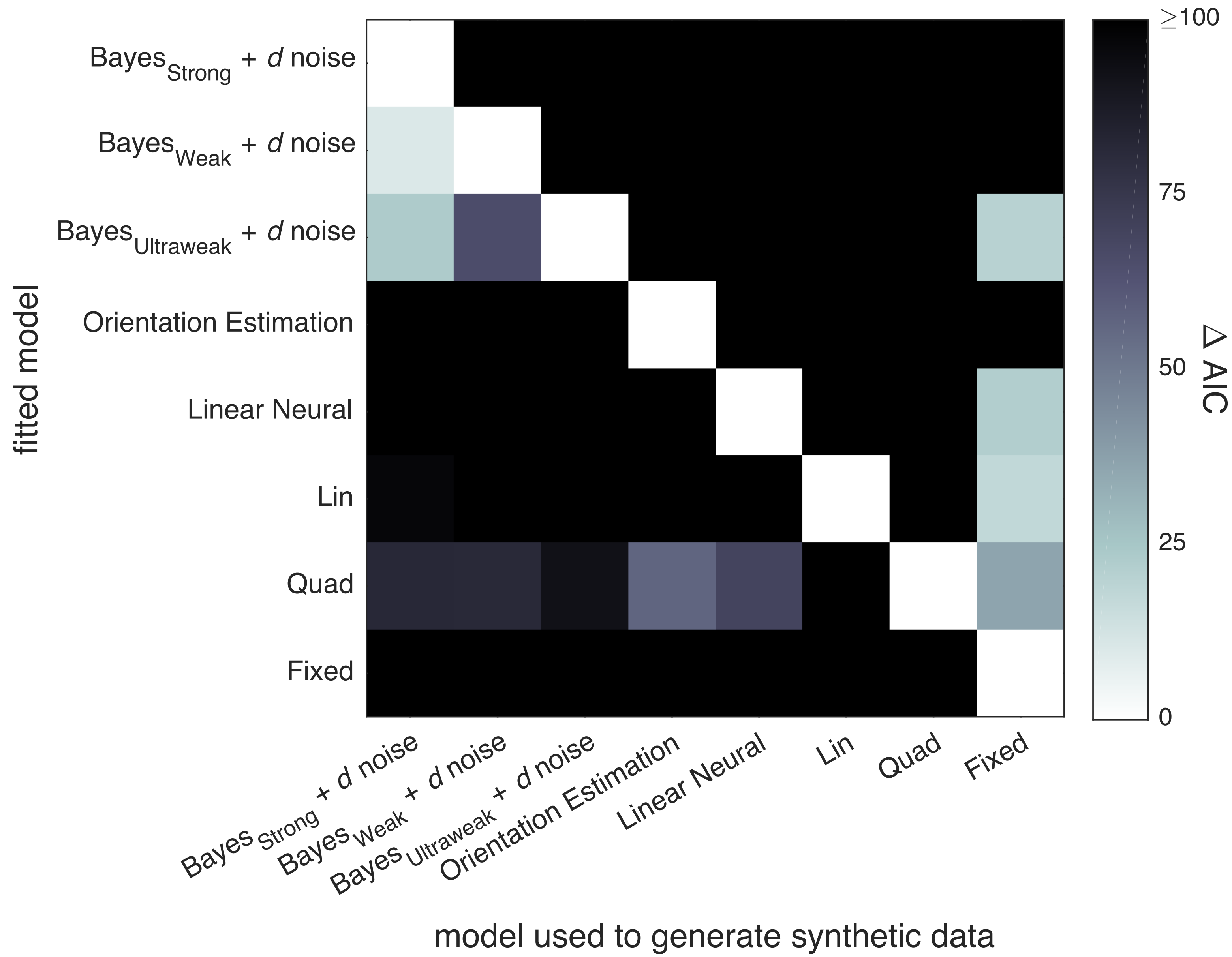
Fitted model

Data generation model



Devkar, Wright, Ma, Journal of Vision, in press

Model recovery



Adler and Ma, PLoS Comp Bio 2018

Challenge: how to avoid “handpicking” models?

c. Factorial model comparison

c. Factorial model comparison

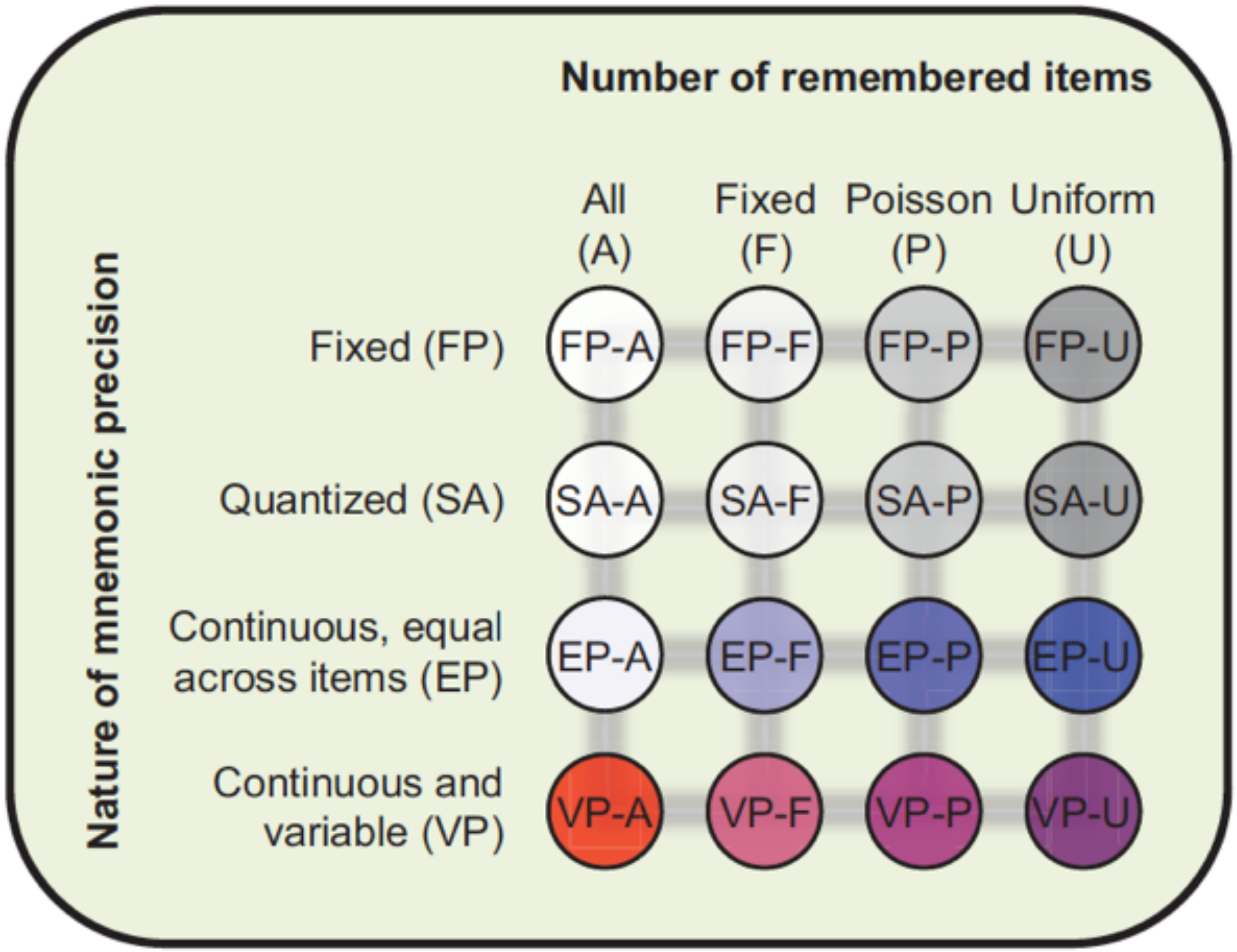
- Models often have many “moving parts”, components that can be in or out
- Similar to factorial design of experiments, one can mix and match these moving parts.
- Similar to stepwise regression
- References:
  - Acerbi, Vijayakumar, Wolpert 2014
  - Van den Berg, Awh, Ma 2014
  - Shen and Ma, 2017

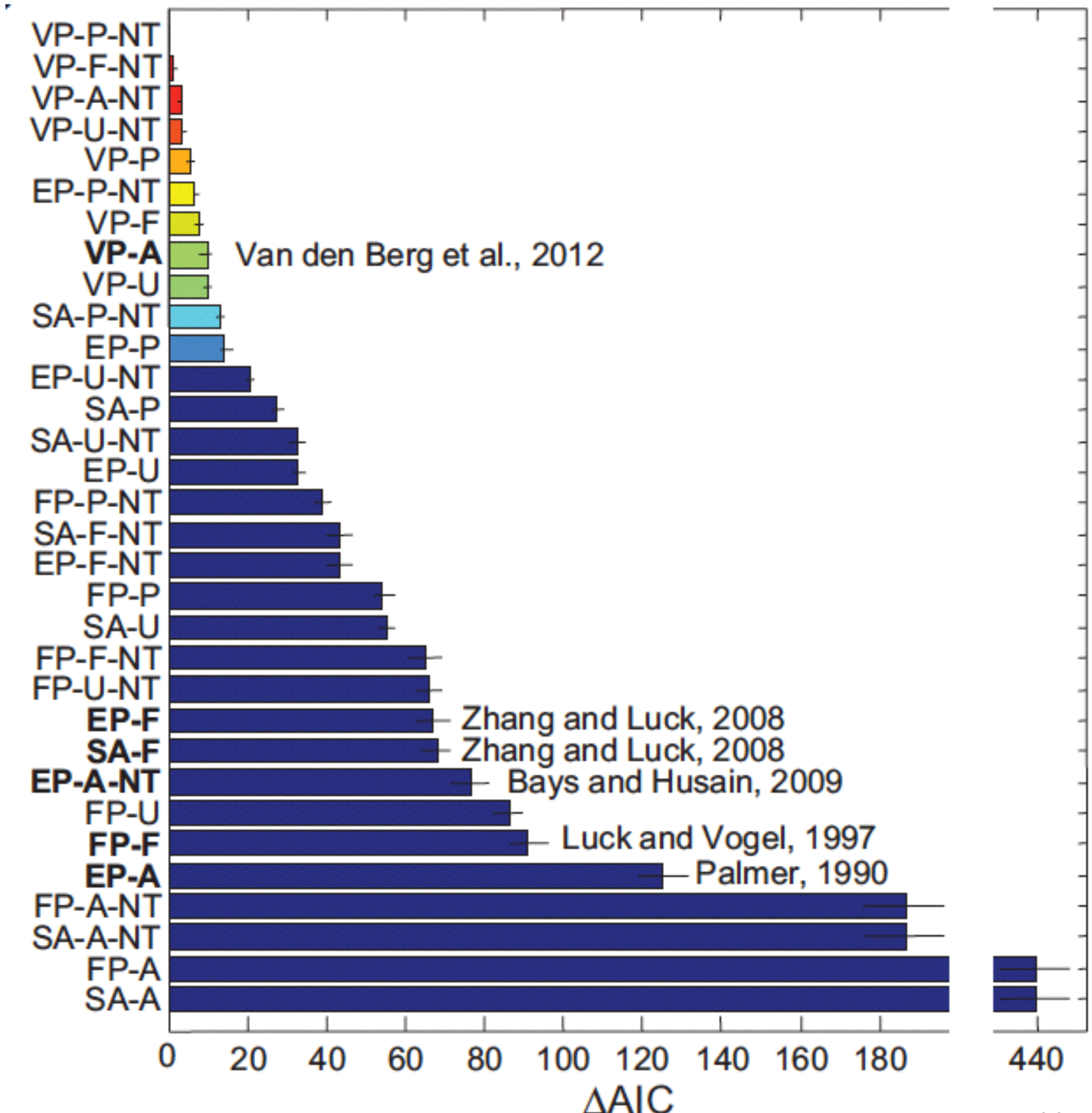
Factorial Comparison of Working Memory Models

Ronald van den Berg
University of Cambridge and Baylor College of Medicine

Edward Awh
University of Oregon

Wei Ji Ma
New York University and Baylor College of Medicine





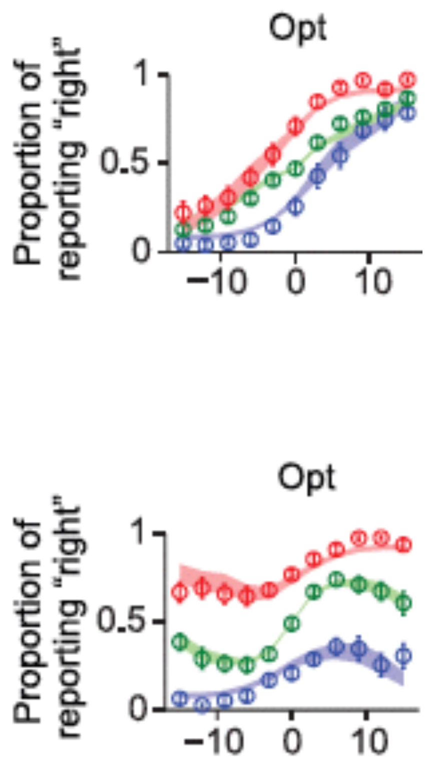
Van den Berg, Awh, Ma 2014

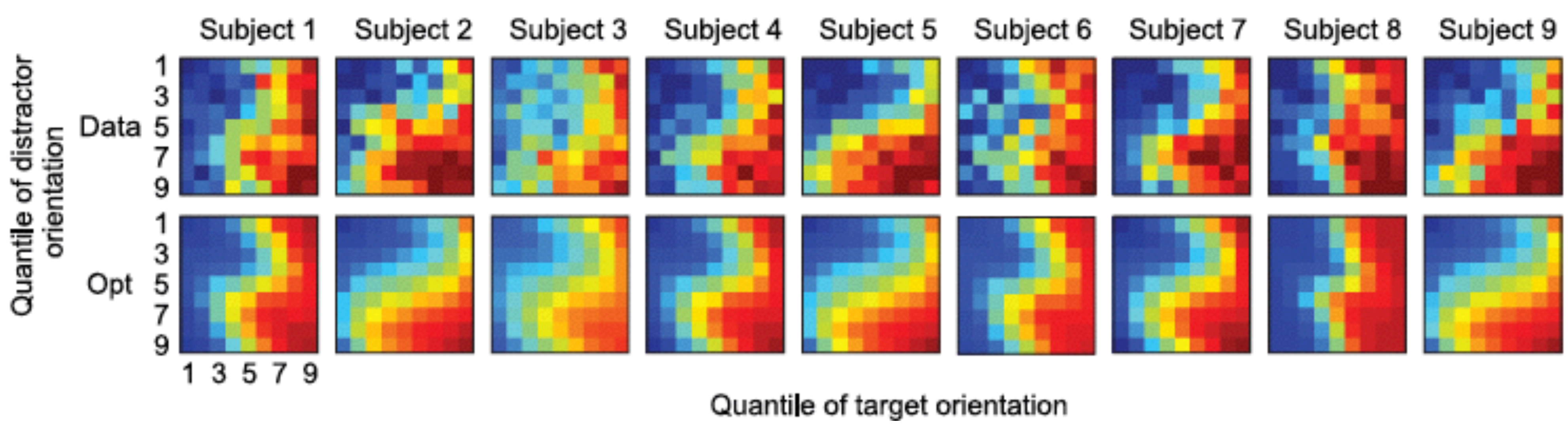
Challenge: the *best* model is not necessarily a *good* model.

d. Absolute goodness of fit

Absolute goodness of fit

- How close is the *best* model to the data?
- Method 1: Visual inspection (model checking)





Shen and Ma, 2016

d. Absolute goodness of fit

- Method 2: Deviance / negative entropy
  - There is irreducible, unexplainable variation in the data
  - This sets an upper limit on the goodness of fit of *any* model: negative entropy
  - How far away is a model from this upper bound?
  - Wichmann and Hill (2001)
  - Shen and Ma (2016)

Challenge: what if different subjects follow different models?
(heterogeneity in the population)

e. Hierarchical model selection

Consider all possible partitions of your population

Bayesian model selection for group studies

Klaas Enno Stephan [a,b,*], Will D. Penny [a], Jean Daunizeau [a], Rosalyn J. Moran [a], Karl J. Friston [a]

Neuroimage, 2009

Bayesian model selection for group studies — Revisited

L. Rigoux [a], K.E. Stephan [b,c], K.J. Friston [b], J. Daunizeau [a,b,*]

Neuroimage, 2014

- Returns probability that each model is the most common one in a population
- Returns posterior probability for each model
- Matlab code available online!

Take-home messages model comparison

- There are many metrics for model comparison.
  - Specialized lab meetings / reading club?
  - Do due diligence to prevent your conclusions from being metric-dependent.
- Do model recovery
- Consider doing factorial model comparison
- Quantify absolute goodness of fit if possible
- Heterogeneity in population? Hierarchical model selection

Schedule for today

		Concept
12:10-13:10	• Why Bayesian modeling	
	• Bayesian explanations for illusions	priors
	• Case 1: Gestalt perception	likelihoods
	• Case 2: Motion sickness	prior/likelihood interplay
13:30-14:40	• Case 3: Color perception	nuisance parameters
	• Case 4: Sound localization	measurement noise
	• Case 5: Change point detection	hierarchical inference
15:00-16:00	• Model fitting and model comparison	
	• $\boxed{\text{Critiques of Bayesian modeling}}$	

Critique of Bayesian models:

- **Prior is hard to get**
- **Inference intractable**
- *Behavior might not be Bayesian*
- Hard to apply to neural data / make connection to neural representation
- Parametric assumptions in distributions
- **Learning the structure of the Bayesian model**
- **Scaling to large data sets**
- Weirdness in distribution (non-Gaussian)
- **Dynamics**

Please help me thank our amazing teaching assistants!

Anna Kutschireiter
Anne-Lene Sax
Jenn Laura Lee
Jorge Menéndez
Julie Lee
Lucy Lai
Sashank Pisupati

Organizers: Il Memming Park, Leslie Weekes

Good job everyone!!