MENTÁLIS MODELLEK BAYES-I MEGKÖZELÍTÉSBEN
A BAYESIAN APPROACH TO INTERNAL MODELS

LENGYEL MÁTÉ

Computational and Biological Learning Lab
Department of Engineering
University of Cambridge

Department of Cognitive Science
Central European University

BAYESIAN INFERENCE FOR DATA ANALYSIS

BAYESIAN INFERENCE FOR DATA ANALYSIS

data, x

BAYESIAN INFERENCE FOR DATA ANALYSIS

parameters, y

↓

data, x

BAYESIAN INFERENCE FOR DATA ANALYSIS

parameters, y

↓

data, x $\mathrm{P}(x|y)$

BAYESIAN INFERENCE FOR DATA ANALYSIS

parameters, y $\quad$ $\mathrm{P}(y)$

↓

data, x $\quad$ $\mathrm{P}(x|y)$

BAYESIAN INFERENCE FOR DATA ANALYSIS

parameters, y — $\mathrm{P}(y)$

↑

data, x — $\mathrm{P}(x|y)$

$$\mathrm{P}(y|x) = \frac{\mathrm{P}(x|y)\ \mathrm{P}(y)}{\mathrm{P}(x)}$$

$$\mathrm{P}(x) = \sum_{y} \mathrm{P}(x|y)\ \mathrm{P}(y)$$

BAYESIAN INFERENCE FOR NEURAL DATA ANALYSIS

parameters, y $\qquad \mathrm{P}(y)$

↑

data, x $\qquad \mathrm{P}(x|y)$

$$\mathrm{P}(y|x) = \frac{\mathrm{P}(x|y)\ \mathrm{P}(y)}{\mathrm{P}(x)}$$

$$\mathrm{P}(x) = \sum_y \mathrm{P}(x|y)\ \mathrm{P}(y)$$

BAYESIAN INFERENCE FOR NEURAL DATA ANALYSIS

network connectivity

parameters, y $\quad P(y)$

↑

data, x $\quad P(x|y)$

neural activity

$$P(y|x) = \frac{P(x|y)\ P(y)}{P(x)}$$

$$P(x) = \sum_y P(x|y)\ P(y)$$

BAYESIAN INFERENCE FOR NEURAL DATA ANALYSIS

network connectivity

neural activity

parameters, y — $\mathrm{P}(y)$

↑

data, x — $\mathrm{P}(x|y)$

neural activity

imaging measurements

$$\mathrm{P}(y|x) = \frac{\mathrm{P}(x|y)\ \mathrm{P}(y)}{\mathrm{P}(x)}$$

$$\mathrm{P}(x) = \sum_{y} \mathrm{P}(x|y)\ \mathrm{P}(y)$$

BAYESIAN INFERENCE FOR NEURAL DATA ANALYSIS

network connectivity

neural activity

cognitive processes

parameters, y — $\mathrm{P}(y)$

↑

data, x — $\mathrm{P}(x|y)$

$$\mathrm{P}(y|x) = \frac{\mathrm{P}(x|y)\ \mathrm{P}(y)}{\mathrm{P}(x)}$$

$$\mathrm{P}(x) = \sum_y \mathrm{P}(x|y)\ \mathrm{P}(y)$$

neural activity

imaging measurements

behaviour

NEURAL COMPUTATION

How do neural circuits solve challenging computational tasks?

NEURAL COMPUTATION

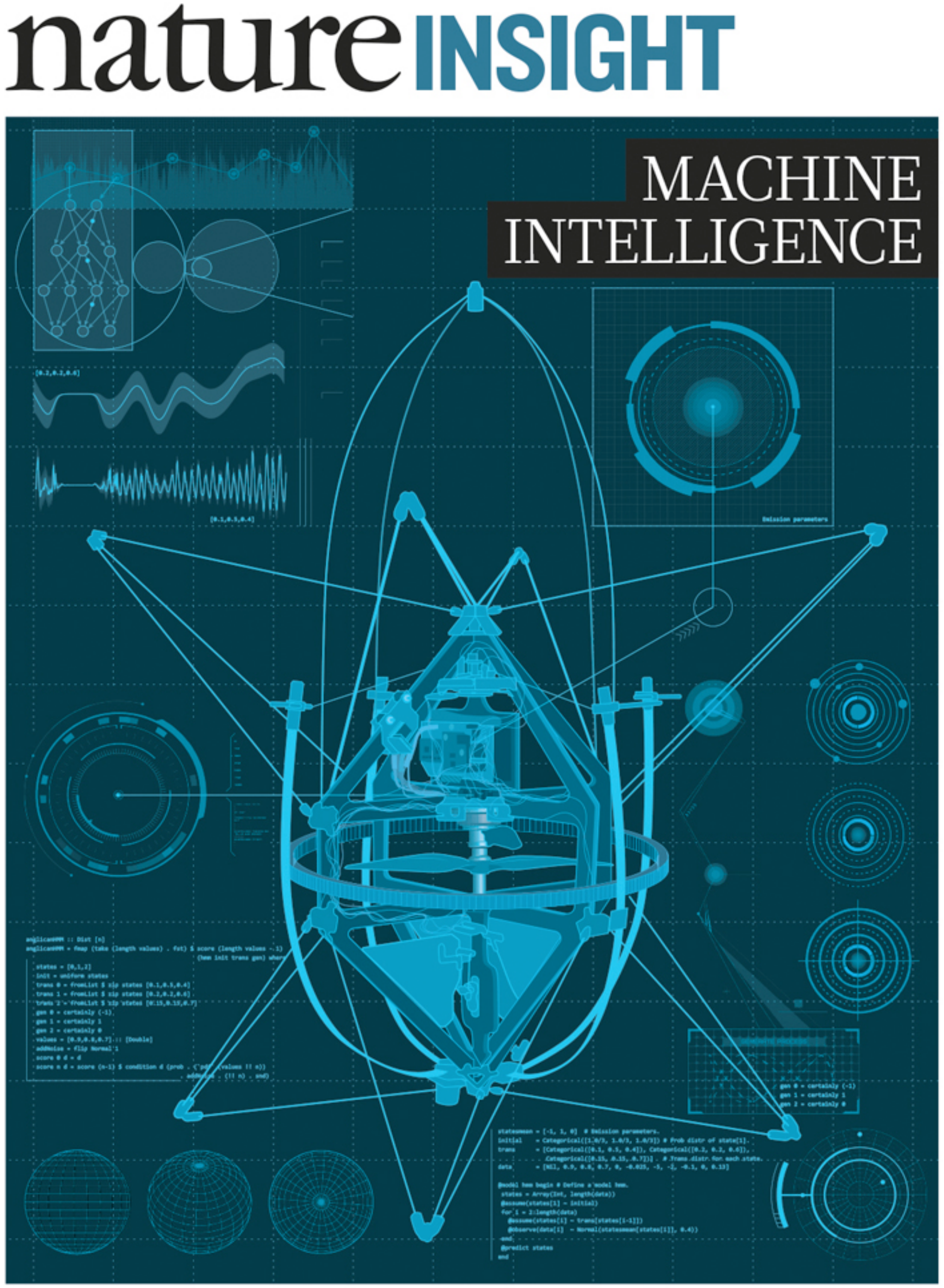
How do neural circuits solve challenging computational tasks?

What is the computational role of elements of neural circuits?

NEURAL COMPUTATION

How do neural circuits solve challenging computational tasks?

What is the computational role of elements of neural circuits?



REVIEW

doi:10.1038/nature14539

Deep learning

Yann LeCun[1,2], Yoshua Bengio[3] & Geoffrey Hinton[4,5]

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

REVIEW

doi:10.1038/nature14541

Probabilistic machine learning and artificial intelligence

Zoubin Ghahramani[1]

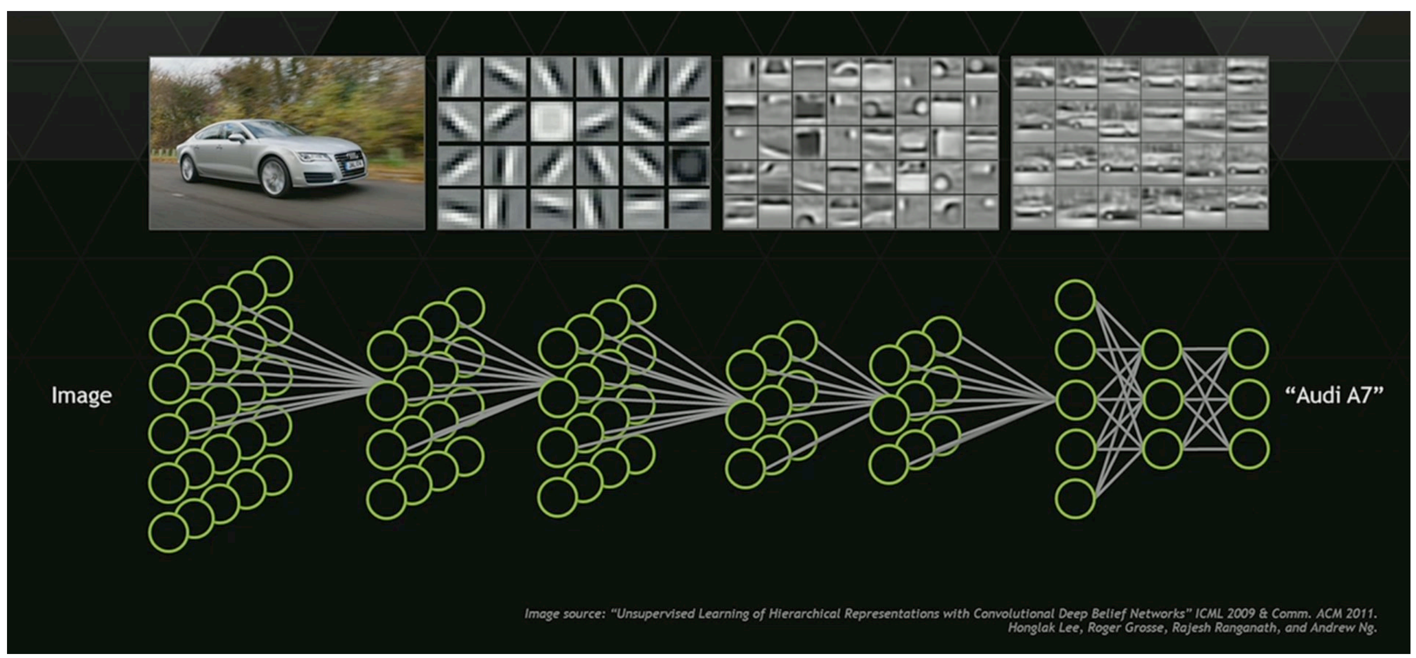
How can a machine learn from experience? Probabilistic modelling provides a framework for understanding what learning is, and has therefore emerged as one of the principal theoretical and practical approaches for designing machines that learn from data acquired through experience. The probabilistic framework, which describes how to represent and manipulate uncertainty about models and predictions, has a central role in scientific data analysis, machine learning, robotics, cognitive science and artificial intelligence. This Review provides an introduction to this framework, and discusses some of the state-of-the-art advances in the field, namely, probabilistic programming, Bayesian optimization, data compression and automatic model discovery.

NEURAL COMPUTATION

deep learning algorithms

NEURAL COMPUTATION

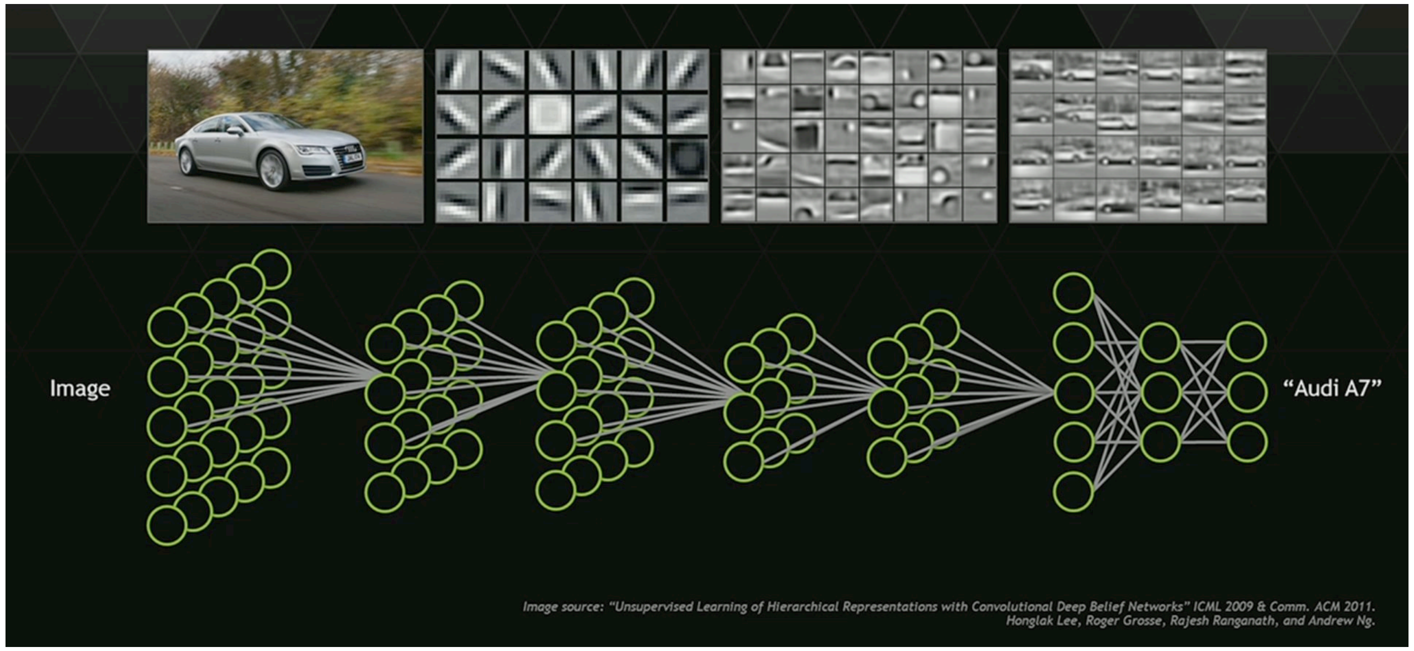
deep learning algorithms



NEURAL COMPUTATION

deep learning algorithms

+ perform well on difficult tasks
+ scale up to large data sets / parameter spaces
+ end-to-end training

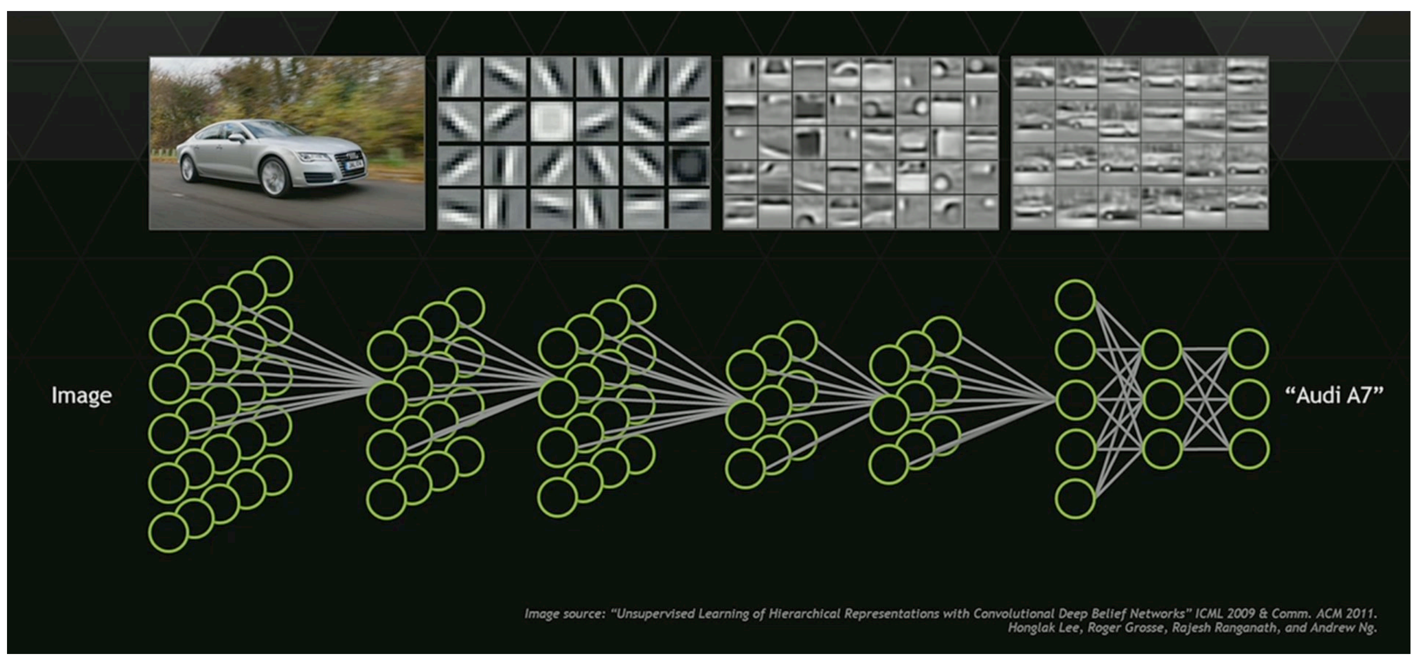


NEURAL COMPUTATION

deep learning algorithms

+ perform well on difficult tasks
+ scale up to large data sets / parameter spaces
+ end-to-end training

+ they are **neural** networks
+ realistic receptive fields



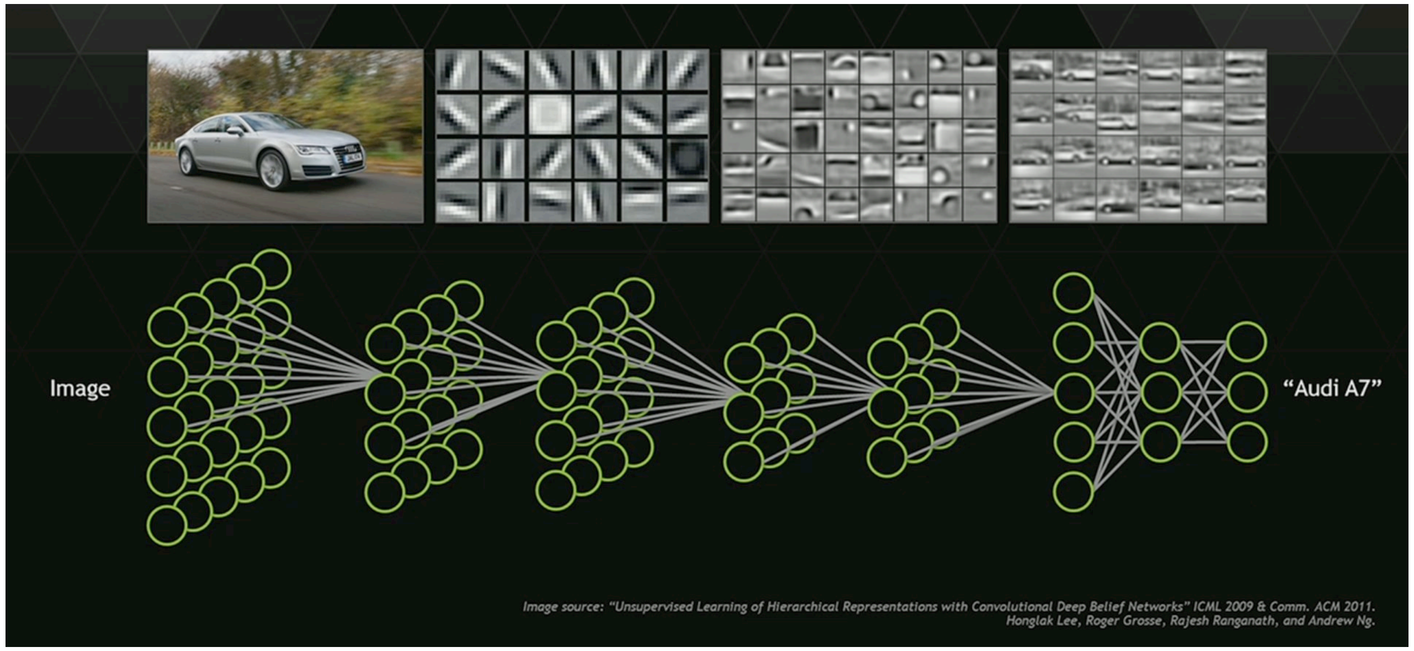
NEURAL COMPUTATION

deep learning algorithms

+ perform well on difficult tasks
+ scale up to large data sets / parameter spaces
+ end-to-end training

- supervised learning
- need lots of data
- susceptible to perturbations
  adversarial samples
- do not represent uncertainty
  or only at the final stage of computation

+ they are **neural** networks
+ realistic receptive fields



NEURAL COMPUTATION

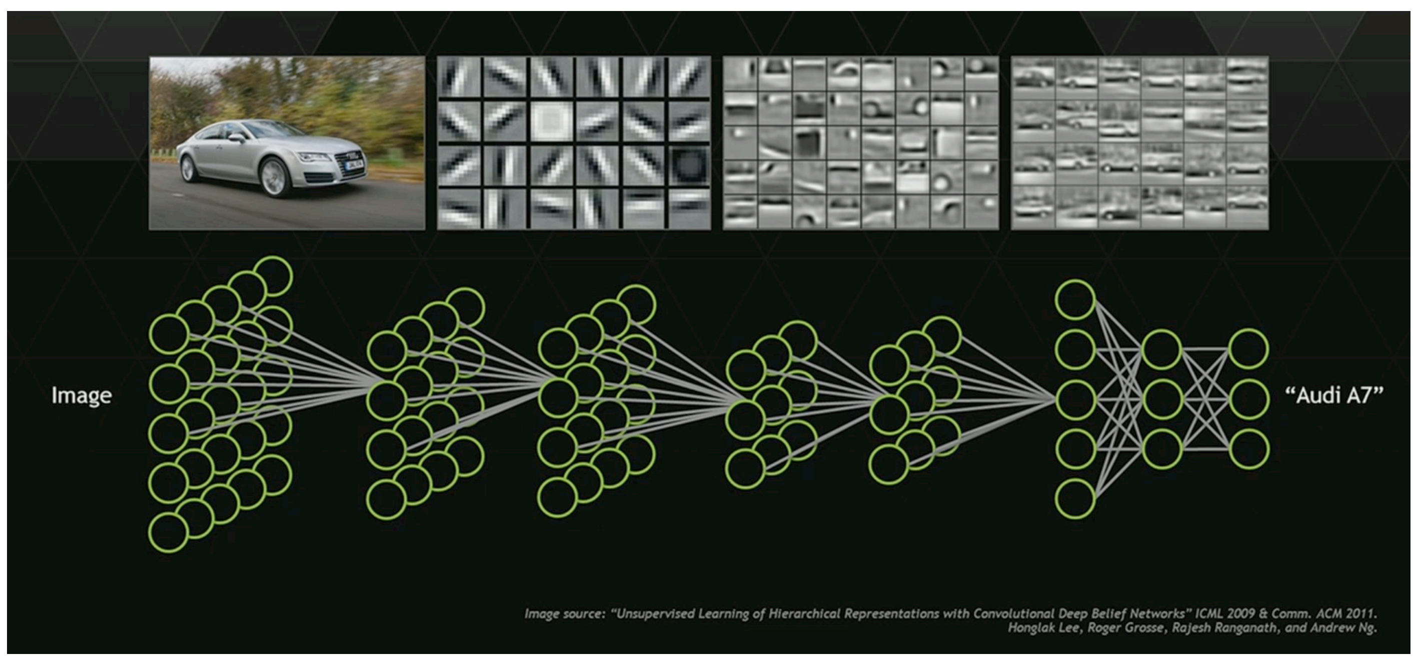
deep learning algorithms

+ perform well on difficult tasks
+ scale up to large data sets / parameter spaces
+ end-to-end training

- supervised learning
- need lots of data
- susceptible to perturbations
  adversarial samples
- do not represent uncertainty
  or only at the final stage of computation

+ they are **neural** networks
+ realistic receptive fields

- feed-forward
- mixed E and I
- saturating or threshold-linear units
- no dynamics, trial-average behaviour



NEURAL COMPUTATION

deep learning algorithms

+ perform well on difficult tasks
+ scale up to large data sets / parameter spaces
+ end-to-end training

- supervised learning
- need lots of data
- susceptible to perturbations
  adversarial samples
- do not represent uncertainty
  or only at the final stage of computation

+ they are **neural** networks
+ realistic receptive fields

- feed-forward
- mixed E and I
- saturating or threshold-linear units
- no dynamics, trial-average behaviour

probabilistic algorithms

NEURAL COMPUTATION

deep learning algorithms

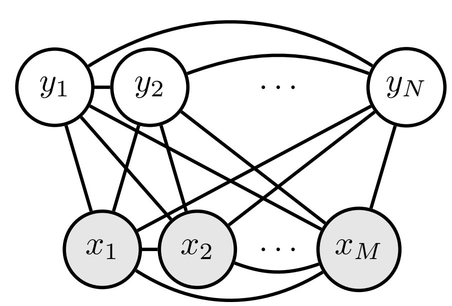
+ perform well on difficult tasks
+ scale up to large data sets / parameter spaces
+ end-to-end training

- supervised learning
- need lots of data
- susceptible to perturbations
  adversarial samples
- do not represent uncertainty
  or only at the final stage of computation

+ they are **neural** networks
+ realistic receptive fields

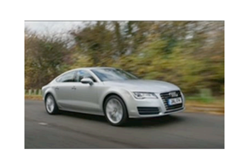
- feed-forward
- mixed E and I
- saturating or threshold-linear units
- no dynamics, trial-average behaviour

probabilistic algorithms



latent causes “Audi A7”

observations



NEURAL COMPUTATION

deep learning algorithms

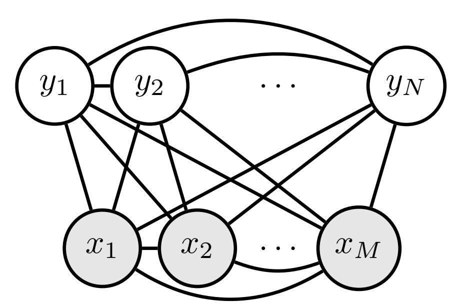
+ perform well on difficult tasks
+ scale up to large data sets / parameter spaces
+ end-to-end training

- supervised learning
- need lots of data
- susceptible to perturbations
  adversarial samples
- do not represent uncertainty
  or only at the final stage of computation

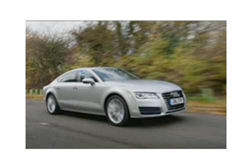
+ they are **neural** networks
+ realistic receptive fields

- feed-forward
- mixed E and I
- saturating or threshold-linear units
- no dynamics, trial-average behaviour

probabilistic algorithms



latent causes “Audi A7”

observations 

noise
ambiguity
partial knowledge

NEURAL COMPUTATION

deep learning algorithms

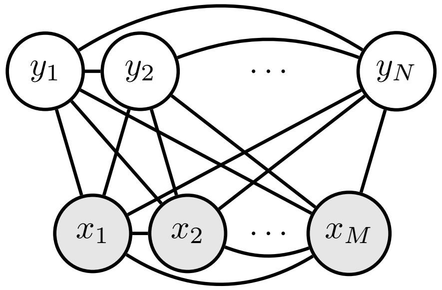
+ perform well on difficult tasks
+ scale up to large data sets / parameter spaces
+ end-to-end training

- supervised learning
- need lots of data
- susceptible to perturbations
  adversarial samples
- do not represent uncertainty
  or only at the final stage of computation

+ they are **neural** networks
+ realistic receptive fields

- feed-forward
- mixed E and I
- saturating or threshold-linear units
- no dynamics, trial-average behaviour

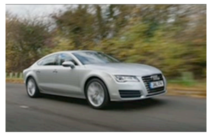
probabilistic algorithms



latent causes "Audi A7"

$\mathrm{P}(x, y)$

observations



noise
ambiguity
partial knowledge

NEURAL COMPUTATION

deep learning algorithms

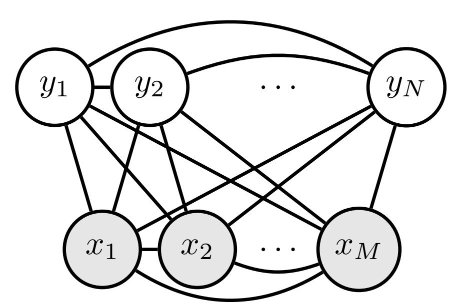
+ perform well on difficult tasks
+ scale up to large data sets / parameter spaces
+ end-to-end training

- supervised learning
- need lots of data
- susceptible to perturbations
  adversarial samples
- do not represent uncertainty
  or only at the final stage of computation

+ they are **neural** networks
+ realistic receptive fields

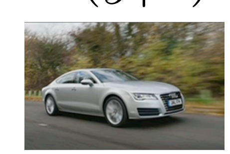
- feed-forward
- mixed E and I
- saturating or threshold-linear units
- no dynamics, trial-average behaviour

probabilistic algorithms



latent causes

observations



$$P(y|x) \propto \frac{P(x, y)}{\int P(x, y')\, dy'}$$

noise
ambiguity
partial knowledge

NEURAL COMPUTATION

deep learning algorithms

+ perform well on difficult tasks
+ scale up to large data sets / parameter spaces
+ end-to-end training

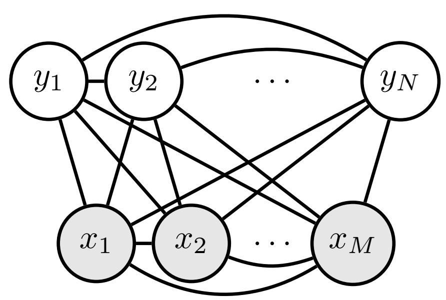
- supervised learning
- need lots of data
- susceptible to perturbations
  adversarial samples
- do not represent uncertainty
  or only at the final stage of computation

+ they are **neural** networks
+ realistic receptive fields

- feed-forward
- mixed E and I
- saturating or threshold-linear units
- no dynamics, trial-average behaviour

probabilistic algorithms

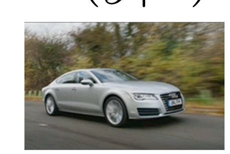
+ perform well on some tasks
+ unsupervised learning
+ generalise from few data points
  one shot learning
+ robust to noisy and missing data
+ represent **uncertainty**
  at every stage stage of computation



latent causes

observations

"Audi A7"



$$P(y|x) \propto \frac{P(x,y)}{\int P(x,y')\,dy'}$$

noise
ambiguity
partial knowledge

NEURAL COMPUTATION

deep learning algorithms

+ perform well on difficult tasks
+ scale up to large data sets / parameter spaces
+ end-to-end training

- supervised learning
- need lots of data
- susceptible to perturbations
  adversarial samples
- do not represent uncertainty
  or only at the final stage of computation

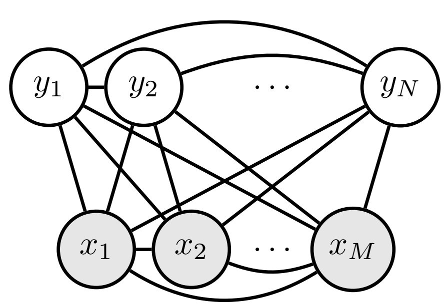
+ they are **neural** networks
+ realistic receptive fields

- feed-forward
- mixed E and I
- saturating or threshold-linear units
- no dynamics, trial-average behaviour

probabilistic algorithms

+ perform well on some tasks
+ unsupervised learning
+ generalise from few data points
  one shot learning
+ robust to noisy and missing data
+ represent **uncertainty**
  at every stage stage of computation

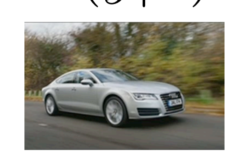
- fewer success stories for performance
- computationally intensive
- do not scale well
- end-to-end training difficult



latent causes

observations

“Audi A7”



$$P(y|x) \propto \frac{P(x,y)}{\int P(x,y')\,dy'}$$

noise
ambiguity
partial knowledge

NEURAL COMPUTATION

deep learning algorithms

+ perform well on difficult tasks
+ scale up to large data sets / parameter spaces
+ end-to-end training

- supervised learning
- need lots of data
- susceptible to perturbations
  adversarial samples
- do not represent uncertainty
  or only at the final stage of computation

+ they are **neural** networks
+ realistic receptive fields

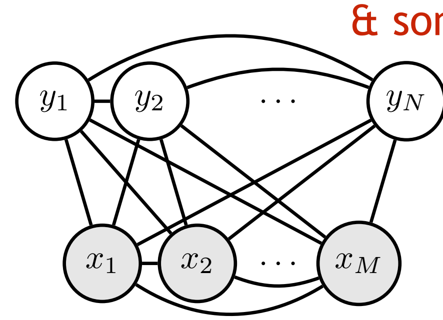
- feed-forward
- mixed E and I
- saturating or threshold-linear units
- no dynamics, trial-average behaviour

probabilistic algorithms

+ perform well on some tasks
+ unsupervised learning
+ generalise from few data points
  one shot learning
+ robust to noisy and missing data
+ represent **uncertainty**
  at every stage stage of computation

- fewer success stories for performance
- computationally intensive
- do not scale well
- end-to-end training difficult

+ behavioural evidence
  & some neural evidence



latent causes

observations

noise
ambiguity
partial knowledge

"Audi A7"

$$P(y|x) \propto \frac{P(x,y)}{\int P(x,y')\,dy'}$$

NEURAL COMPUTATION

deep learning algorithms

+ perform well on difficult tasks
+ scale up to large data sets / parameter spaces
+ end-to-end training

- supervised learning
- need lots of data
- susceptible to perturbations
  adversarial samples
- do not represent uncertainty
  or only at the final stage of computation

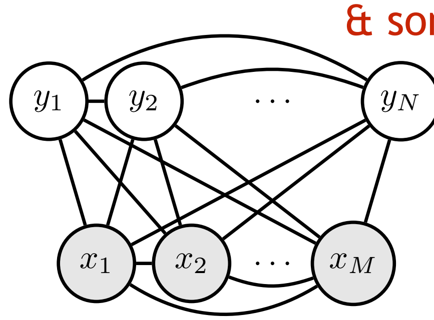
+ they are **neural** networks
+ realistic receptive fields

- feed-forward
- mixed E and I
- saturating or threshold-linear units
- no dynamics, trial-average behaviour

probabilistic algorithms

+ perform well on some tasks
+ unsupervised learning
+ generalise from few data points
  one shot learning
+ robust to noisy and missing data
+ represent **uncertainty**
  at every stage stage of computation

- fewer success stories for performance
- computationally intensive
- do not scale well
- end-to-end training difficult



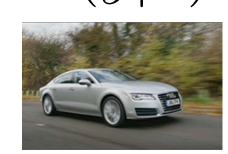
noise
ambiguity
partial knowledge

+ behavioural evidence
  & some neural evidence

latent causes

observations

"Audi A7"



$$P(y|x) \propto \frac{P(x,y)}{\int P(x,y')\,dy'}$$

- **not neural**-like

UNCONSCIOUS INFERENCES

UNCONSCIOUS INFERENCES

which words?

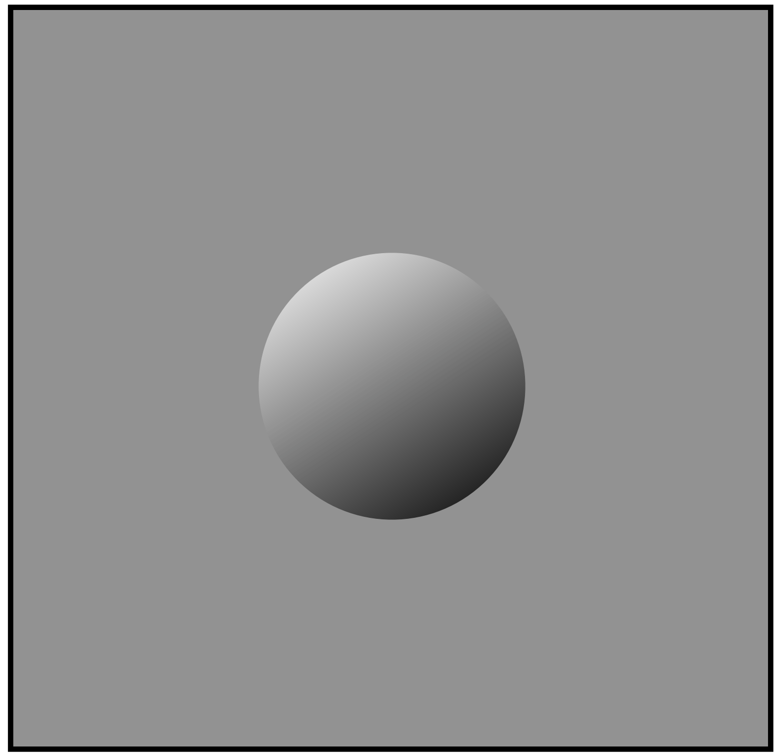
UNCONSCIOUS INFERENCES

which words?

UNCONSCIOUS INFERENCES

which words?

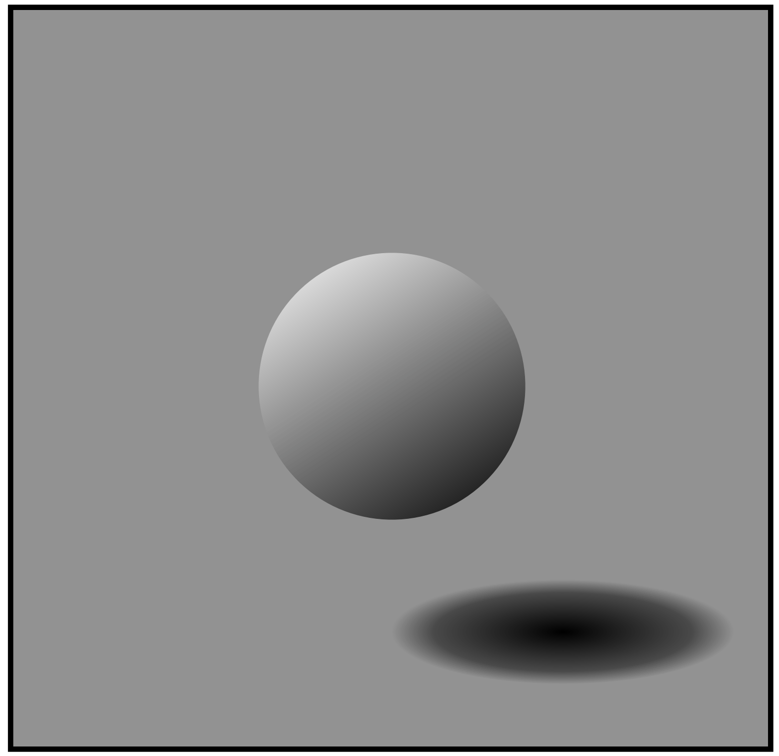
bubble or dimple?



UNCONSCIOUS INFERENCES

which words?

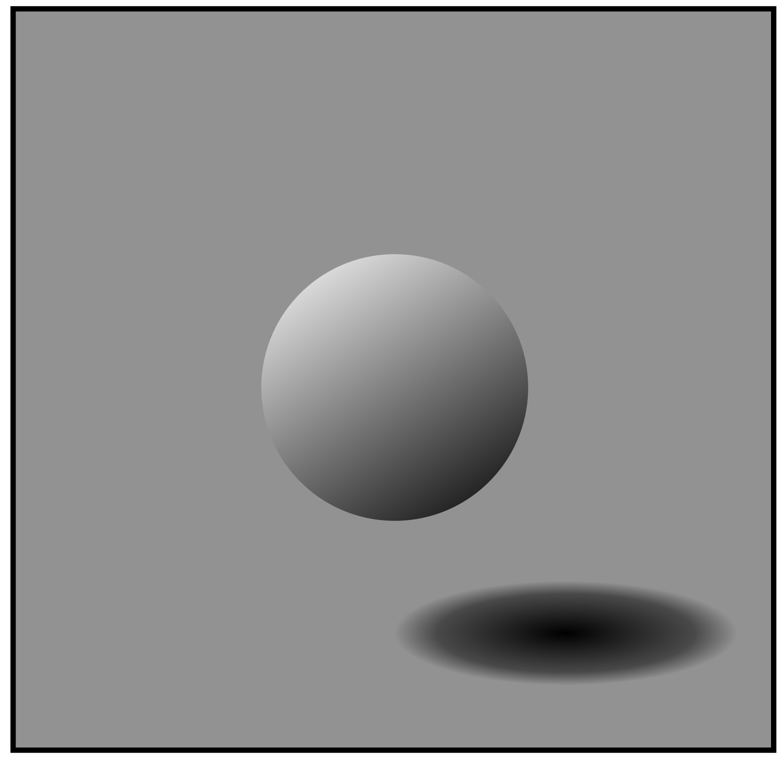
bubble or dimple?



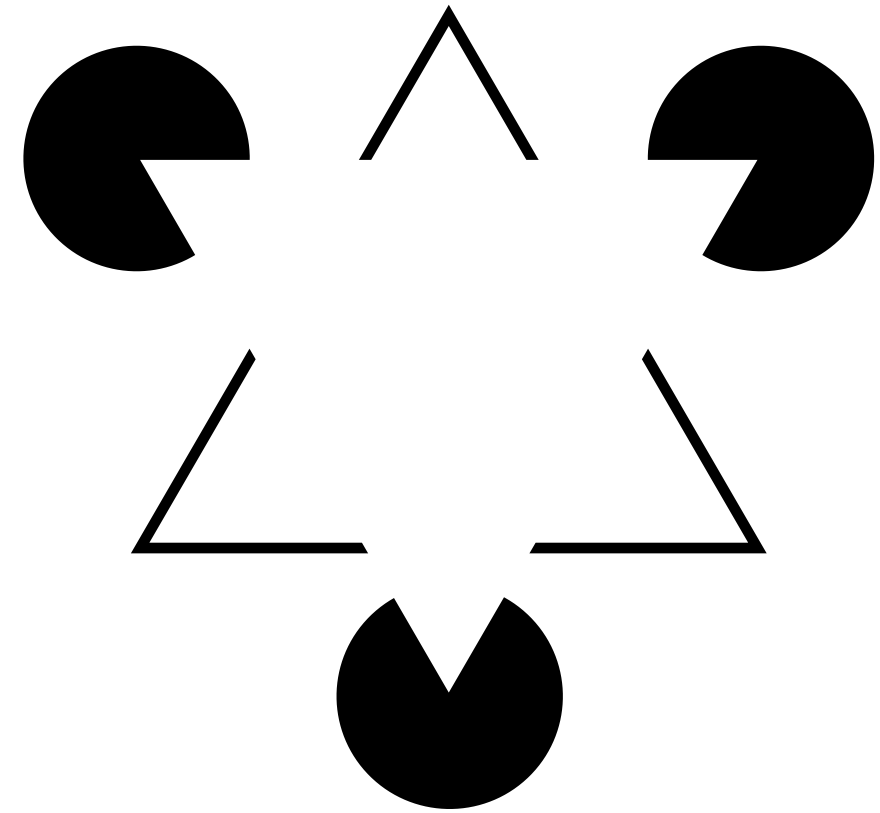
UNCONSCIOUS INFERENCES

which words?

bubble or dimple?



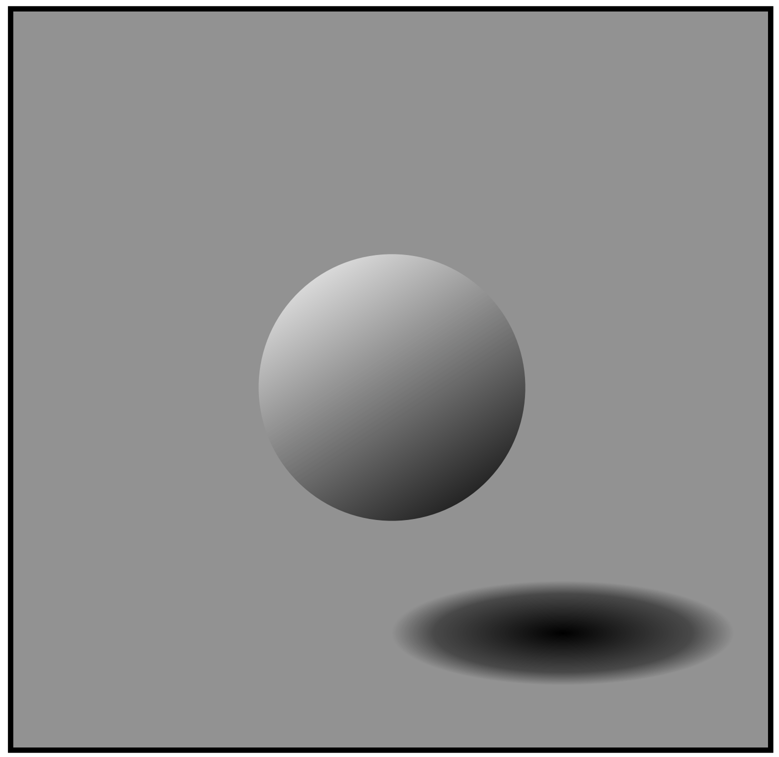
how many triangles?



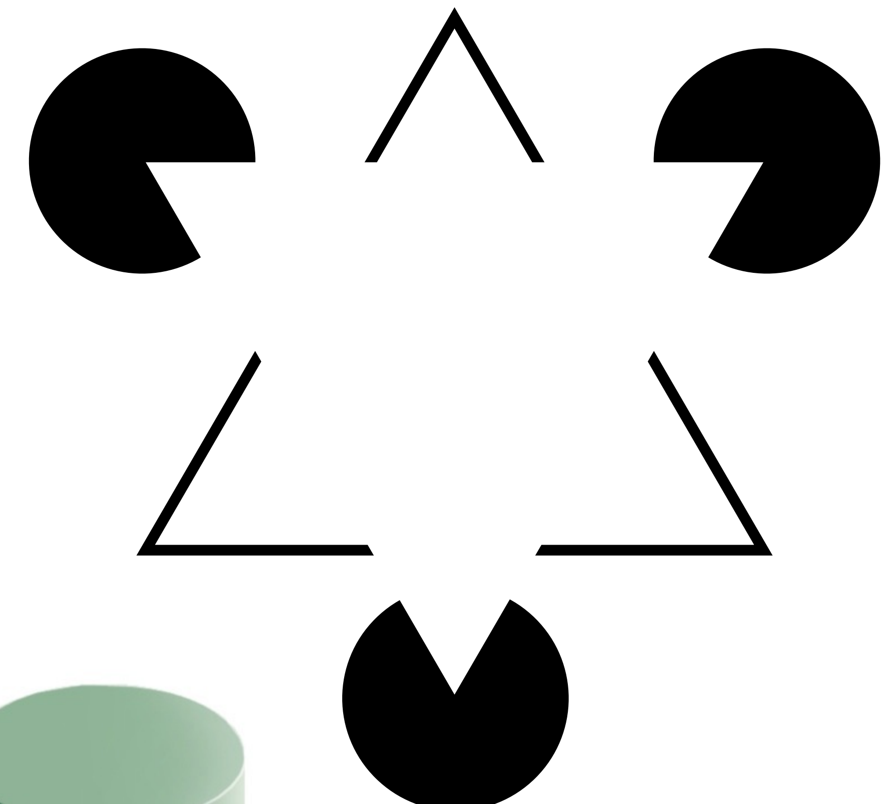
UNCONSCIOUS INFERENCES

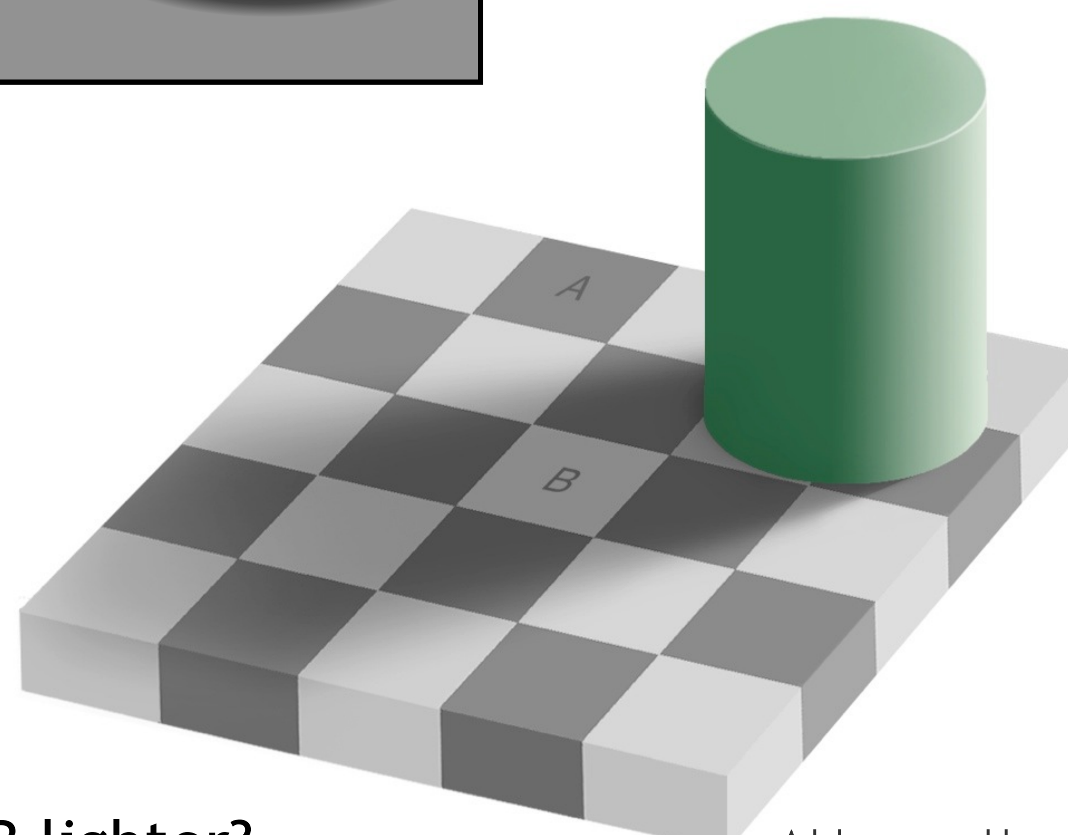
which words?

bubble or dimple?



how many triangles?





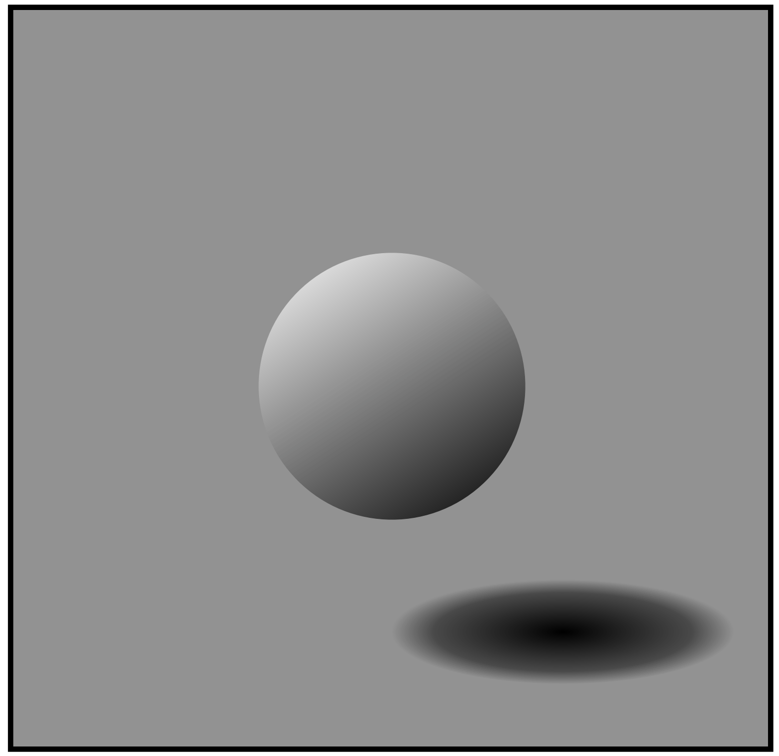
is A or B lighter?

Adelson, unpubl

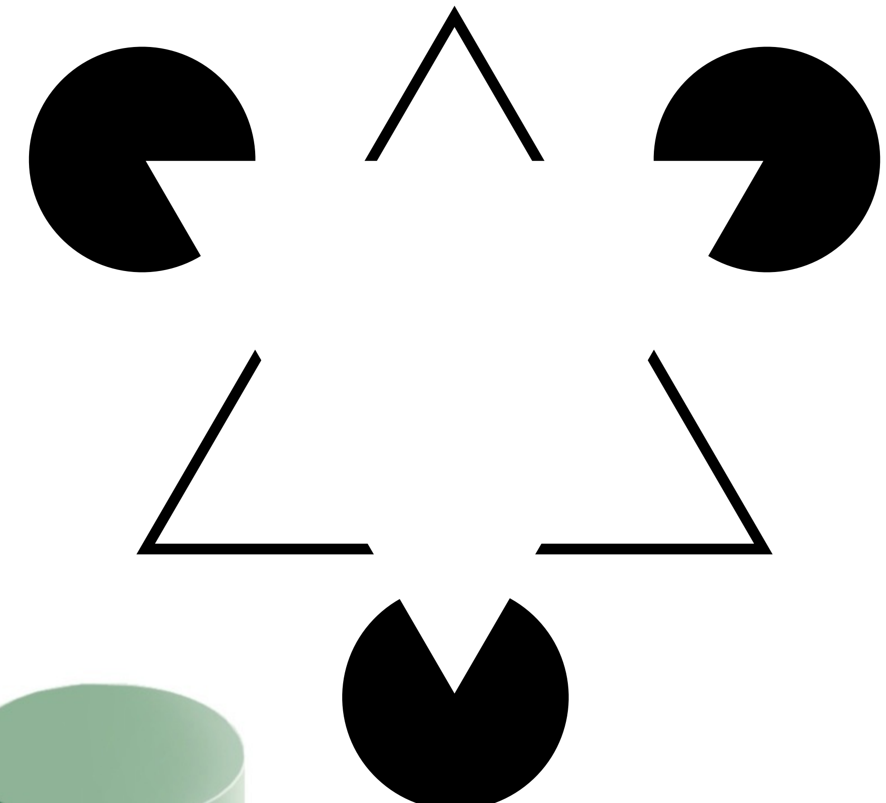
UNCONSCIOUS INFERENCES

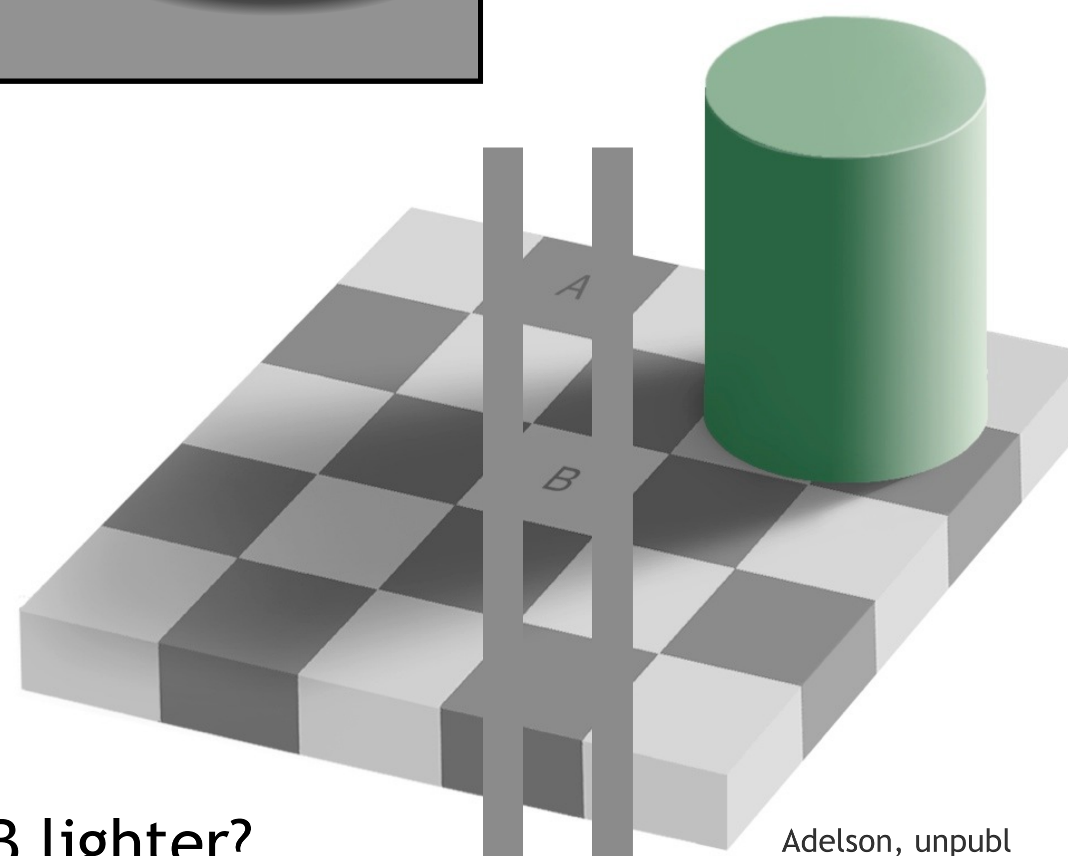
which words?

bubble or dimple?



how many triangles?





is A or B lighter?

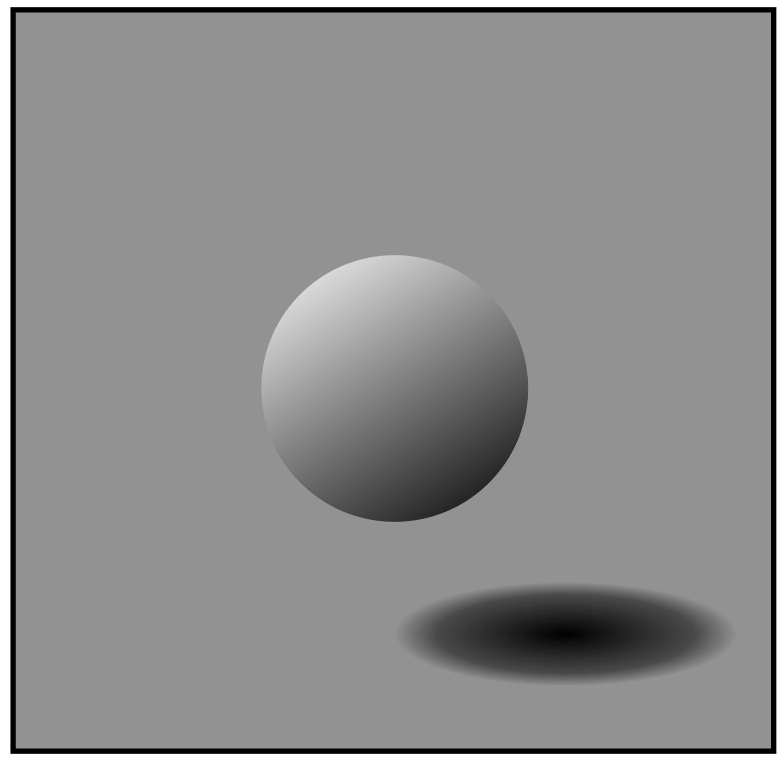
Adelson, unpubl

UNCONSCIOUS INFERENCES

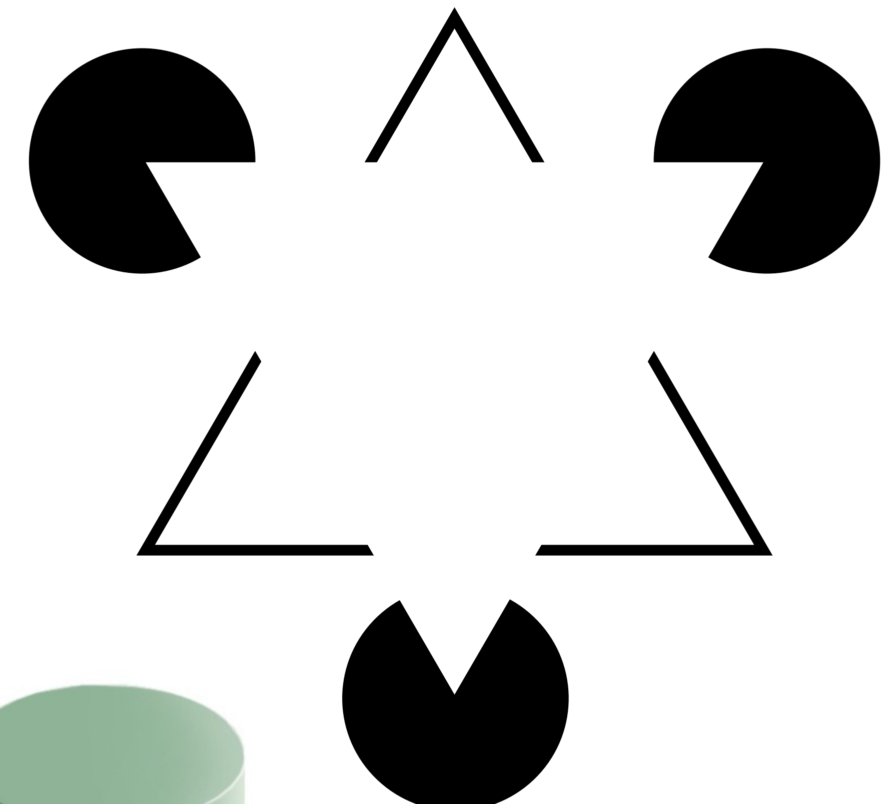
which words?

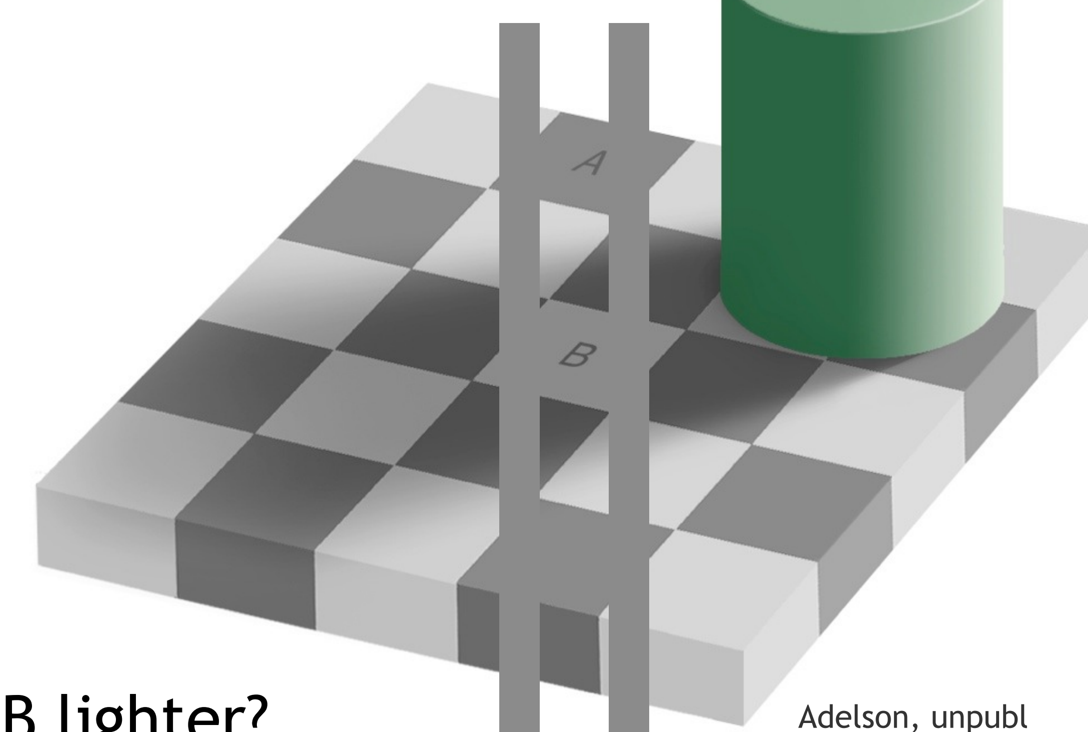
seat

bubble or dimple?



how many triangles?





is A or B lighter?

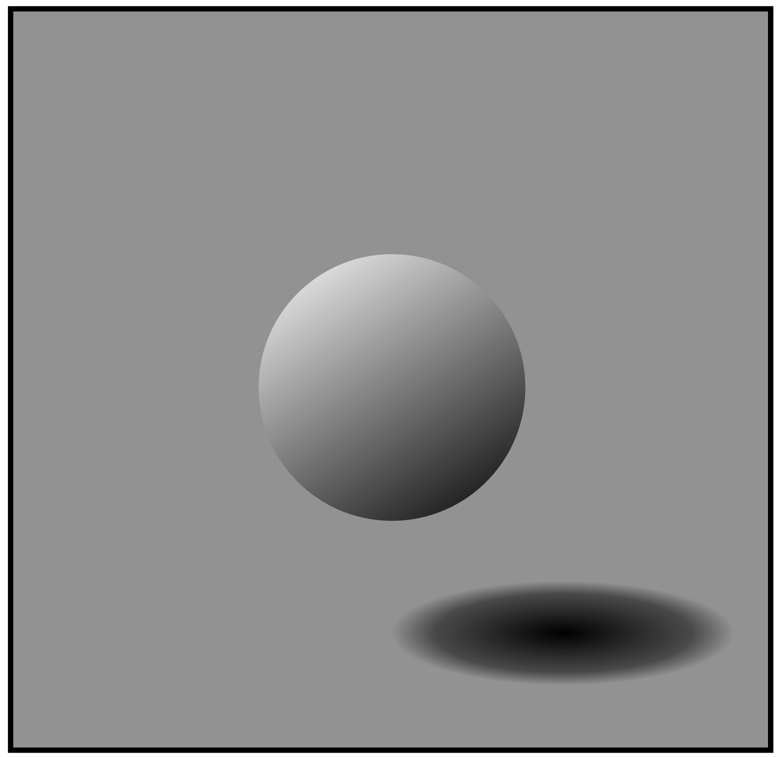
Adelson, unpubl

UNCONSCIOUS INFERENCES

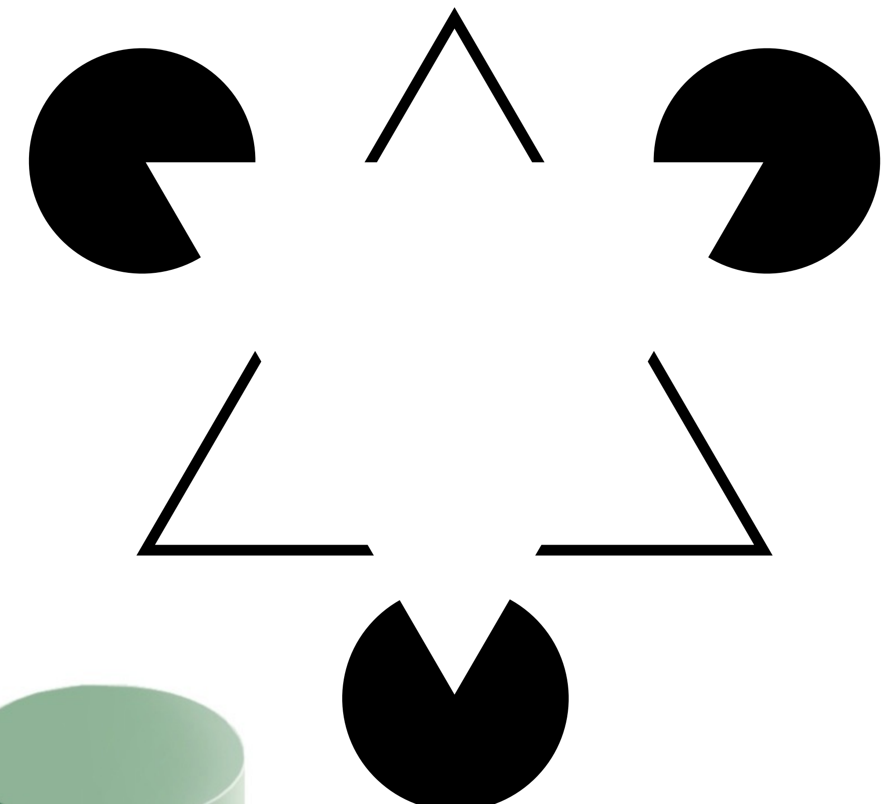
which words?

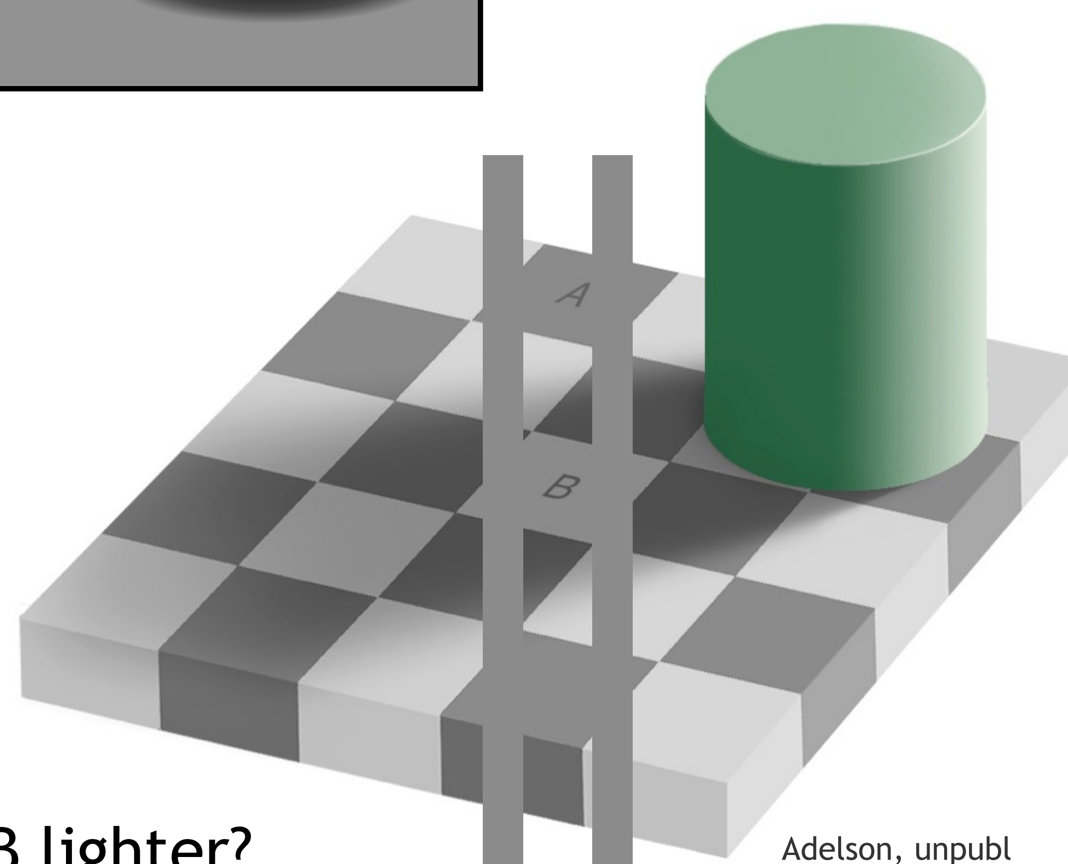
✓ seat

bubble or dimple?



how many triangles?





is A or B lighter?

Adelson, unpubl

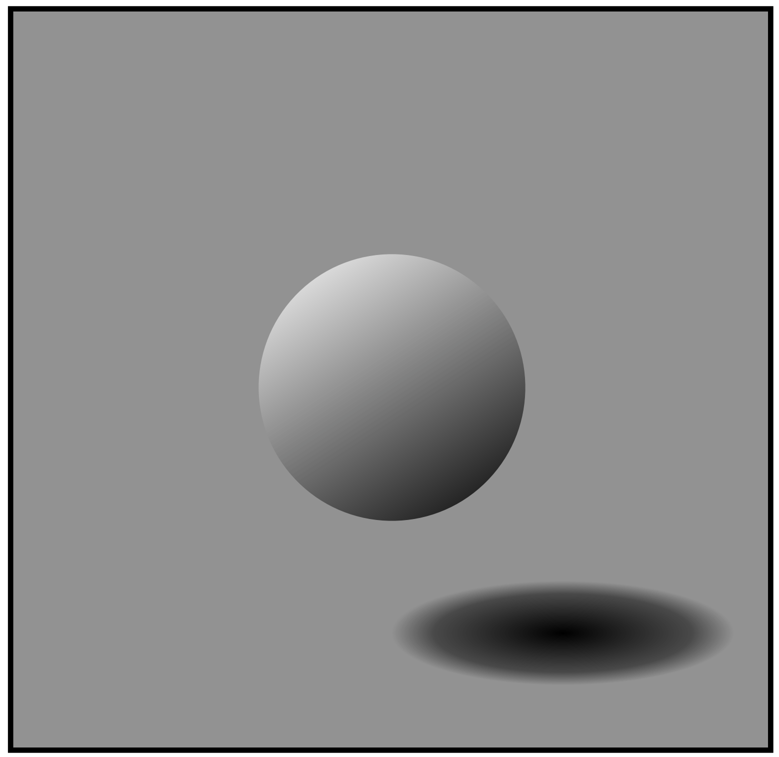
UNCONSCIOUS INFERENCES

which words?

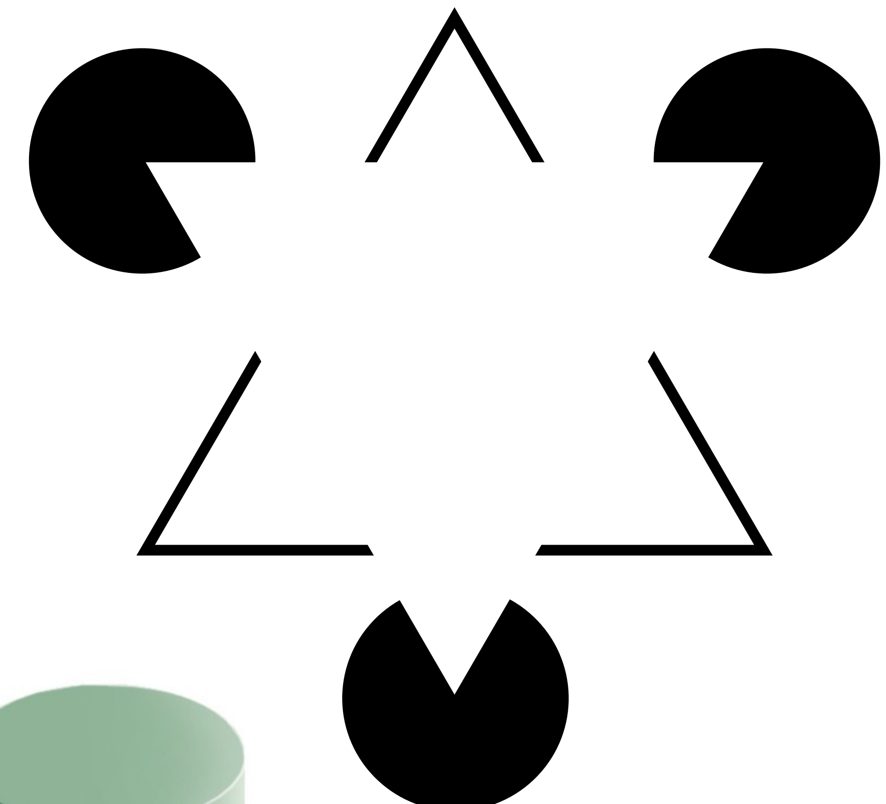
✓ seat

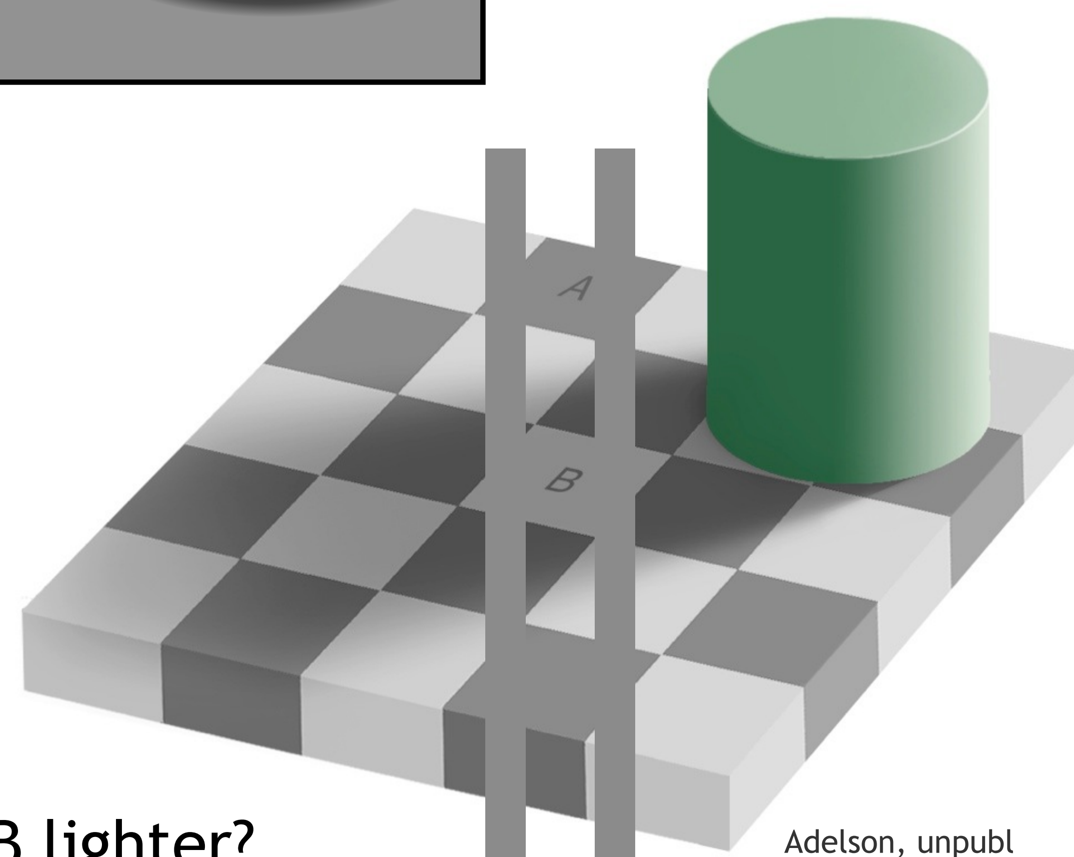
radio

bubble or dimple?



how many triangles?





is A or B lighter?

Adelson, unpubl

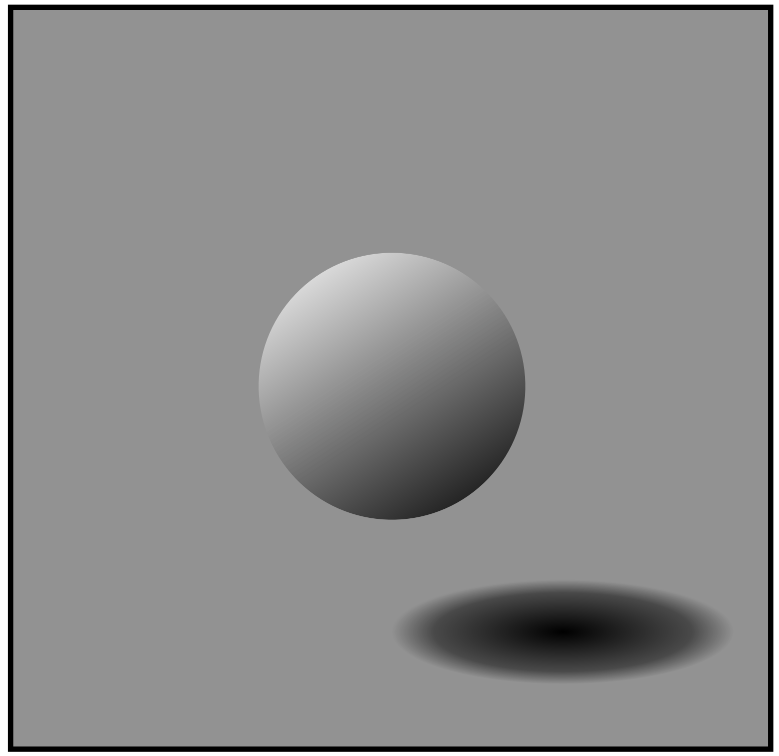
UNCONSCIOUS INFERENCES

which words?

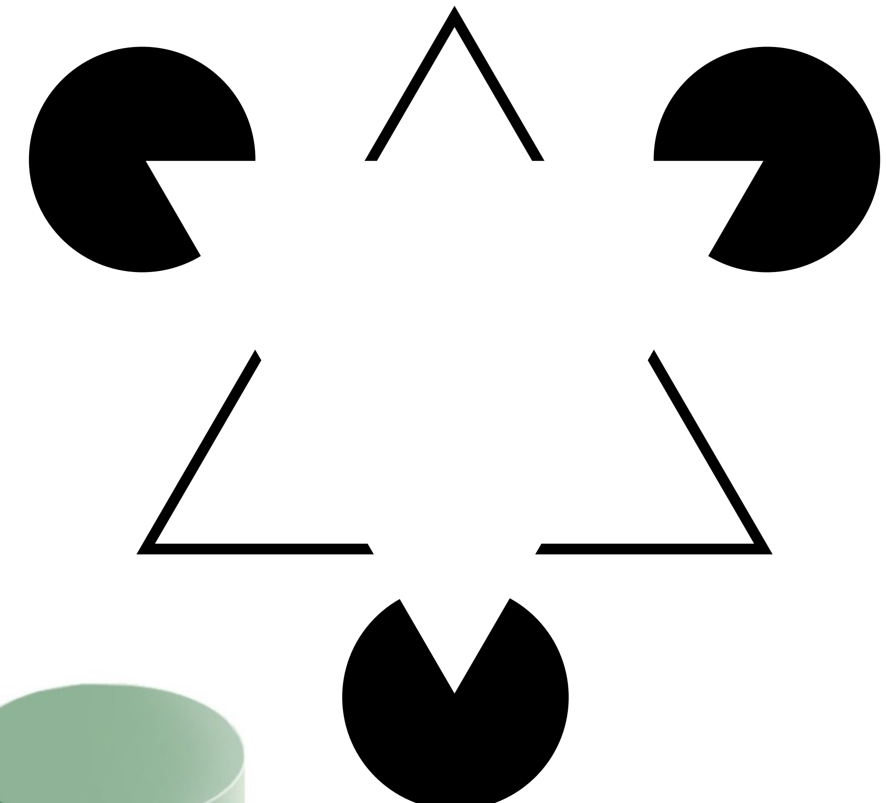
✓ seat

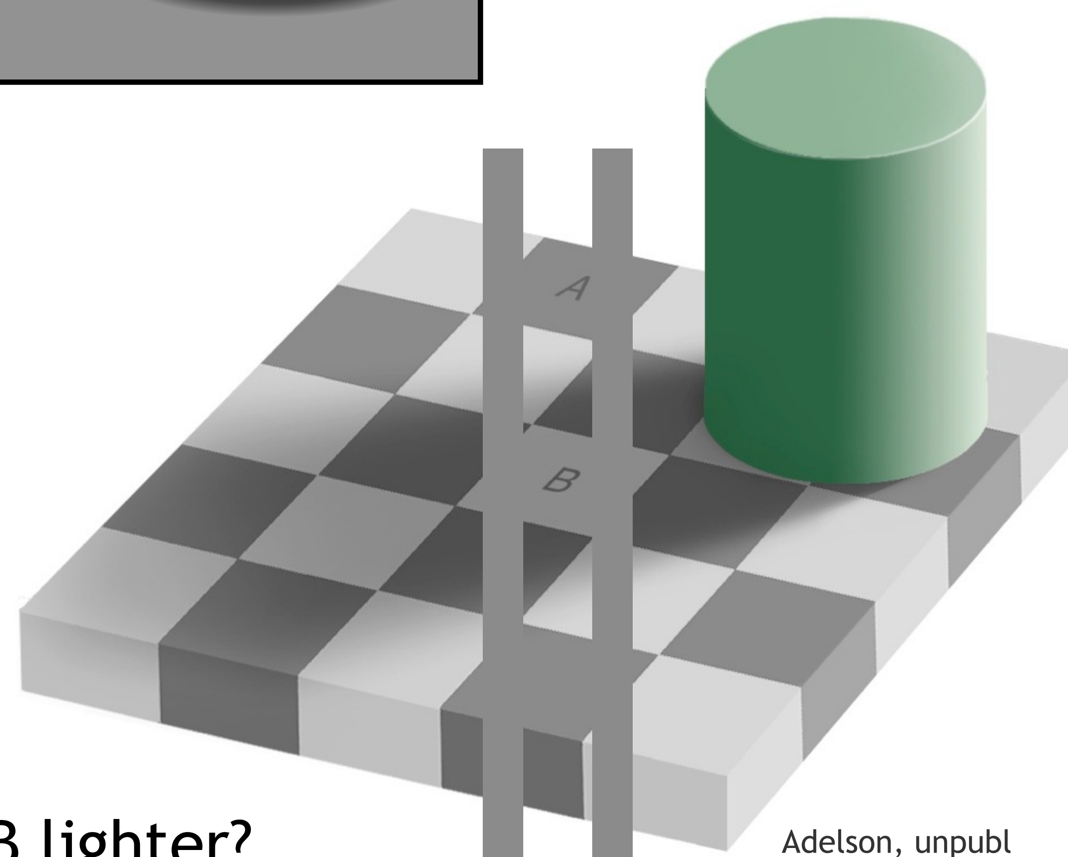
✗ radio

bubble or dimple?



how many triangles?





is A or B lighter?

Adelson, unpubl

UNCONSCIOUS INFERENCES

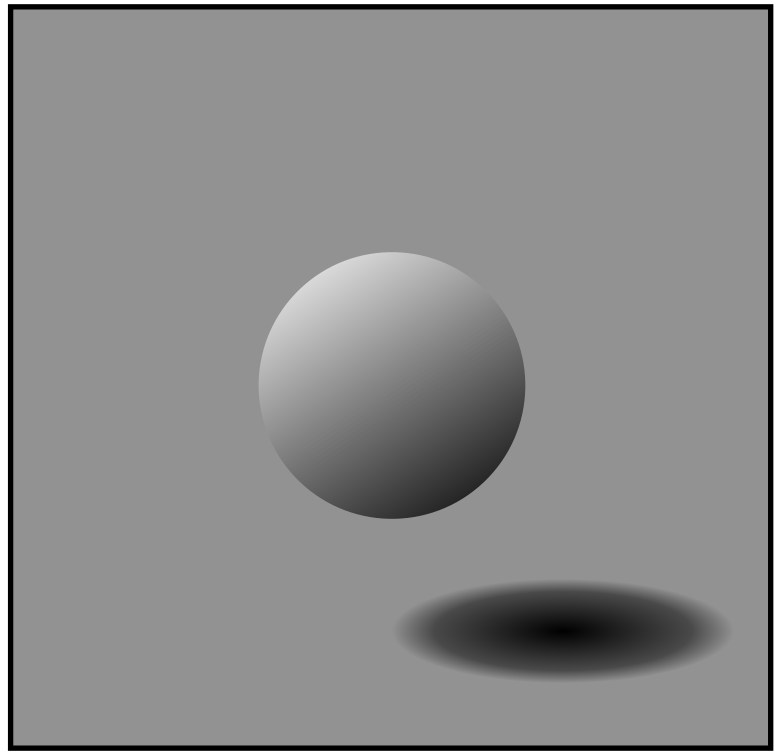
which words?

✓ seat

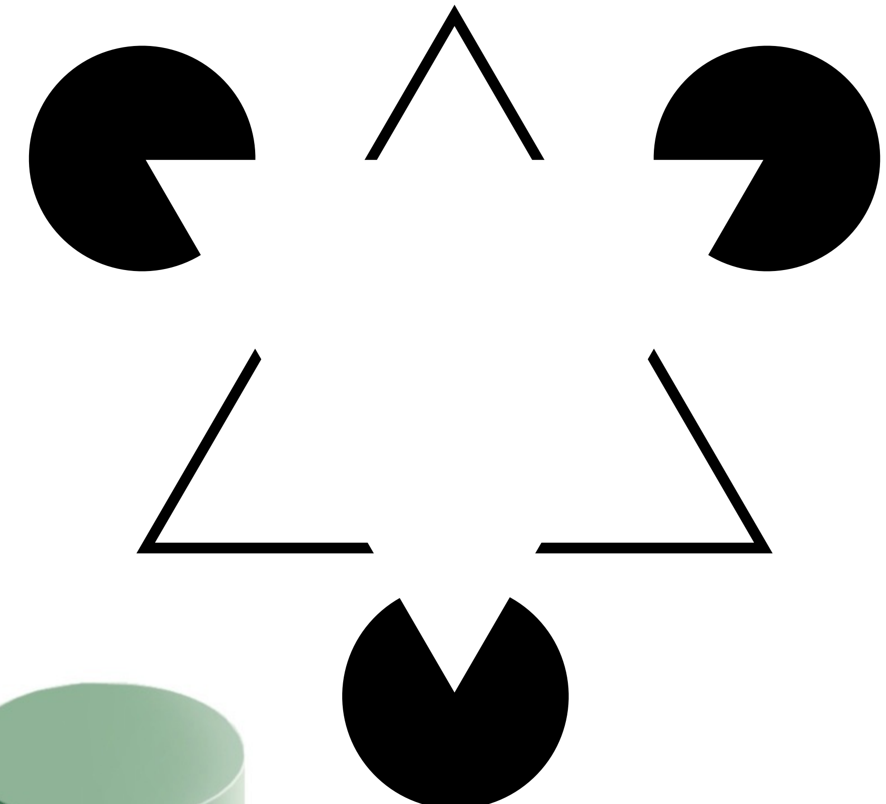
✗ radio

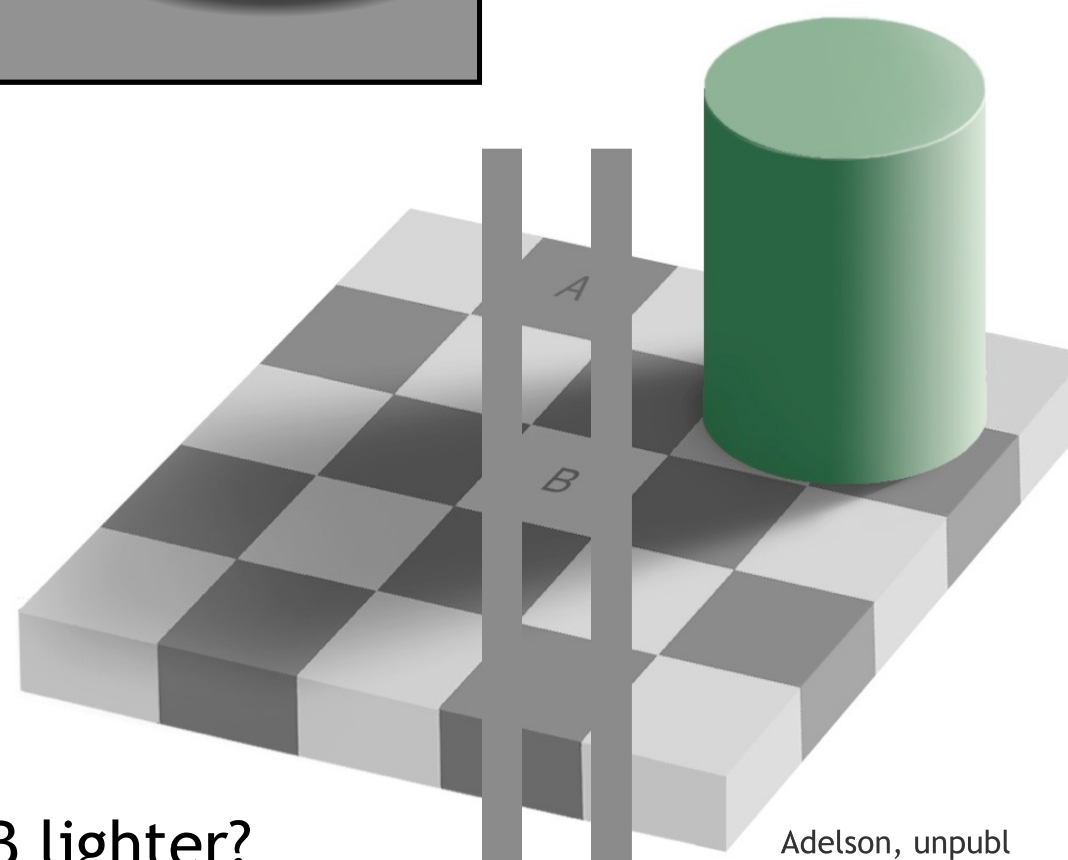
table

bubble or dimple?



how many triangles?





is A or B lighter?

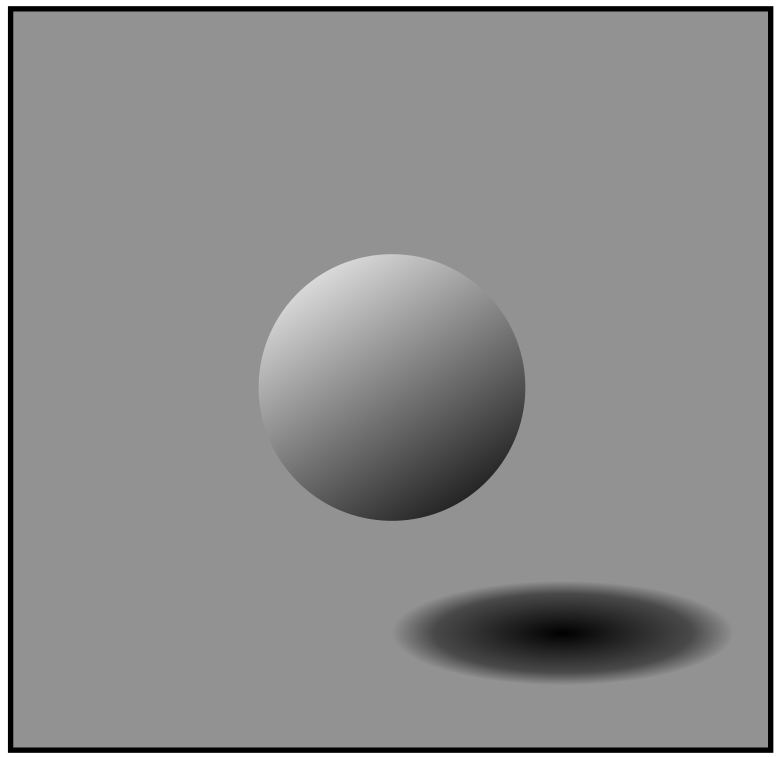
Adelson, unpubl

UNCONSCIOUS INFERENCES

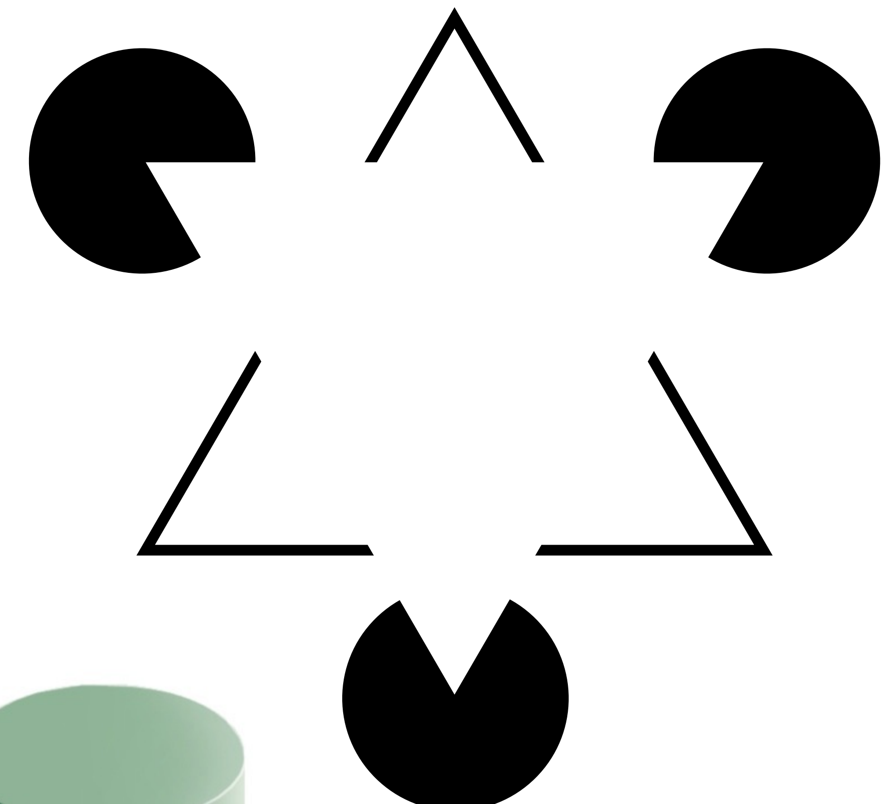
which words?

- ✓ seat
- ✗ radio
- ✓ table

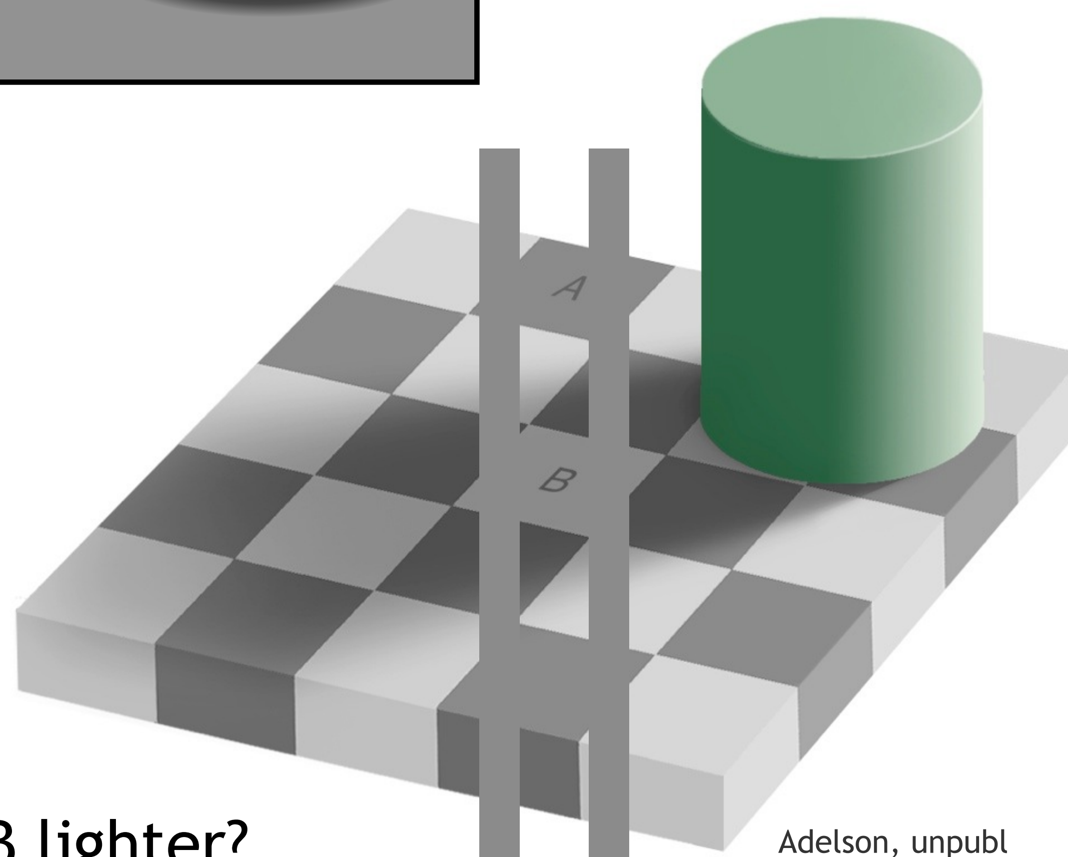
bubble or dimple?



how many triangles?



is A or B lighter?



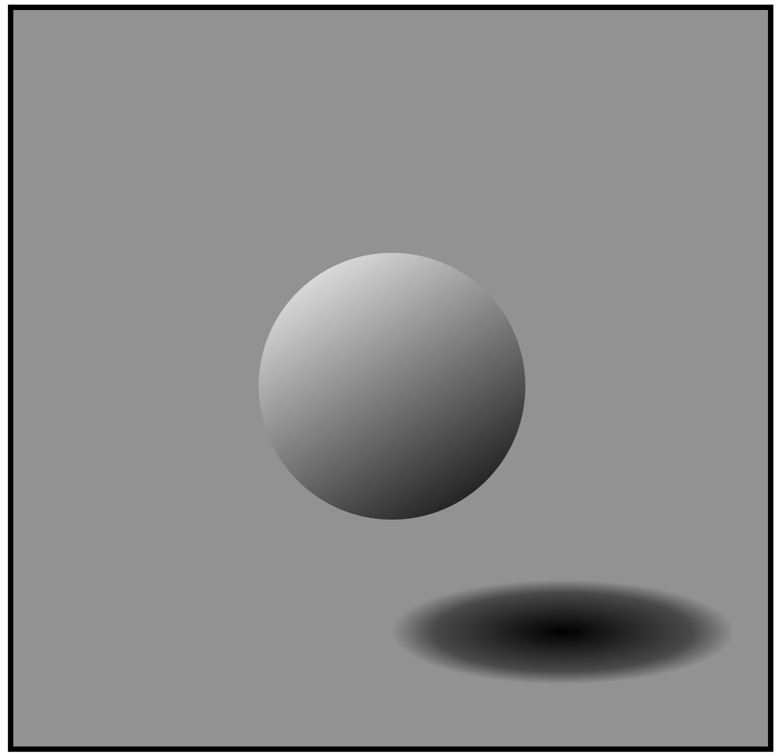
Adelson, unpubl

UNCONSCIOUS INFERENCES

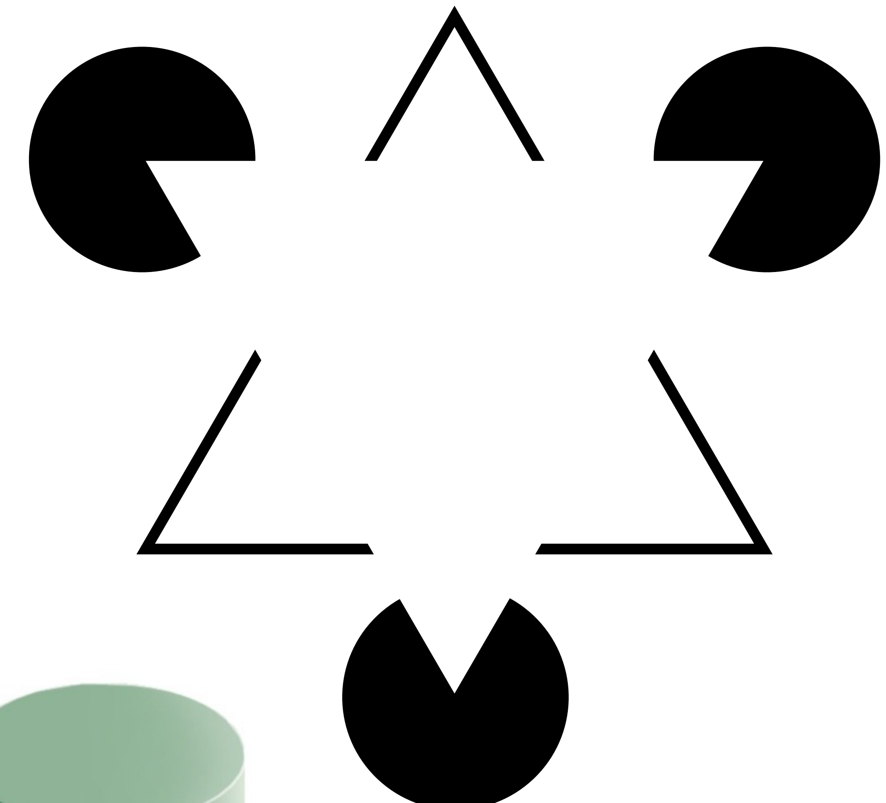
which words?

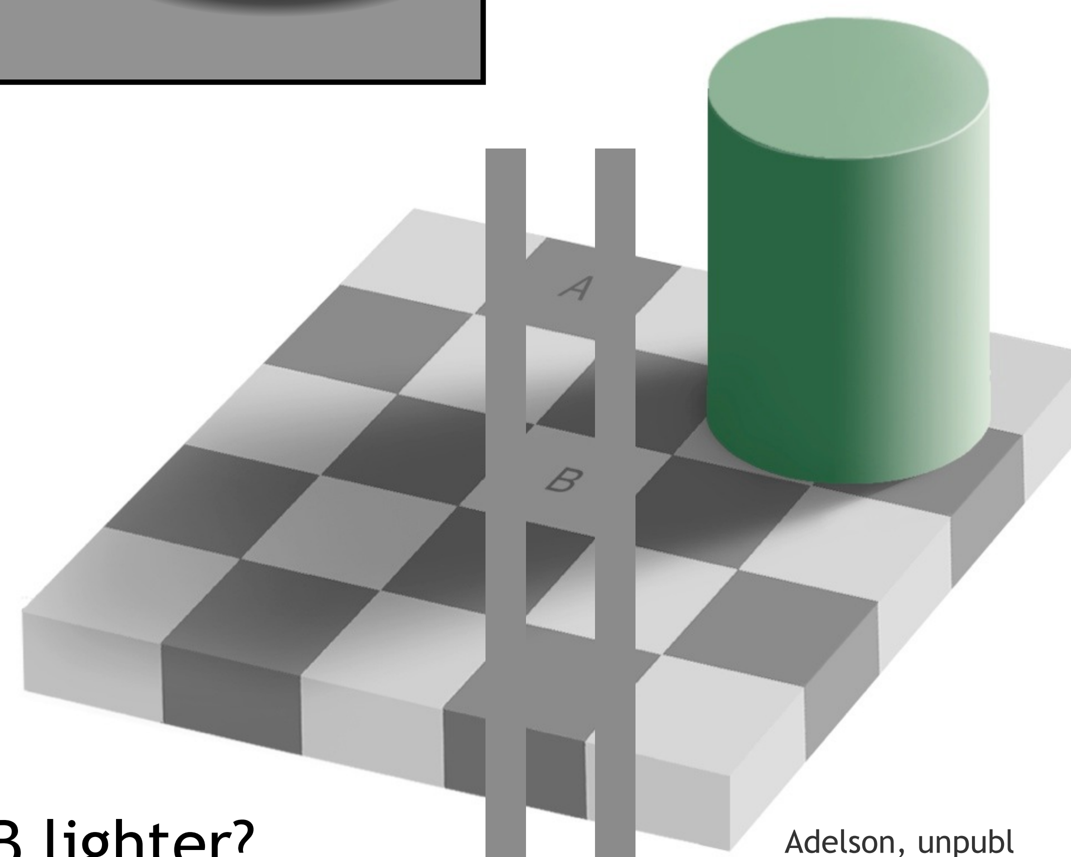
- ✓ seat
- ✗ radio
- ✓ table
- rocking

bubble or dimple?



how many triangles?





is A or B lighter?

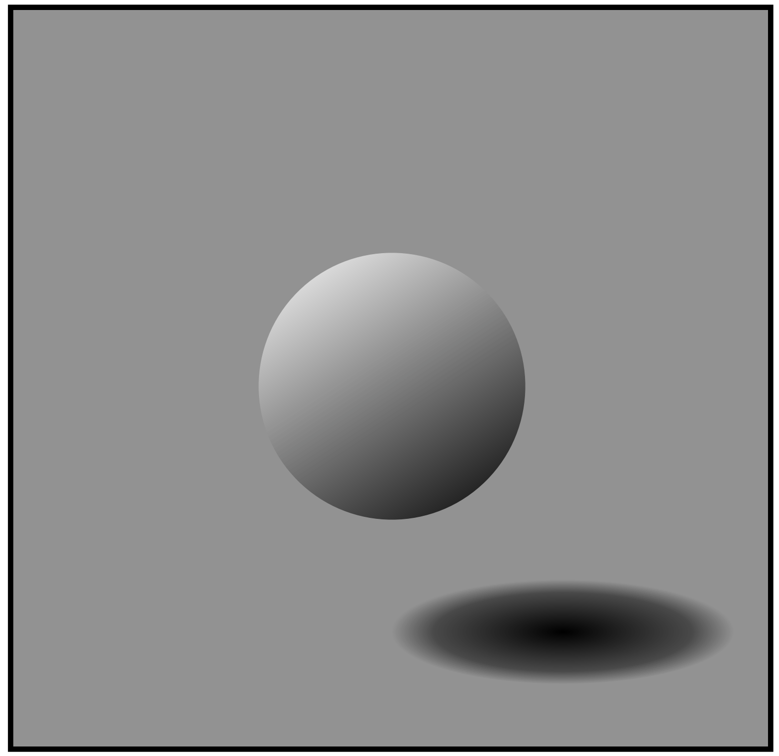
Adelson, unpubl

UNCONSCIOUS INFERENCES

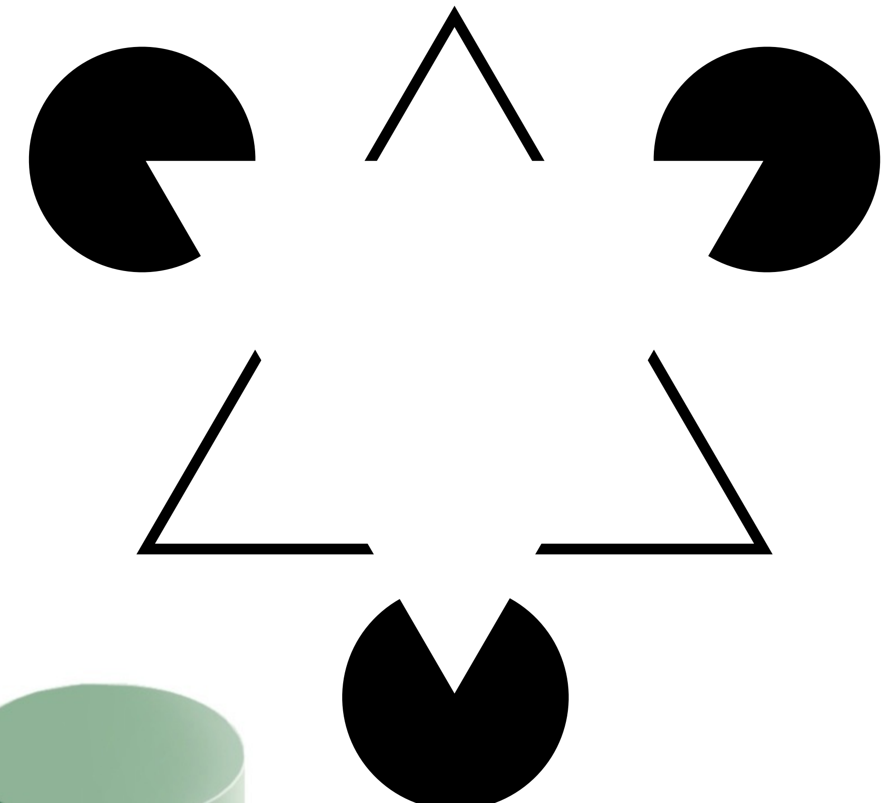
which words?

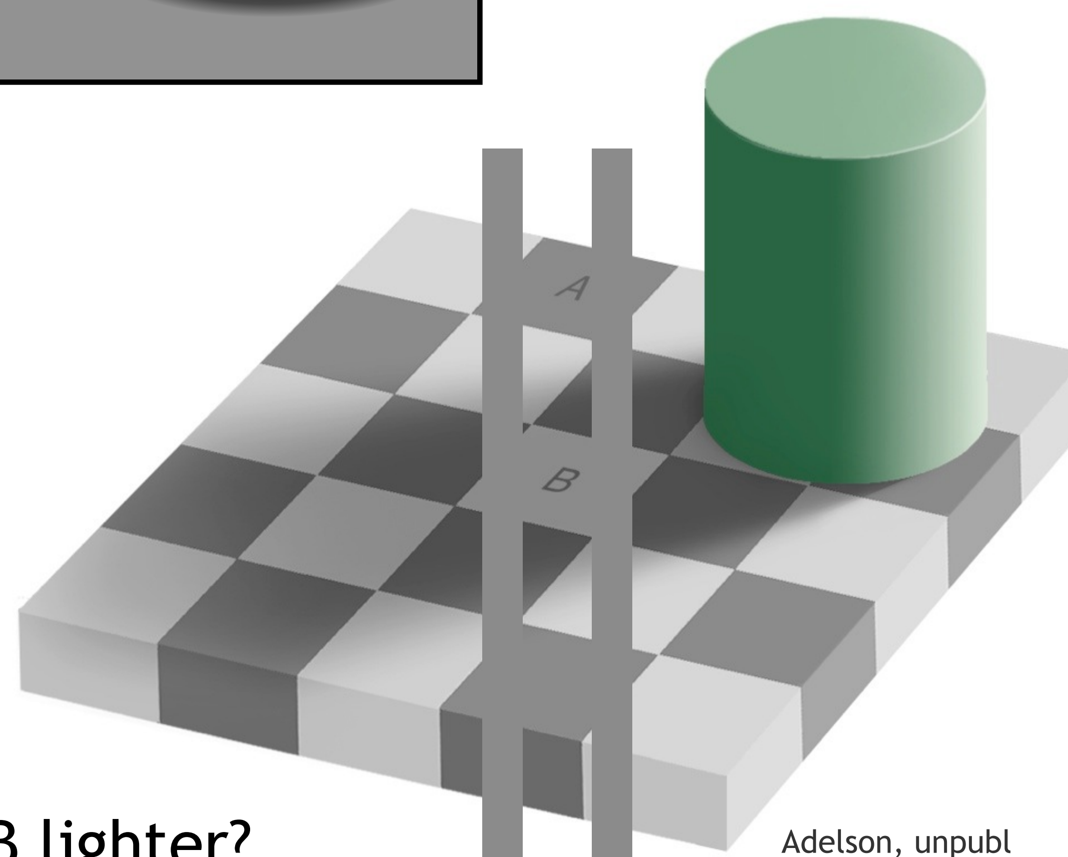
- ✓ seat
- ✗ radio
- ✓ table
- ✗ rocking

bubble or dimple?



how many triangles?





is A or B lighter?

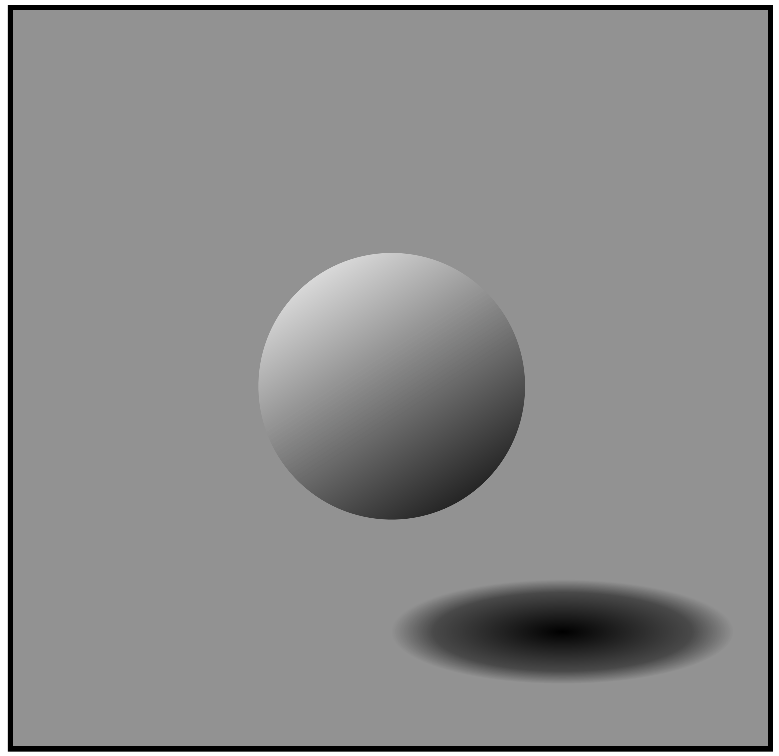
Adelson, unpubl

UNCONSCIOUS INFERENCES

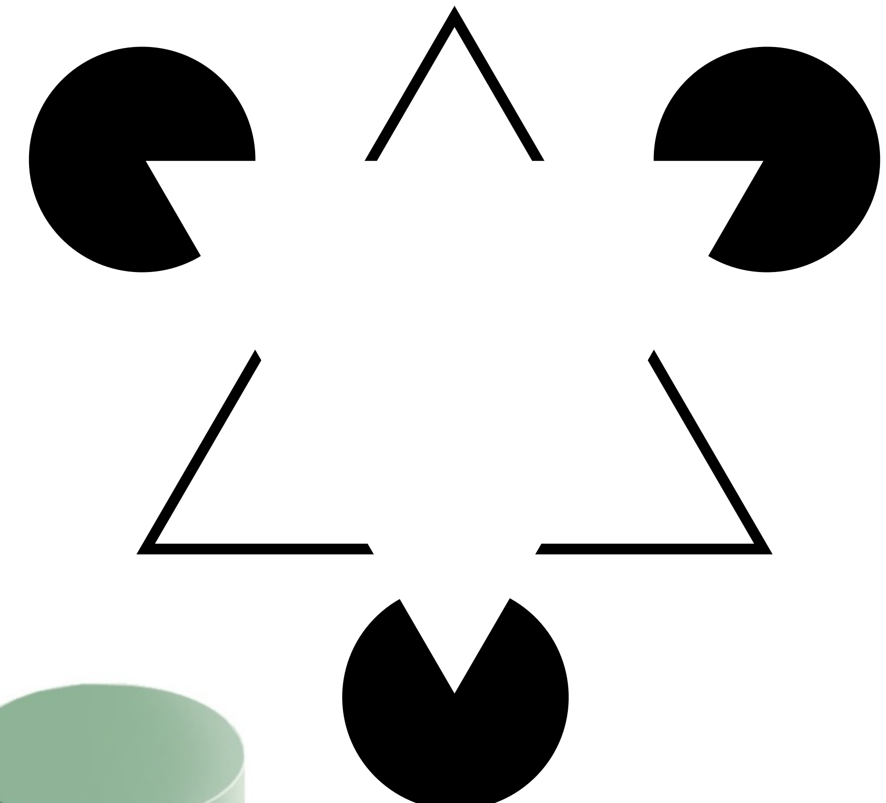
which words?

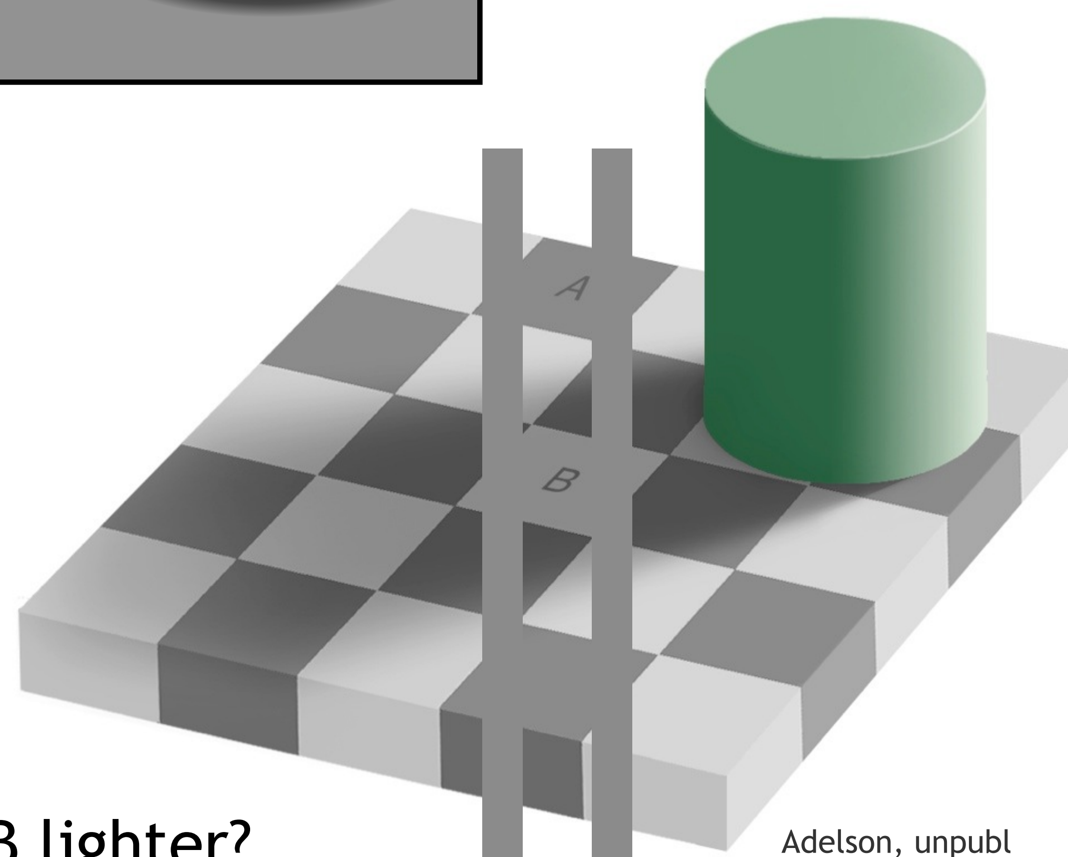
- ✓ seat
- ✗ radio
- ✓ table
- ✗ rocking
- bench

bubble or dimple?



how many triangles?





is A or B lighter?

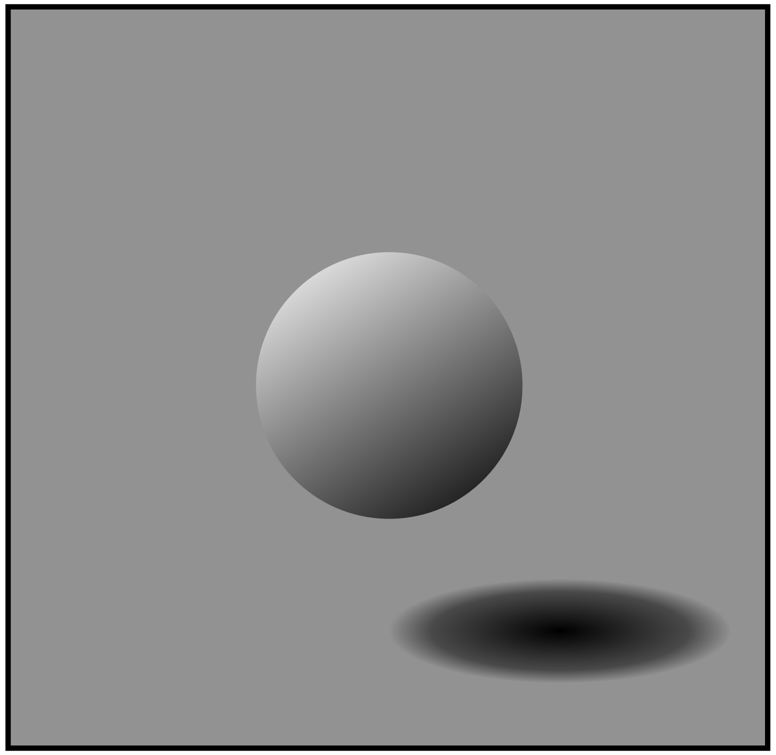
Adelson, unpubl

UNCONSCIOUS INFERENCES

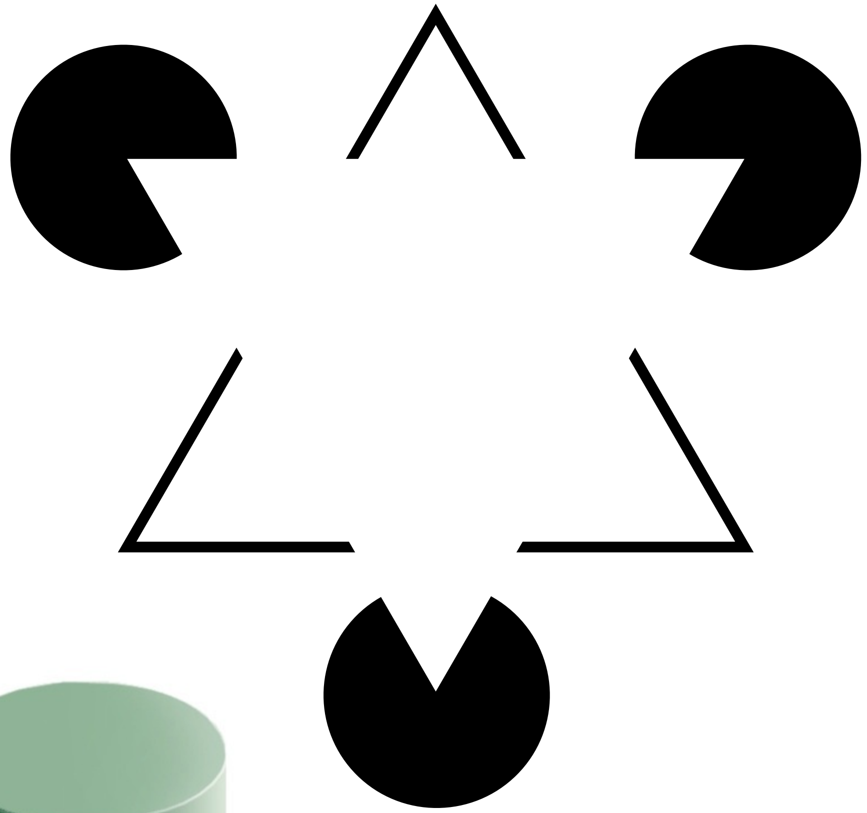
which words?

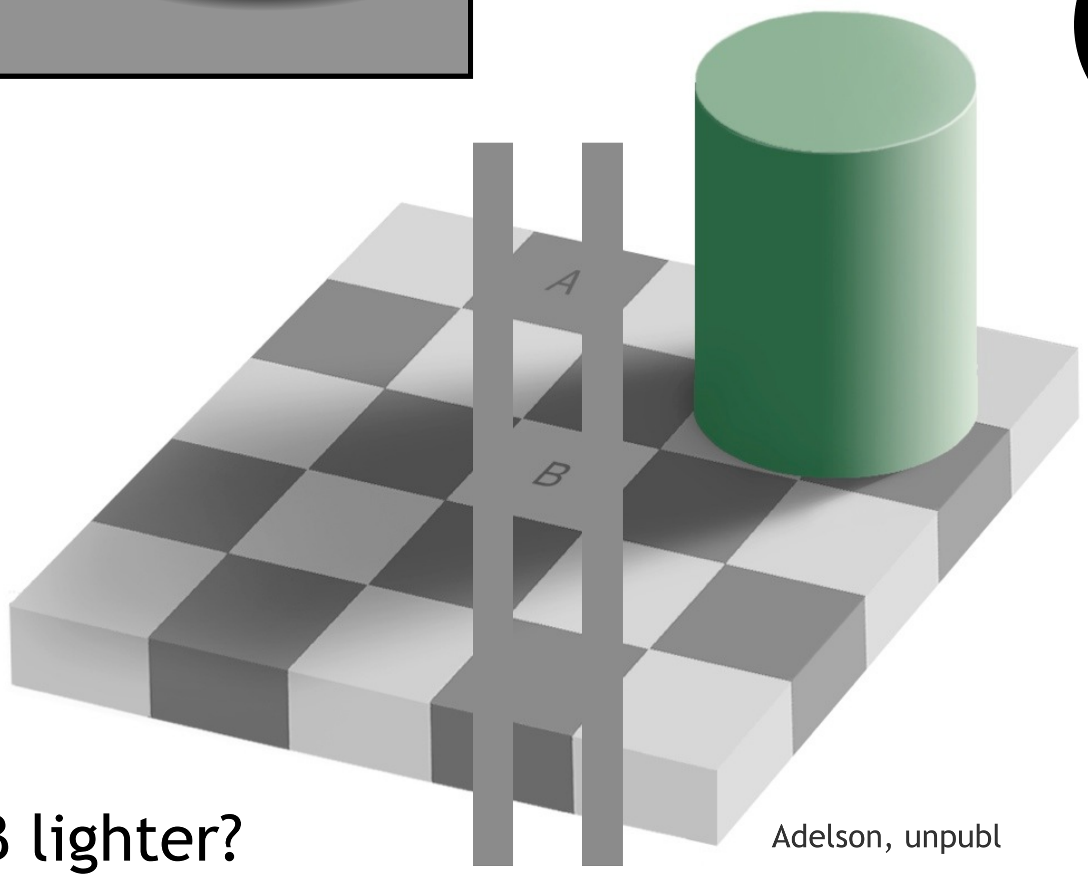
- ✓ seat
- ✗ radio
- ✓ table
- ✗ rocking
- ✗ bench

bubble or dimple?



how many triangles?





is A or B lighter?

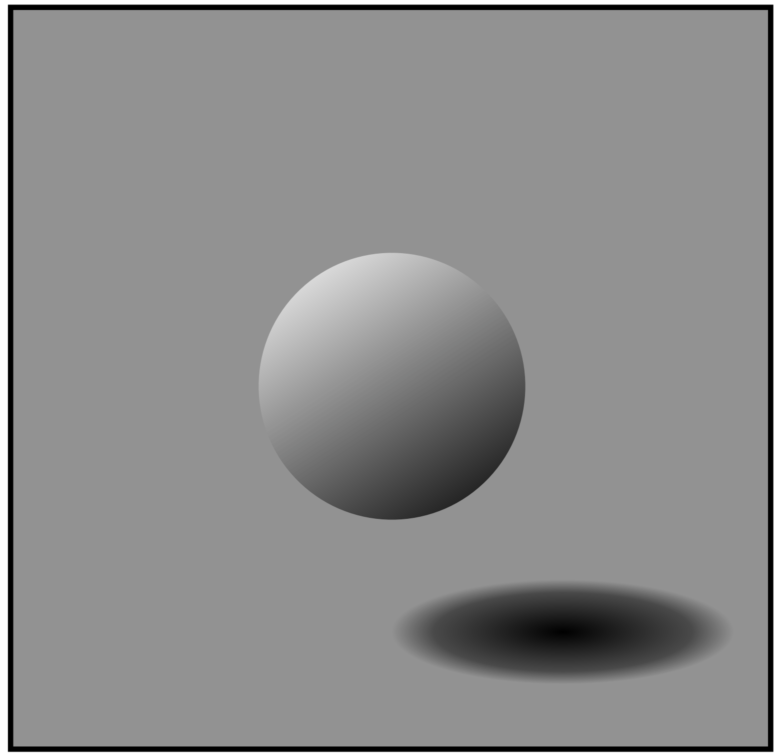
Adelson, unpubl

UNCONSCIOUS INFERENCES

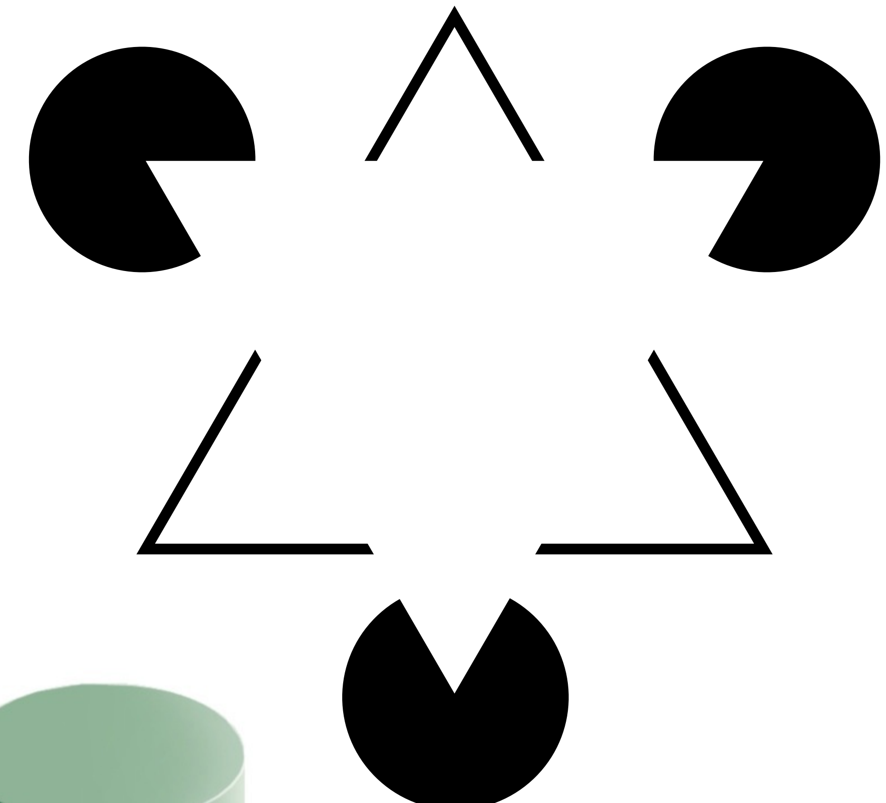
which words?

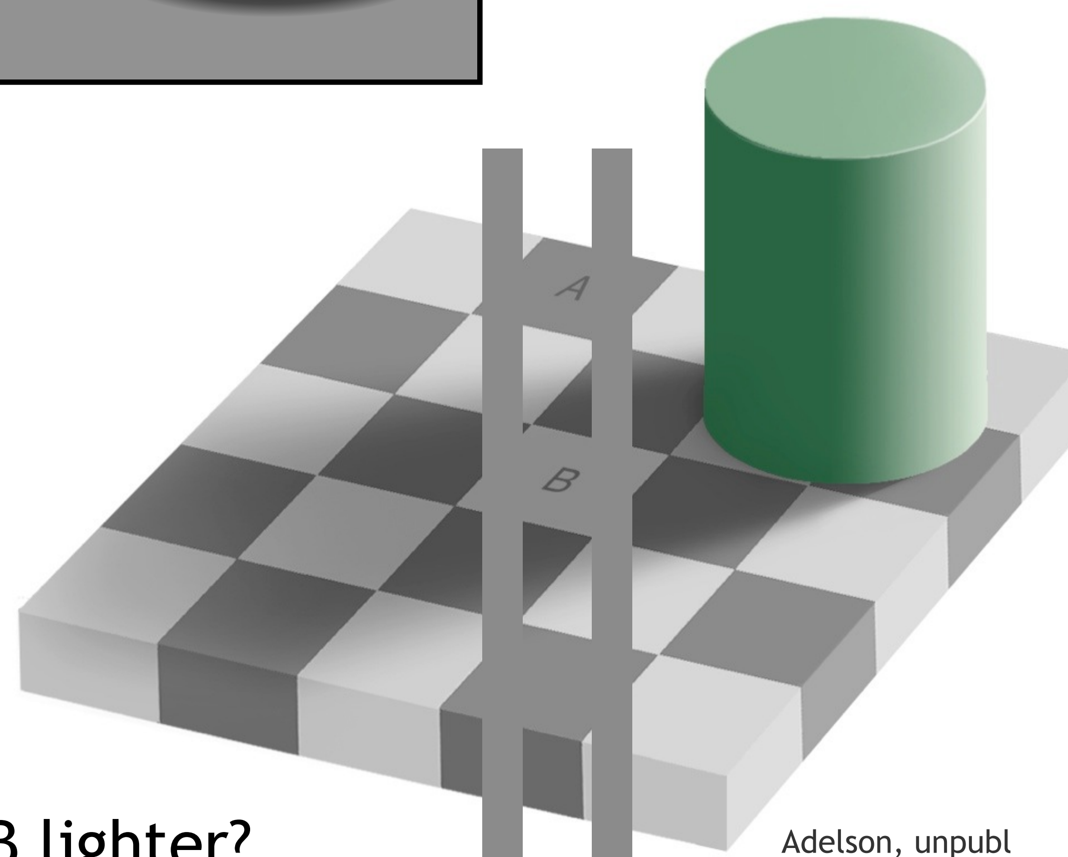
- ✓ seat
- ✗ radio
- ✓ table
- ✗ rocking
- ✗ bench
- boat

bubble or dimple?



how many triangles?





is A or B lighter?

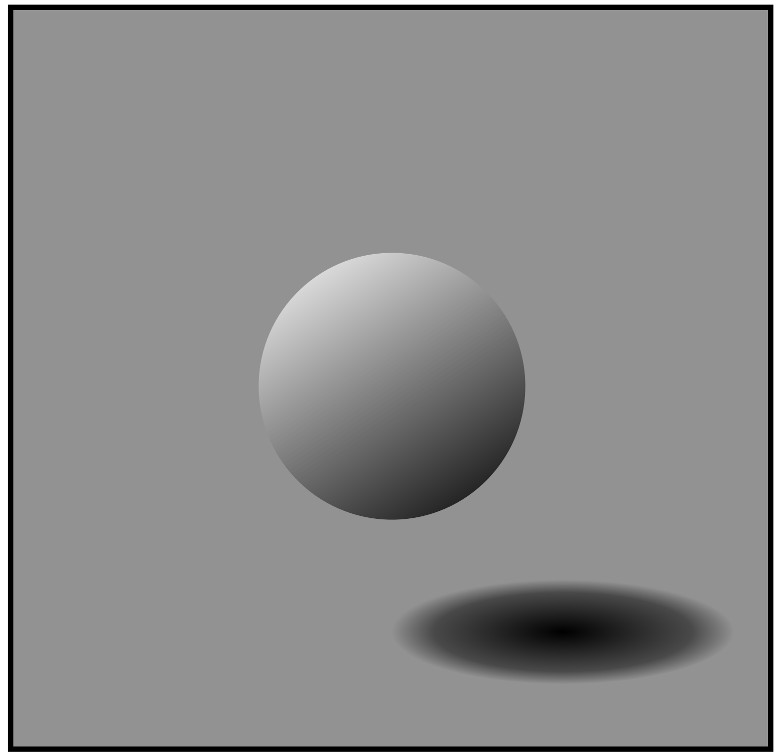
Adelson, unpubl

UNCONSCIOUS INFERENCES

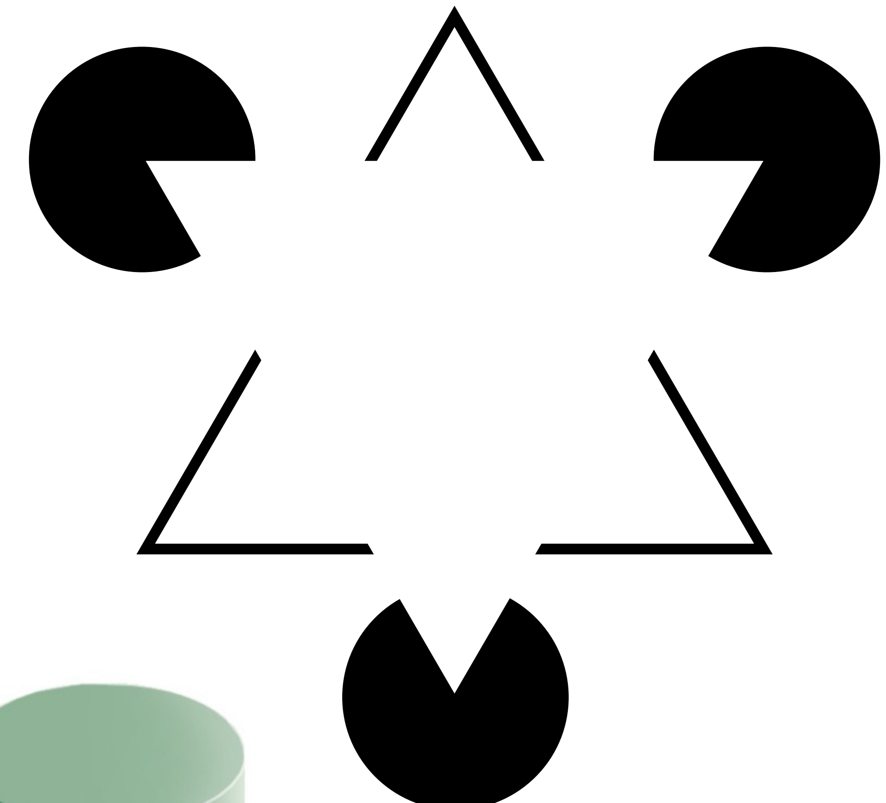
which words?

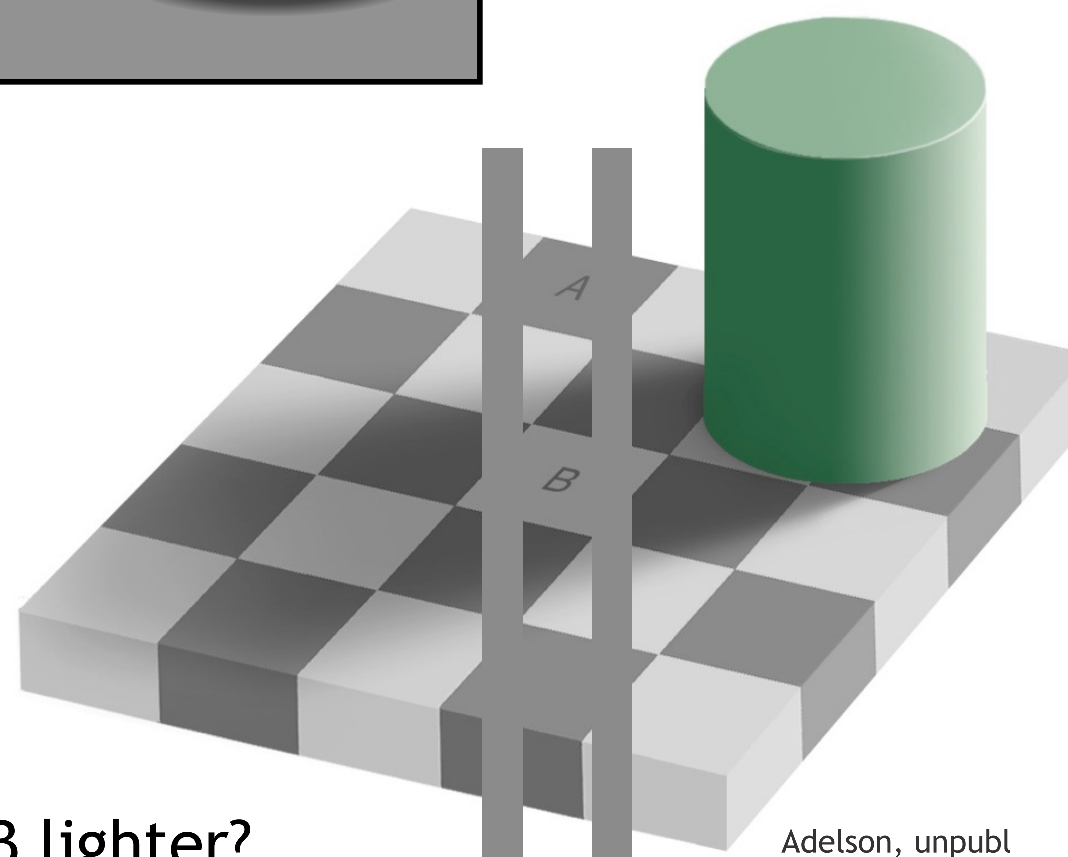
✓ seat
✗ radio
✓ table
✗ rocking
✗ bench
✗ boat
chair

bubble or dimple?



how many triangles?





is A or B lighter?

Adelson, unpubl

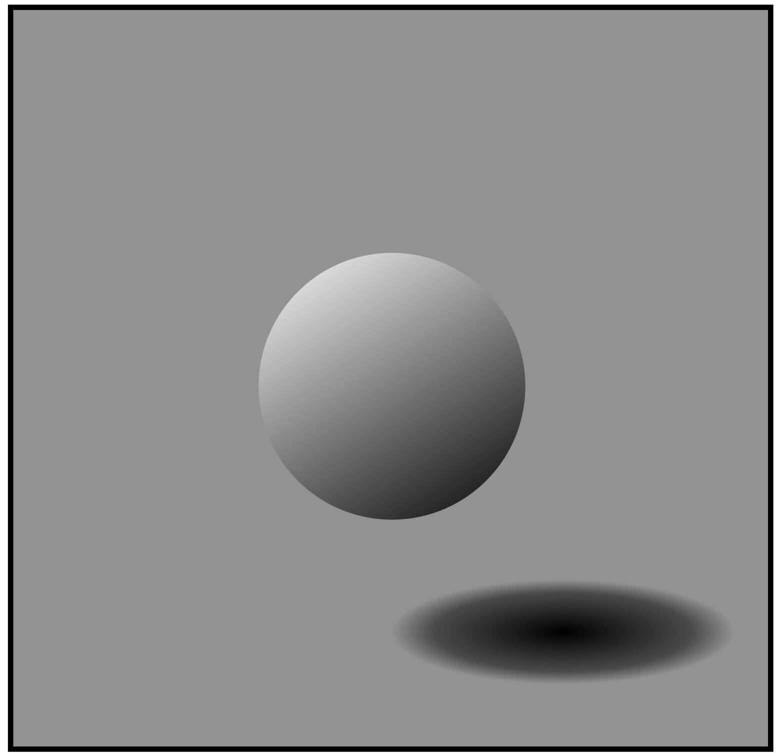
UNCONSCIOUS INFERENCES

which words?

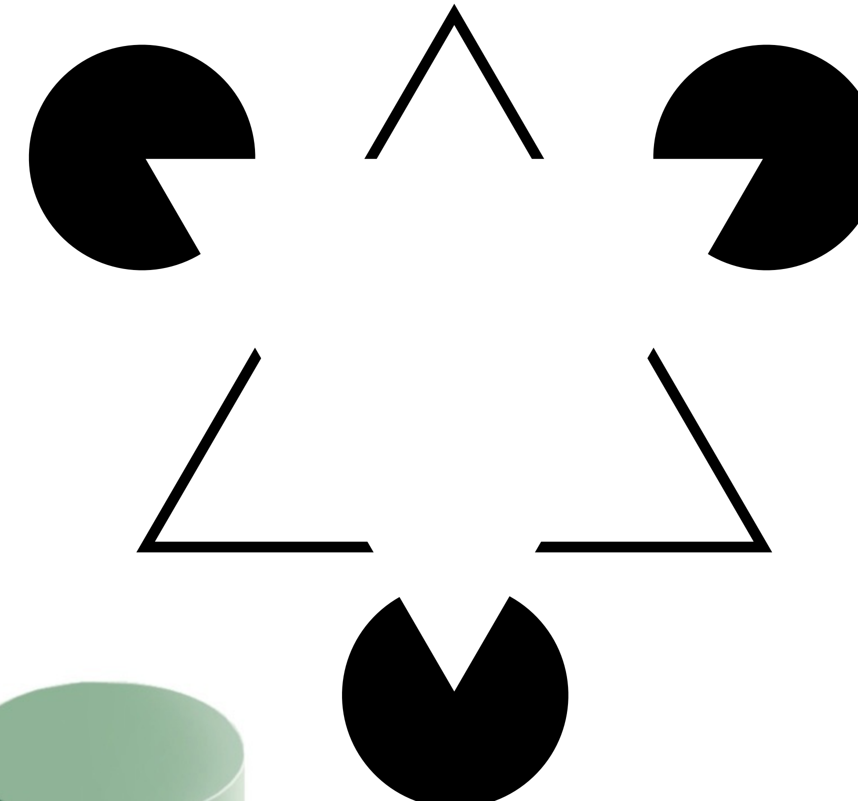
- ✓ seat
- ✗ radio
- ✓ table
- ✗ rocking
- ✗ bench
- ✗ boat
- ✗ chair

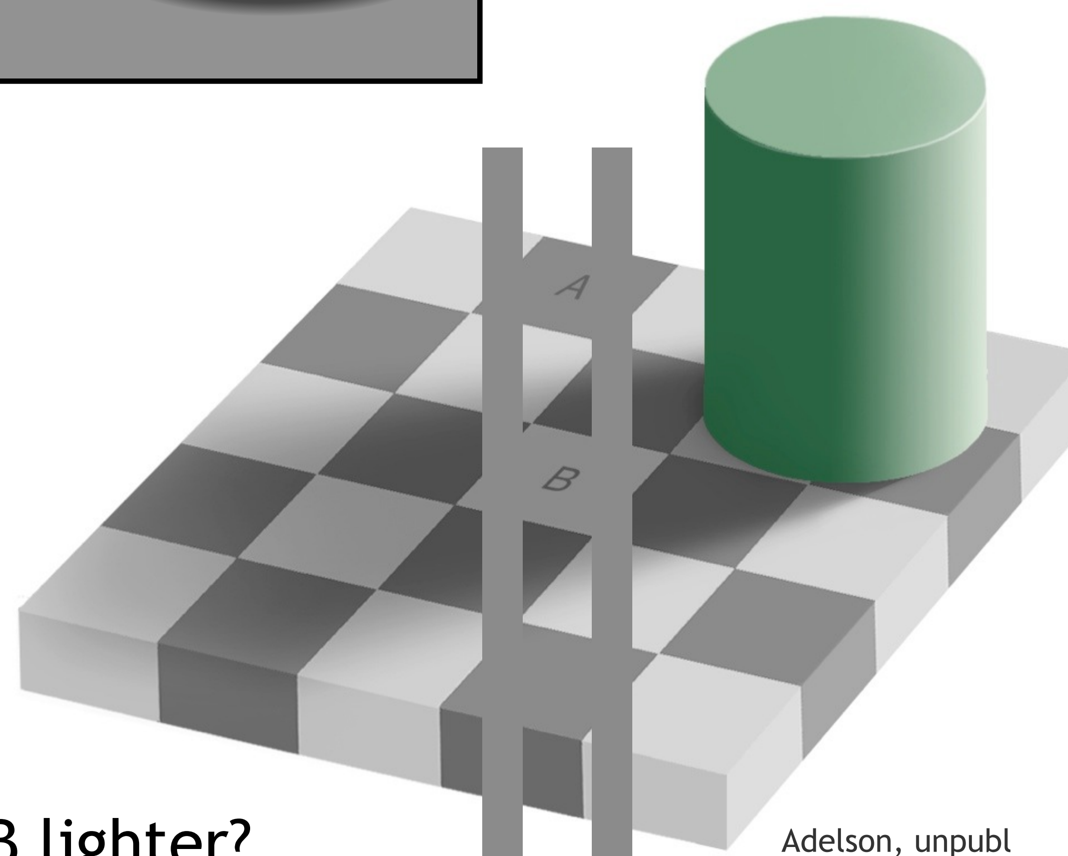
Roediger & McDermott, 1995

bubble or dimple?



how many triangles?

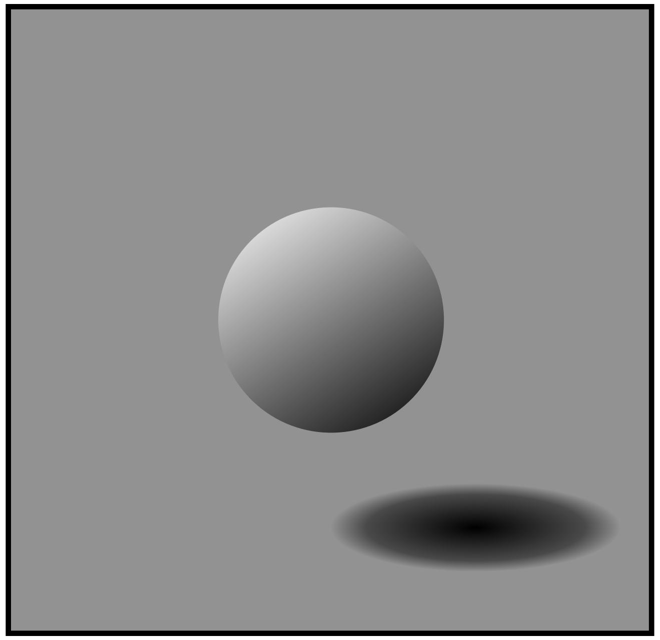




is A or B lighter?

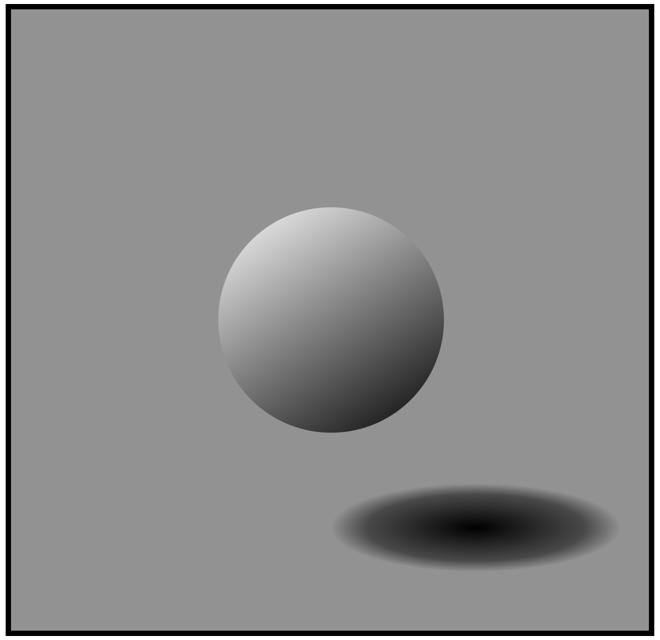
Adelson, unpubl

THE BAYESIAN BRAIN



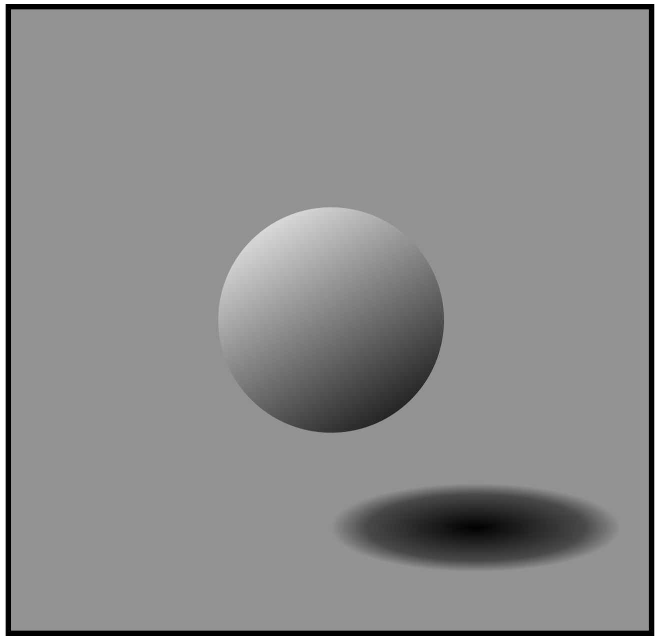
THE BAYESIAN BRAIN

"perception is unconscious inference"



Hermann von Helmholtz
1821-1894

THE BAYESIAN BRAIN

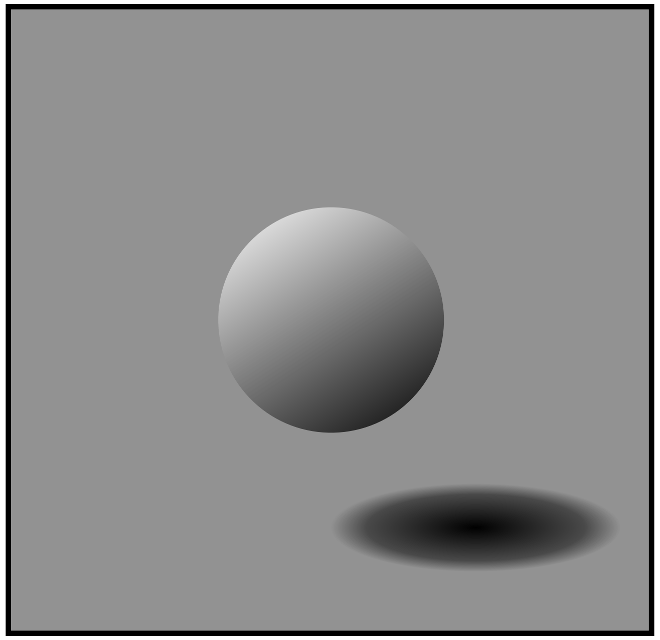


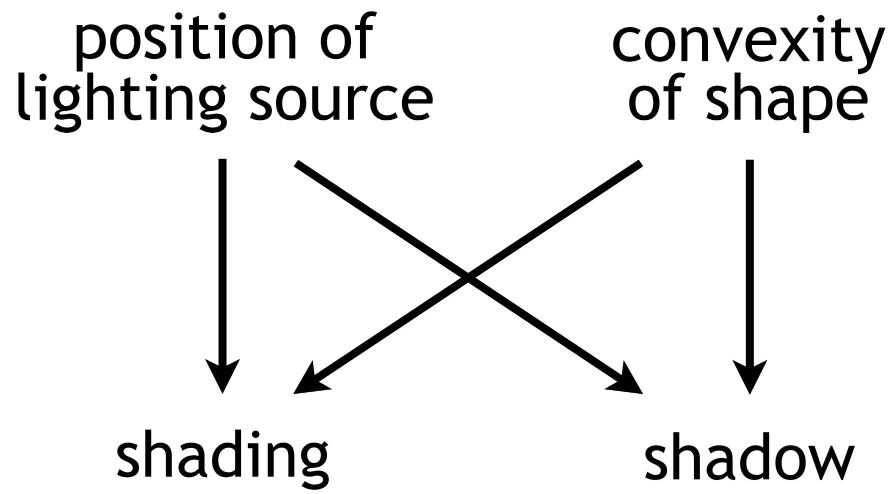
“perception is unconscious inference”
& memory & learning & …

Hermann von Helmholtz
1821-1894

THE BAYESIAN BRAIN

“perception is unconscious inference”
& memory & learning & …

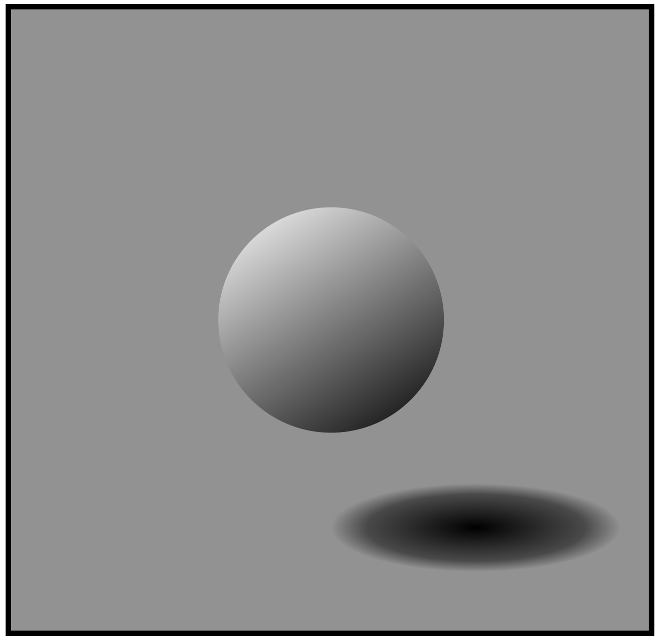


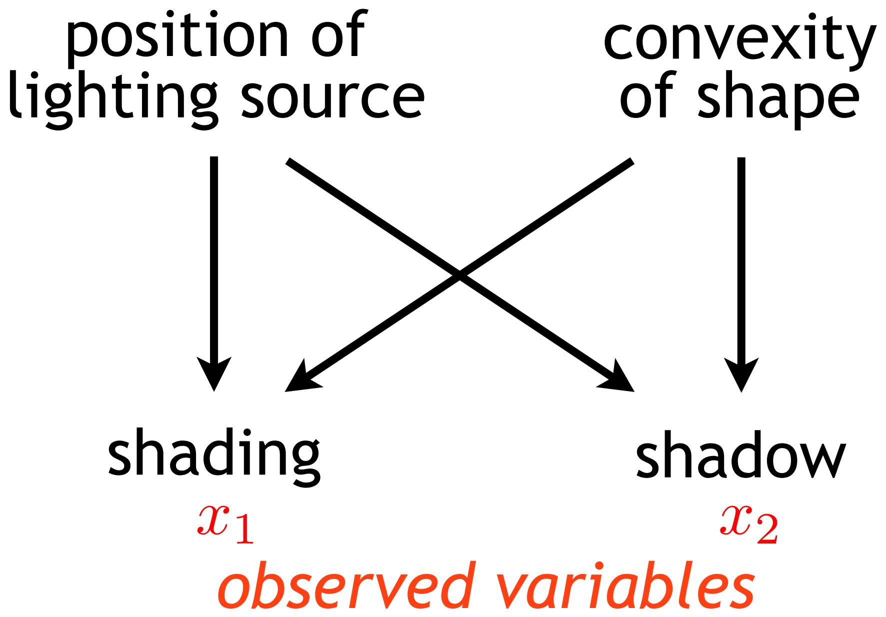


Hermann von Helmholtz
1821-1894

THE BAYESIAN BRAIN

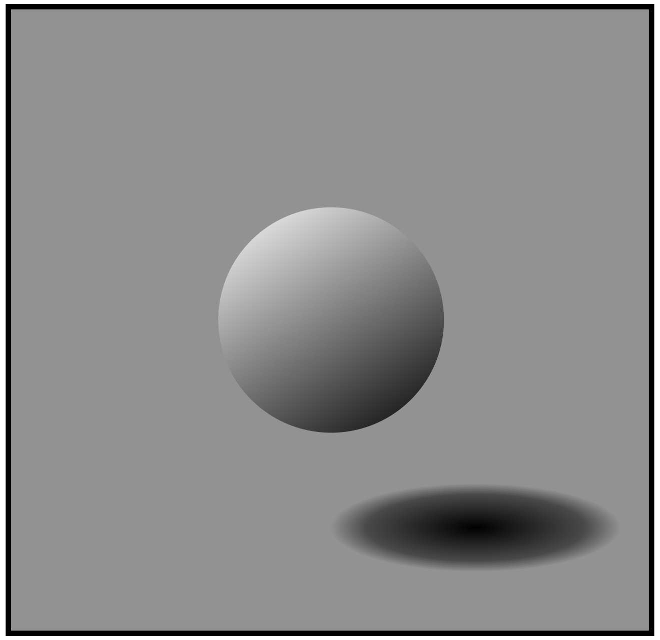
“perception is unconscious inference”
& memory & learning & …





Hermann von Helmholtz
1821-1894

THE BAYESIAN BRAIN



“perception is unconscious inference”
& memory & learning & …

hidden variables

y_1 position of lighting source

y_2 convexity of shape

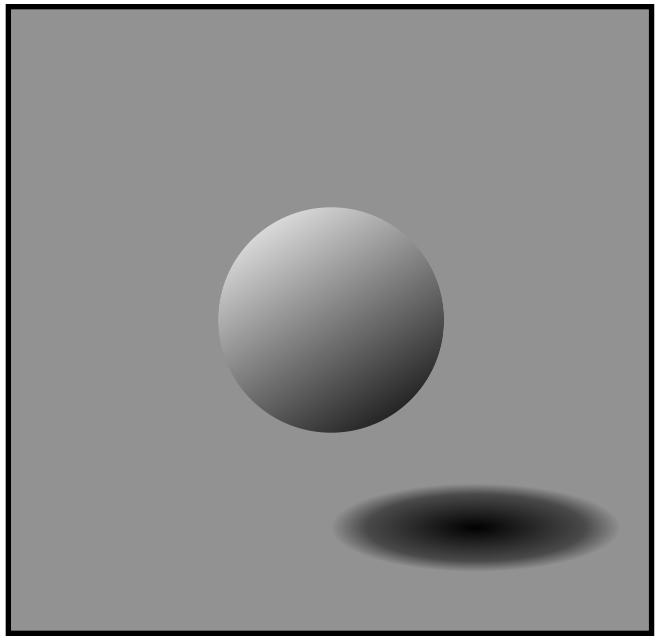
shading x_1

shadow x_2

observed variables

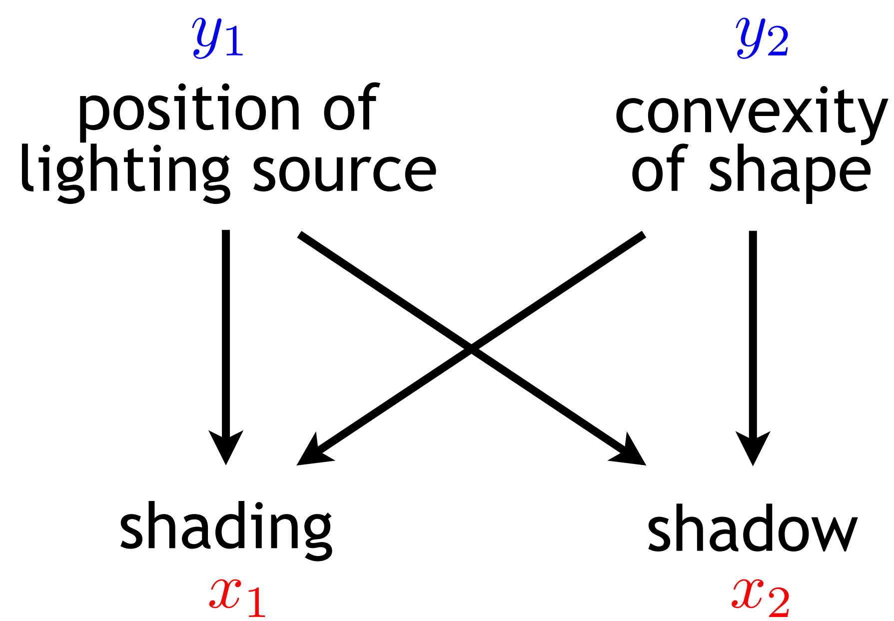
Hermann von Helmholtz
1821-1894

THE BAYESIAN BRAIN



“perception is unconscious inference”
& memory & learning & …

hidden variables

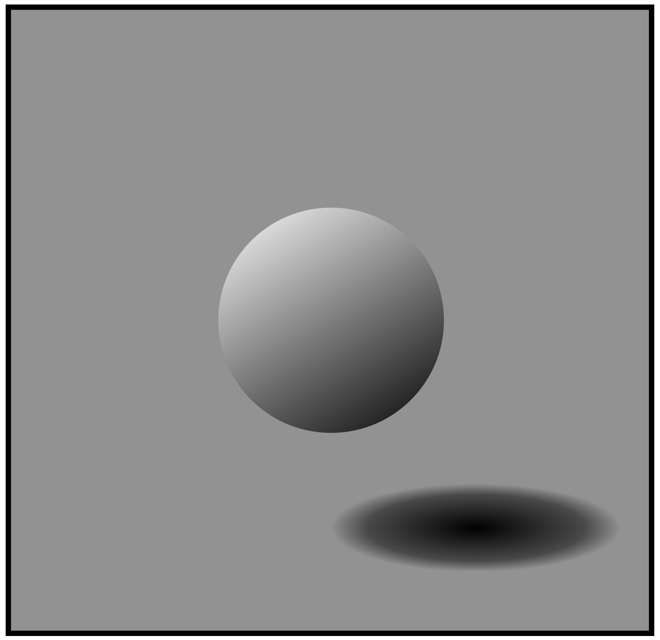


observed variables

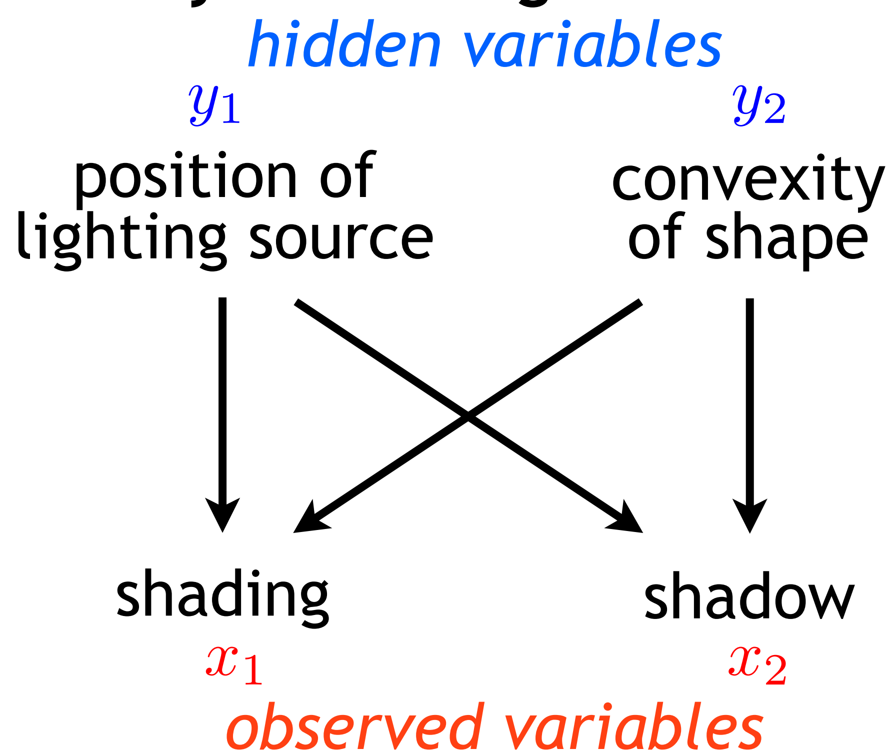
Hermann von Helmholtz
1821-1894

“There are things known and there are things unknown, and between are the doors of perception” *Jim Morrison*

THE BAYESIAN BRAIN



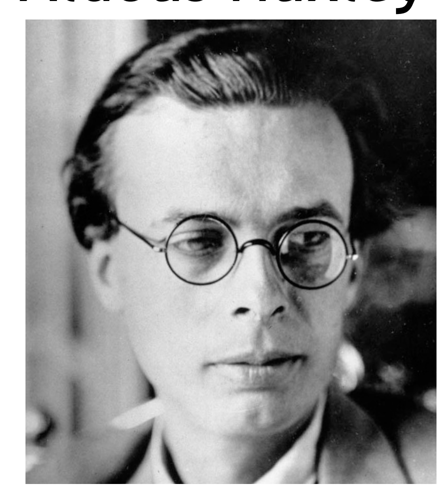
"perception is unconscious inference"
& memory & learning & …



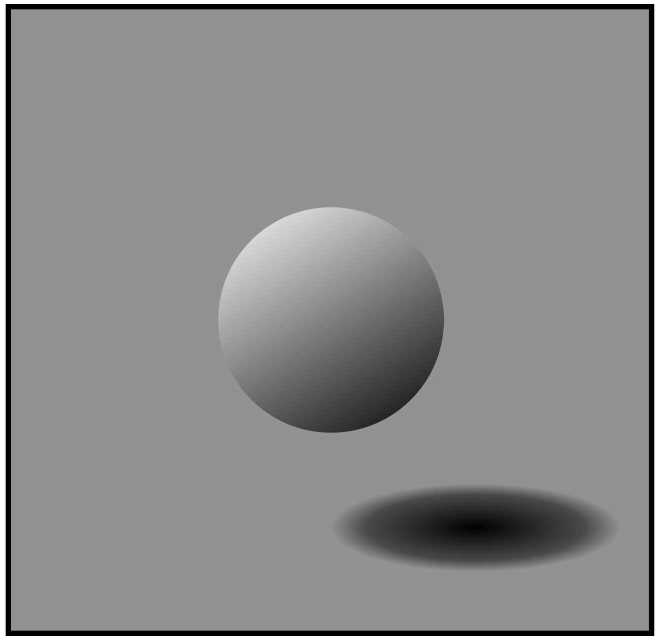
Hermann von Helmholtz
1821-1894

"There are things known and there are things unknown, and between are the doors of perception"

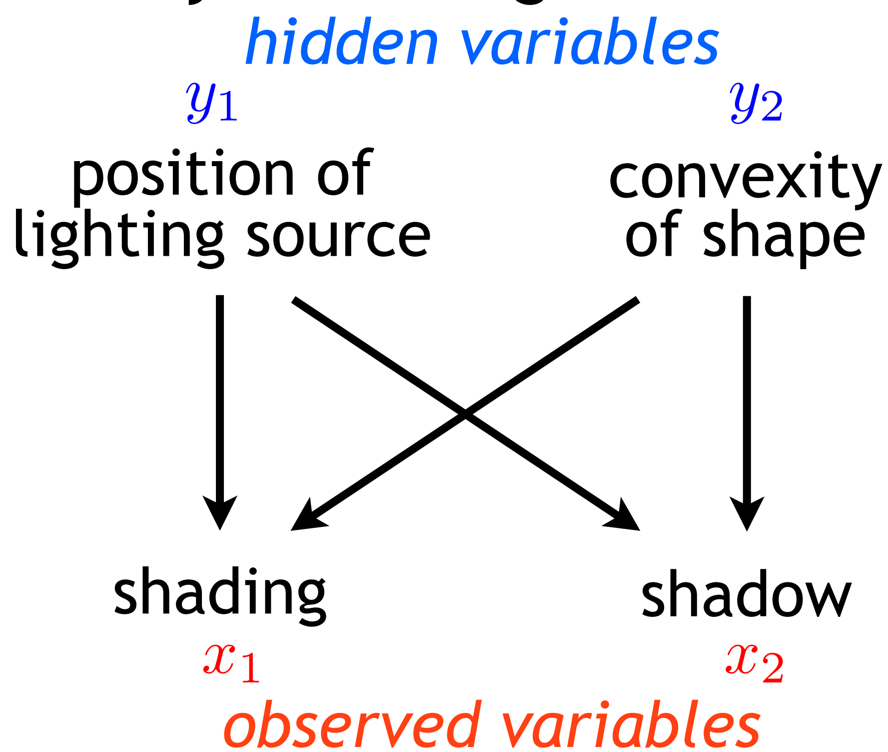
Aldous Huxley



THE BAYESIAN BRAIN



“perception is unconscious inference”
& memory & learning & …

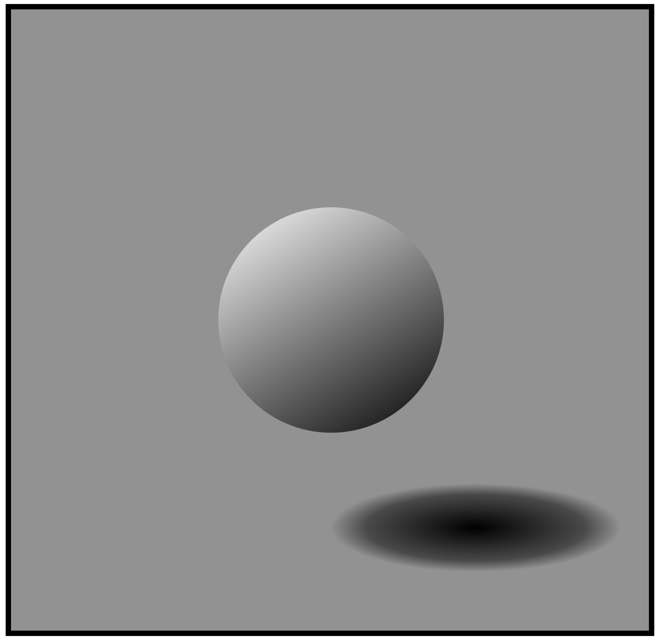


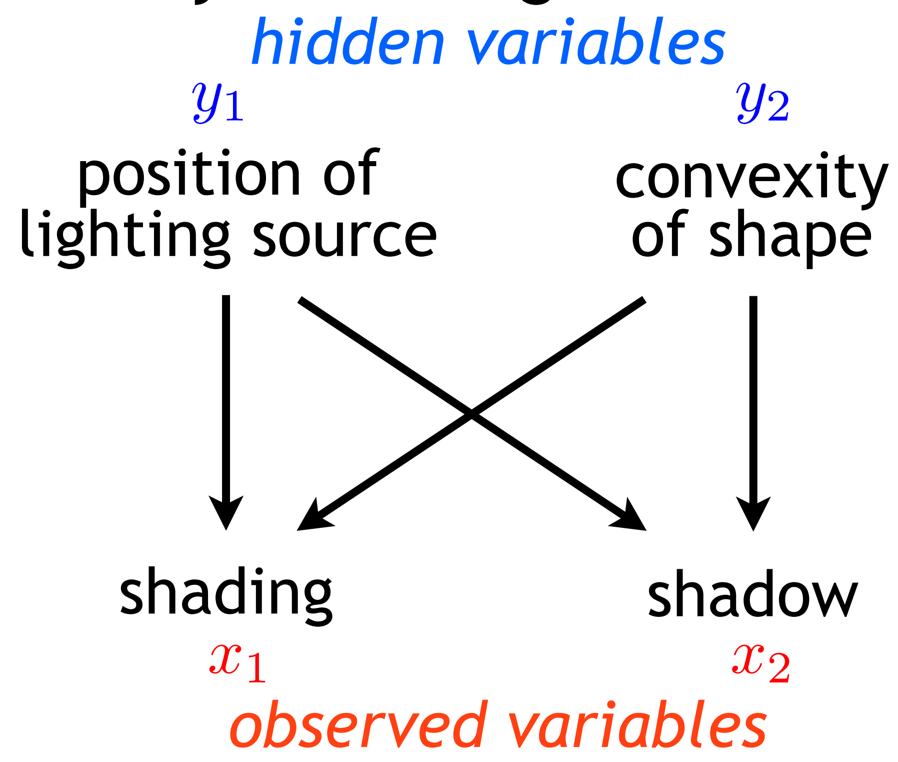
Hermann von Helmholtz
1821-1894

There are things known and there are things unknown, and between are
the rules of probability

THE BAYESIAN BRAIN

“perception is unconscious inference”
& memory & learning & …



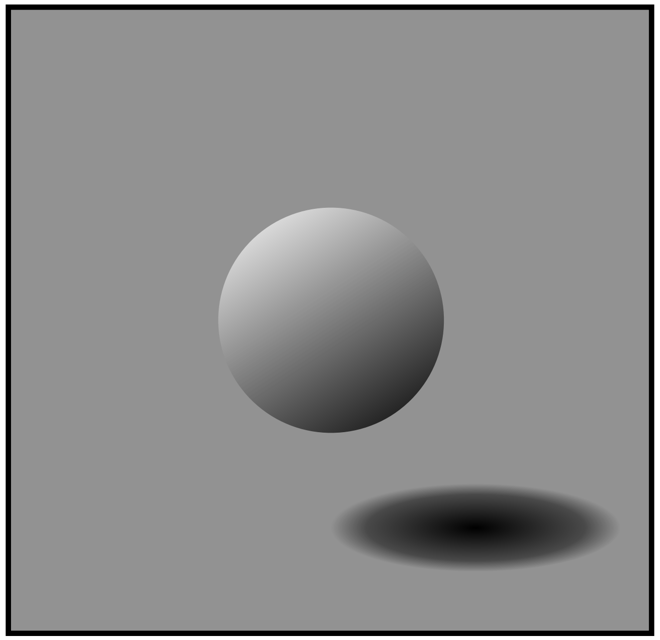


Hermann von Helmholtz
1821-1894

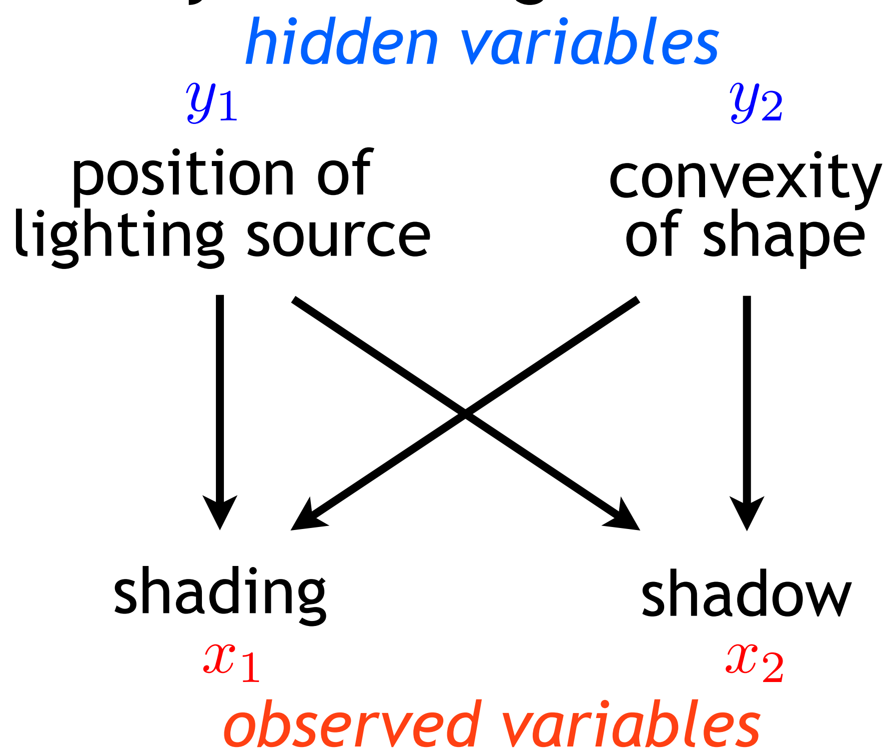
There are things known and there are things unknown, and between are **the rules of probability**

product: $\mathrm{P}(X,Y) = \mathrm{P}(Y,X) = \mathrm{P}(X|Y)\ \mathrm{P}(Y)$

THE BAYESIAN BRAIN



"perception is unconscious inference"
& memory & learning & ...



Hermann von Helmholtz
1821-1894

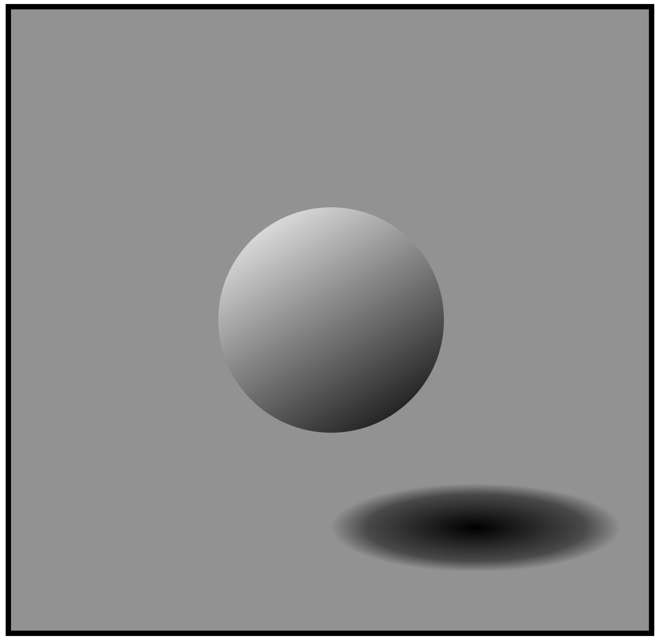
There are things known and there are things unknown, and between are
the rules of probability

product: $\mathrm{P}(X,Y) = \mathrm{P}(Y,X) = \mathrm{P}(X|Y)\ \mathrm{P}(Y)$

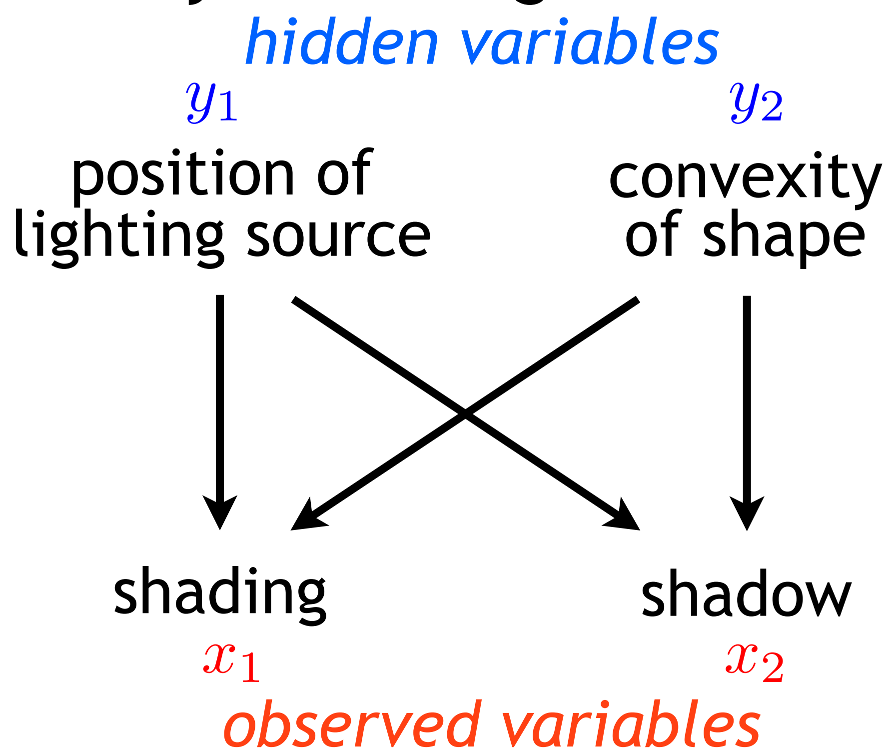
⇒ Bayes' rule: $\mathrm{P}(Y|X) = \dfrac{\mathrm{P}(X|Y)\ \mathrm{P}(Y)}{\mathrm{P}(X)}$

Rev. Thomas Bayes
1702-1761

THE BAYESIAN BRAIN



“perception is unconscious inference”
& memory & learning & …



Hermann von Helmholtz
1821-1894

There are things known and there are things unknown, and between are **the rules of probability**

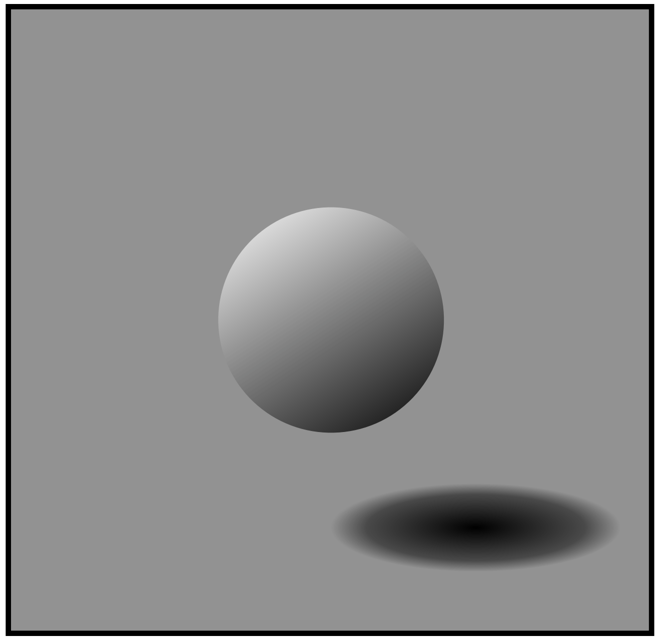
product: $\mathrm{P}(X,Y) = \mathrm{P}(Y,X) = \mathrm{P}(X|Y)\ \mathrm{P}(Y)$

Bayes’ rule: $\mathrm{P}(Y|X) = \dfrac{\mathrm{P}(X|Y)\ \mathrm{P}(Y)}{\mathrm{P}(X)}$

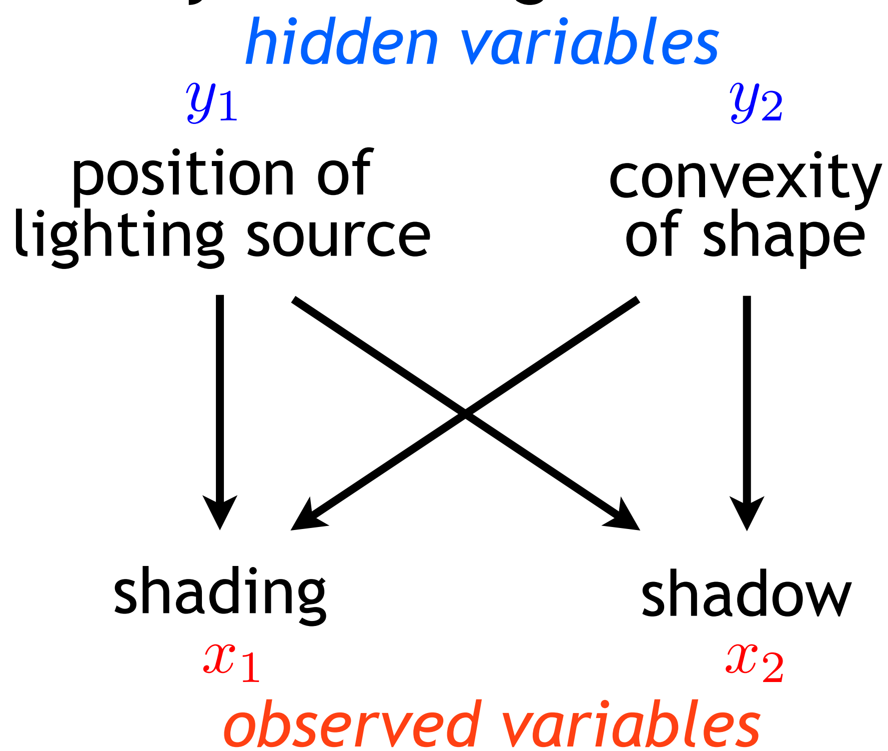
$\mathrm{P}(X,Y) = \mathrm{P}(X)\ \mathrm{P}(Y)$ iff X and Y are independent!

Rev. Thomas Bayes
1702-1761

THE BAYESIAN BRAIN



"perception is unconscious inference"
& memory & learning & ...



Hermann von Helmholtz
1821-1894

There are things known and there are things unknown, and between are **the rules of probability**

product: $\mathrm{P}(X, Y) = \mathrm{P}(Y, X) = \mathrm{P}(X|Y)\ \mathrm{P}(Y)$

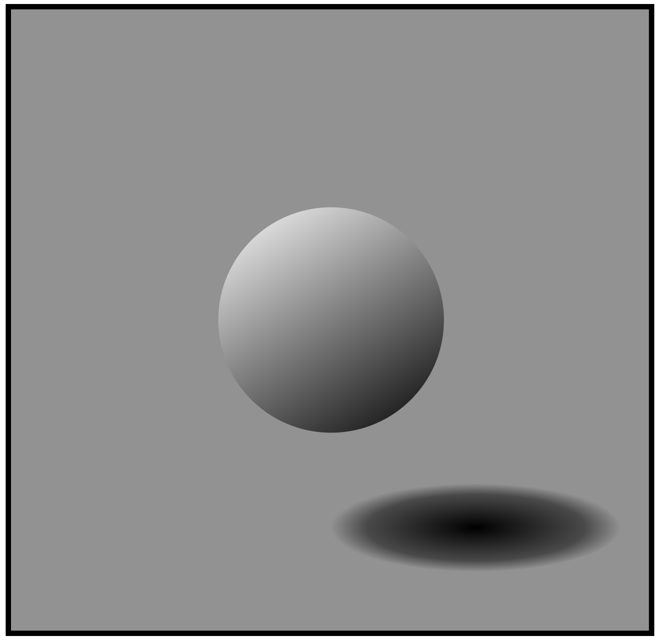
Bayes' rule: $\mathrm{P}(Y|X) = \dfrac{\mathrm{P}(X|Y)\ \mathrm{P}(Y)}{\mathrm{P}(X)}$

$\mathrm{P}(X, Y) = \mathrm{P}(X)\ \mathrm{P}(Y)$ iff X and Y are independent!

sum:
(marginalisation)
$$\mathrm{P}(X) = \sum_Y \mathrm{P}(X, Y)$$

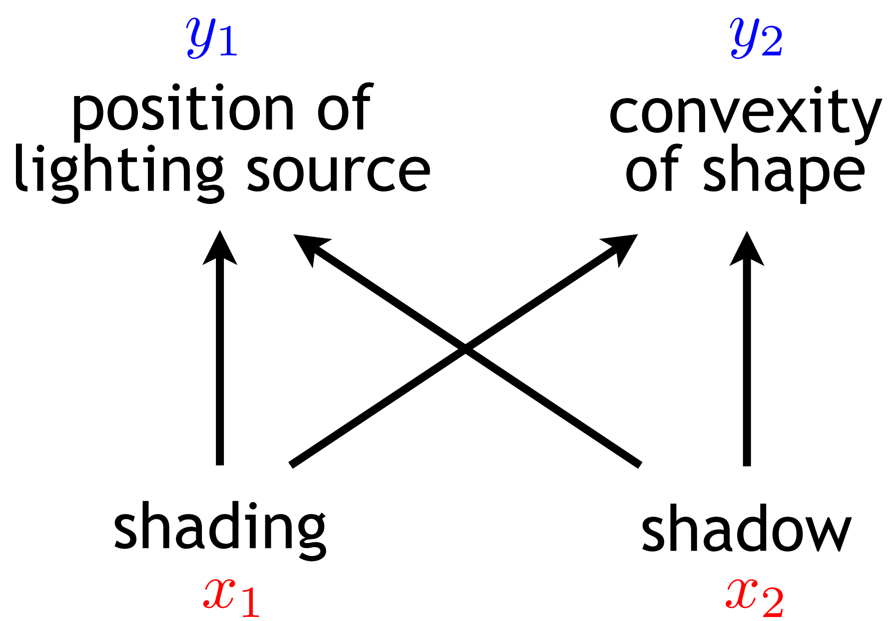
Rev. Thomas Bayes
1702-1761

THE BAYESIAN BRAIN



“perception is unconscious inference”
& memory & learning & …

hidden variables



observed variables

Hermann von Helmholtz
1821-1894

There are things known and there are things unknown, and between are **the rules of probability**

product: $\mathrm{P}(X,Y) = \mathrm{P}(Y,X) = \mathrm{P}(X|Y)\ \mathrm{P}(Y)$

⇨ Bayes’ rule: $\mathrm{P}(Y|X) = \dfrac{\mathrm{P}(X|Y)\ \mathrm{P}(Y)}{\mathrm{P}(X)}$

$\mathrm{P}(X,Y) = \mathrm{P}(X)\ \mathrm{P}(Y)$ iff X and Y are independent!

sum: (marginalisation) $\mathrm{P}(X) = \sum_Y \mathrm{P}(X,Y)$

Rev. Thomas Bayes
1702-1761

BAYESIAN DECISION THEORY
(and how to make point estimates)

BAYESIAN DECISION THEORY
(and how to make point estimates)

loss function

state of the world

action	y_1	y_2	y_3	…
a_1	$L(a_1,y_1)$	$L(a_1,y_2)$	$L(a_1,y_3)$	
a_2	$L(a_2,y_1)$	$L(a_2,y_2)$	$L(a_2,y_3)$	
a_3	$L(a_3,y_1)$	$L(a_3,y_2)$	$L(a_3,y_3)$	
⋮				

BAYESIAN DECISION THEORY
(and how to make point estimates)

loss function

state of the world

action		y_1	y_2	y_3	…
	a_1	$L(a_1,y_1)$	$L(a_1,y_2)$	$L(a_1,y_3)$	
	a_2	$L(a_2,y_1)$	$L(a_2,y_2)$	$L(a_2,y_3)$	
	a_3	$L(a_3,y_1)$	$L(a_3,y_2)$	$L(a_3,y_3)$	
	⋮				

action to choose:

$$a^*(x) = \operatorname*{argmin}_a \sum_y L(a,y)\,\mathrm{P}(y \mid x)$$

note: a and y need not live in the same space

BAYESIAN DECISION THEORY
(and how to make point estimates)

state of the world

loss function

action		y_1	y_2	y_3	…
	a_1	$L(a_1,y_1)$	$L(a_1,y_2)$	$L(a_1,y_3)$	
	a_2	$L(a_2,y_1)$	$L(a_2,y_2)$	$L(a_2,y_3)$	
	a_3	$L(a_3,y_1)$	$L(a_3,y_2)$	$L(a_3,y_3)$	
	⋮				

action to choose:

$$a^*(x) = \underset{a}{\operatorname{argmin}} \sum_y L(a,y)\, \mathrm{P}(y \mid x)$$

note: a and y need not live in the same space

special cases when $a = \hat{y}$ and y do live in the same space

BAYESIAN DECISION THEORY
(and how to make point estimates)

loss function

state of the world

action		y_1	y_2	y_3	…
	a_1	$L(a_1,y_1)$	$L(a_1,y_2)$	$L(a_1,y_3)$	
	a_2	$L(a_2,y_1)$	$L(a_2,y_2)$	$L(a_2,y_3)$	
	a_3	$L(a_3,y_1)$	$L(a_3,y_2)$	$L(a_3,y_3)$	
	⋮				

action to choose:

$$a^*(x) = \underset{a}{\operatorname{argmin}} \sum_y L(a,y)\,\mathrm{P}(y \mid x)$$

note: a and y need not live in the same space

special cases when $a = \hat{y}$ and y do live in the same space

$$L(\hat{y}, y) = (\hat{y} - y)^2$$

BAYESIAN DECISION THEORY
(and how to make point estimates)

loss function

state of the world

action		y_1	y_2	y_3	…
	a_1	$L(a_1,y_1)$	$L(a_1,y_2)$	$L(a_1,y_3)$	
	a_2	$L(a_2,y_1)$	$L(a_2,y_2)$	$L(a_2,y_3)$	
	a_3	$L(a_3,y_1)$	$L(a_3,y_2)$	$L(a_3,y_3)$	
	⋮				

action to choose:

$$a^*(x) = \operatorname*{argmin}_a \sum_y L(a,y)\,\mathrm{P}(y \mid x)$$

note: a and y need not live in the same space

special cases when $a = \hat{y}$ and y do live in the same space

$$L(\hat{y},y) = (\hat{y}-y)^2 \longrightarrow \hat{y}(x) = \sum_y y\,\mathrm{P}(y \mid x)$$ posterior mean

BAYESIAN DECISION THEORY
(and how to make point estimates)

loss function

state of the world

action	y_1	y_2	y_3	…
a_1	$L(a_1,y_1)$	$L(a_1,y_2)$	$L(a_1,y_3)$	
a_2	$L(a_2,y_1)$	$L(a_2,y_2)$	$L(a_2,y_3)$	
a_3	$L(a_3,y_1)$	$L(a_3,y_2)$	$L(a_3,y_3)$	
⋮				

action to choose:

$$a^*(x) = \underset{a}{\operatorname{argmin}} \sum_y L(a,y)\,\mathrm{P}(y \mid x)$$

note: a and y need not live in the same space

special cases when $a = \hat{y}$ and y do live in the same space

$L(\hat{y},y) = (\hat{y}-y)^2$ ⟶ $\hat{y}(x) = \sum_y y\,\mathrm{P}(y \mid x)$ posterior mean

$L(\hat{y},y) = |\hat{y}-y|$

BAYESIAN DECISION THEORY
(and how to make point estimates)

loss function

state of the world

action		y_1	y_2	y_3	...
	a_1	$L(a_1,y_1)$	$L(a_1,y_2)$	$L(a_1,y_3)$	
	a_2	$L(a_2,y_1)$	$L(a_2,y_2)$	$L(a_2,y_3)$	
	a_3	$L(a_3,y_1)$	$L(a_3,y_2)$	$L(a_3,y_3)$	
	⋮				

action to choose: $$a^*(x) = \underset{a}{\operatorname{argmin}} \sum_y L(a,y)\,\mathrm{P}(y \mid x)$$

note: a and y need not live in the same space

special cases when $a = \hat{y}$ and y do live in the same space

$L(\hat{y},y) = (\hat{y}-y)^2$ ⟶ $\hat{y}(x) = \sum_y y\,\mathrm{P}(y \mid x)$ posterior mean

$L(\hat{y},y) = |\hat{y}-y|$ ⟶ $\sum_{y=-\infty}^{\hat{y}(x)} \mathrm{P}(y \mid x) = \frac{1}{2}$ posterior median

BAYESIAN DECISION THEORY
(and how to make point estimates)

loss function

state of the world

action

	y_1	y_2	y_3	…
a_1	$L(a_1,y_1)$	$L(a_1,y_2)$	$L(a_1,y_3)$	
a_2	$L(a_2,y_1)$	$L(a_2,y_2)$	$L(a_2,y_3)$	
a_3	$L(a_3,y_1)$	$L(a_3,y_2)$	$L(a_3,y_3)$	
⋮				

action to choose:

$$a^*(x) = \operatorname*{argmin}_a \sum_y L(a,y)\,\mathrm{P}(y \mid x)$$

note: a and y need not live in the same space

special cases when $a = \hat{y}$ and y do live in the same space

$$L(\hat{y},y) = (\hat{y}-y)^2 \longrightarrow \hat{y}(x) = \sum_y y\,\mathrm{P}(y \mid x)$$ posterior mean

$$L(\hat{y},y) = |\hat{y}-y| \longrightarrow \sum_{y=-\infty}^{\hat{y}(x)} \mathrm{P}(y \mid x) = \frac{1}{2}$$ posterior median

$$L(\hat{y},y) = \begin{cases} 0 & \text{if } y{=}\hat{y} \\ 1 & \text{otherwise} \end{cases}$$

BAYESIAN DECISION THEORY
(and how to make point estimates)

loss function

state of the world

action		y_1	y_2	y_3	…
	a_1	$L(a_1,y_1)$	$L(a_1,y_2)$	$L(a_1,y_3)$	
	a_2	$L(a_2,y_1)$	$L(a_2,y_2)$	$L(a_2,y_3)$	
	a_3	$L(a_3,y_1)$	$L(a_3,y_2)$	$L(a_3,y_3)$	
	⋮				

action to choose:

$$a^*(x) = \operatorname*{argmin}_a \sum_y L(a,y)\,\mathrm{P}(y \mid x)$$

note: a and y need not live in the same space

special cases when $a = \hat{y}$ and y do live in the same space

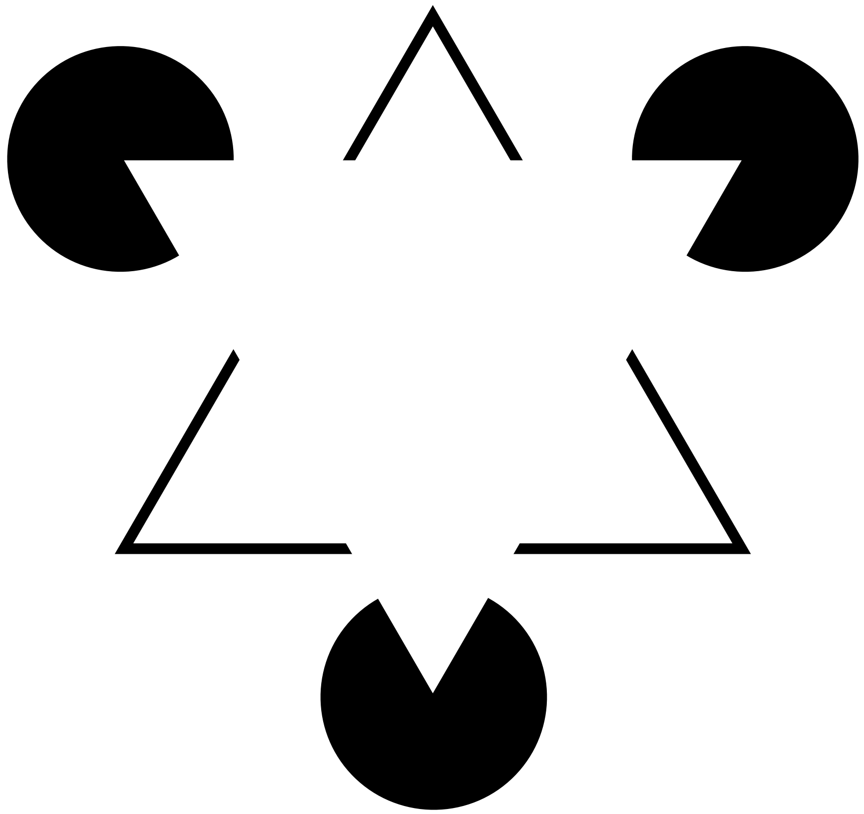
$$L(\hat{y},y) = (\hat{y}-y)^2 \longrightarrow \hat{y}(x) = \sum_y y\,\mathrm{P}(y \mid x)$$ posterior mean

$$L(\hat{y},y) = |\hat{y}-y| \longrightarrow \sum_{y=-\infty}^{\hat{y}(x)} \mathrm{P}(y \mid x) = \frac{1}{2}$$ posterior median

$$L(\hat{y},y) = \begin{cases} 0 & \text{if } y=\hat{y} \\ 1 & \text{otherwise} \end{cases} \longrightarrow \hat{y}(x) = \operatorname*{argmax}_y \mathrm{P}(y \mid x)$$ maximum a posteriori (MAP)

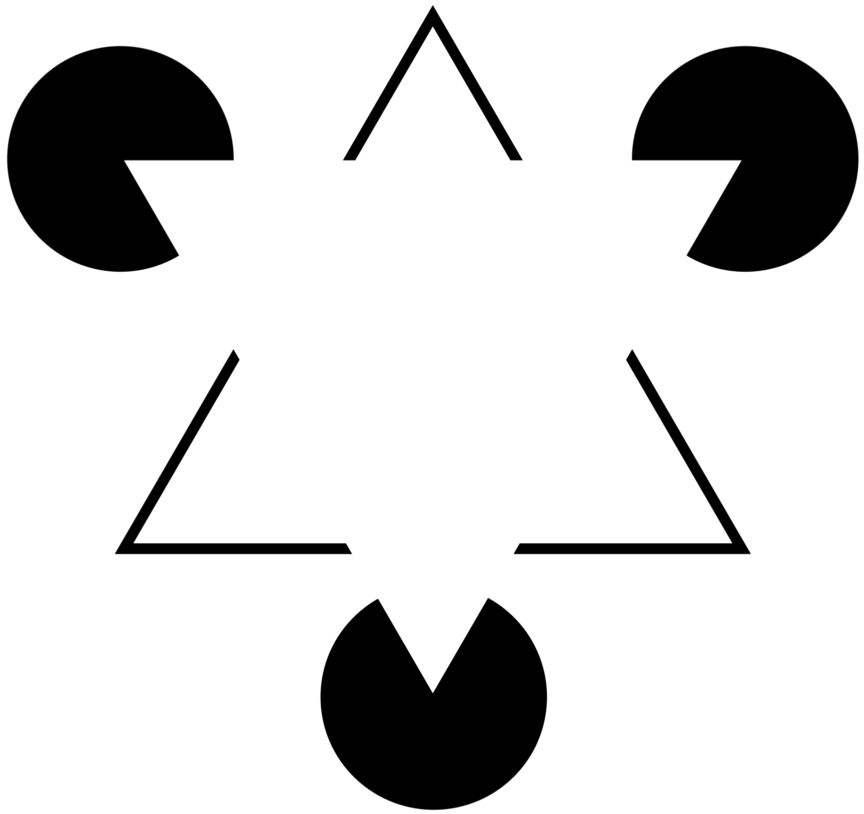
STATISTICAL ADAPTATION IN LOW LEVEL VISION

adaptation to natural image statistics?



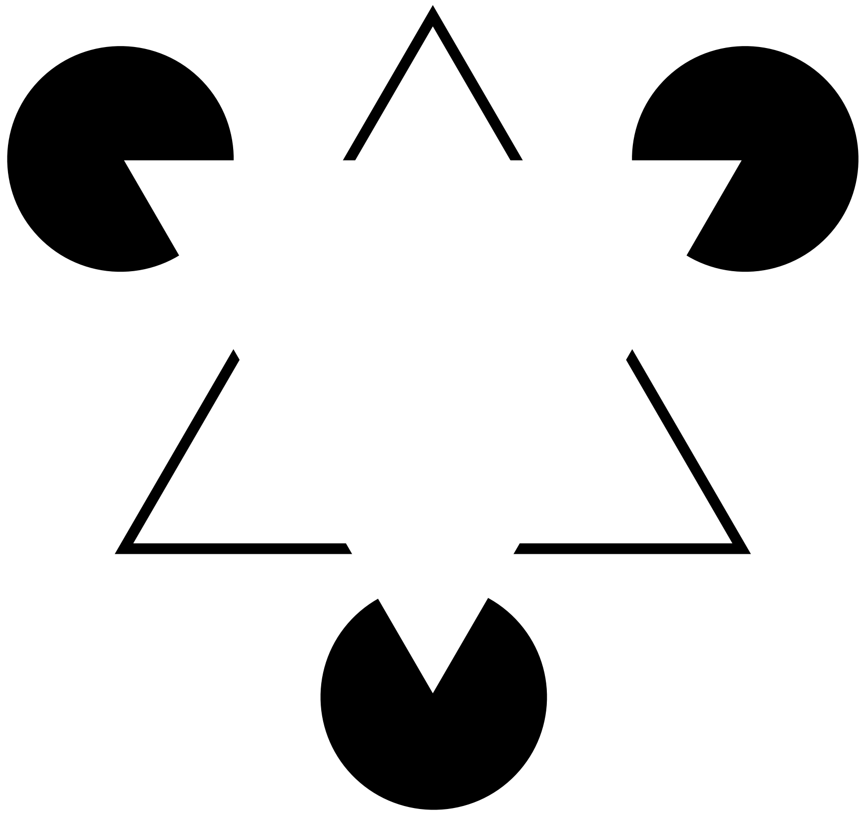
STATISTICAL ADAPTATION IN LOW LEVEL VISION

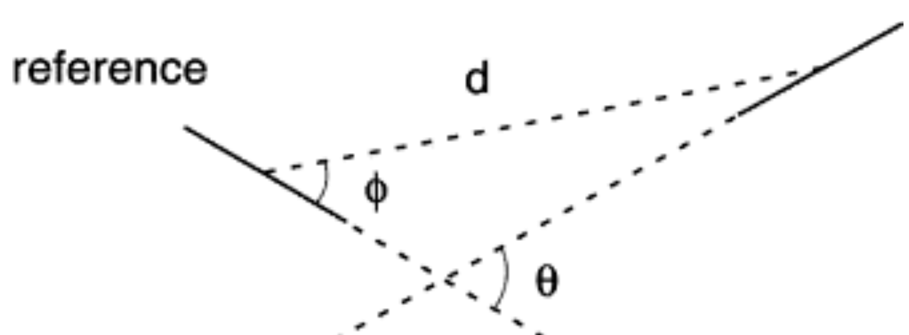
adaptation to natural image statistics?



STATISTICAL ADAPTATION IN LOW LEVEL VISION

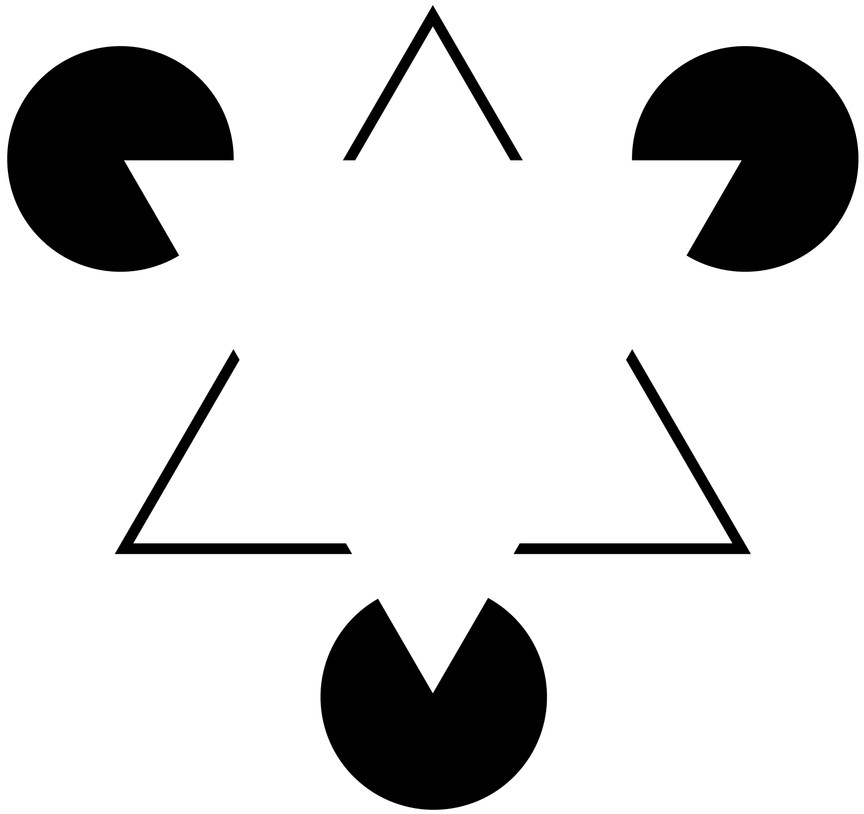
adaptation to natural image statistics?

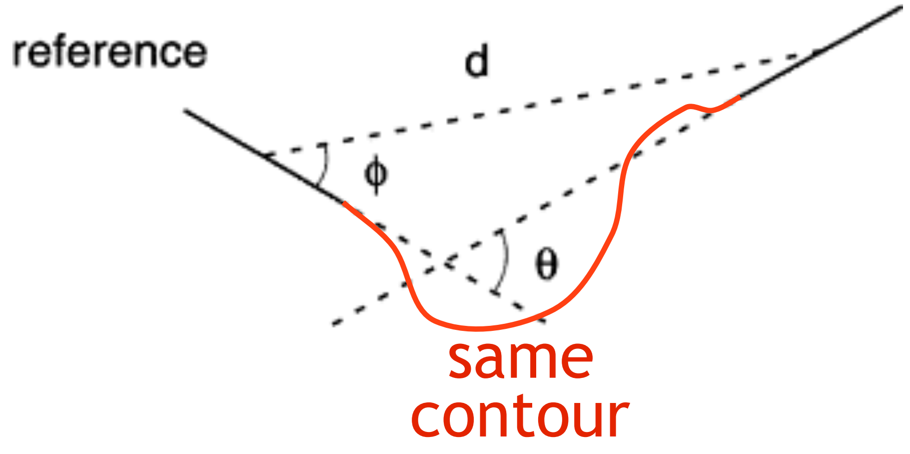




STATISTICAL ADAPTATION IN LOW LEVEL VISION

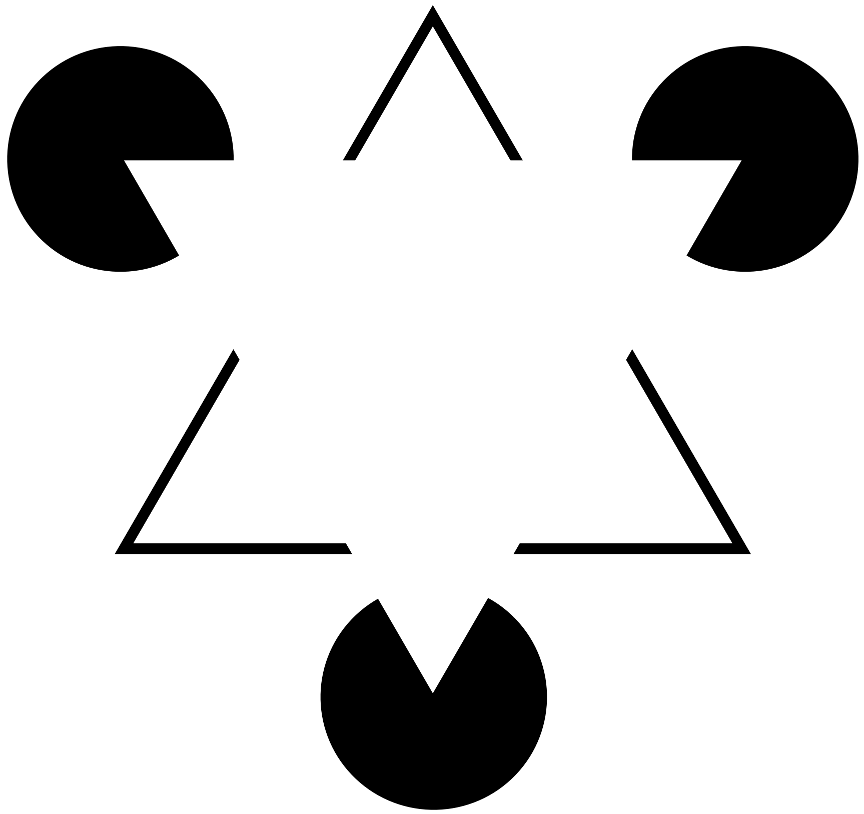
adaptation to natural image statistics?

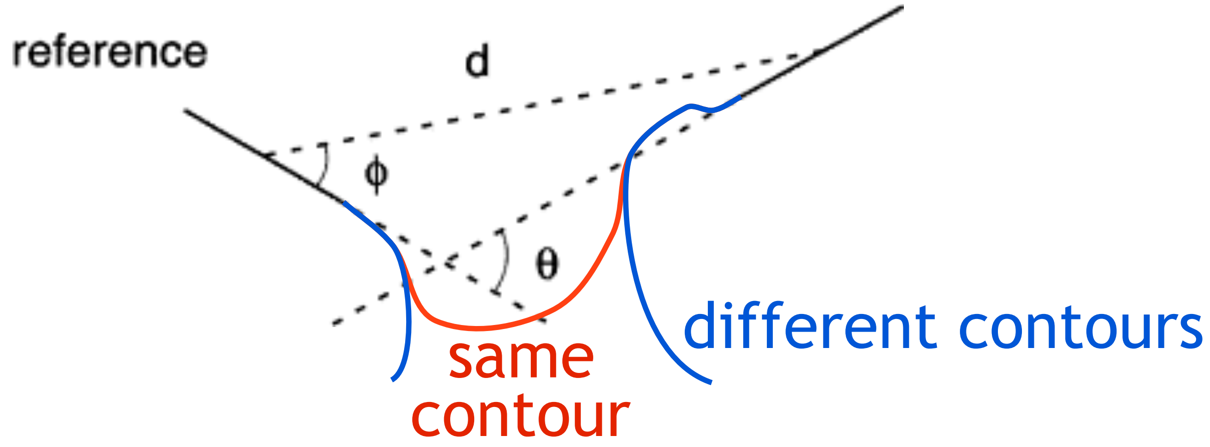




STATISTICAL ADAPTATION IN LOW LEVEL VISION

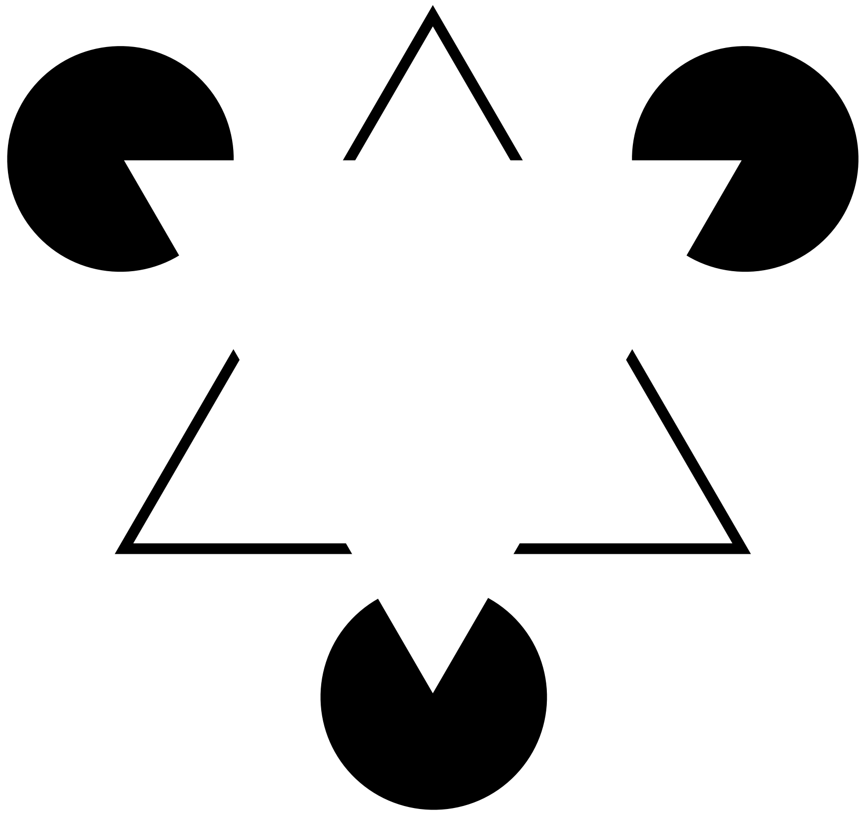
adaptation to natural image statistics?

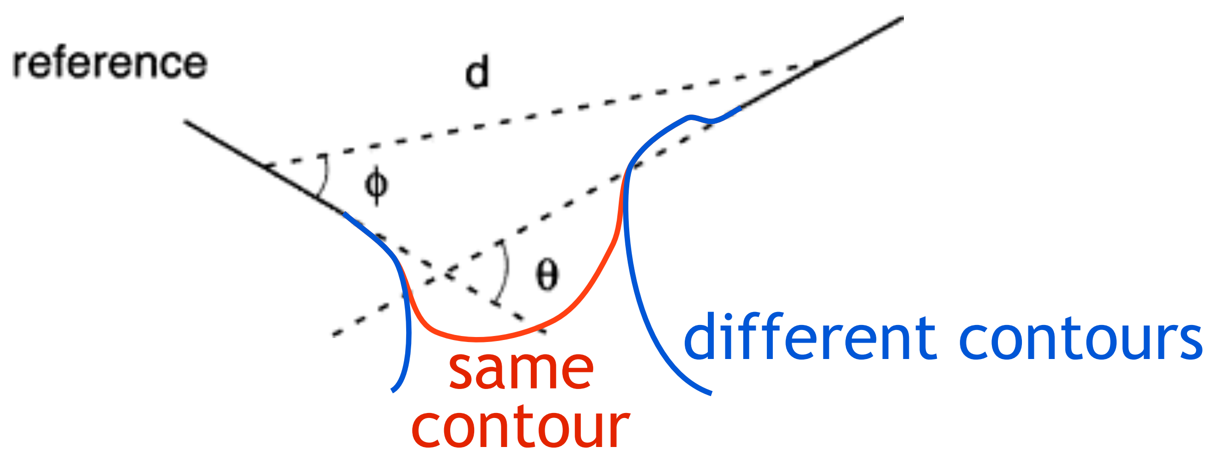


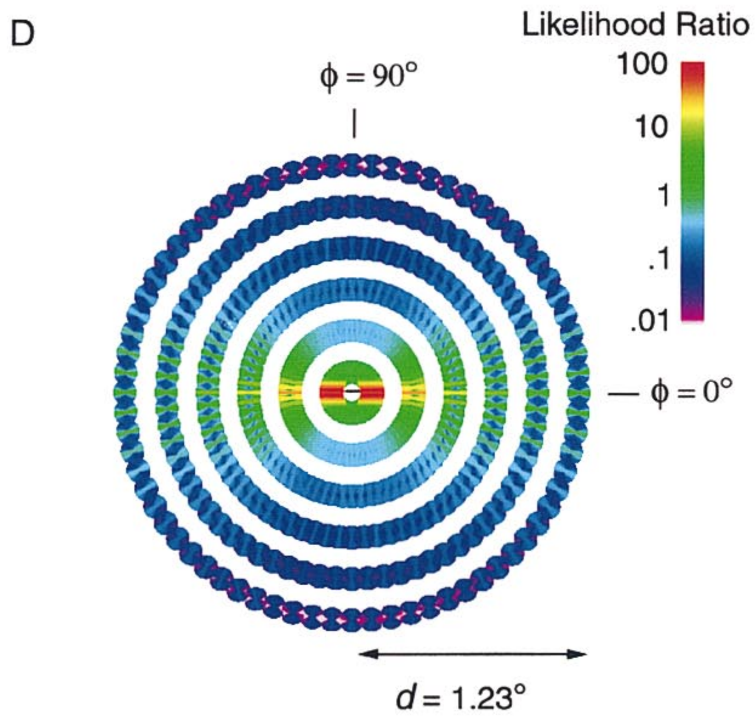


STATISTICAL ADAPTATION IN LOW LEVEL VISION

adaptation to natural image statistics?





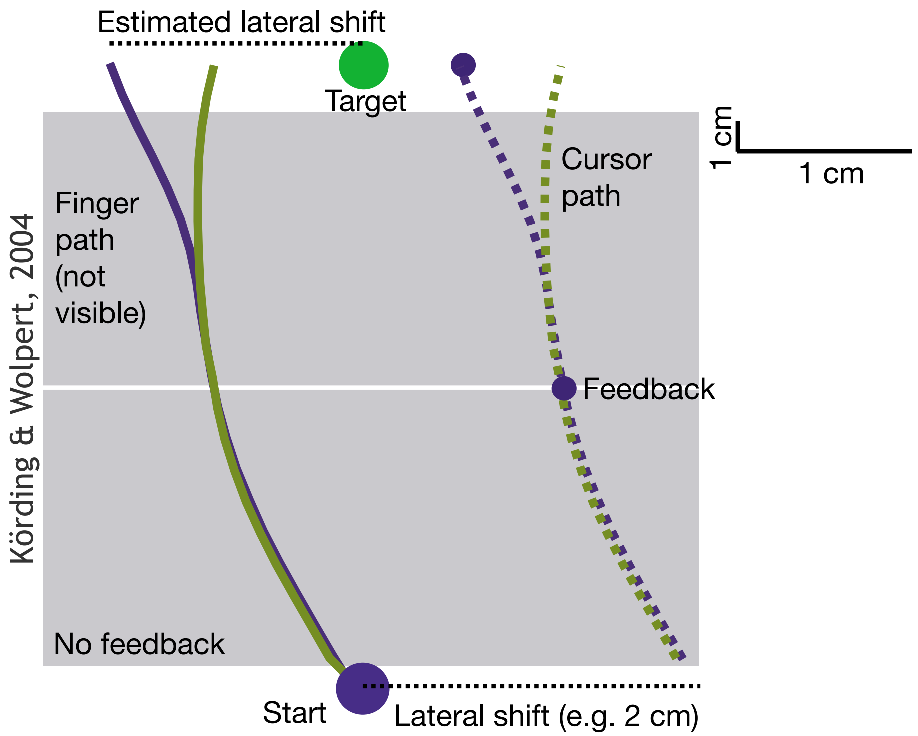


$$\frac{P(d, \phi, \theta | \text{same})}{P(d, \phi, \theta | \text{different})}$$

Geisler et al, 2001

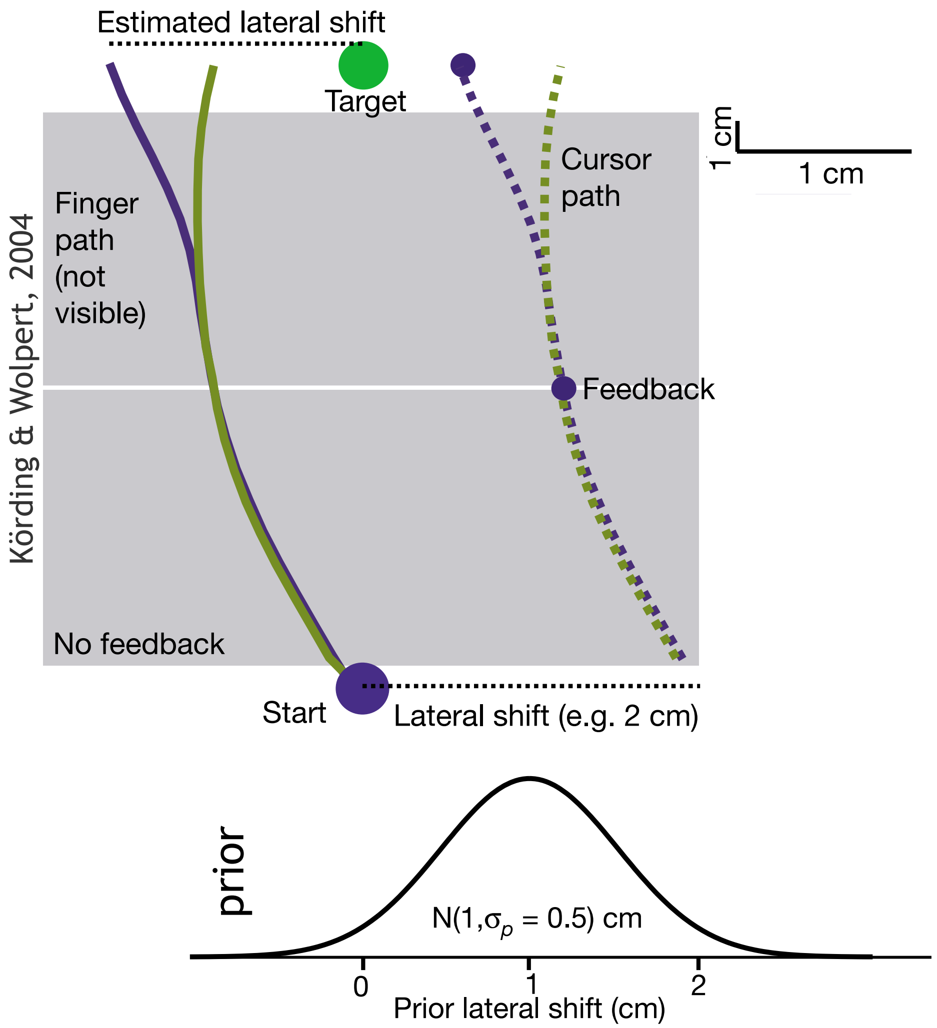
COMBINING PRIORS AND LIKELIHOODS: MOTOR CONTROL

task



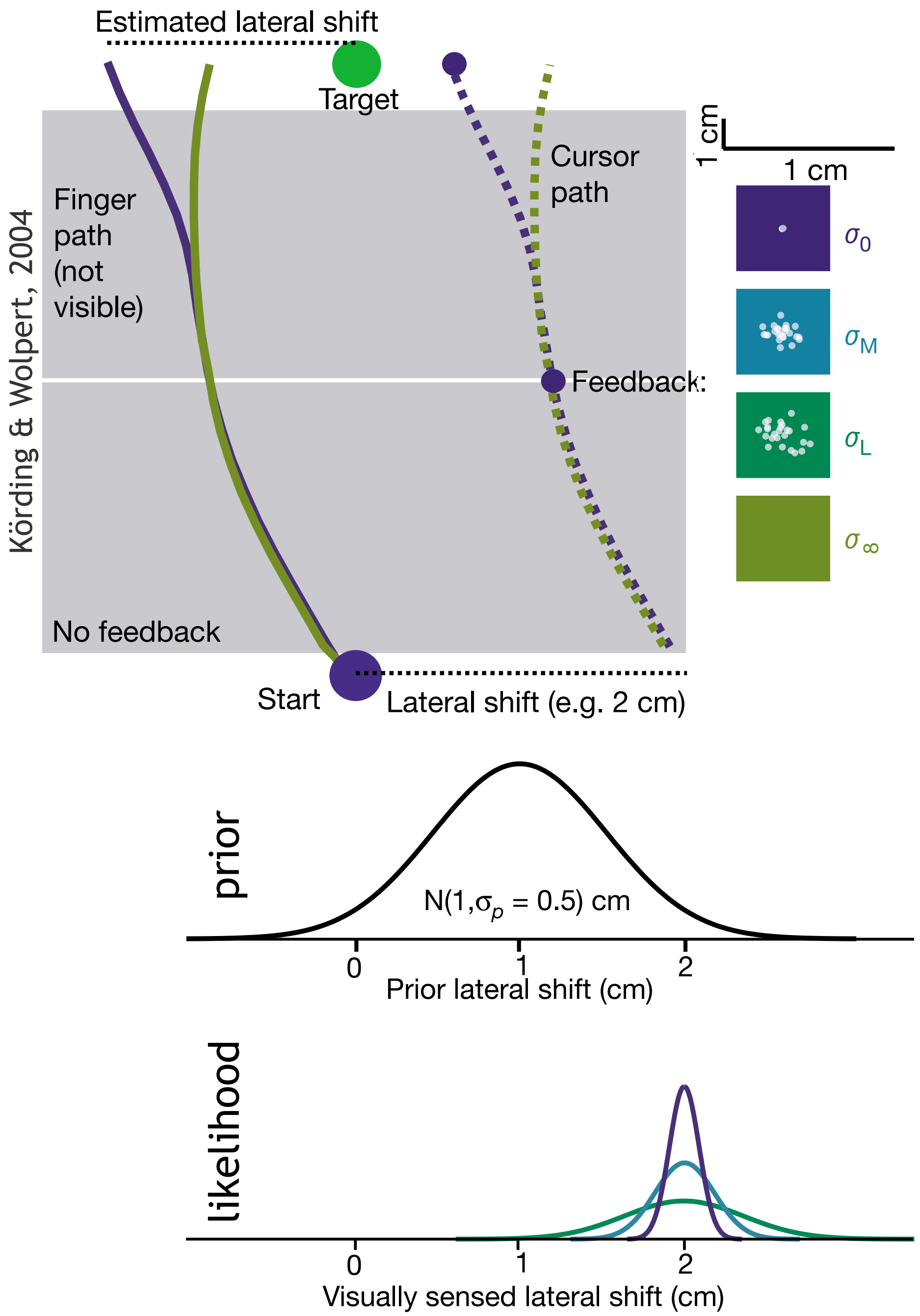
COMBINING PRIORS AND LIKELIHOODS: MOTOR CONTROL

task

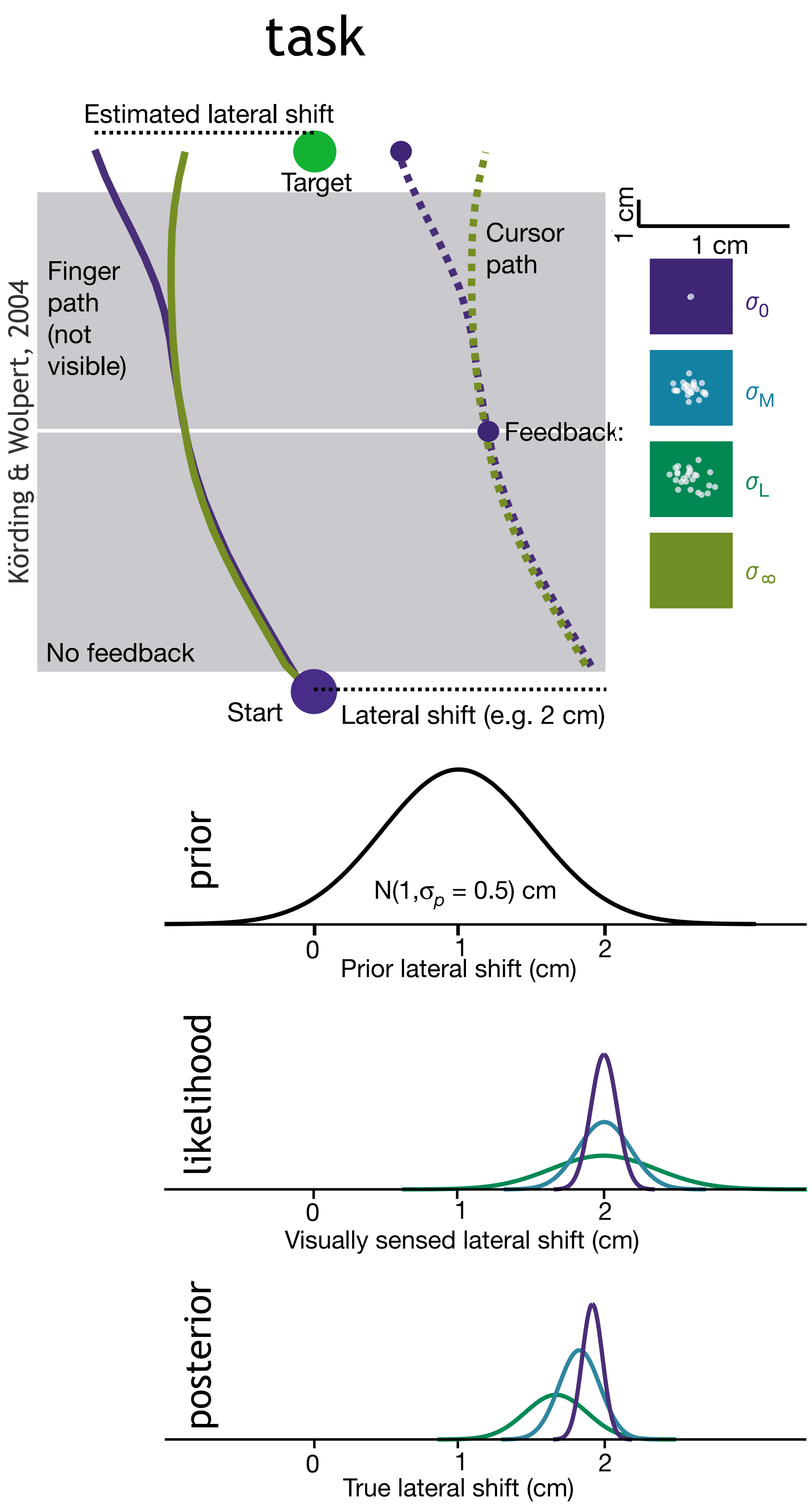


COMBINING PRIORS AND LIKELIHOODS: MOTOR CONTROL

task

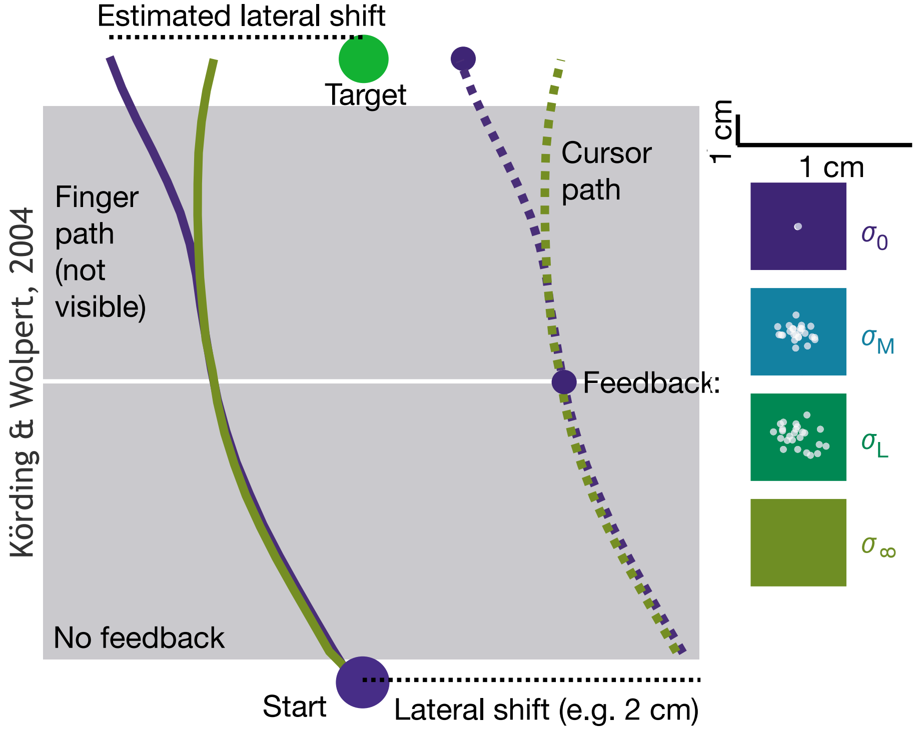


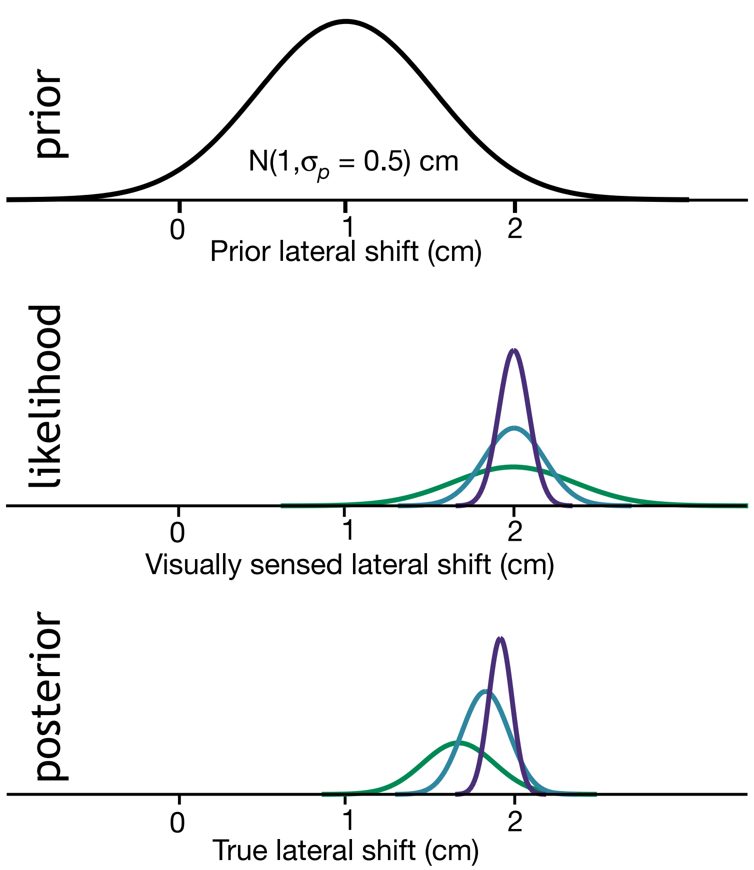
COMBINING PRIORS AND LIKELIHOODS: MOTOR CONTROL



COMBINING PRIORS AND LIKELIHOODS: MOTOR CONTROL

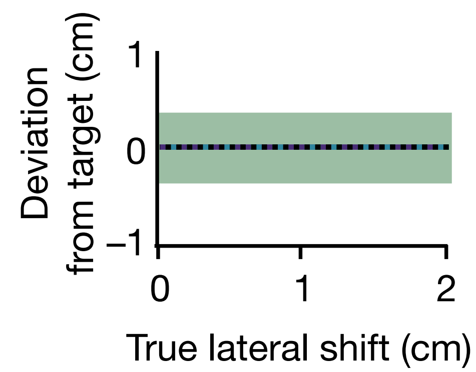
task



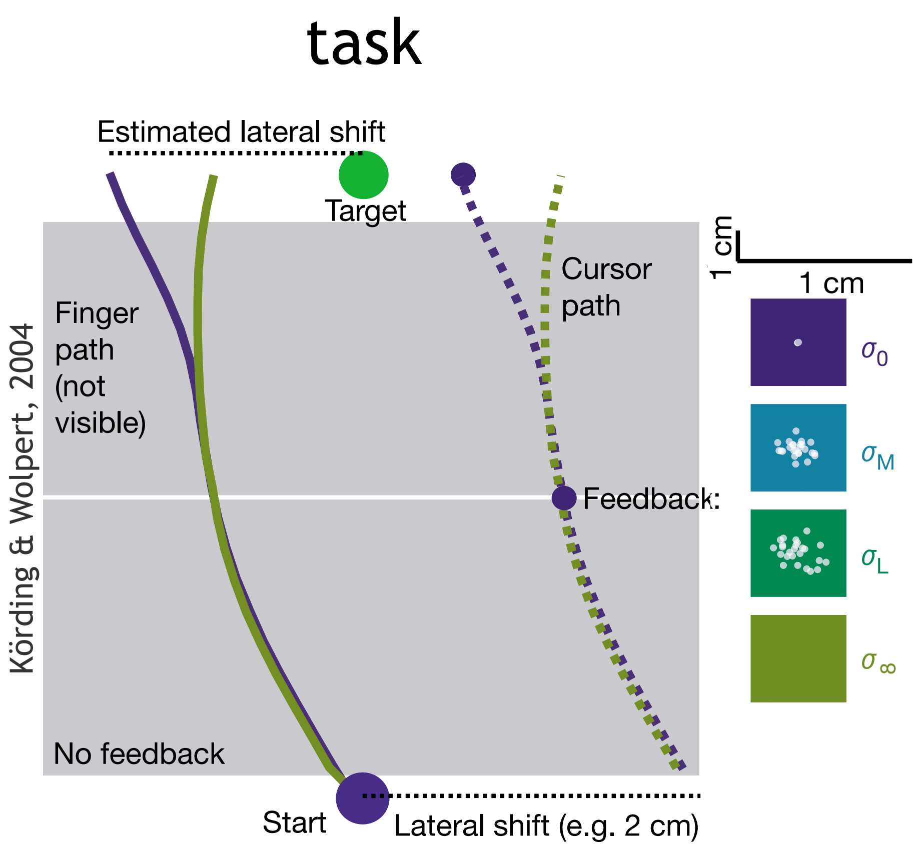


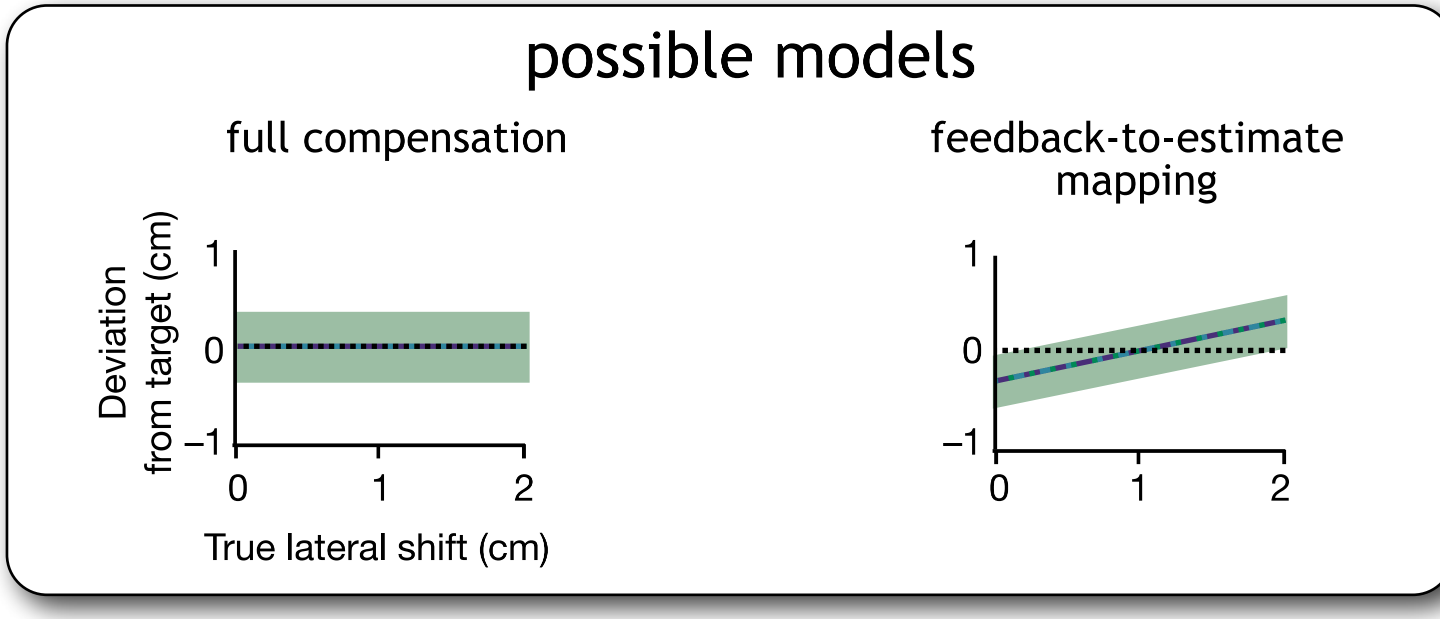
possible models

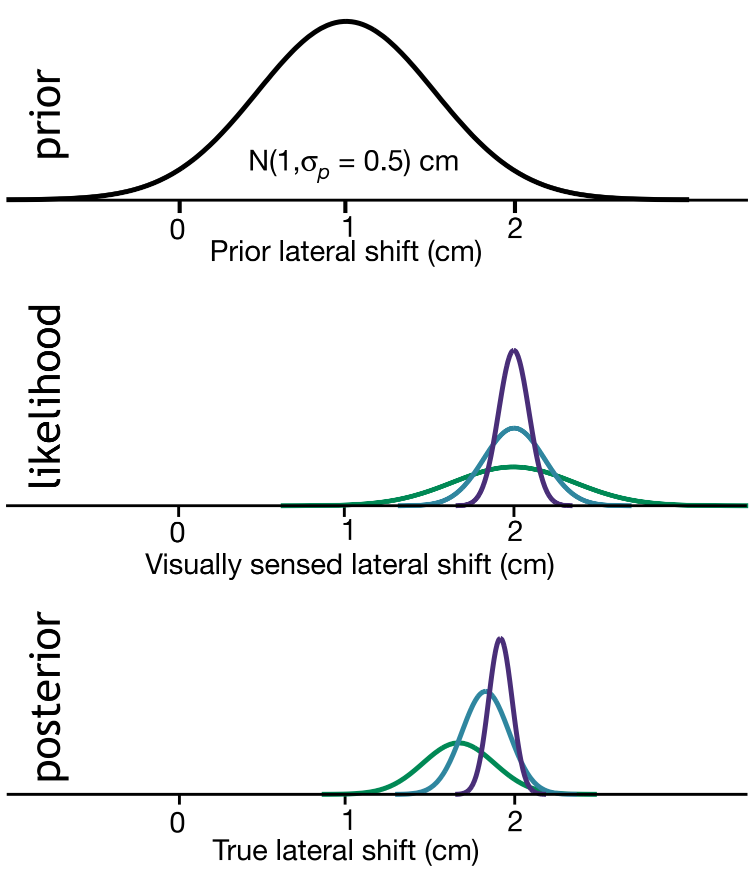
full compensation



COMBINING PRIORS AND LIKELIHOODS: MOTOR CONTROL

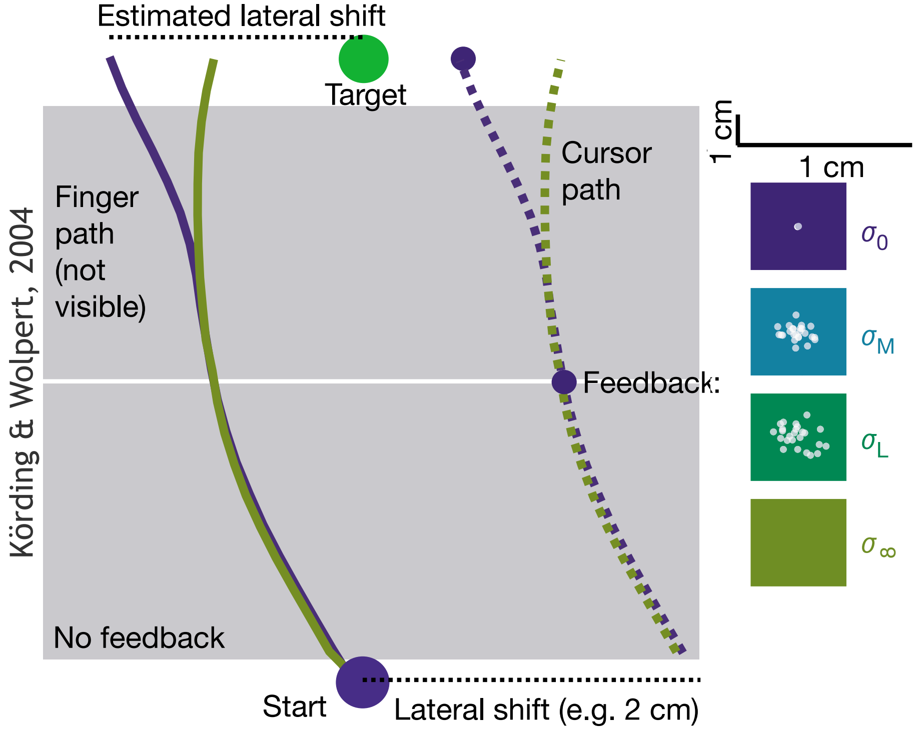


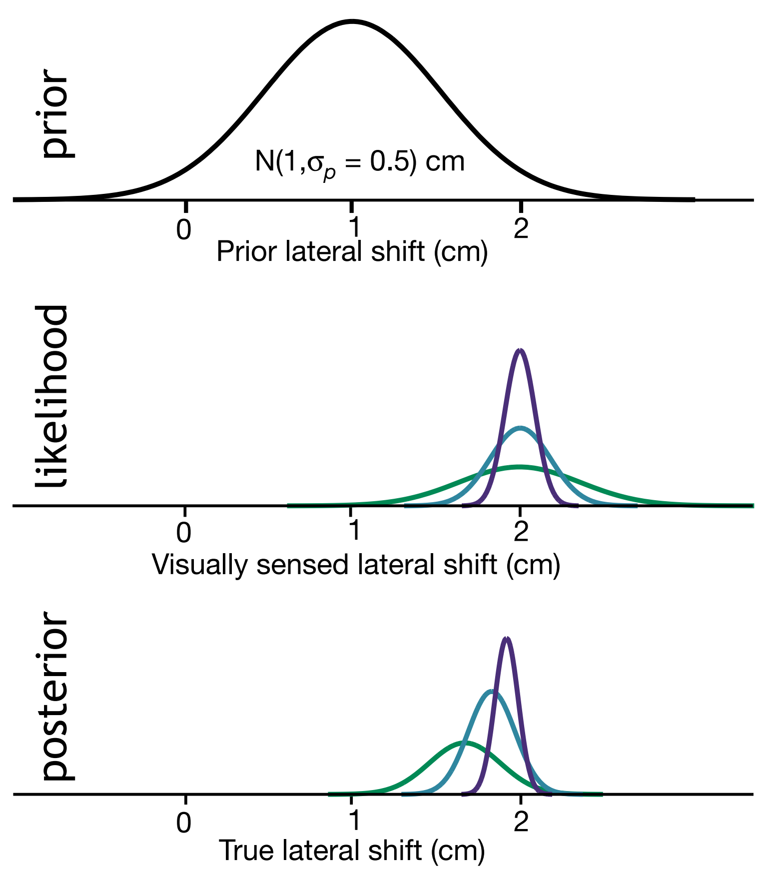


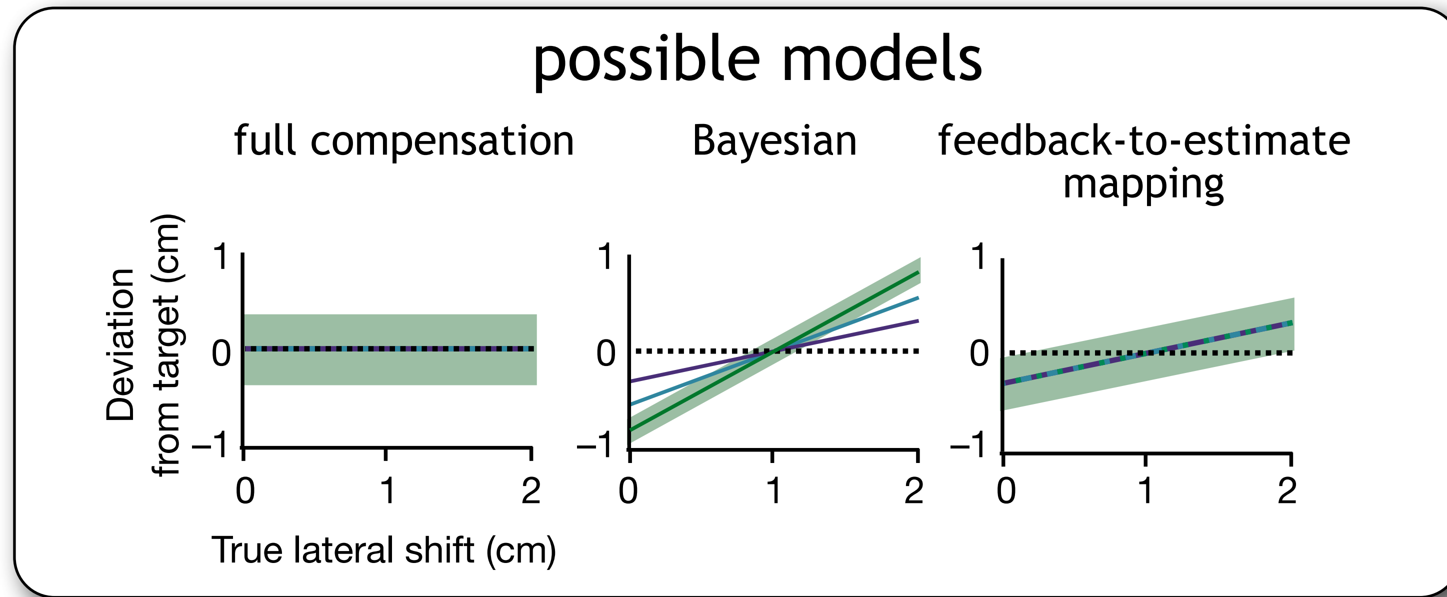


COMBINING PRIORS AND LIKELIHOODS: MOTOR CONTROL

task

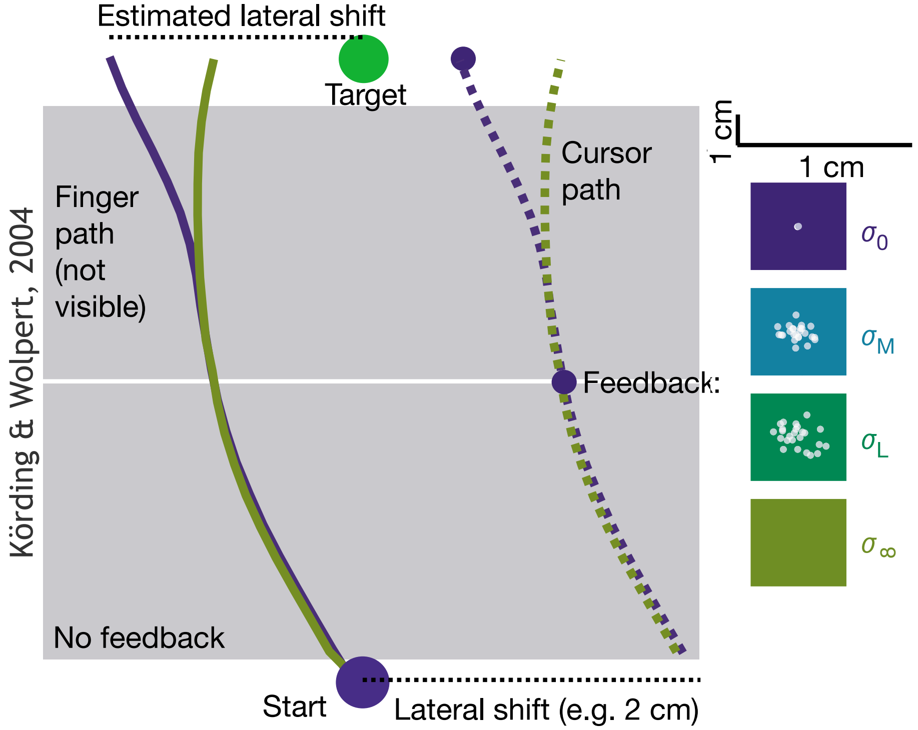


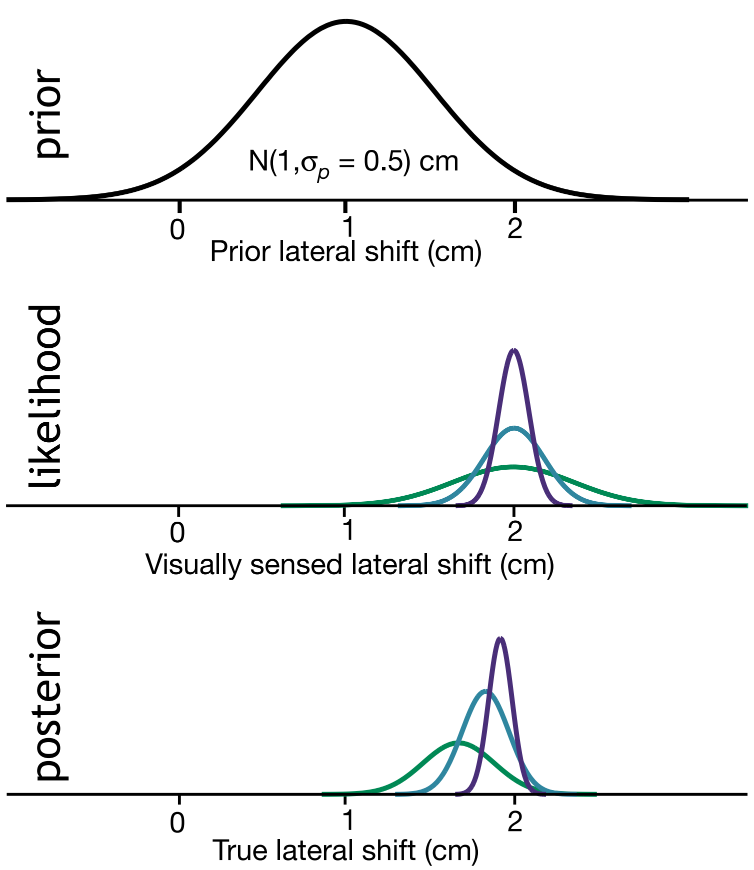




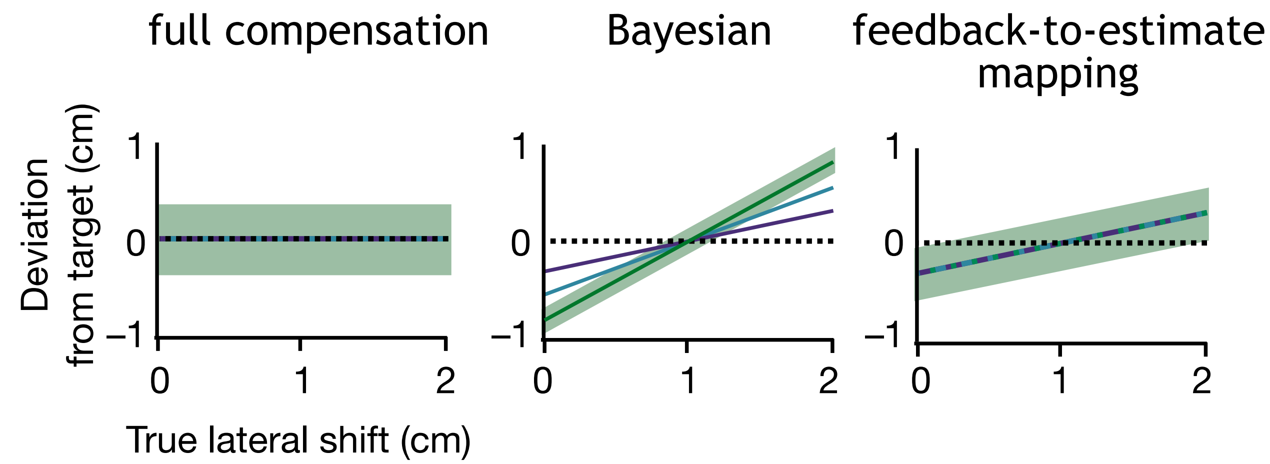
COMBINING PRIORS AND LIKELIHOODS: MOTOR CONTROL

task

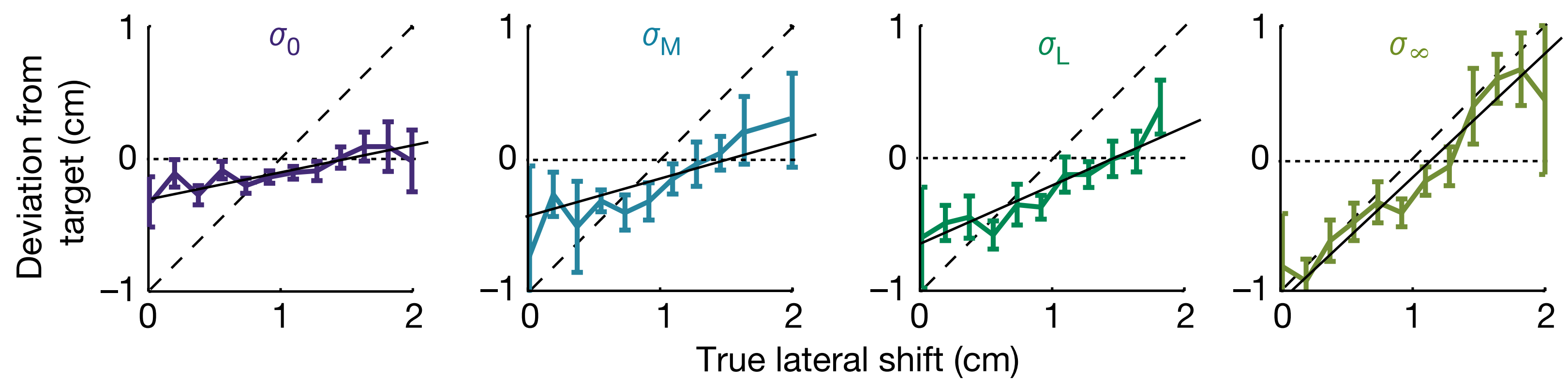




possible models

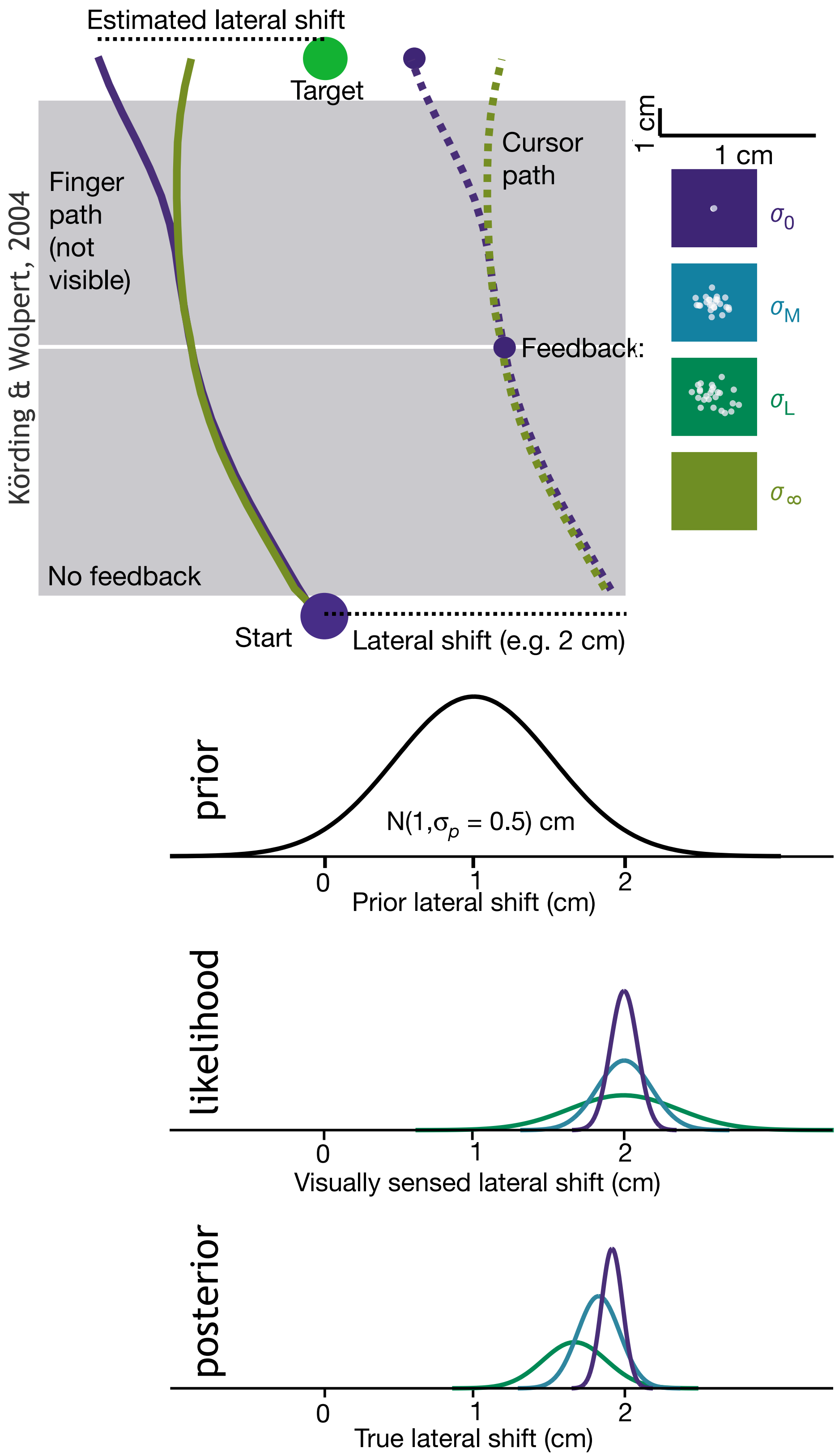


humans

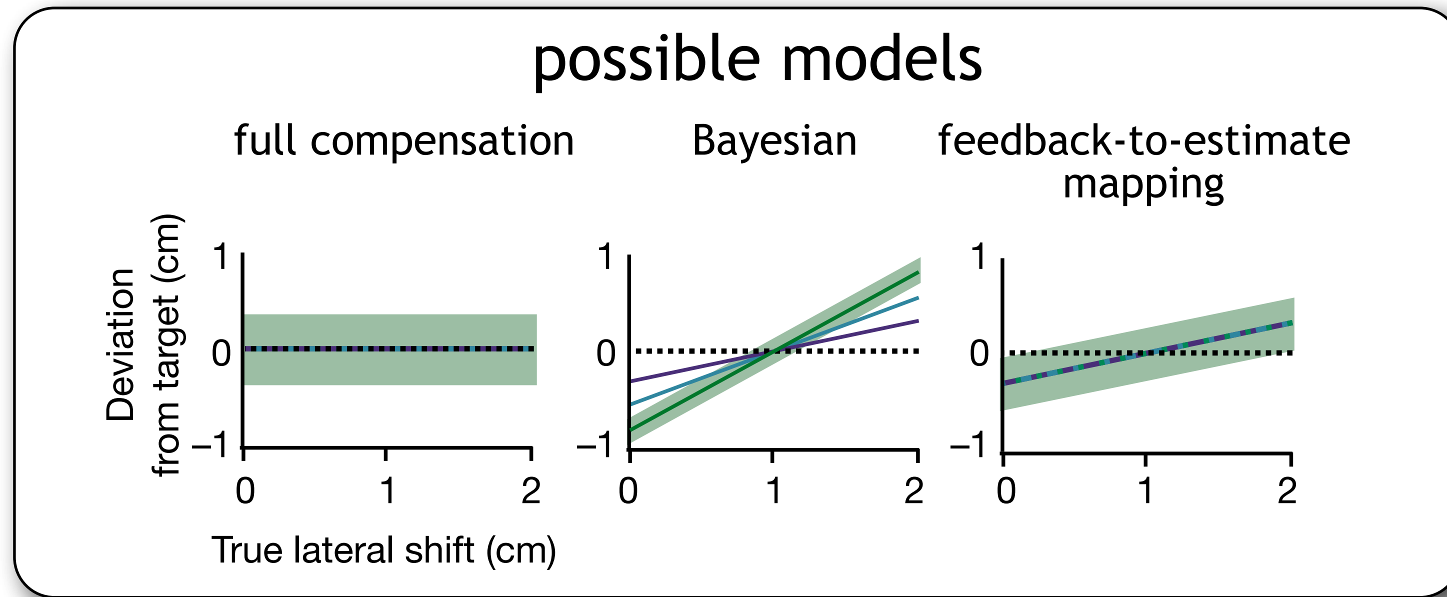


COMBINING PRIORS AND LIKELIHOODS: MOTOR CONTROL

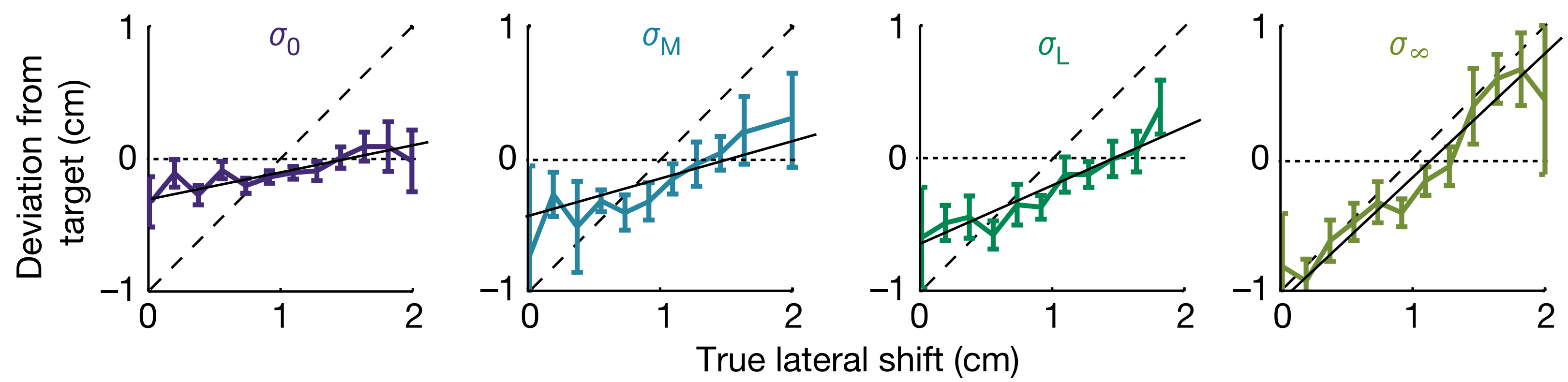
task



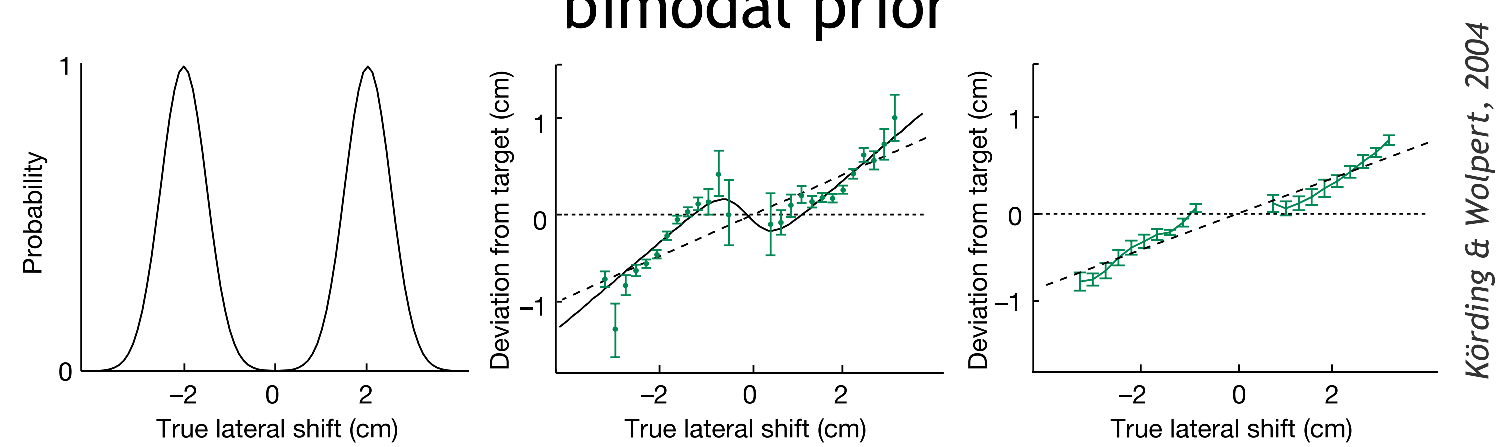
possible models



humans



bimodal prior



COMBINING PRIORS AND LIKELIHOODS: COGNITION

You meet someone who is t years old. What will be his total life span t_{total}?

COMBINING PRIORS AND LIKELIHOODS: COGNITION

You meet someone who is t years old. What will be his total life span t_{total}?

$$\mathrm{P}(t_{\mathrm{total}}|t) \propto \mathrm{P}(t|t_{\mathrm{total}})\ \mathrm{P}(t_{\mathrm{total}})$$

COMBINING PRIORS AND LIKELIHOODS: COGNITION

You meet someone who is t years old. What will be his total life span t_{total}?

$$\mathrm{P}(t_{\mathrm{total}}|t) \propto \mathrm{P}(t|t_{\mathrm{total}})\ \mathrm{P}(t_{\mathrm{total}})$$

the probability that you meet someone at the age of t when s/he will have a total life span of t_{total}

COMBINING PRIORS AND LIKELIHOODS: COGNITION

You meet someone who is t years old. What will be his total life span t_{total}?

$$\mathrm{P}(t_{\text{total}}|t) \propto \mathrm{P}(t|t_{\text{total}})\ \mathrm{P}(t_{\text{total}})$$

the probability that you meet someone at the age of t when s/he will have a total life span of t_{total}

$$\approx \begin{cases} \dfrac{1}{t_{\text{total}}} & \text{if } t < t_{\text{total}} \\ 0 & \text{otherwise} \end{cases}$$

COMBINING PRIORS AND LIKELIHOODS: COGNITION

You meet someone who is t years old. What will be his total life span t_{total}?

$$\mathrm{P}(t_{\text{total}}|t) \propto \mathrm{P}(t|t_{\text{total}})\,\mathrm{P}(t_{\text{total}})$$

$\mathrm{P}(t|t_{\text{total}})$: the probability that you meet someone at the age of t when s/he will have a total life span of t_{total}

$\mathrm{P}(t_{\text{total}})$: prior on life span distribution of people

$$\approx \begin{cases} \dfrac{1}{t_{\text{total}}} & \text{if } t < t_{\text{total}} \\ 0 & \text{otherwise} \end{cases}$$

COMBINING PRIORS AND LIKELIHOODS: COGNITION

You meet someone who is t years old. What will be his total life span t_{total}?

$$\mathrm{P}(t_{\text{total}}|t) \propto \mathrm{P}(t|t_{\text{total}})\,\mathrm{P}(t_{\text{total}})$$

$\mathrm{P}(t|t_{\text{total}})$: the probability that you meet someone at the age of t when s/he will have a total life span of t_{total}

$$\approx \begin{cases} \dfrac{1}{t_{\text{total}}} & \text{if } t < t_{\text{total}} \\ 0 & \text{otherwise} \end{cases}$$

$\mathrm{P}(t_{\text{total}})$: prior on life span distribution of people

\+ decision theory

e.g. report the median of the posterior

COMBINING PRIORS AND LIKELIHOODS: COGNITION

You meet someone who is t years old. What will be his total life span t_{total}?

$$\mathrm{P}(t_{\text{total}}|t) \propto \mathrm{P}(t|t_{\text{total}})\,\mathrm{P}(t_{\text{total}})$$

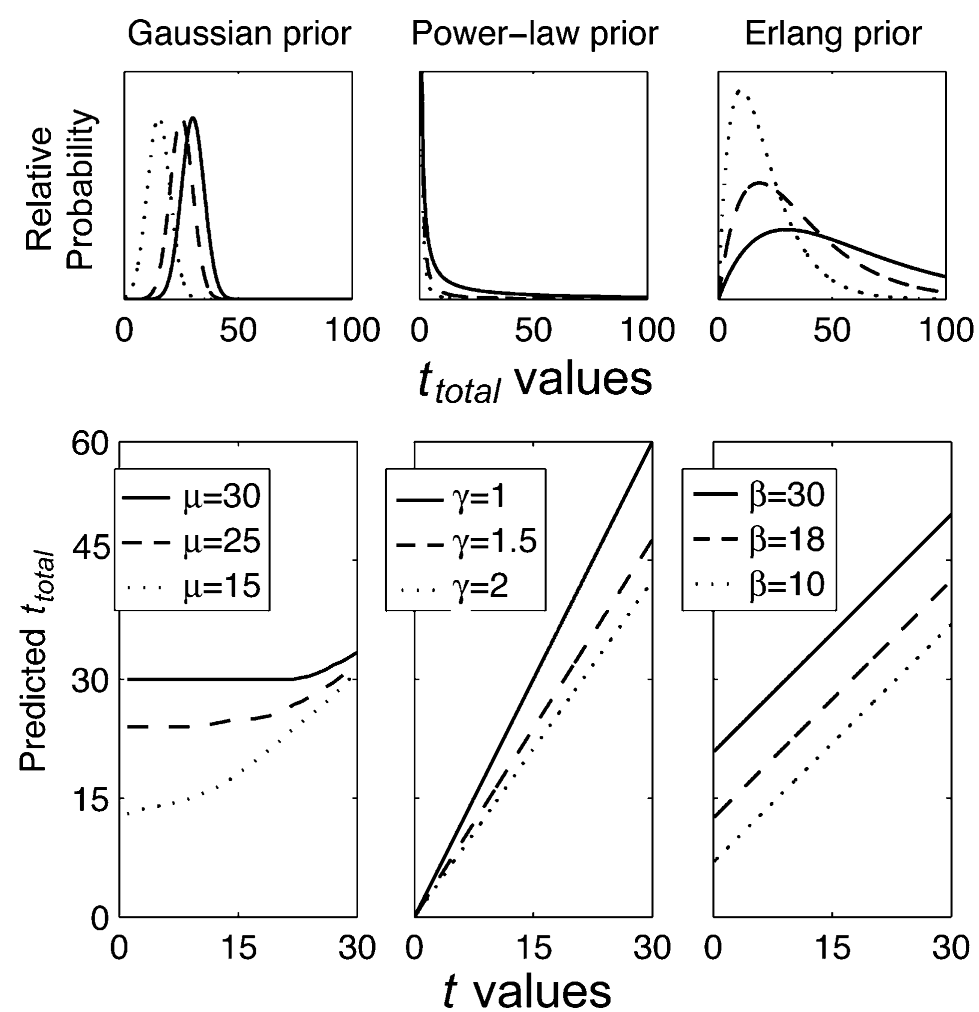
the probability that you meet someone at the age of t when s/he will have a total life span of t_{total}

prior on life span distribution of people

$$\approx \begin{cases} \dfrac{1}{t_{\text{total}}} & \text{if } t < t_{\text{total}} \\ 0 & \text{otherwise} \end{cases}$$

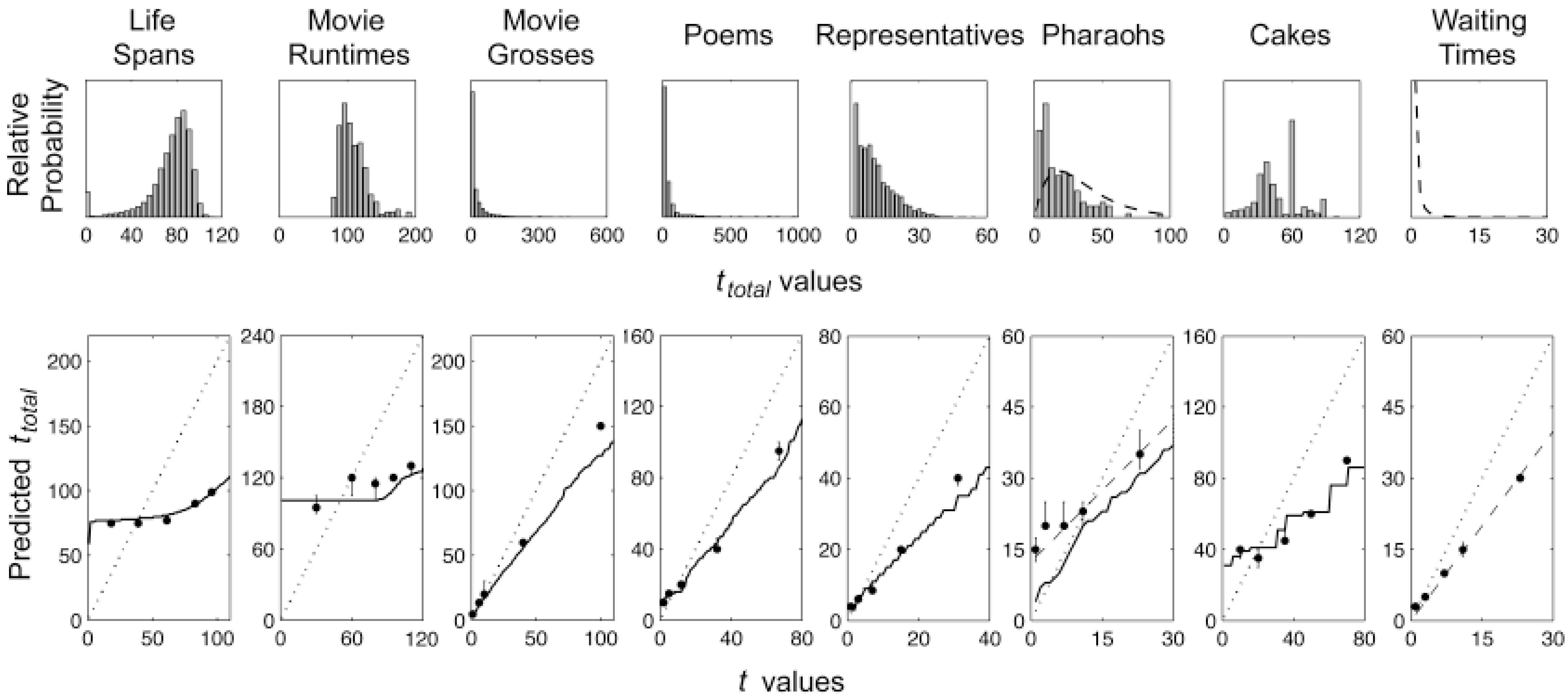
+ decision theory

e.g. report the median of the posterior



Griffiths & Tenenbaum, 2006

ADAPTATION TO ENVIRONMENTAL STATISTICS



Griffiths & Tenenbaum, 2006

WHAT ARE NATURAL PRIORS LIKE?

WHAT ARE NATURAL PRIORS LIKE?

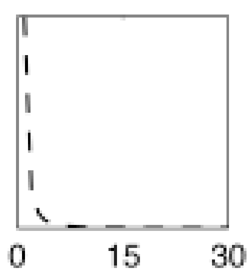
- adapted to ‘natural’ statistics

WHAT ARE NATURAL PRIORS LIKE?

- adapted to 'natural' statistics

box office phone waiting times

Griffiths & Tenenbaum, 2006

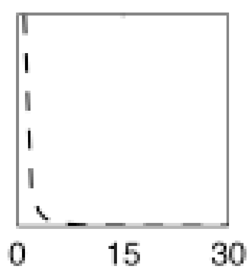


WHAT ARE NATURAL PRIORS LIKE?

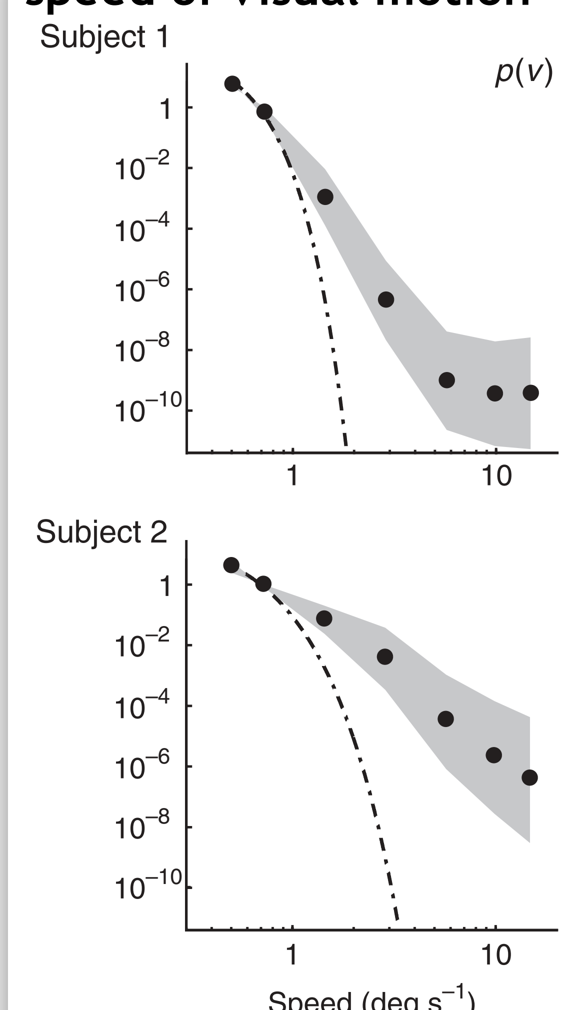
- adapted to 'natural' statistics

box office phone waiting times

Griffiths & Tenenbaum, 2006



speed of visual motion

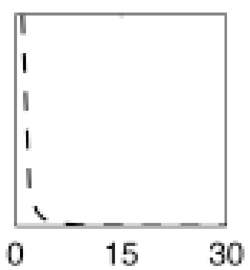


Stocker & Simoncelli, 2006

WHAT ARE NATURAL PRIORS LIKE?

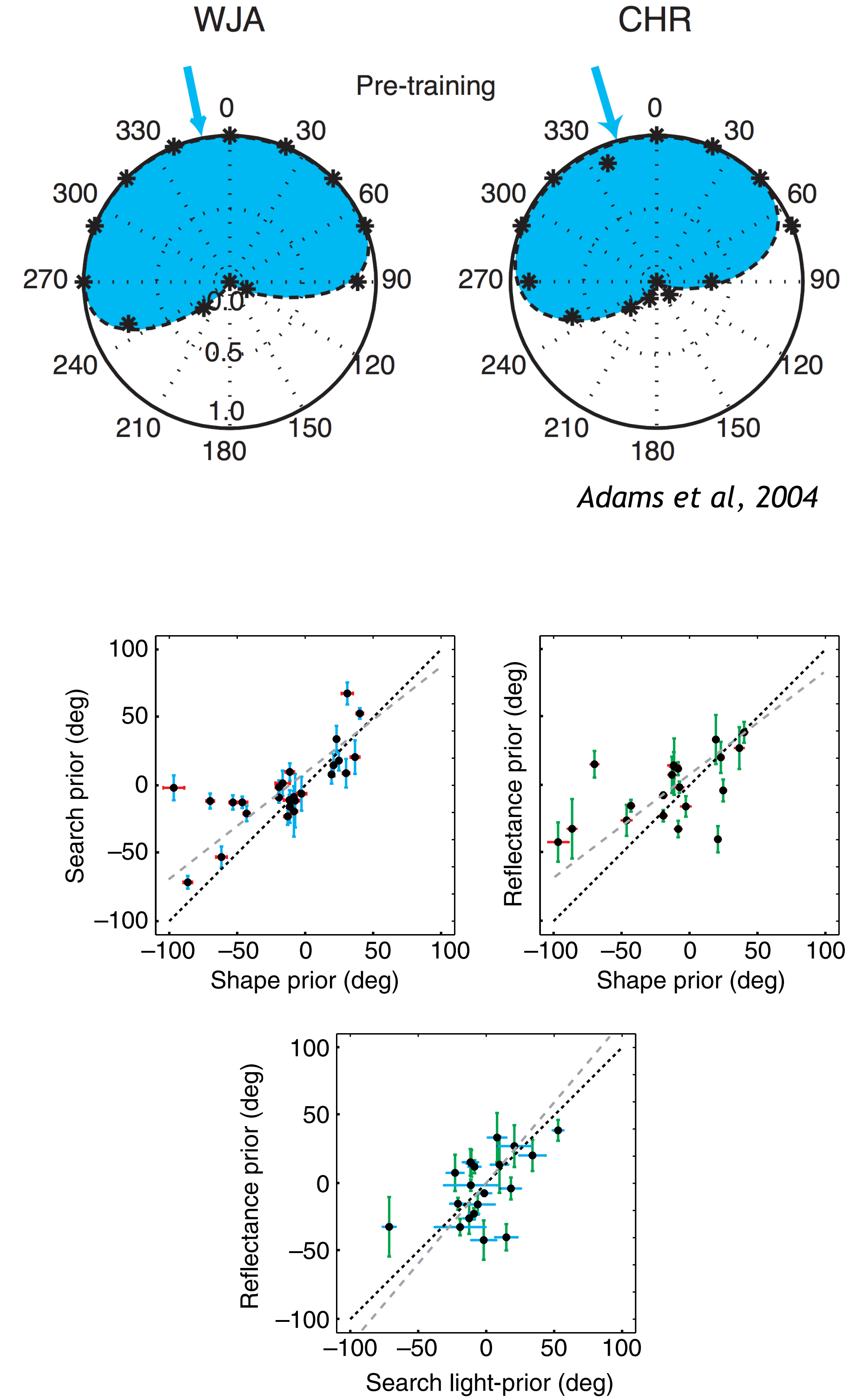
- adapted to 'natural' statistics

box office phone waiting times



Griffiths & Tenenbaum, 2006

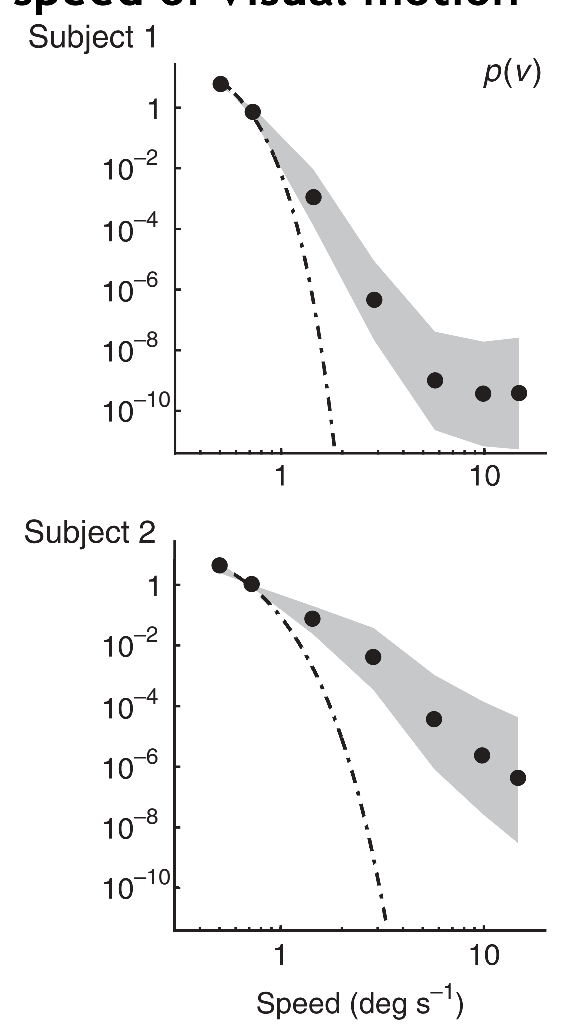
direction of light



Adams et al, 2004

Adams, 2007

speed of visual motion



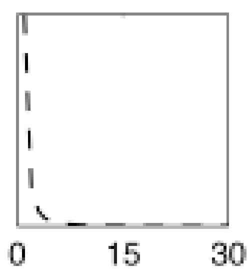
Stocker & Simoncelli, 2006

WHAT ARE NATURAL PRIORS LIKE?

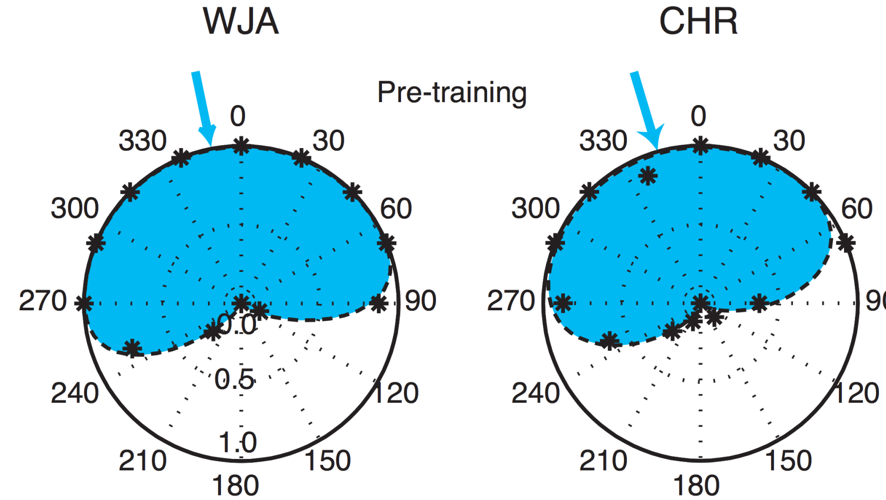
- adapted to 'natural' statistics

box office phone waiting times

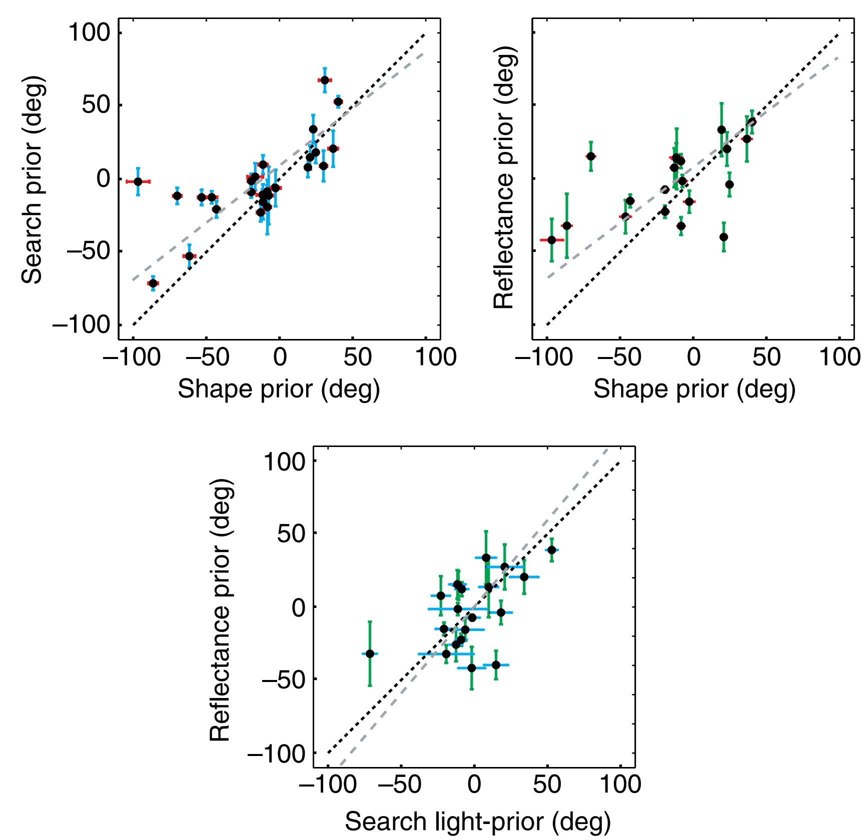
Griffiths & Tenenbaum, 2006



direction of light

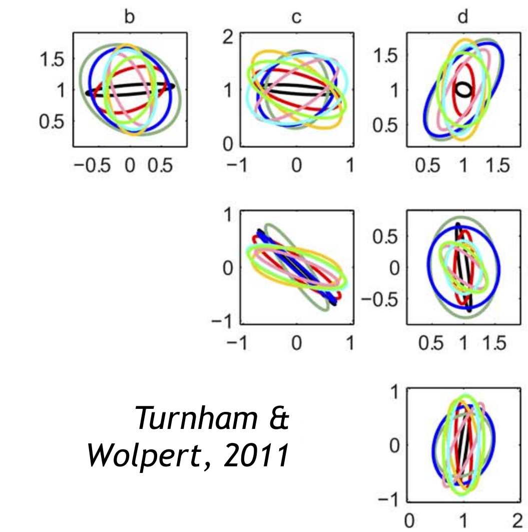


Adams et al, 2004



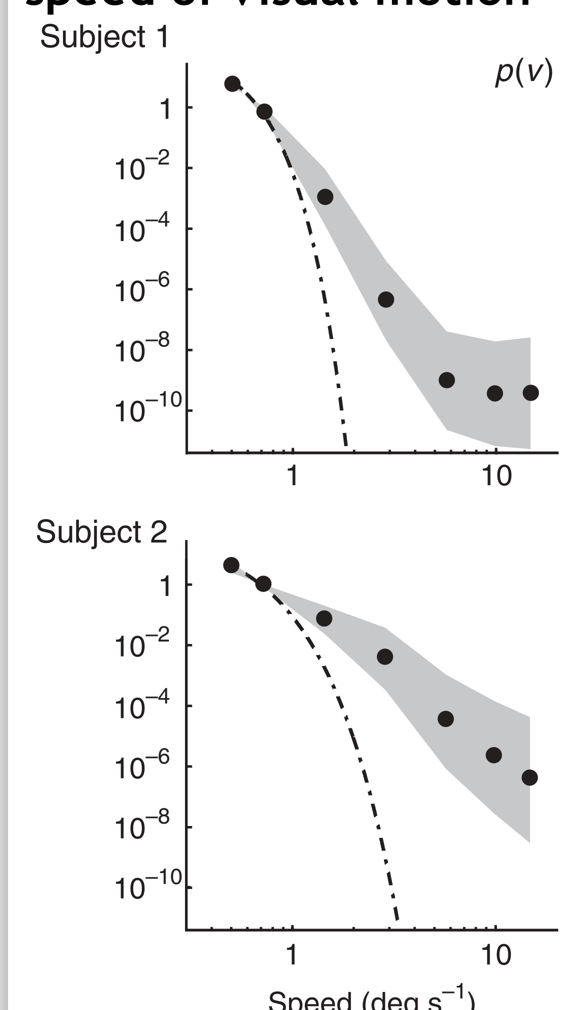
Adams, 2007

visuomotor transformations



Turnham & Wolpert, 2011

speed of visual motion



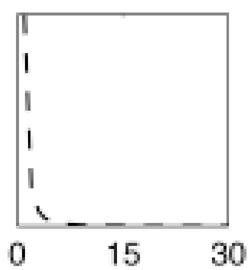
Stocker & Simoncelli, 2006

WHAT ARE NATURAL PRIORS LIKE?

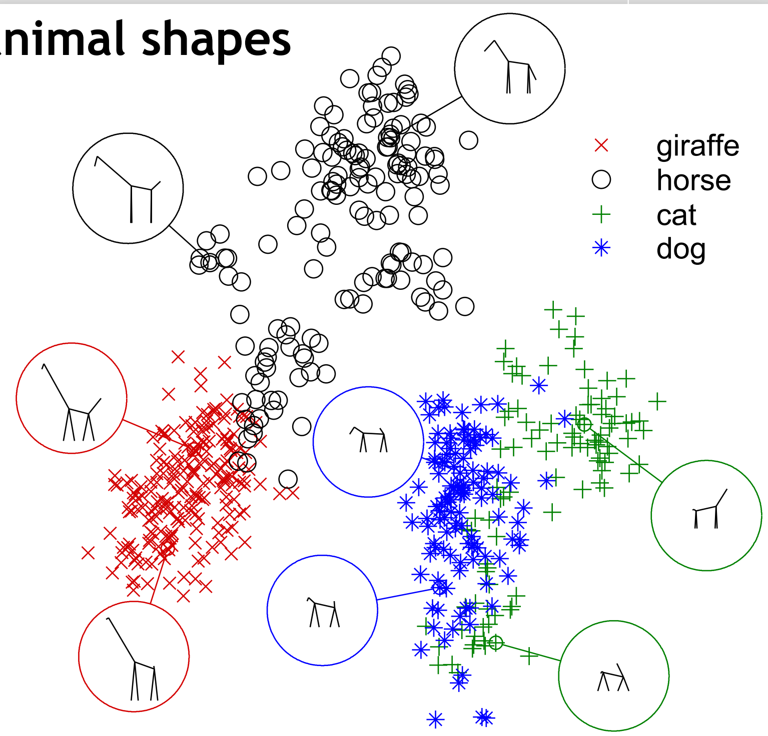
- adapted to ‘natural’ statistics

box office phone waiting times

Griffiths & Tenenbaum, 2006

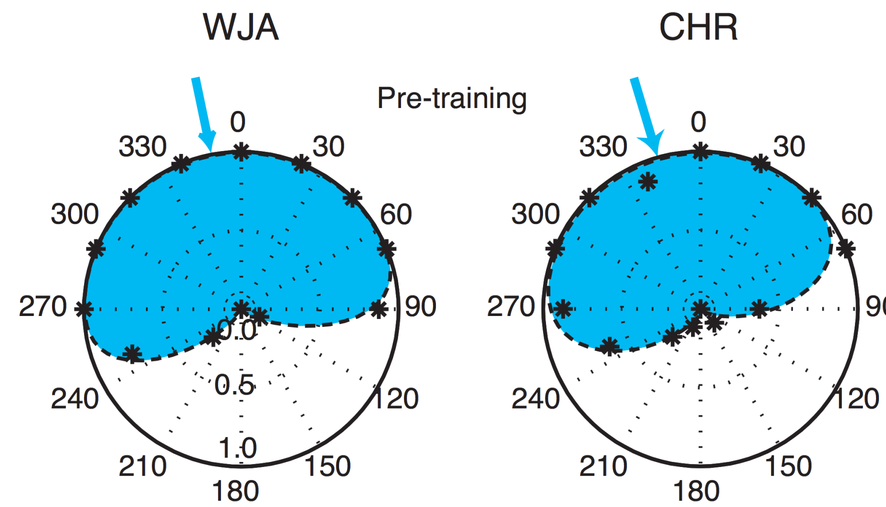


animal shapes

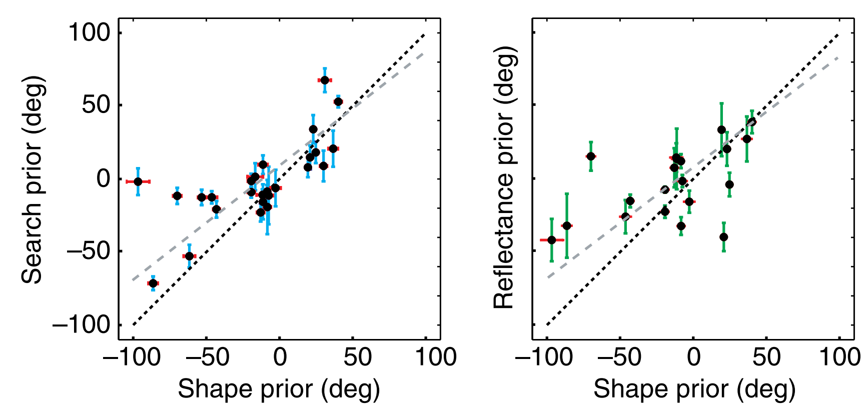


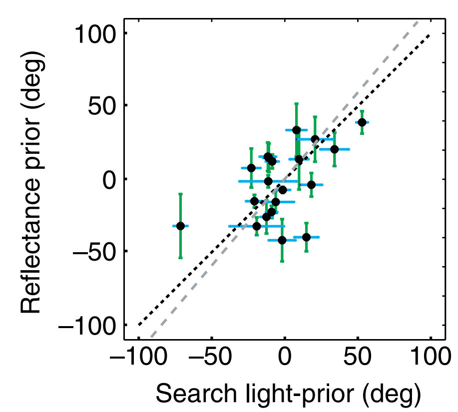
Sanborn & Griffiths, 2008

direction of light



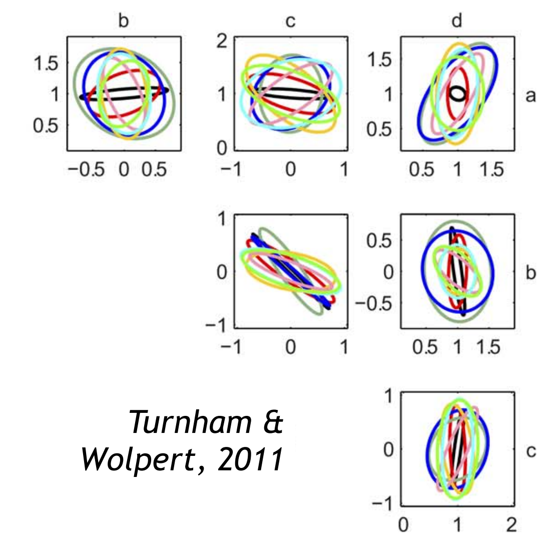
Adams et al, 2004





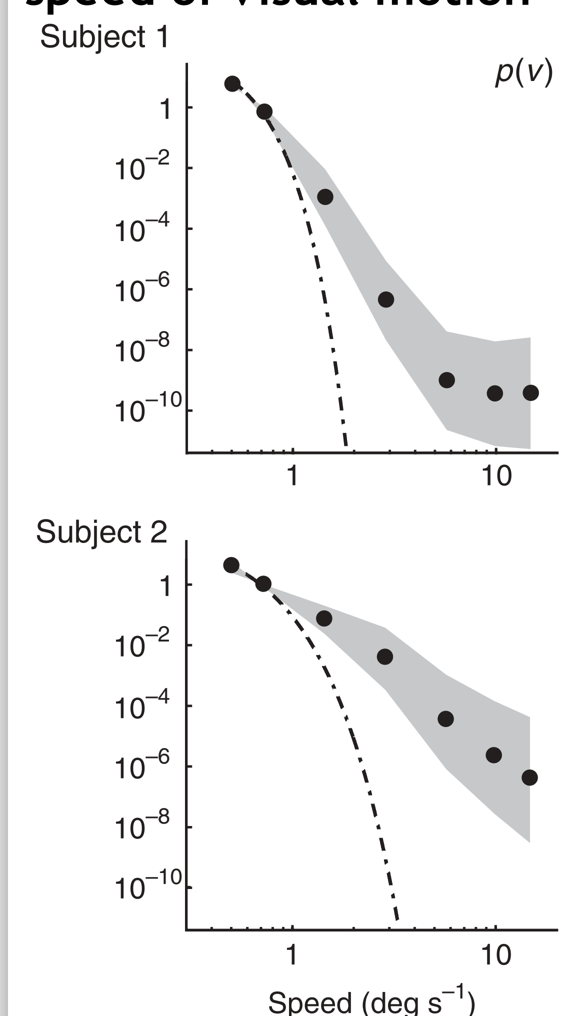
Adams, 2007

visuomotor transformations



Turnham & Wolpert, 2011

speed of visual motion



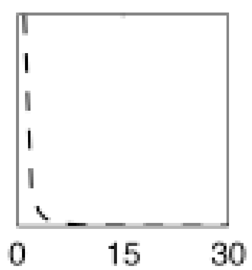
Stocker & Simoncelli, 2006

WHAT ARE NATURAL PRIORS LIKE?

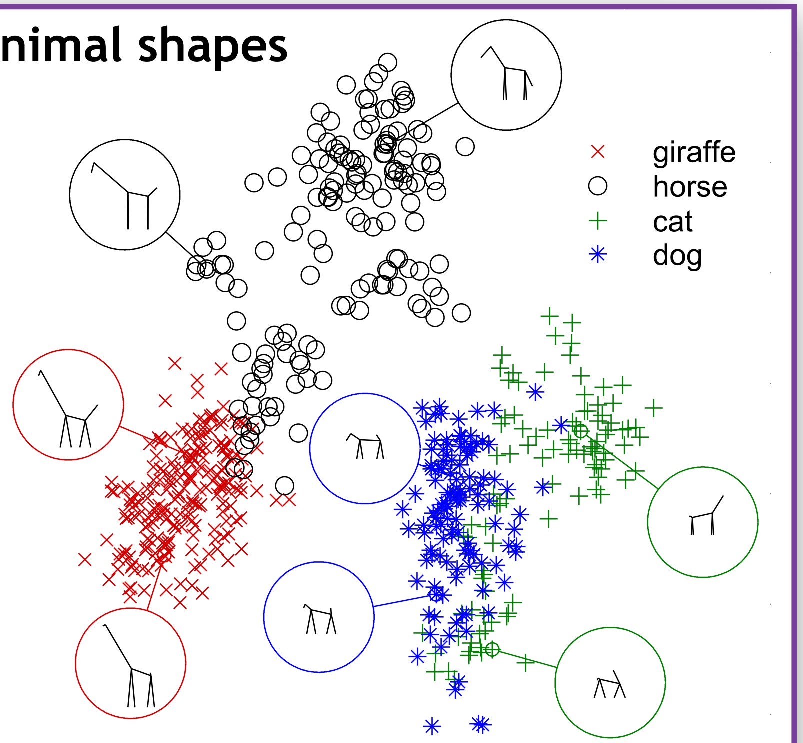
- adapted to 'natural' statistics
- pool data across subjects

box office phone waiting times

Griffiths & Tenenbaum, 2006

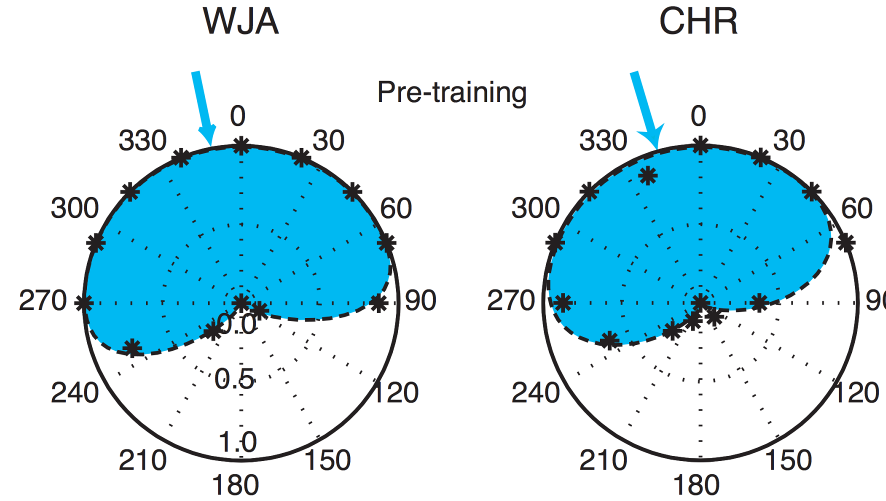


animal shapes

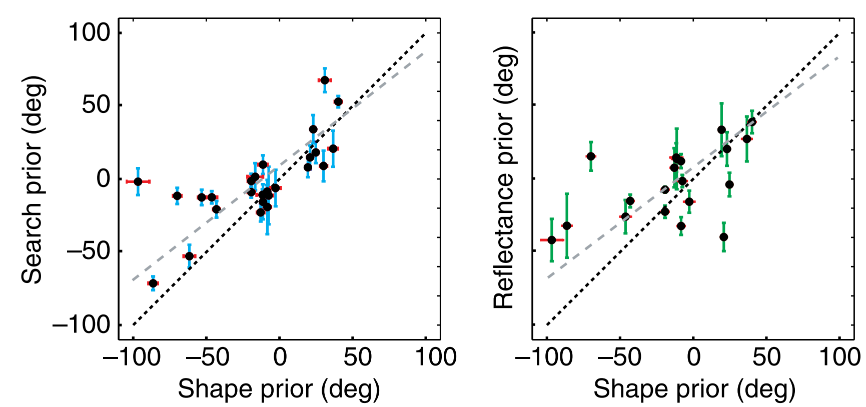


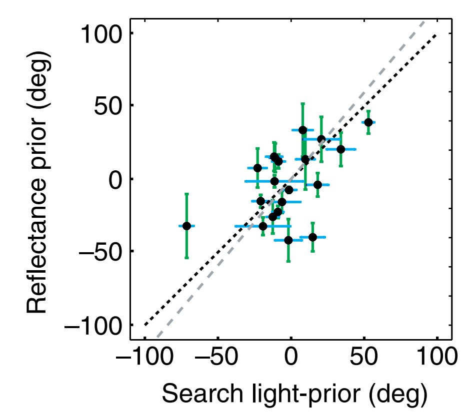
Sanborn & Griffiths, 2008

direction of light



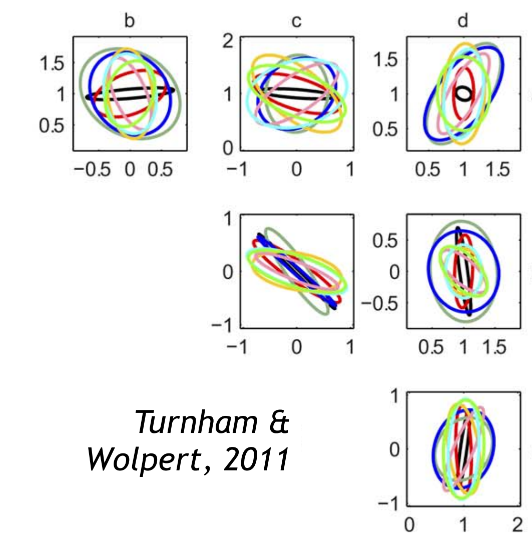
Adams et al, 2004





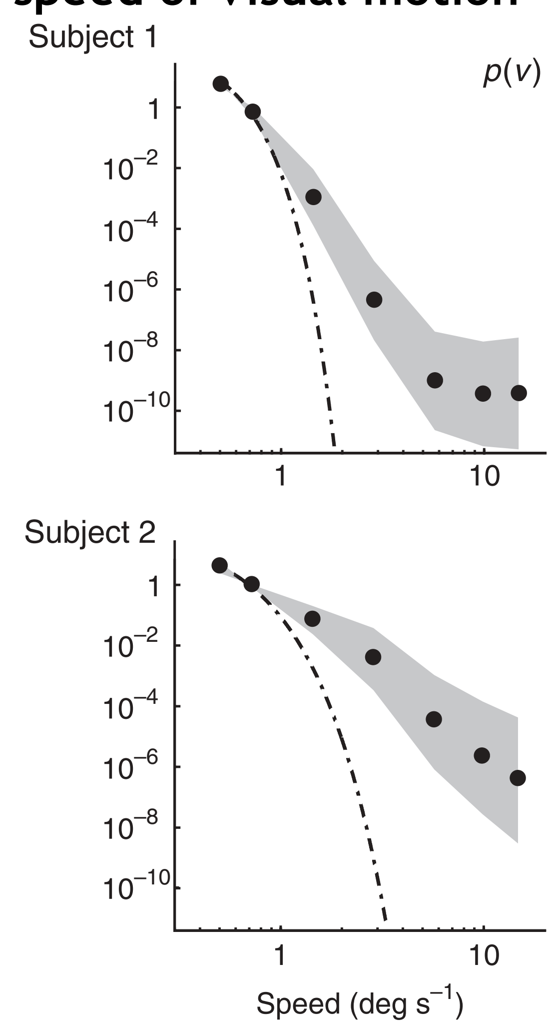
Adams, 2007

visuomotor transformations



Turnham & Wolpert, 2011

speed of visual motion

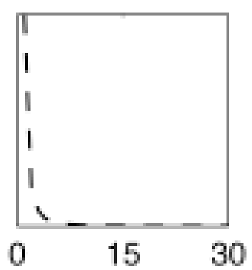


Stocker & Simoncelli, 2006

WHAT ARE NATURAL PRIORS LIKE?

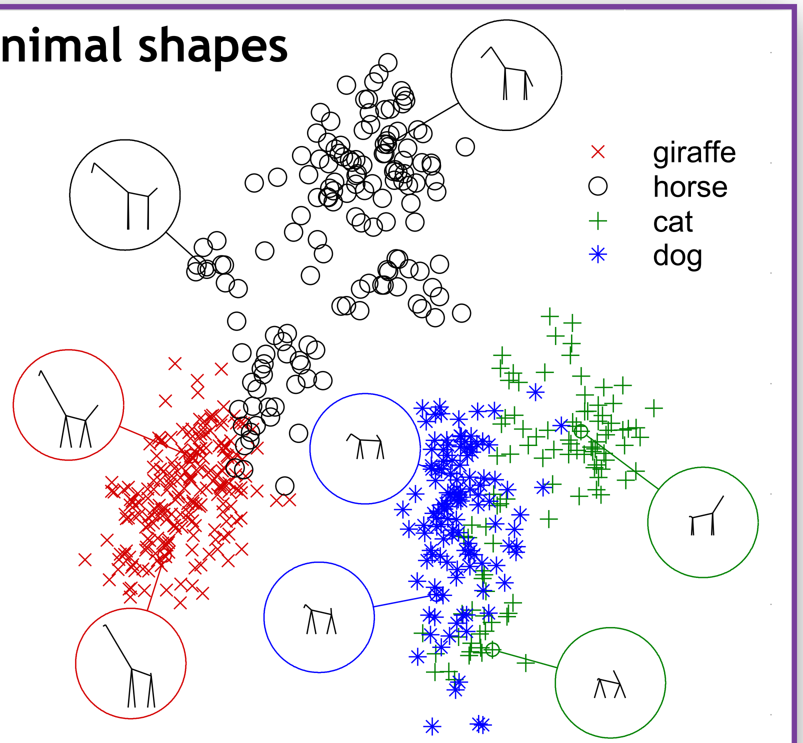
- adapted to 'natural' statistics
- pool data across subjects
- simple scalar measures

box office phone waiting times



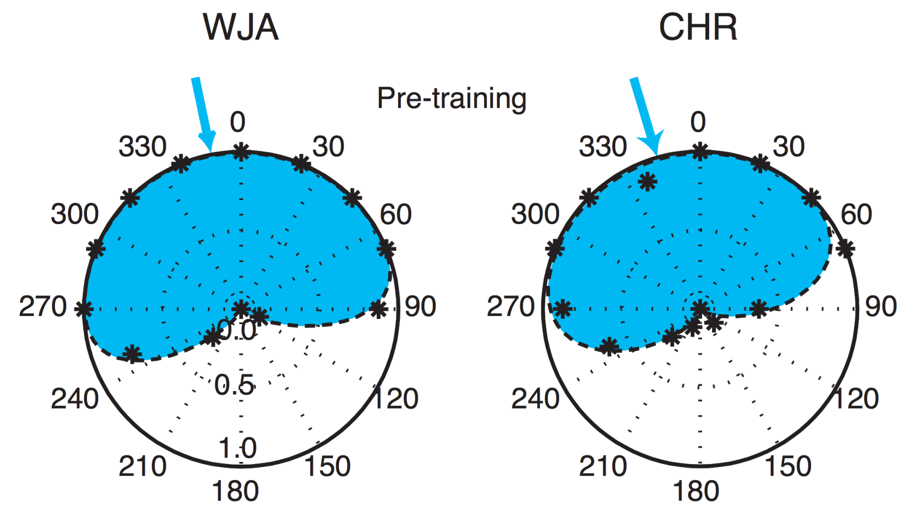
Griffiths & Tenenbaum, 2006

animal shapes

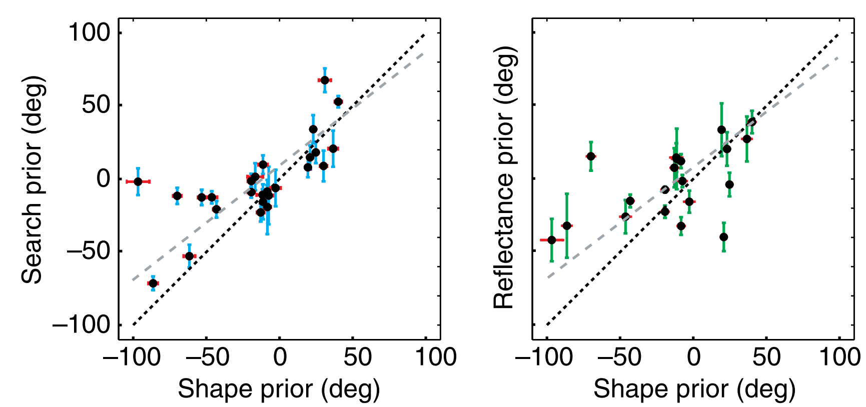


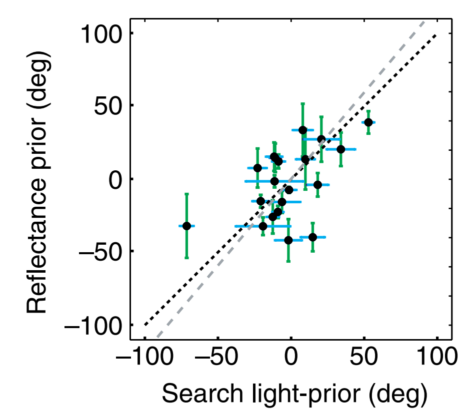
Sanborn & Griffiths, 2008

direction of light



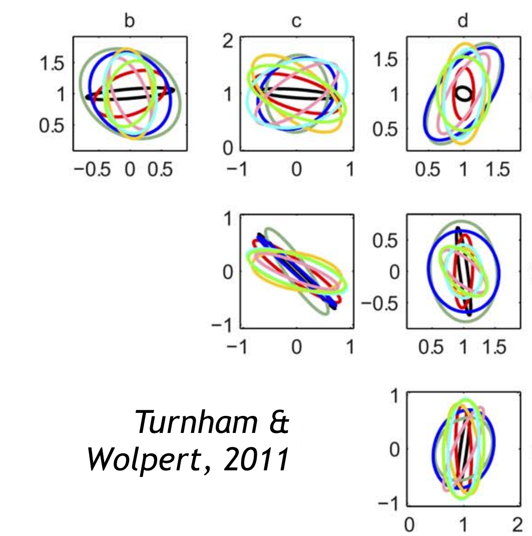
Adams et al, 2004





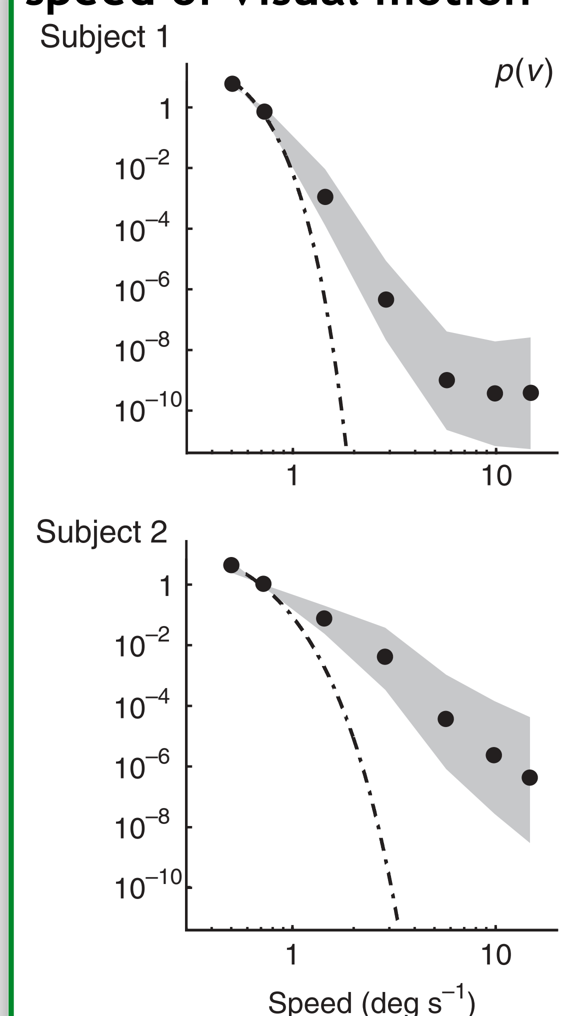
Adams, 2007

visuomotor transformations



Turnham & Wolpert, 2011

speed of visual motion

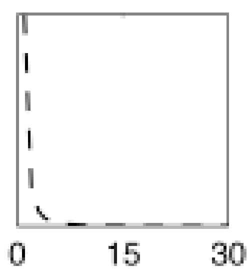


Stocker & Simoncelli, 2006

WHAT ARE NATURAL PRIORS LIKE?

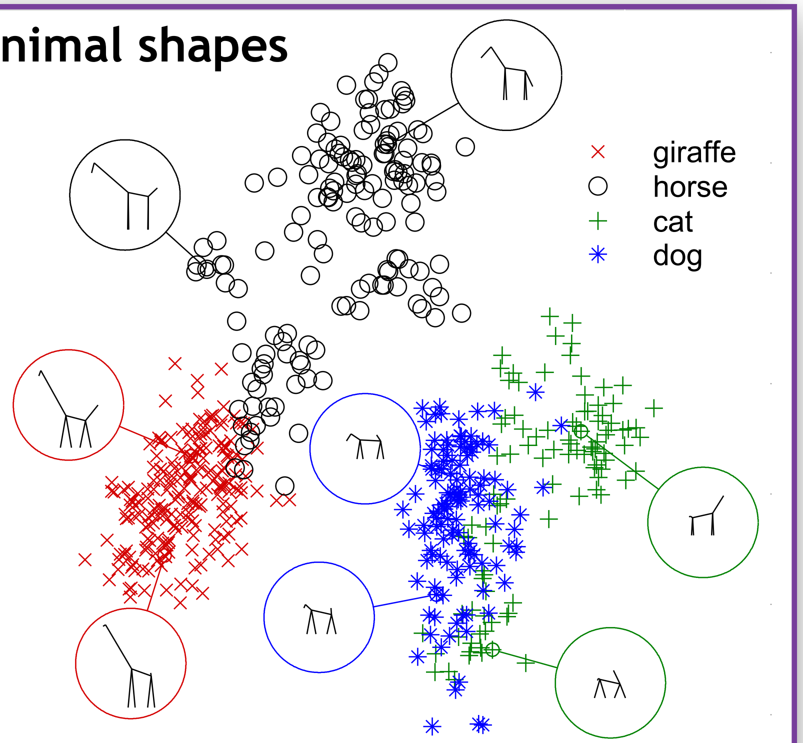
- adapted to ‘natural’ statistics
- pool data across subjects
- simple scalar measures
- simple low-level psychophysics

box office phone waiting times



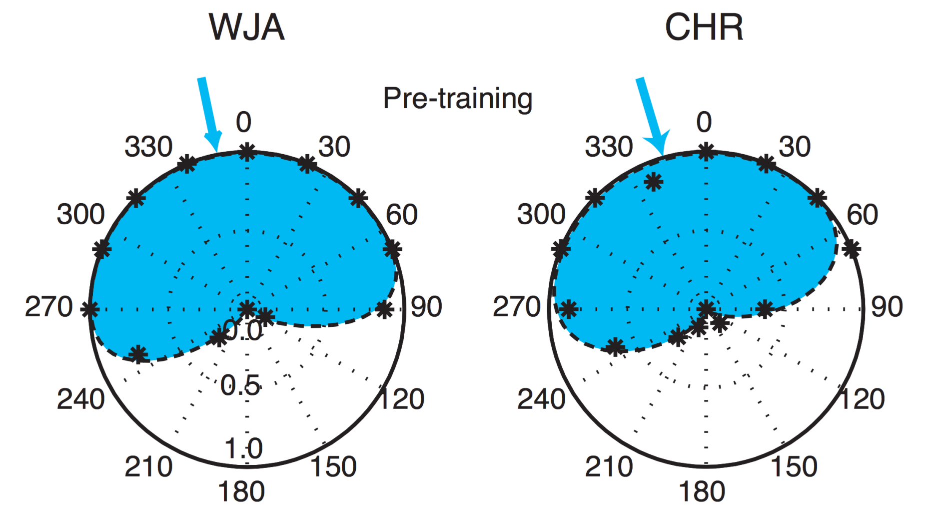
Griffiths & Tenenbaum, 2006

animal shapes

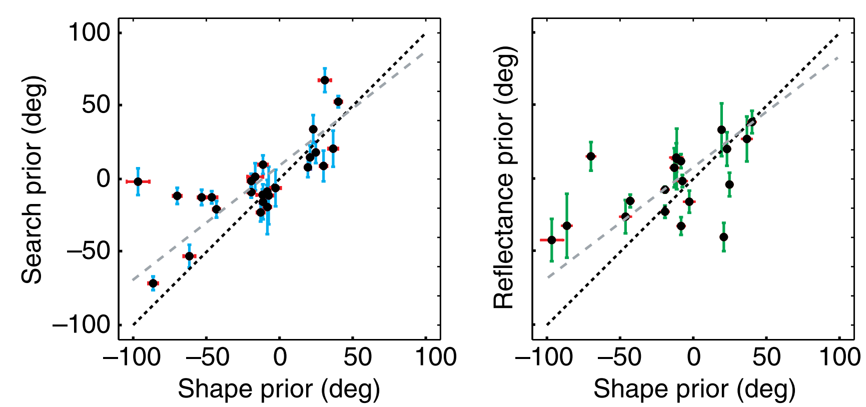


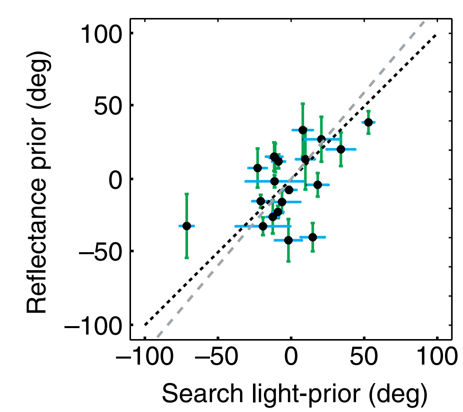
Sanborn & Griffiths, 2008

direction of light



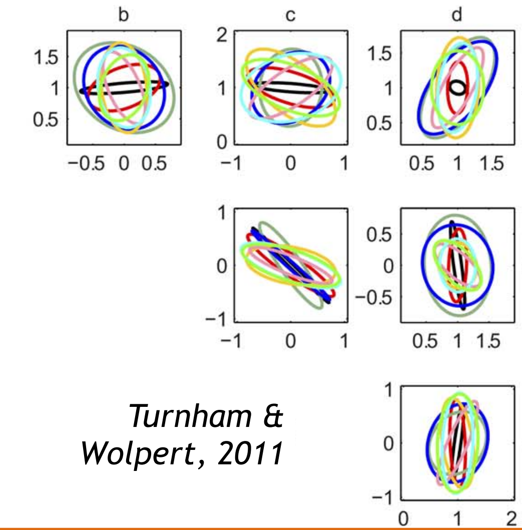
Adams et al, 2004





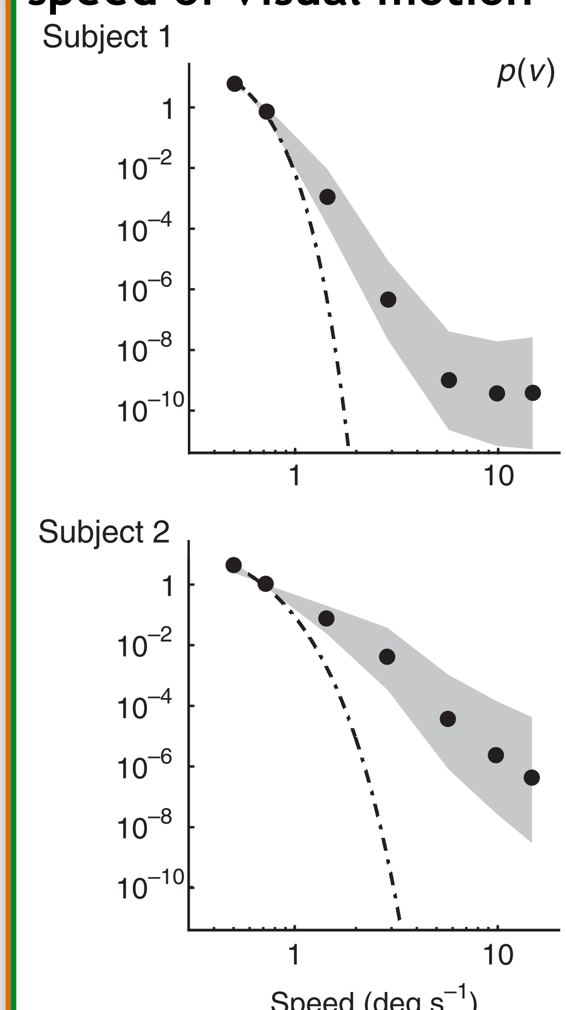
Adams, 2007

visuomotor transformations



Turnham & Wolpert, 2011

speed of visual motion



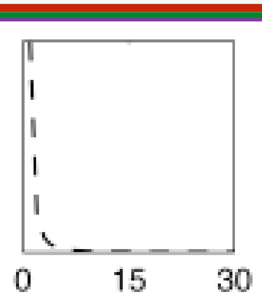
Stocker & Simoncelli, 2006

WHAT ARE NATURAL PRIORS LIKE?

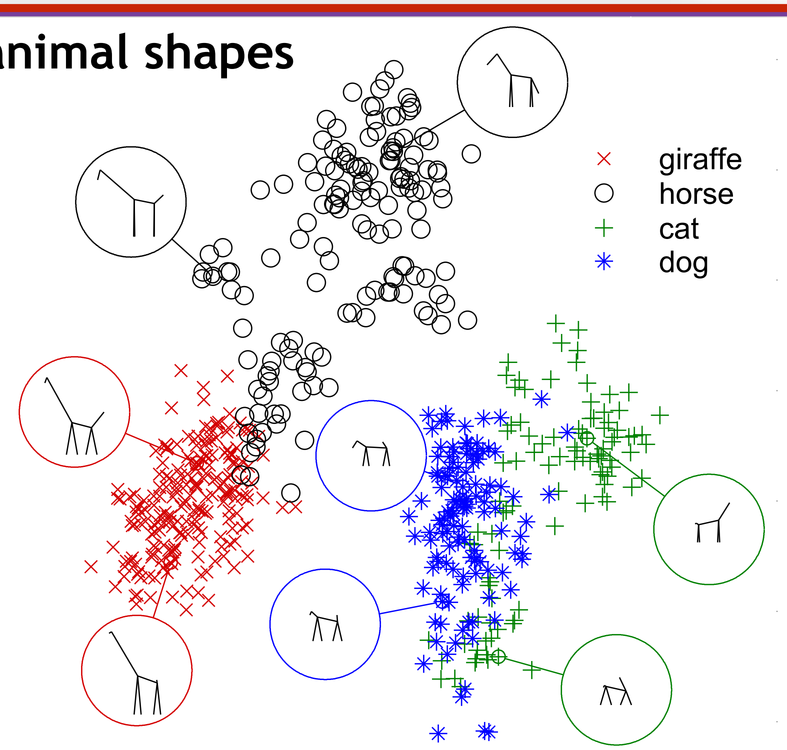
- adapted to ‘natural’ statistics
- pool data across subjects
- simple scalar measures
- simple low-level psychophysics
- limited to one task

box office phone waiting times

Griffiths & Tenenbaum, 2006

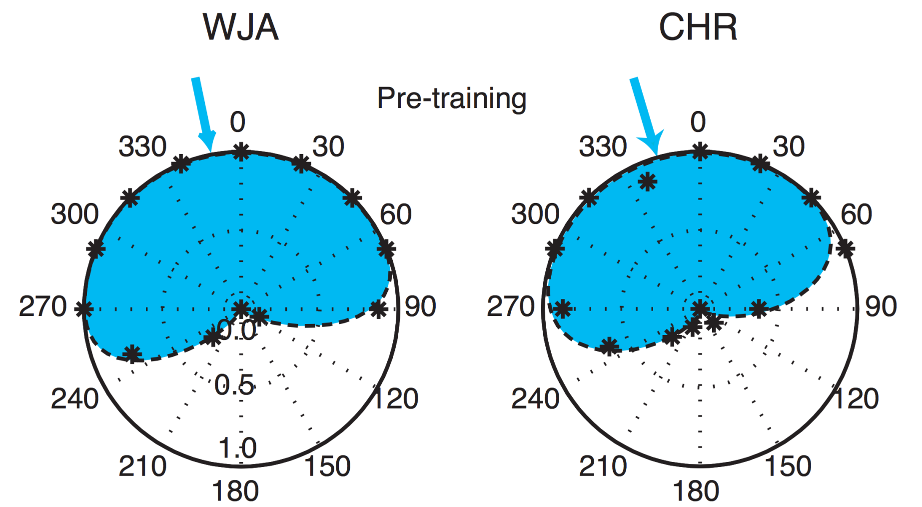


animal shapes

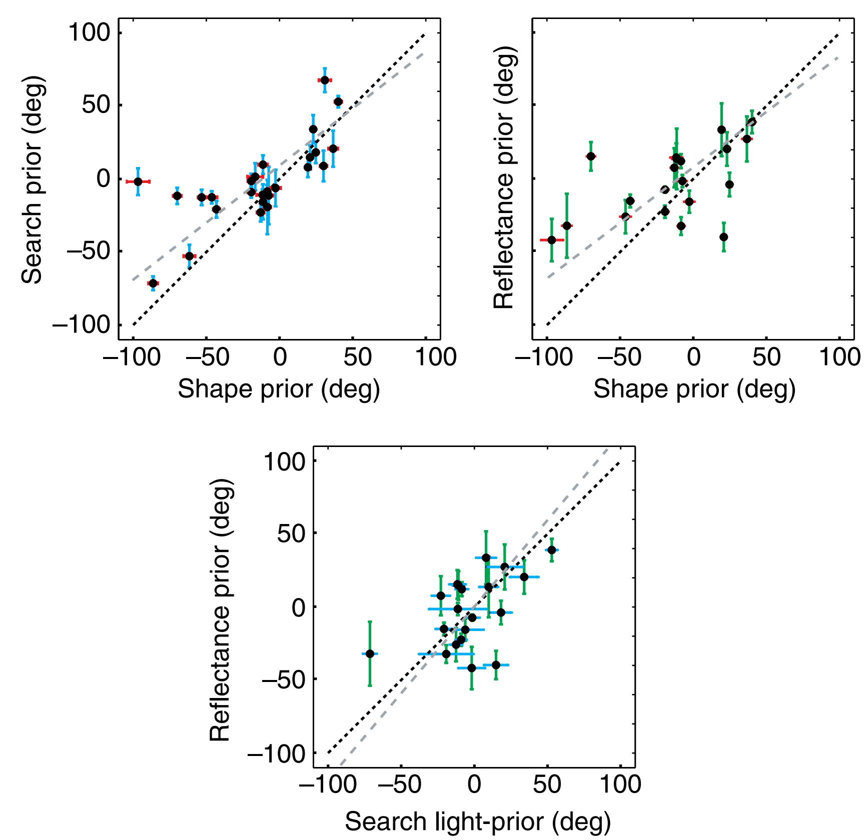


Sanborn & Griffiths, 2008

direction of light

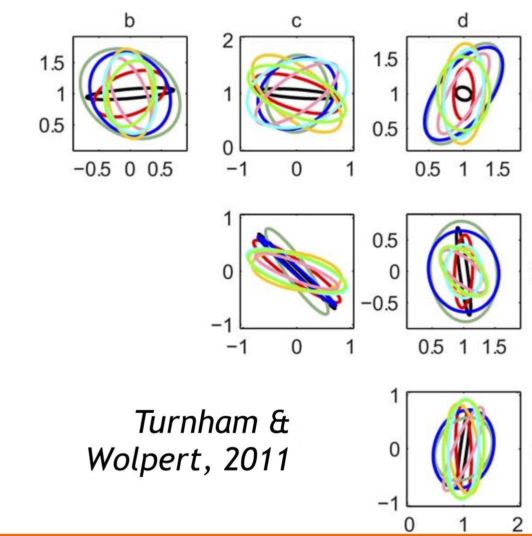


Adams et al, 2004



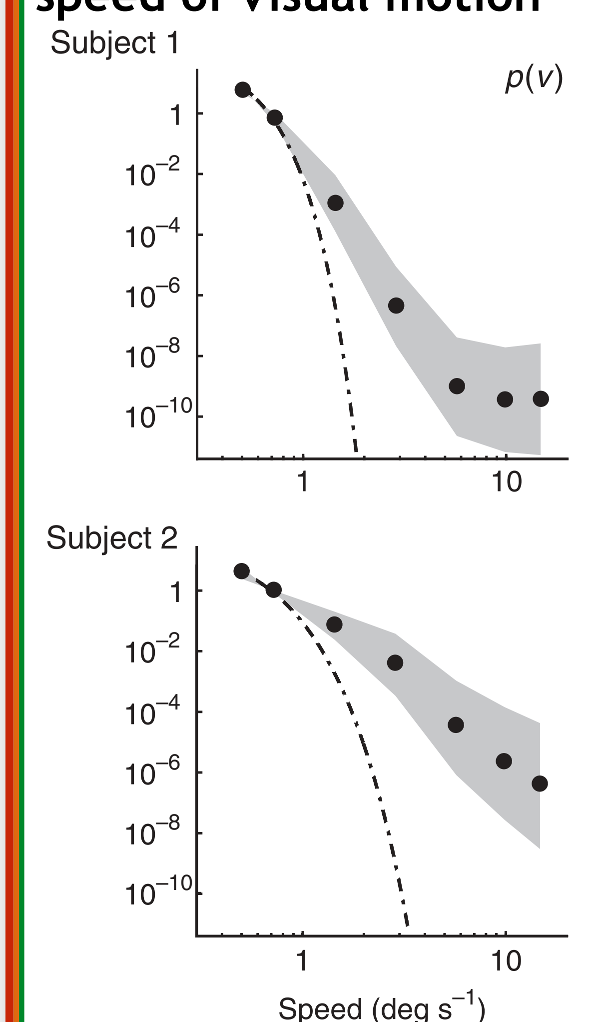
Adams, 2007

visuomotor transformations



Turnham & Wolpert, 2011

speed of visual motion



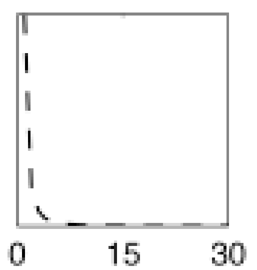
Stocker & Simoncelli, 2006

WHAT ARE NATURAL PRIORS LIKE?

- adapted to 'natural' statistics
- pool data across subjects
- simple scalar measures
- simple low-level psyc
- limited to one task

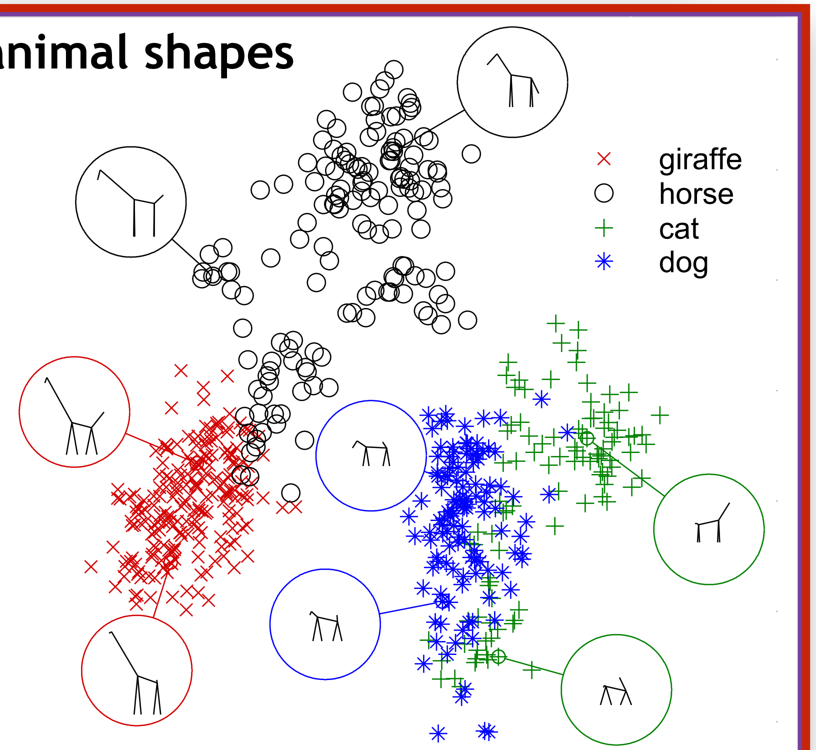
prior = bunch of free parameters → overfitting?

box office phone waiting times



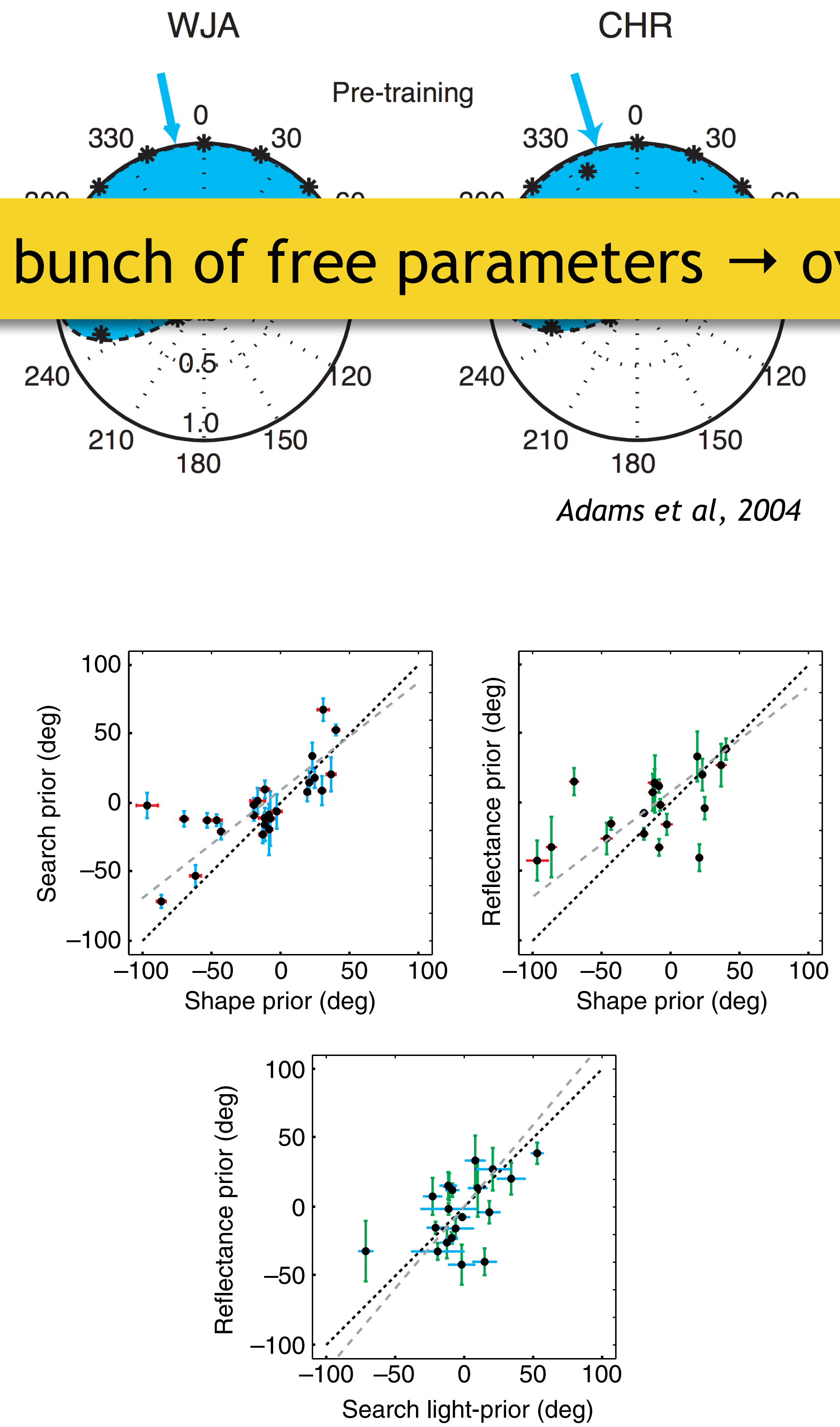
Griffiths & Tenenbaum, 2006

animal shapes



Sanborn & Griffiths, 2008

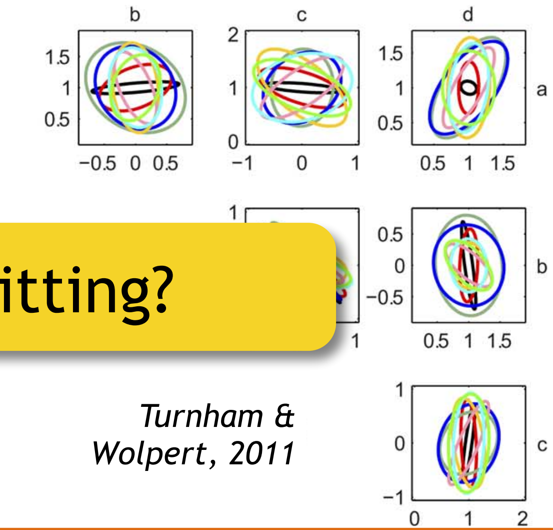
direction of light



Adams et al, 2004

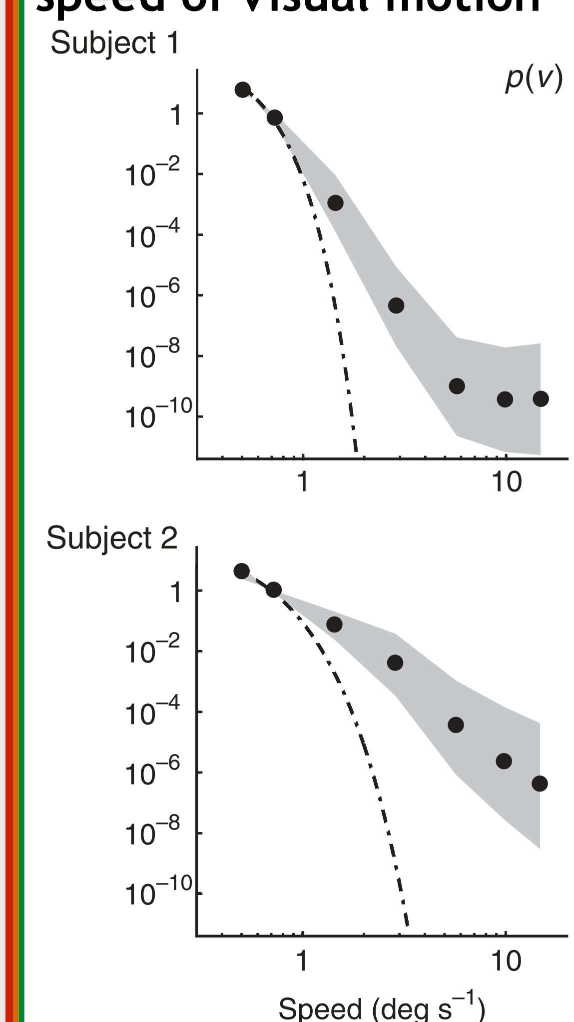
Adams, 2007

visuomotor transformations



Turnham & Wolpert, 2011

speed of visual motion



Stocker & Simoncelli, 2006

WHAT ARE NATURAL PRIORS LIKE?

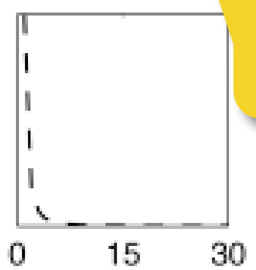
- adapted to ‘natural’ statistics
- pool data across subjects
- simple scalar measures
- simple low-level psyc
- limited to one task

prior = bunch of free parameters → overfitting?

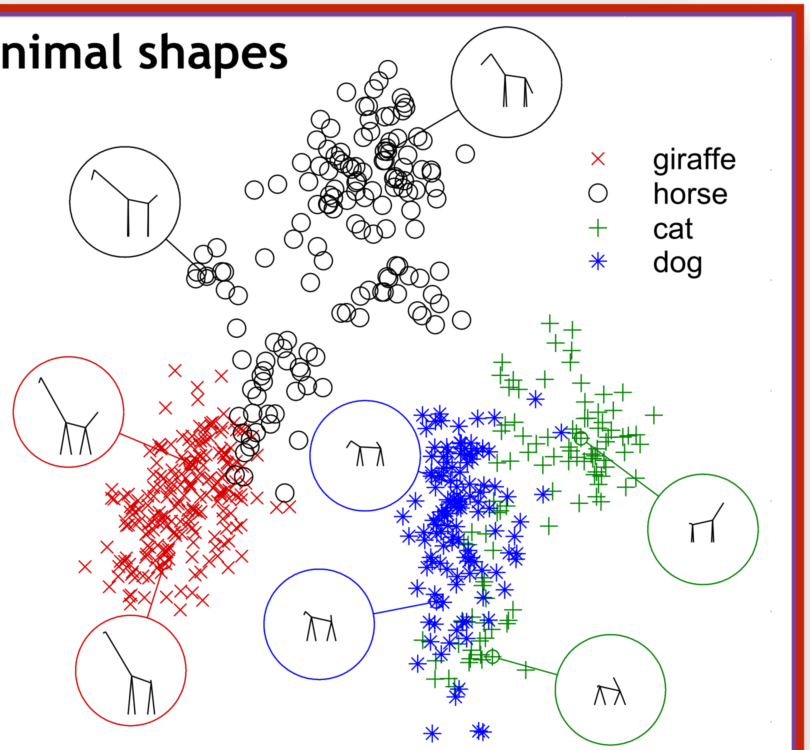
advantage of having a prior: generalisation across tasks

box office phone waiting times

Griffiths & Tenenbaum, 2006

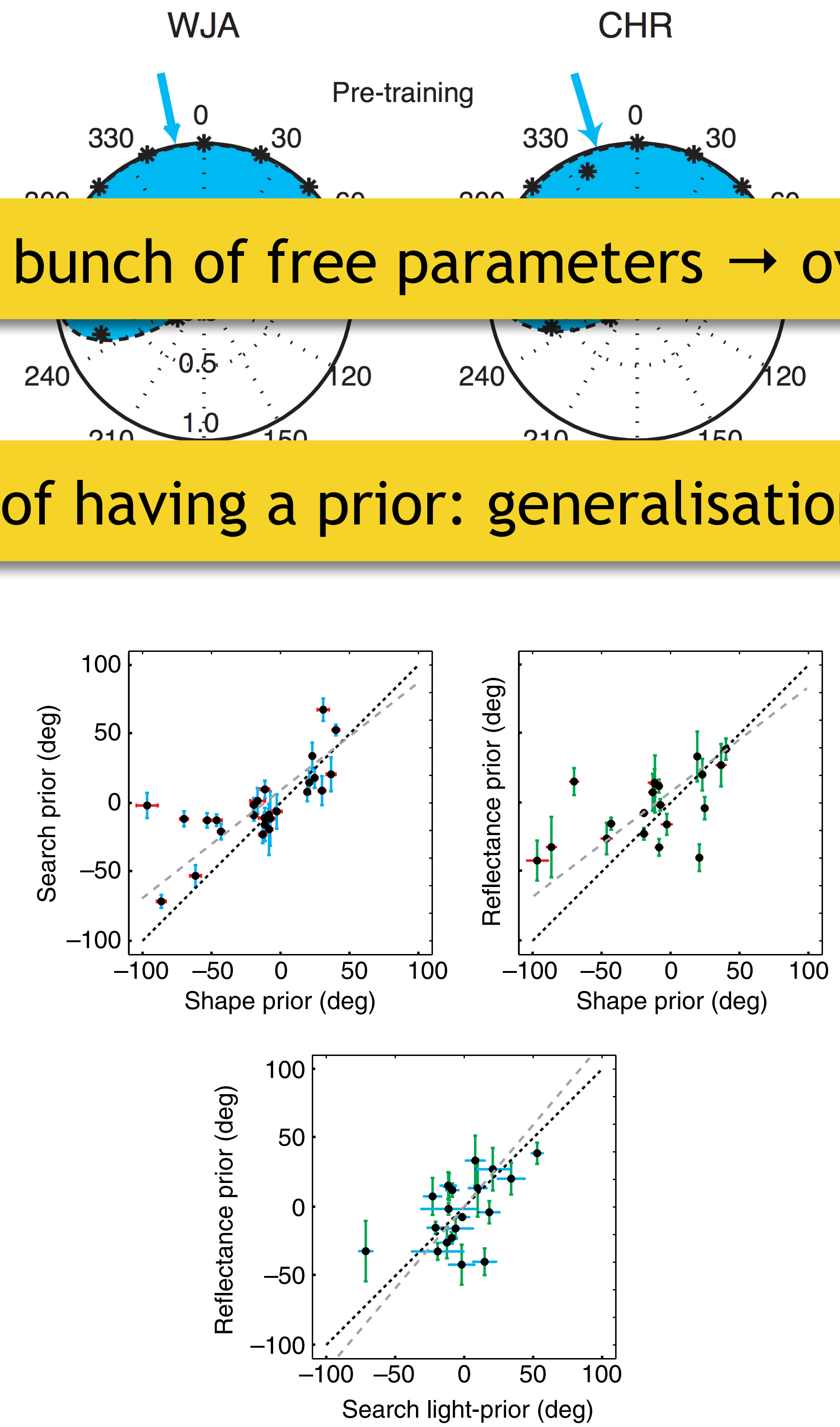


animal shapes



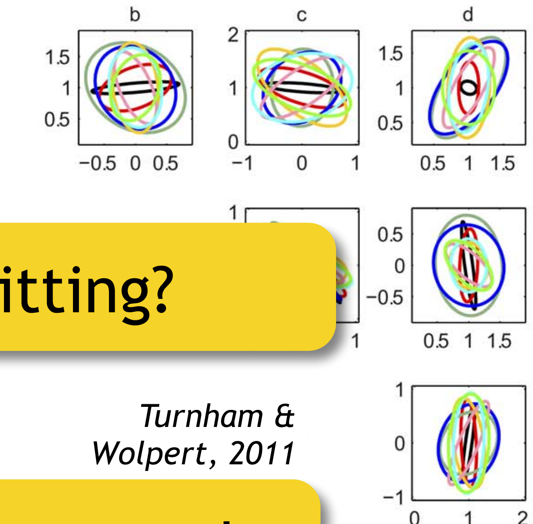
Sanborn & Griffiths, 2008

direction of light



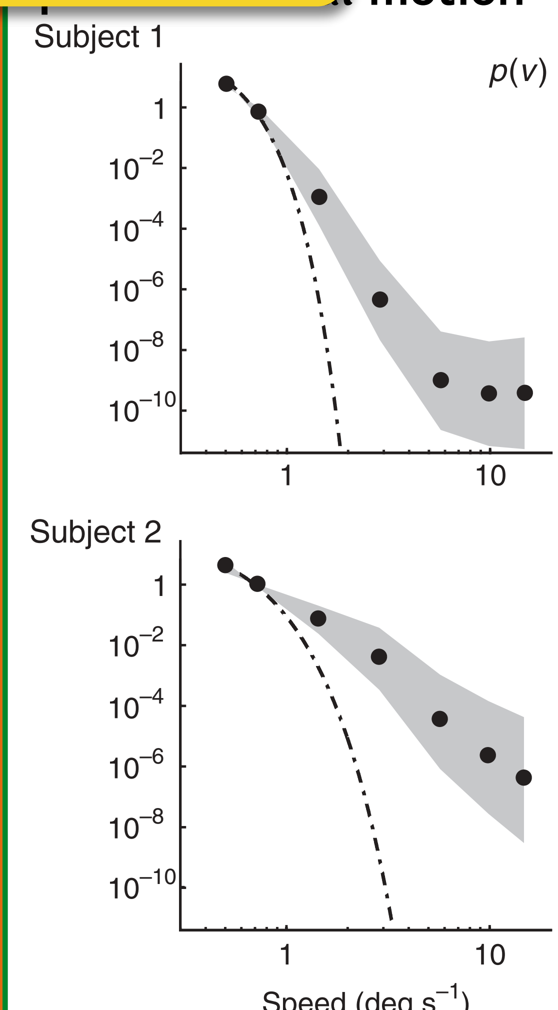
Adams, 2007

visuomotor transformations



Turnham & Wolpert, 2011

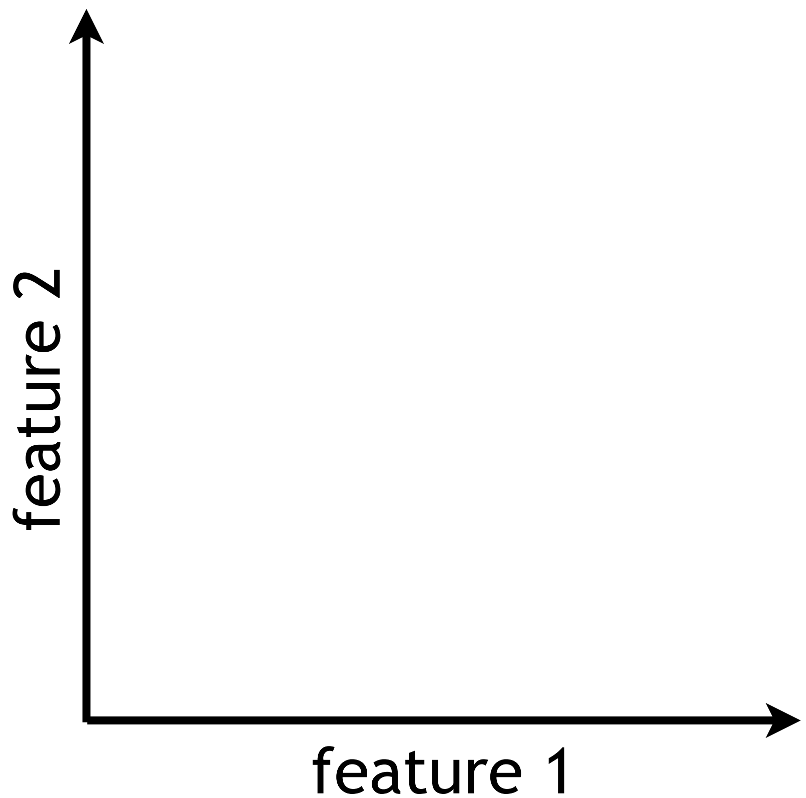
motion



Stocker & Simoncelli, 2006

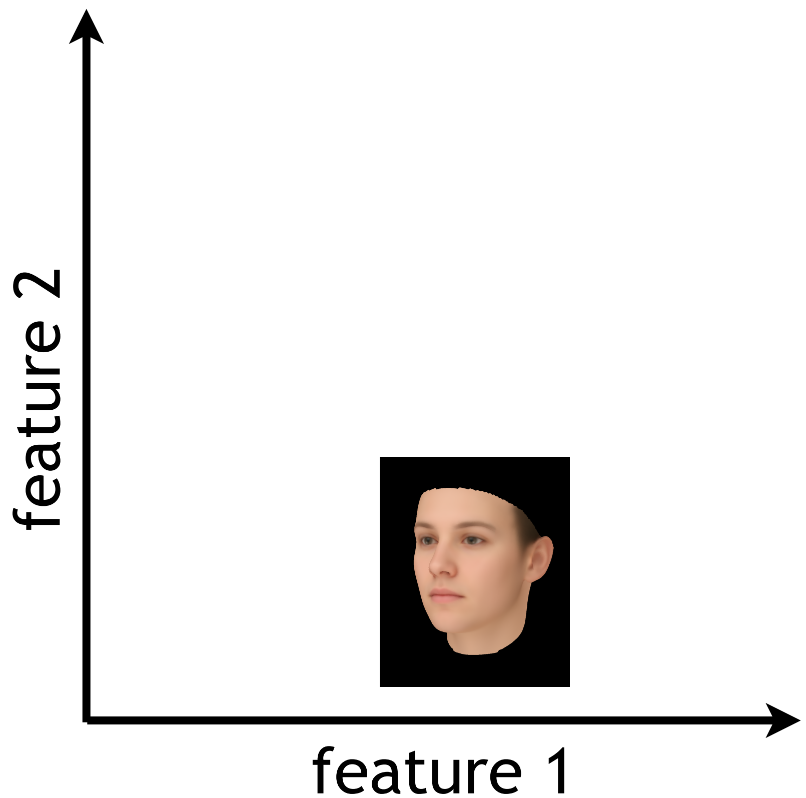
DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE



DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE

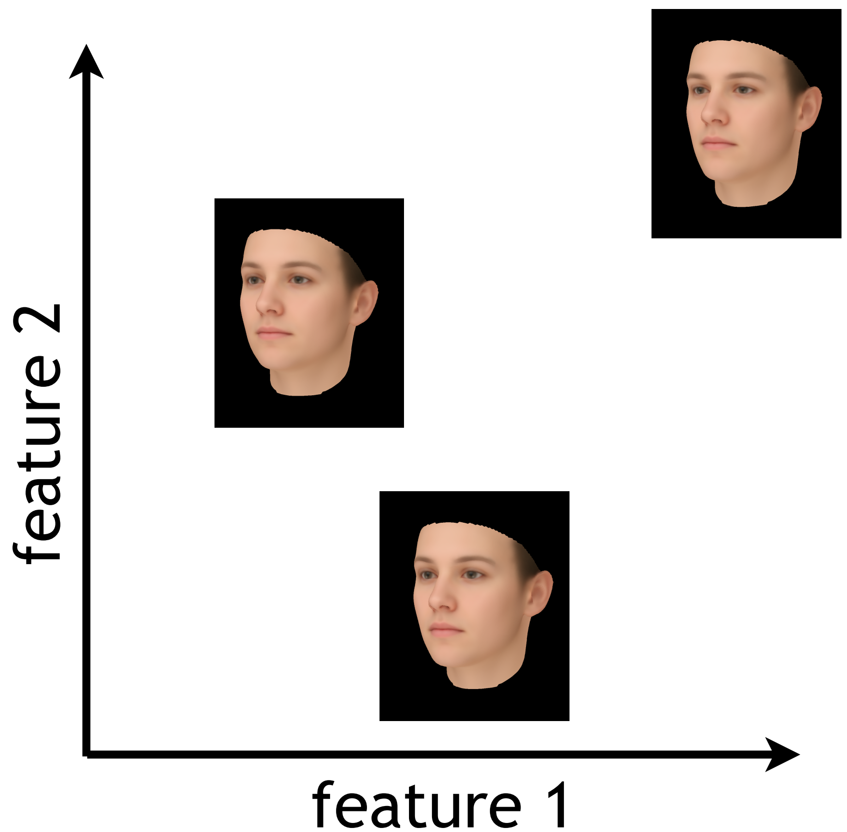


DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE

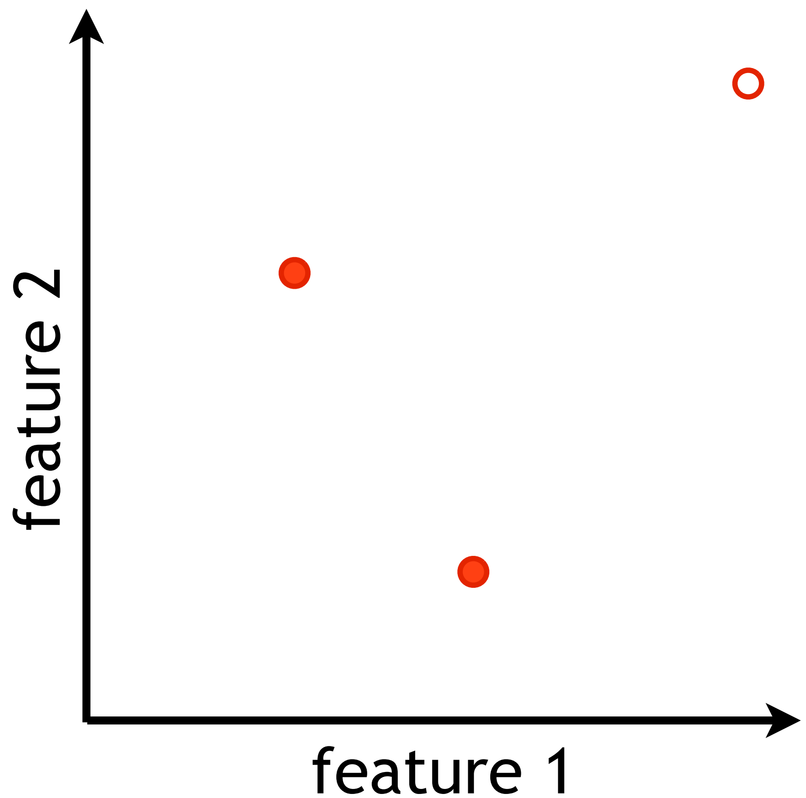
DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE



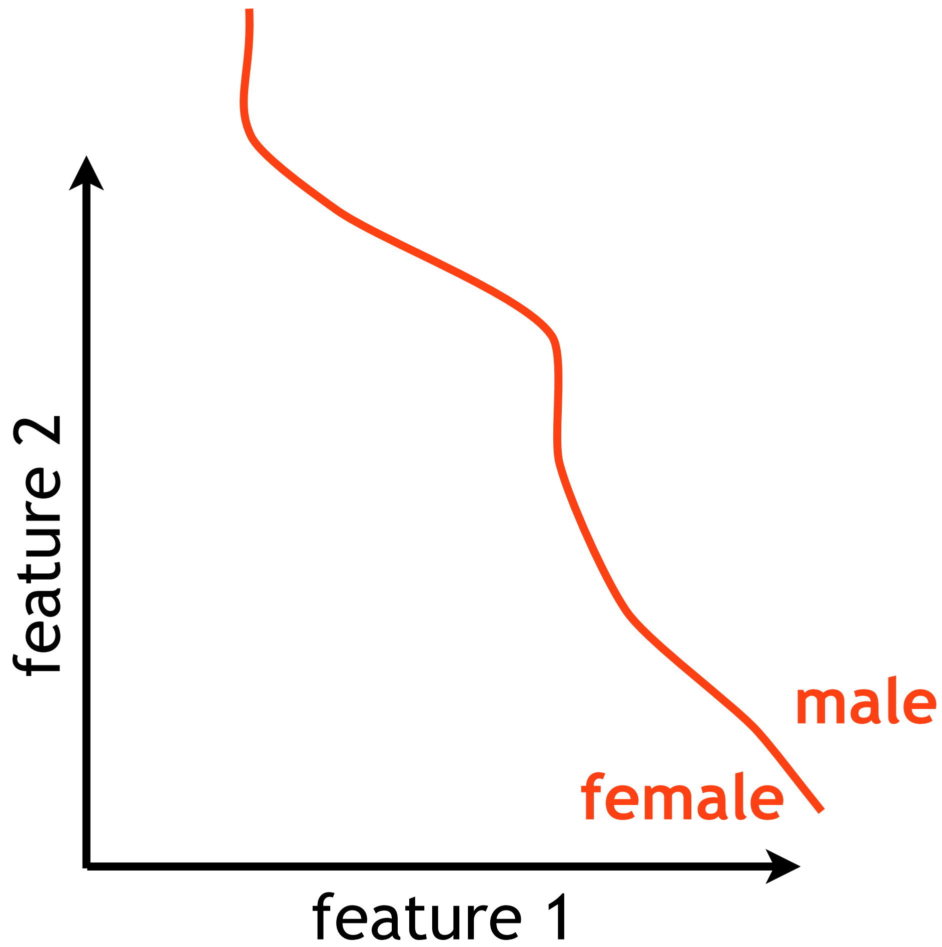
DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE



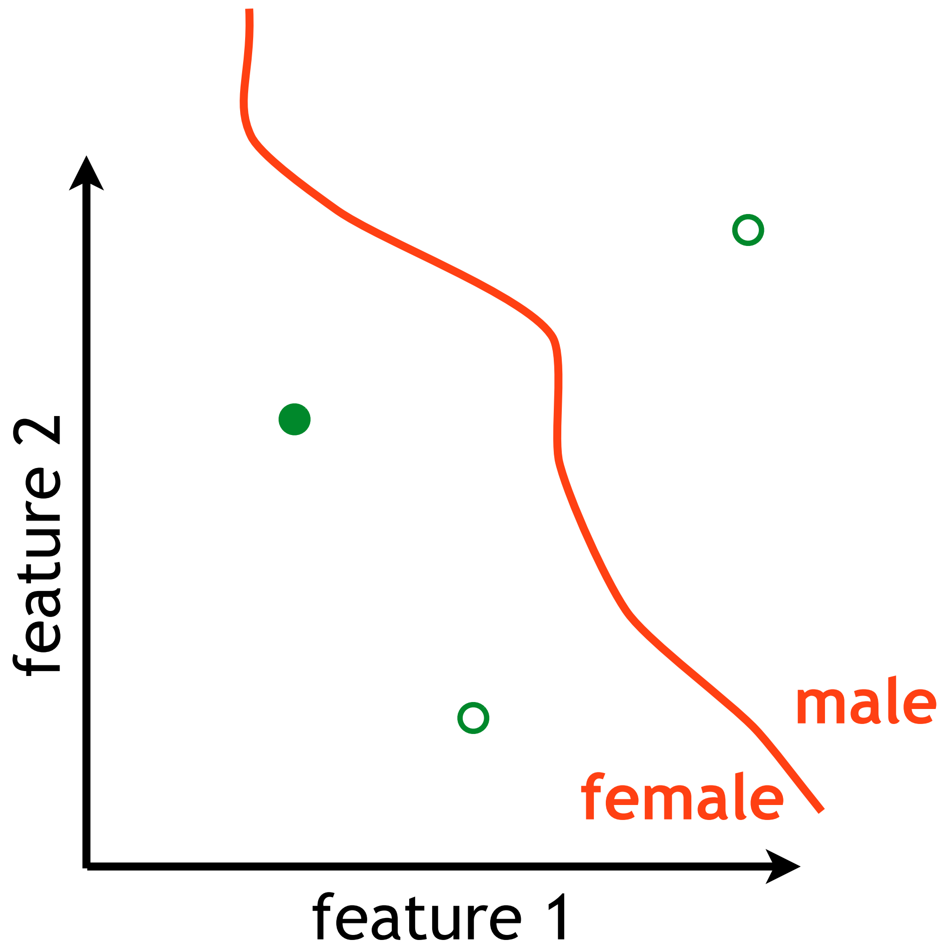
DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE



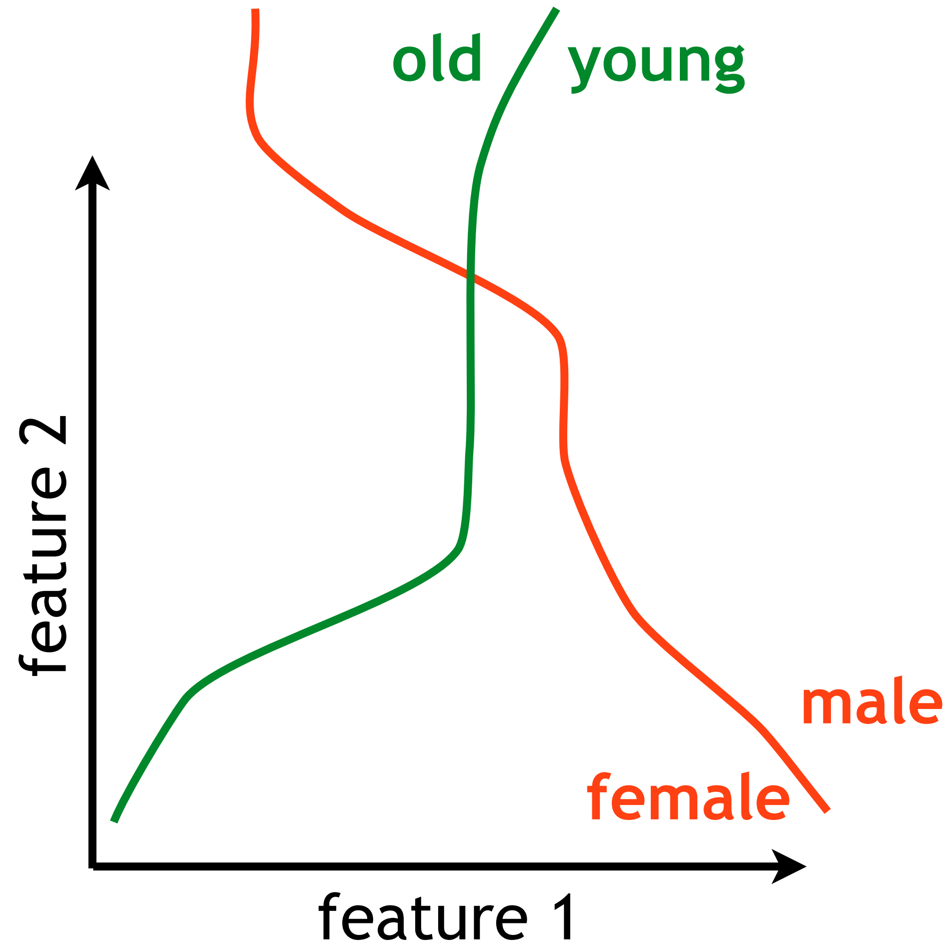
DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE



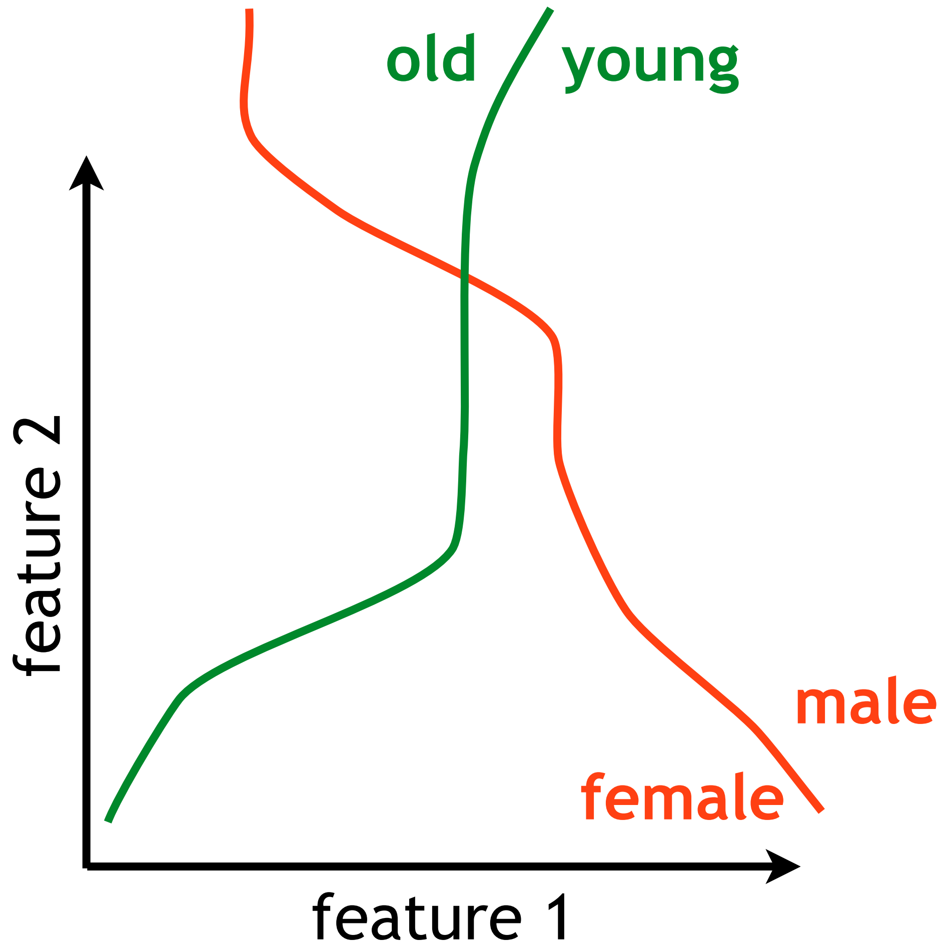
DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE

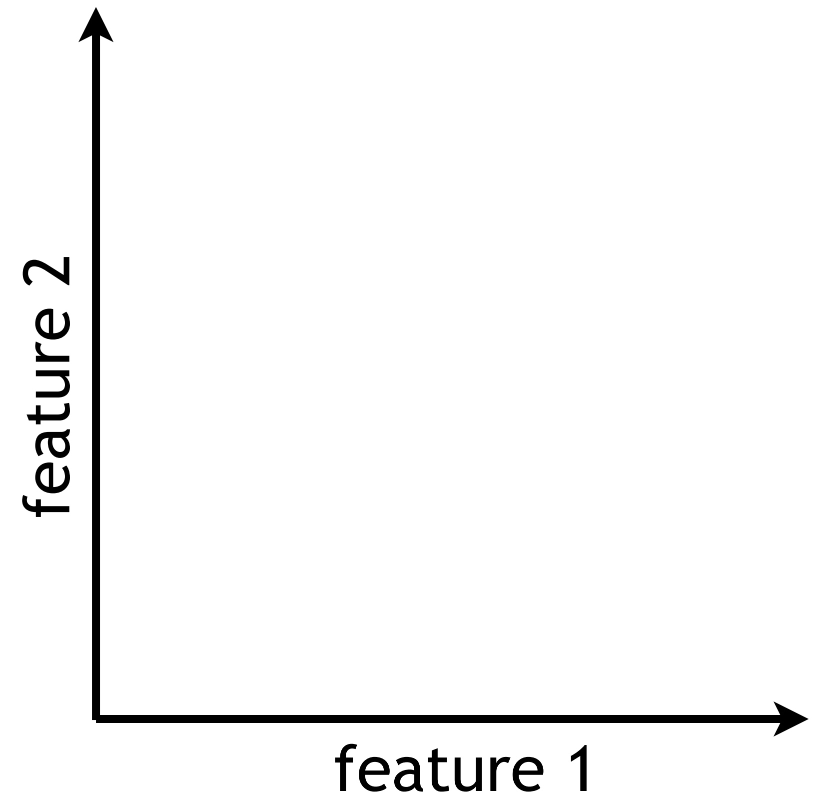


DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE

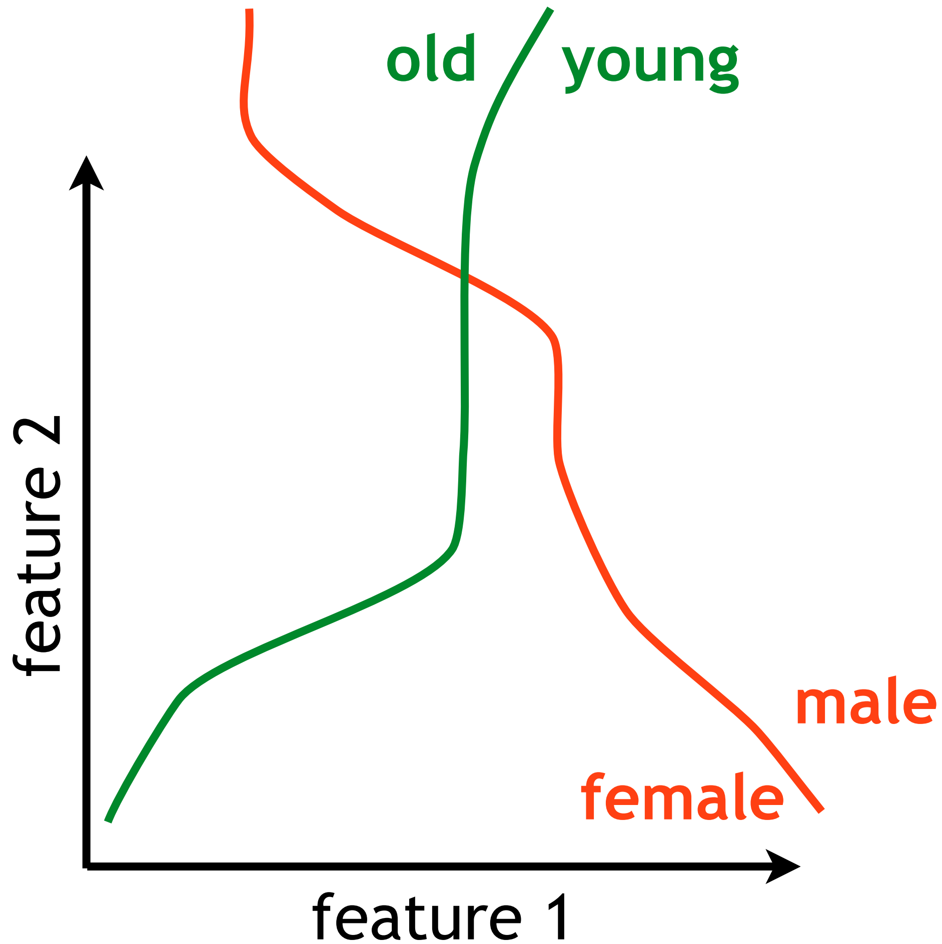


GENERATIVE

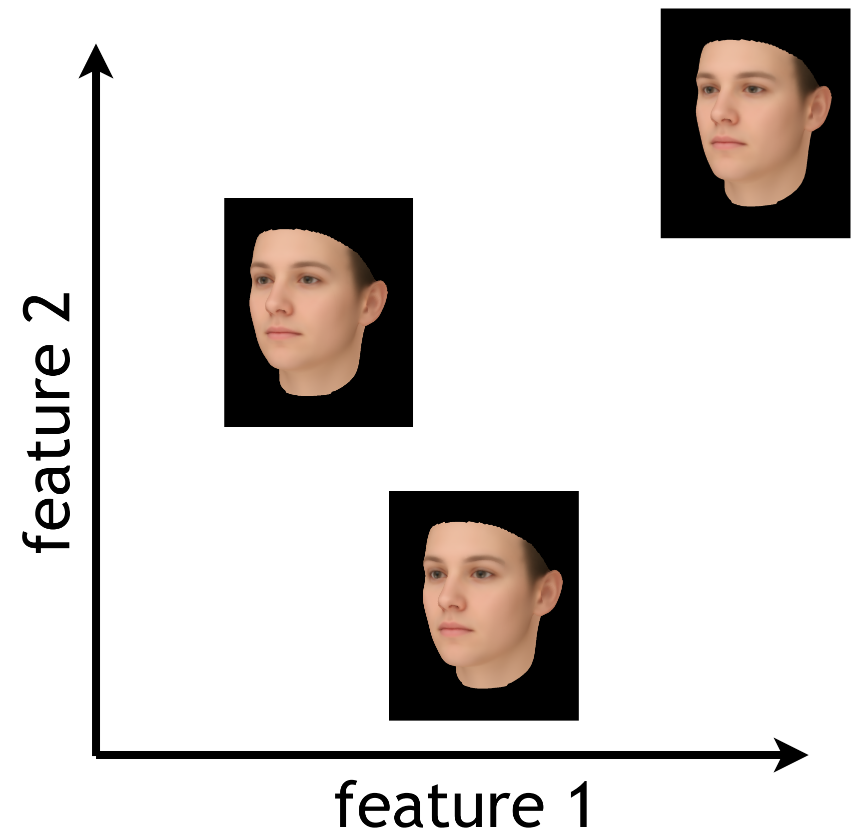


DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE

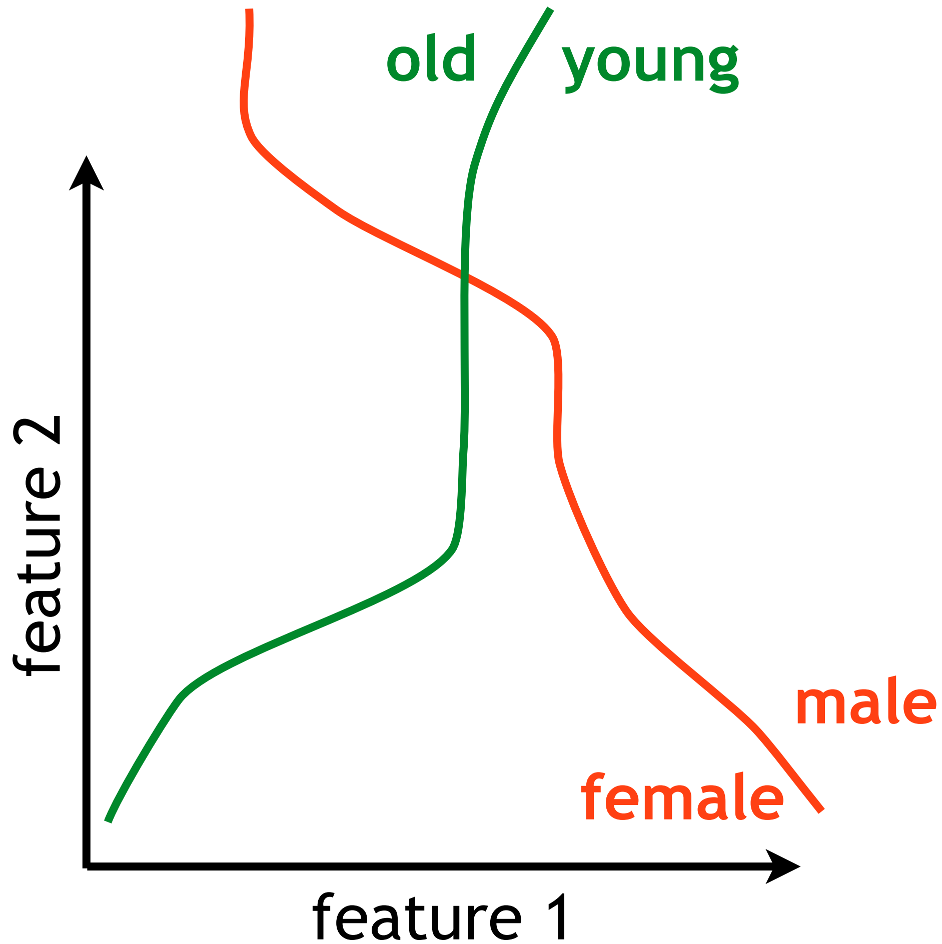


GENERATIVE

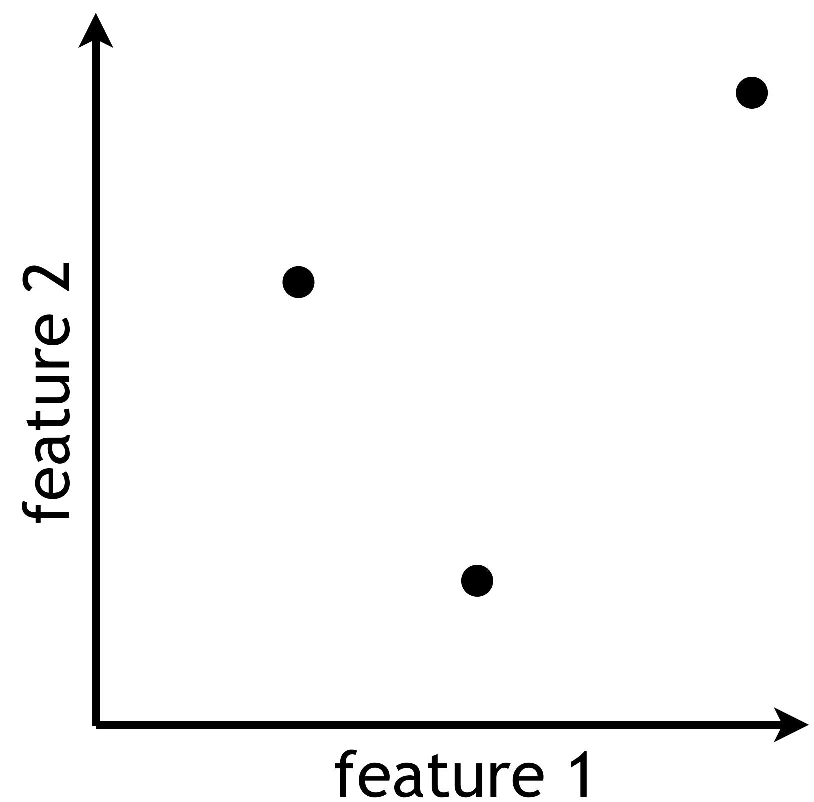


DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE

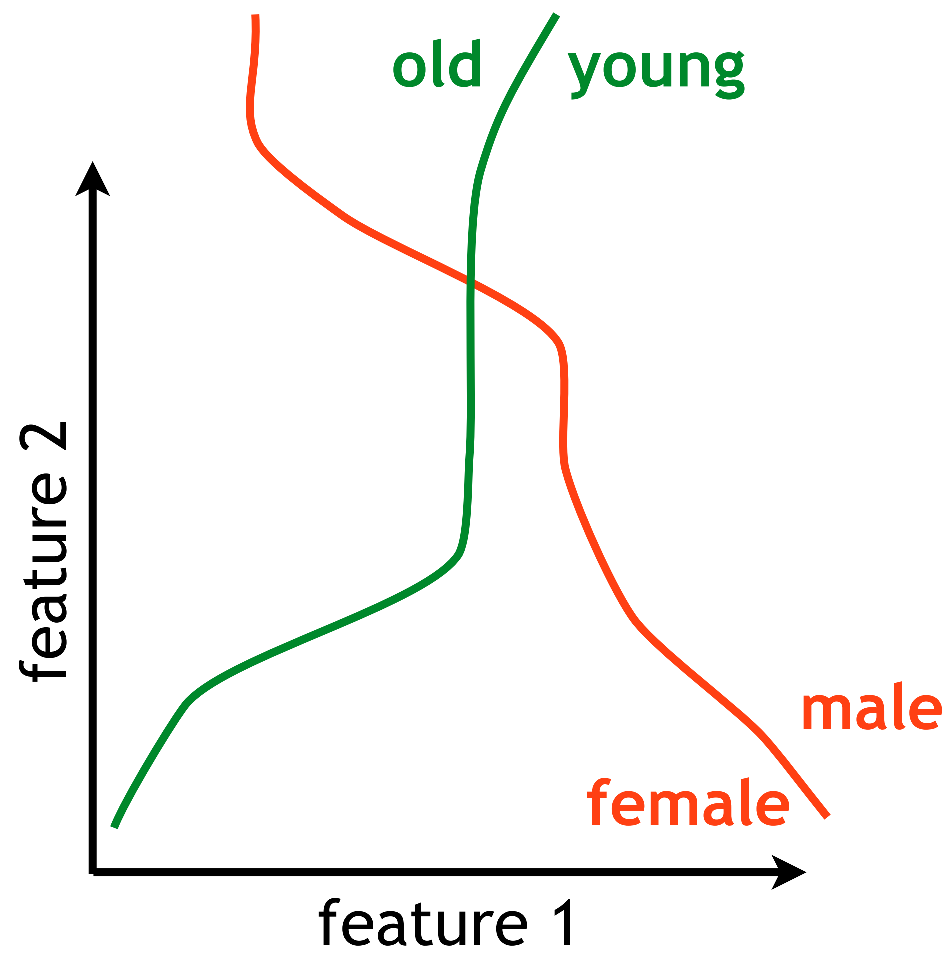


GENERATIVE

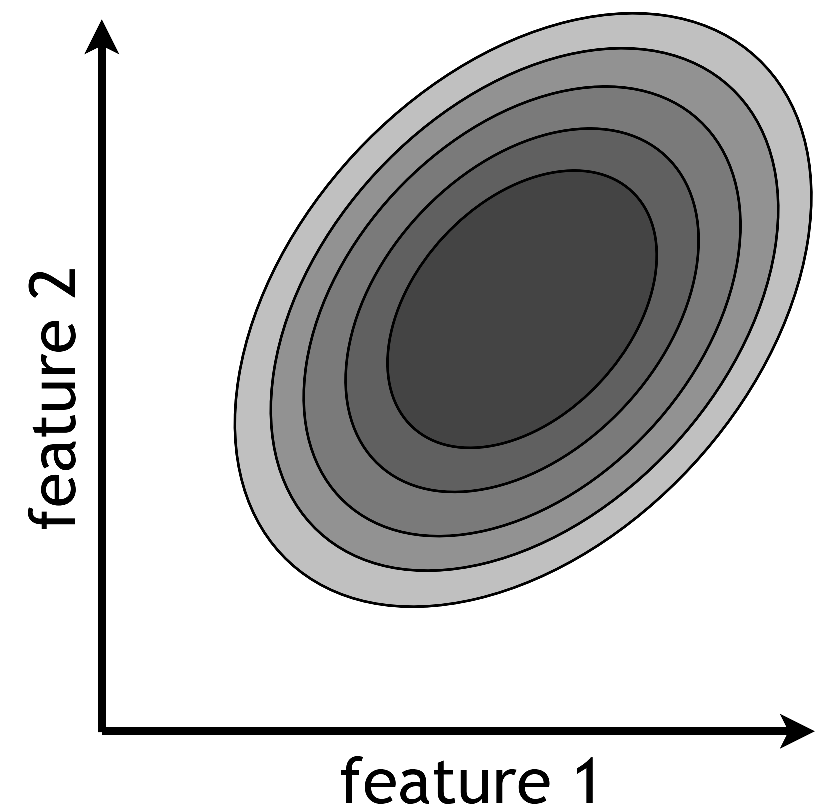


DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE



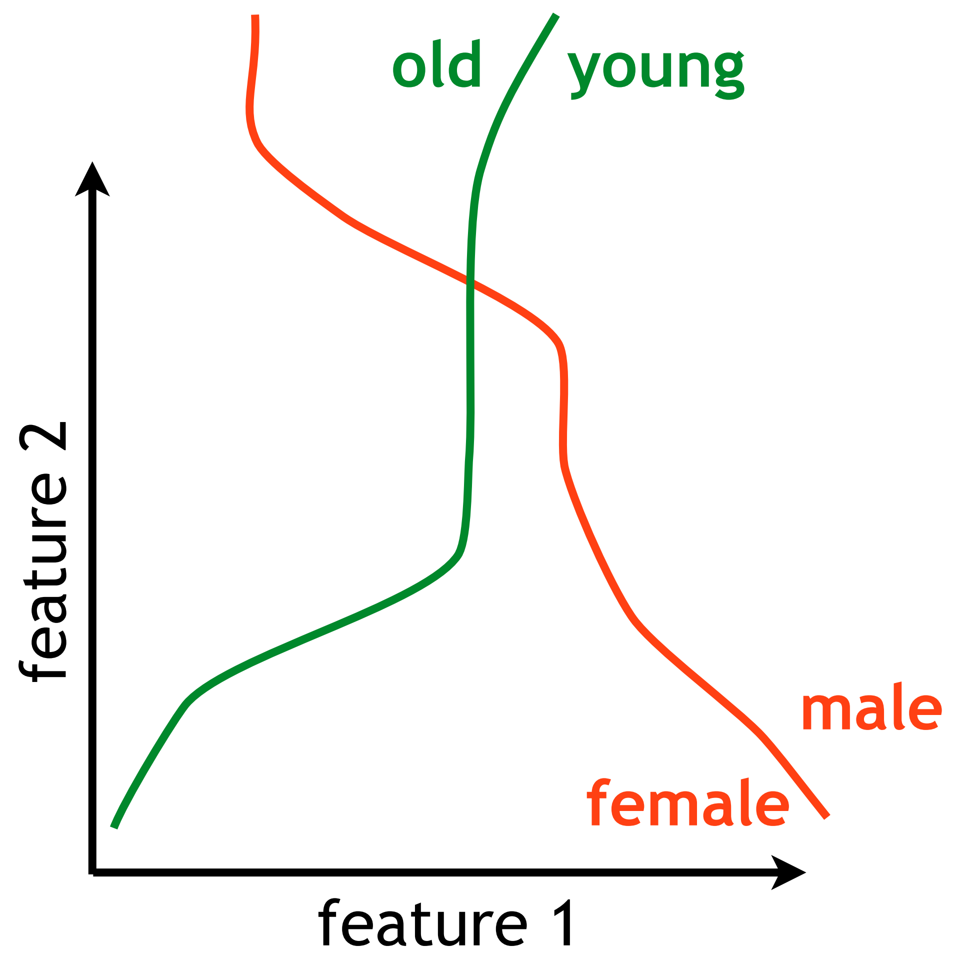
GENERATIVE



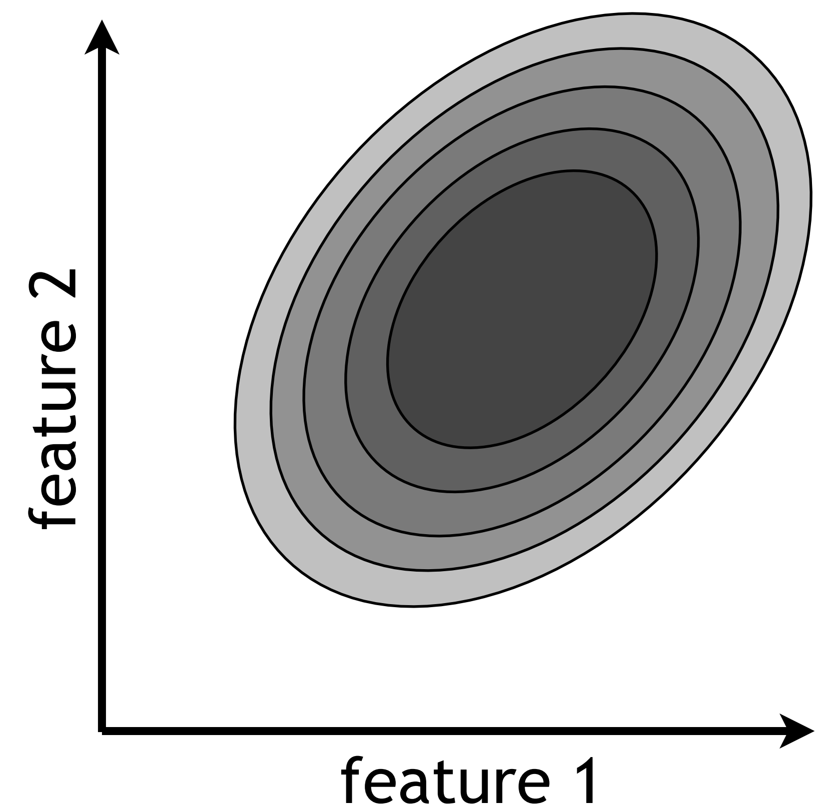
DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE

task-dependent representation

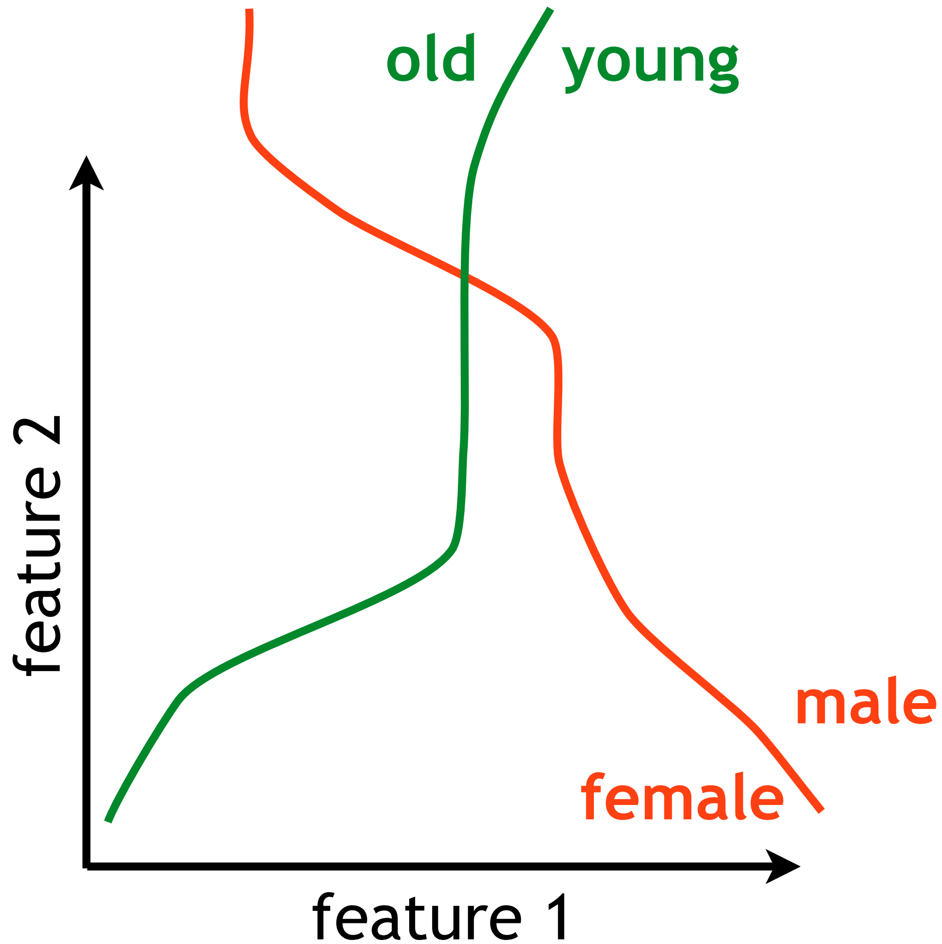


GENERATIVE

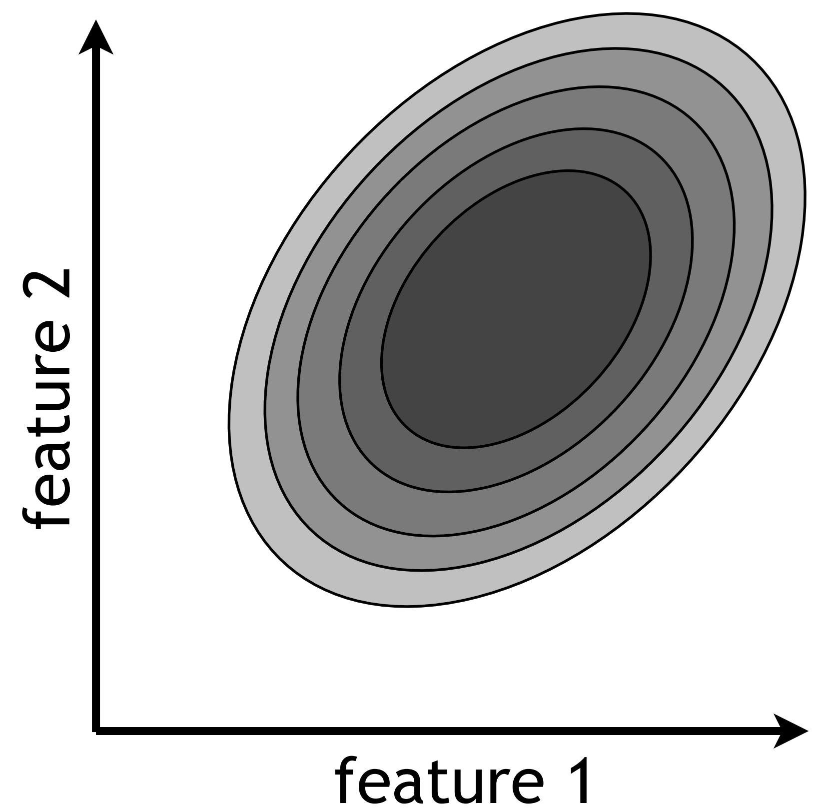


DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE
task-dependent representation



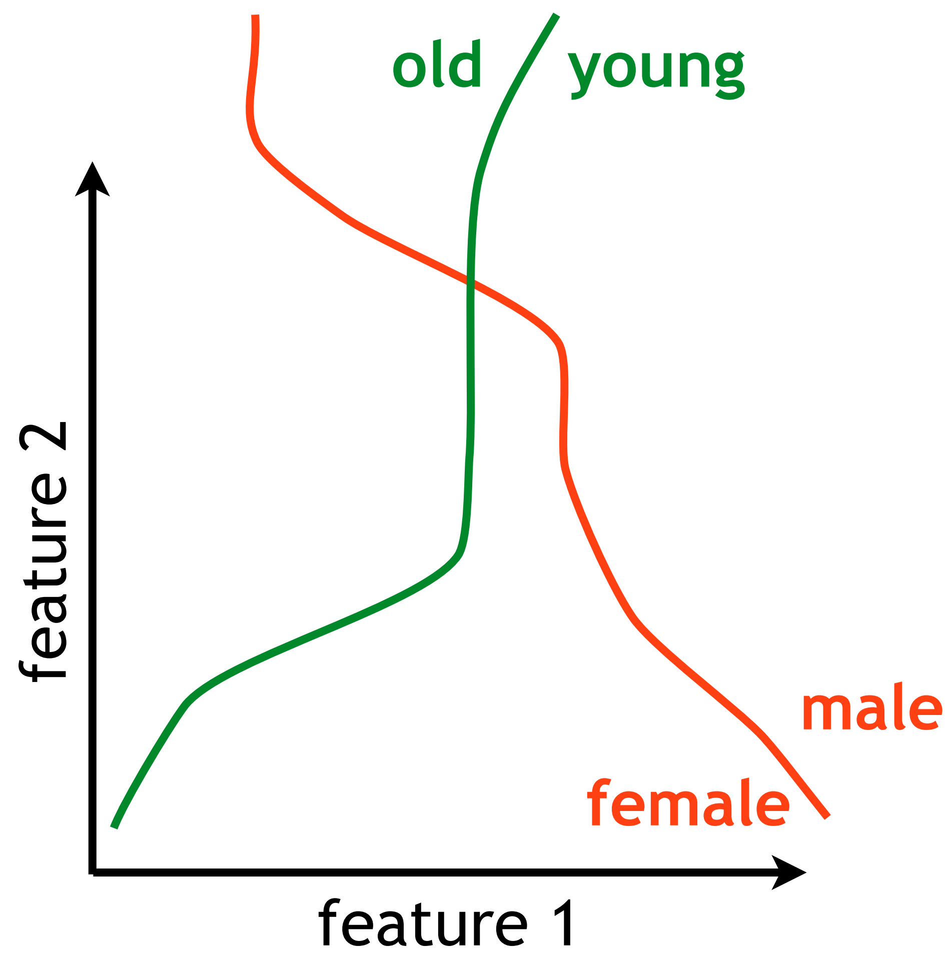
GENERATIVE
task-independent representation



DISCRIMINATIVE VS. GENERATIVE LEARNING

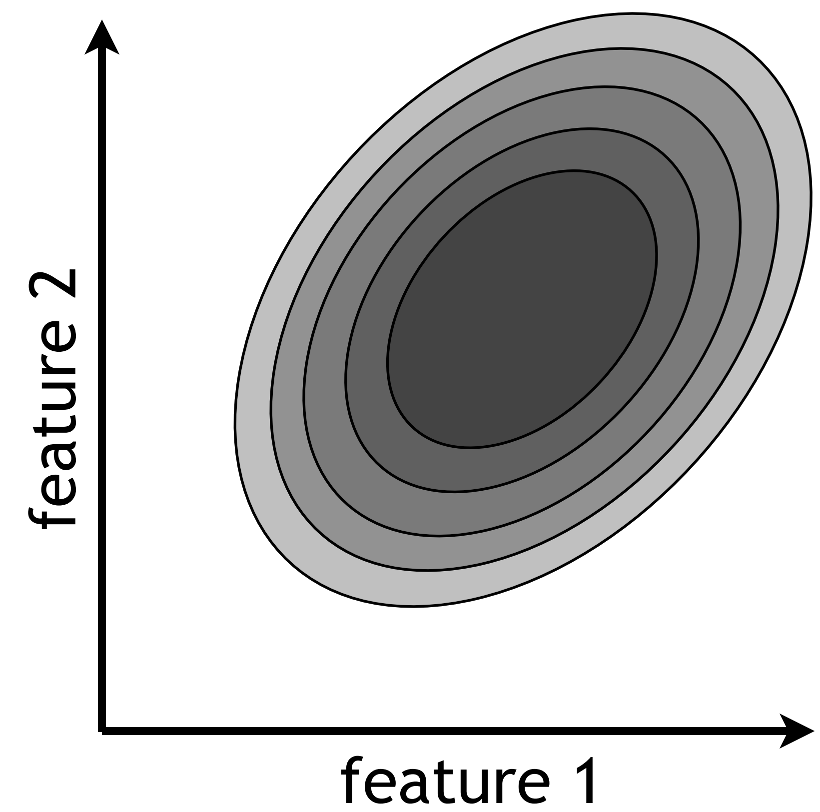
DISCRIMINATIVE

task-dependent representation



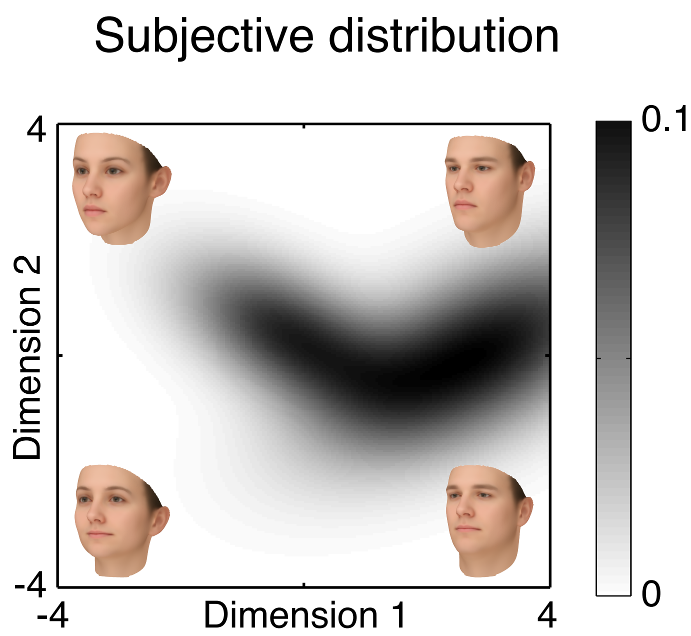
GENERATIVE

task-independent representation

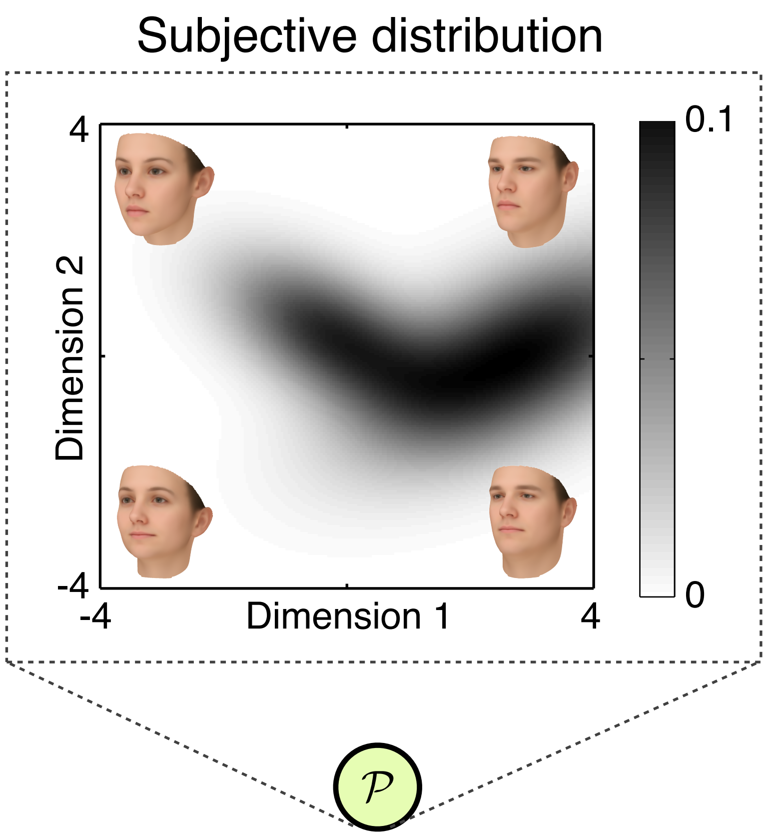


COGNITIVE TOMOGRAPHY

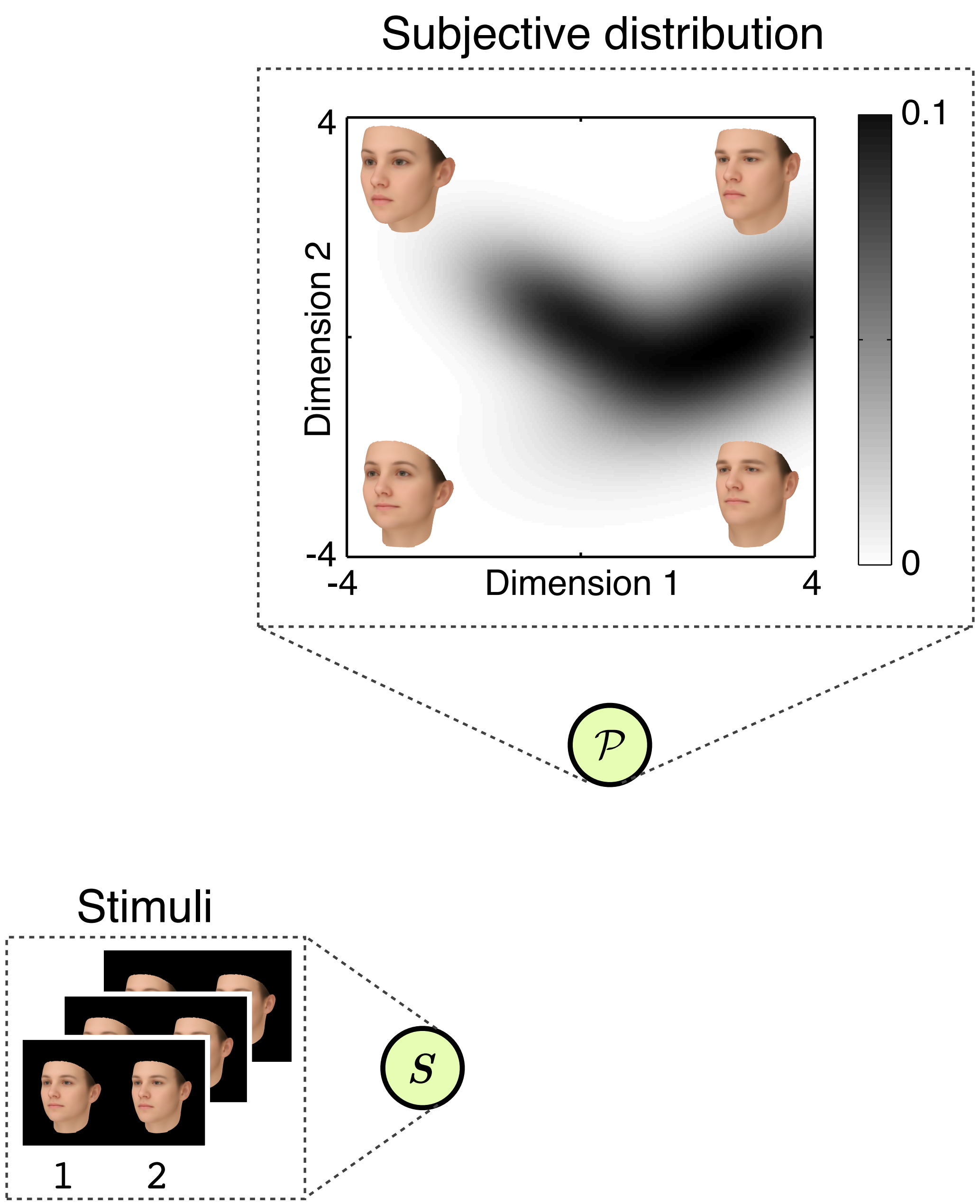
COGNITIVE TOMOGRAPHY



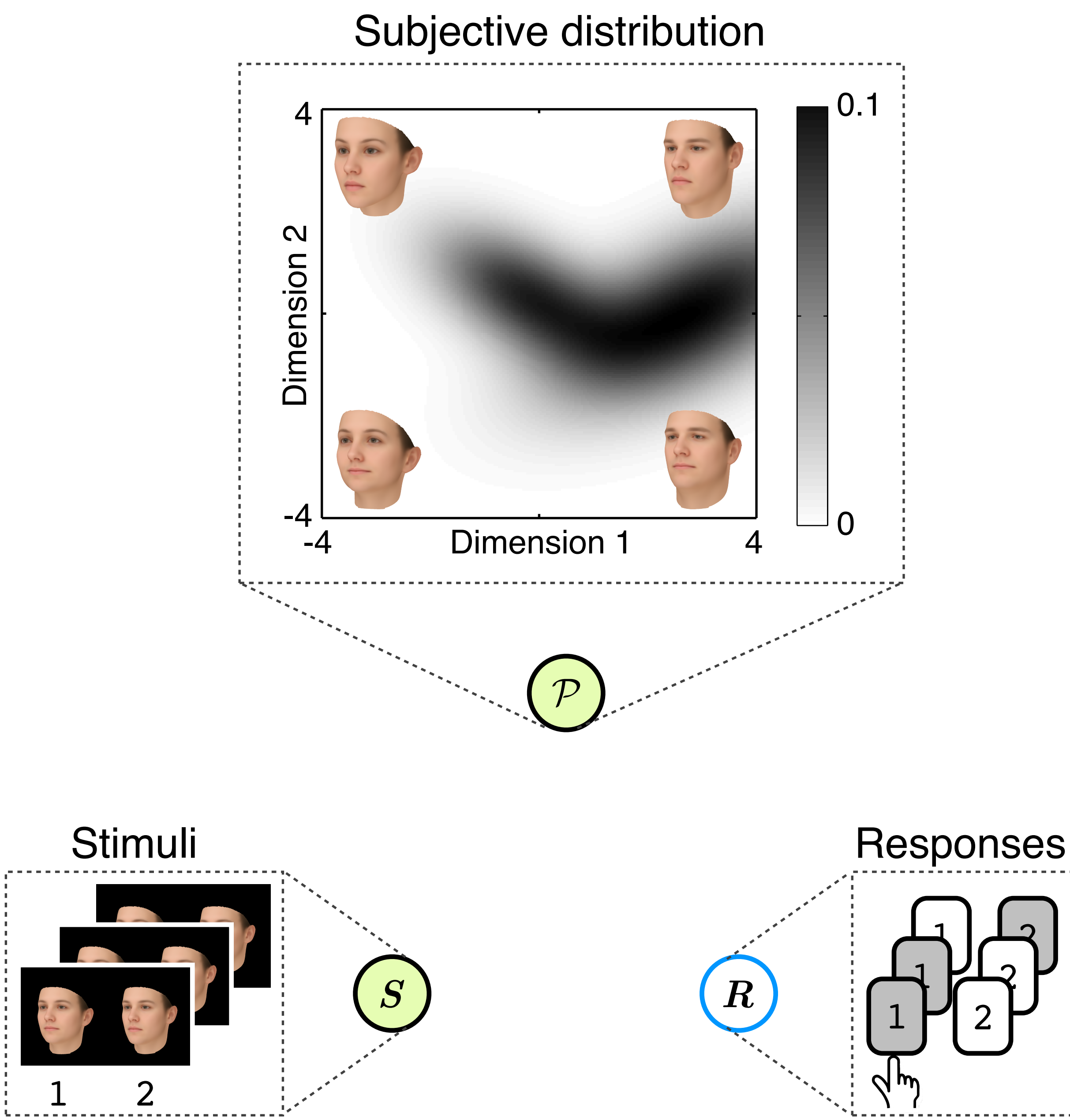
COGNITIVE TOMOGRAPHY



COGNITIVE TOMOGRAPHY

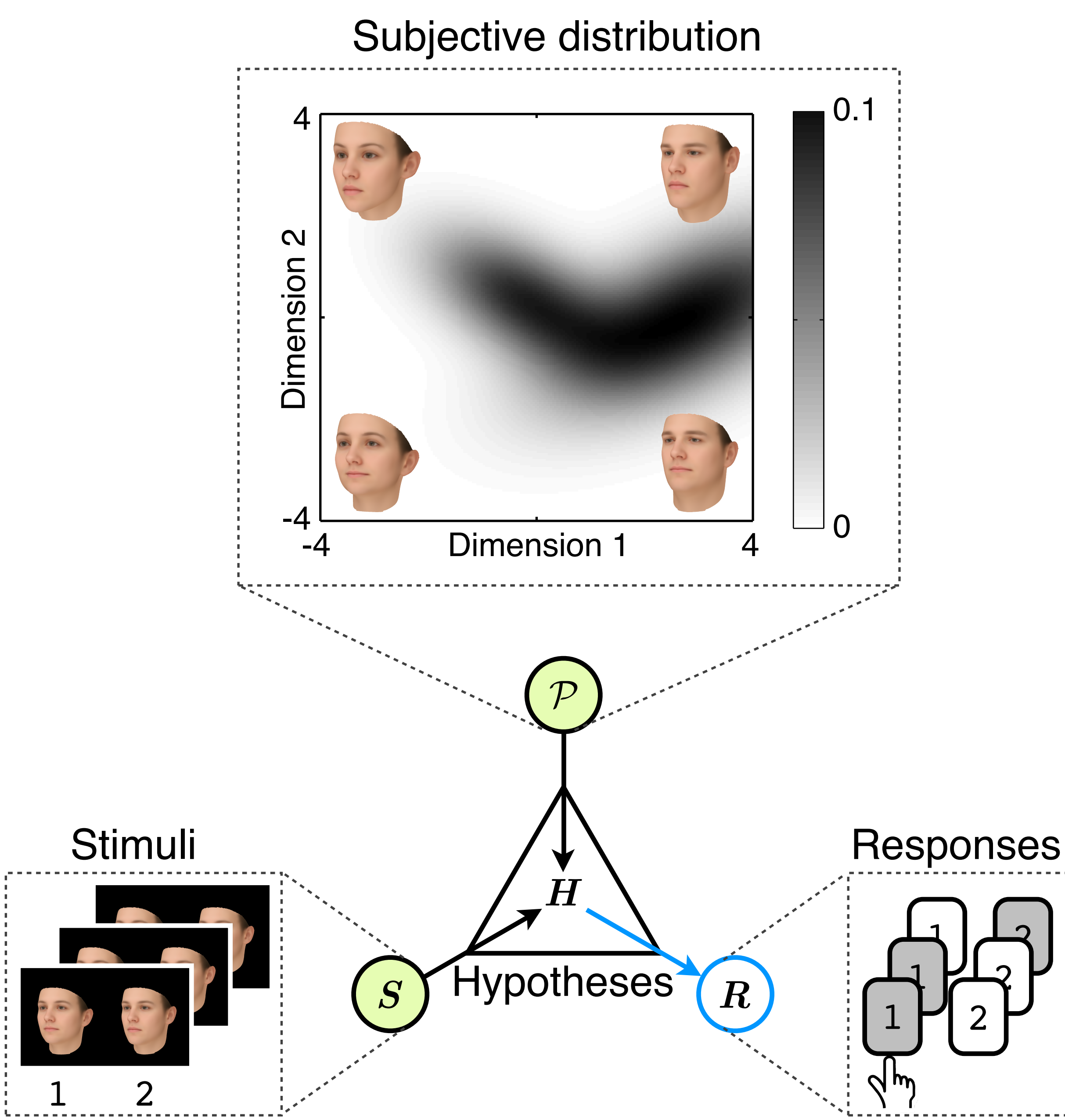


COGNITIVE TOMOGRAPHY



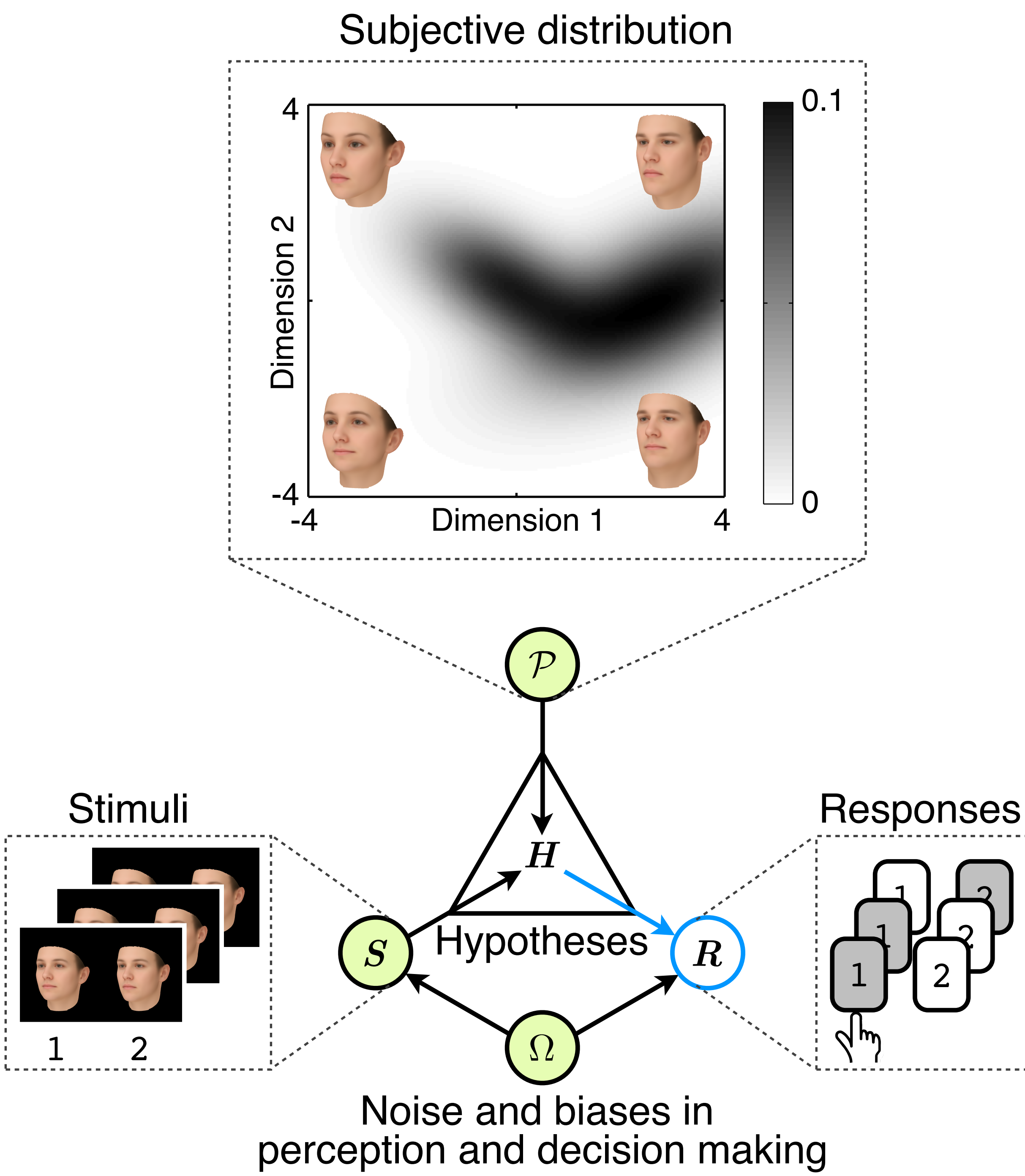
COGNITIVE TOMOGRAPHY

ideal observer model



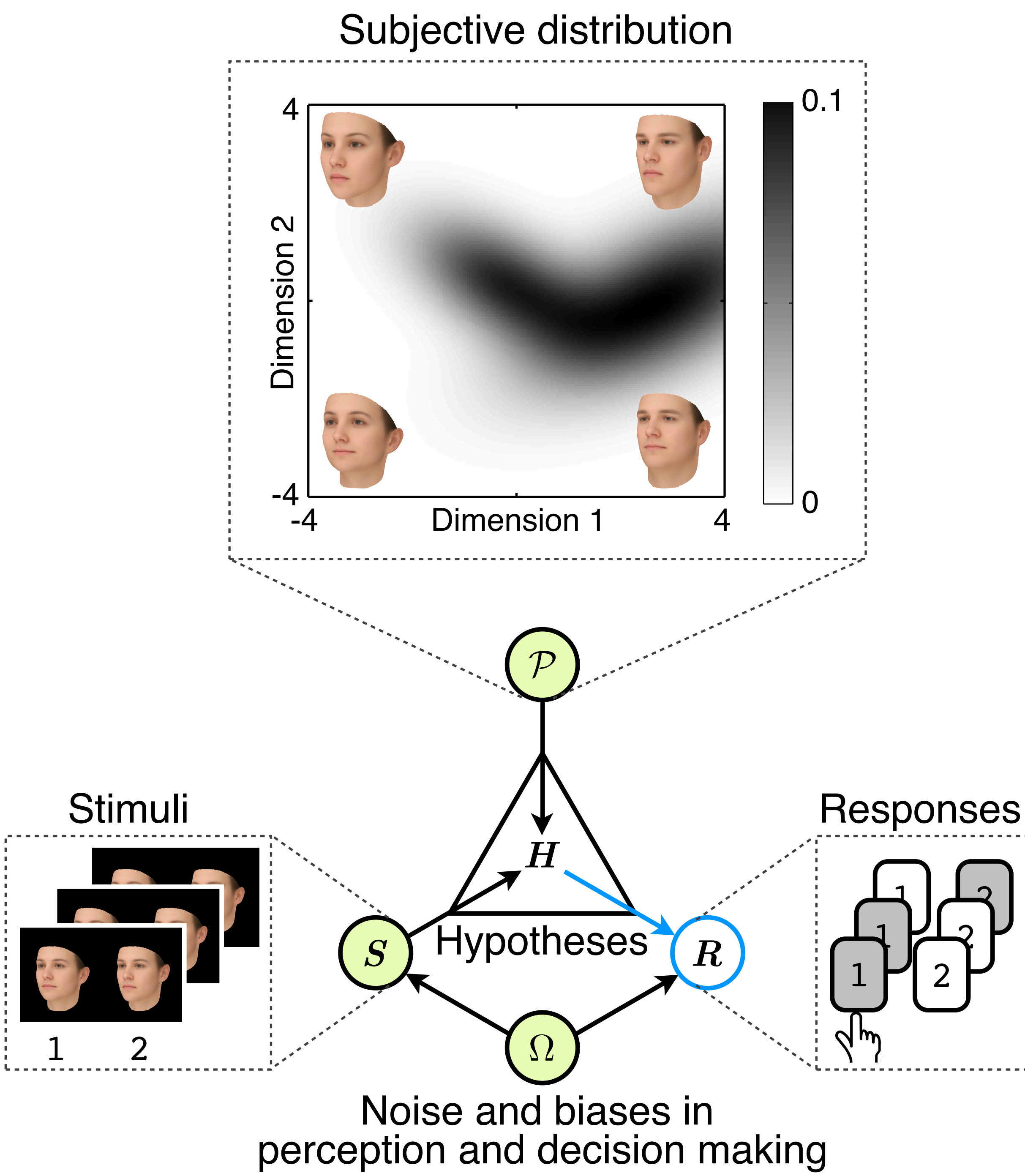
COGNITIVE TOMOGRAPHY

quasi- ideal observer model

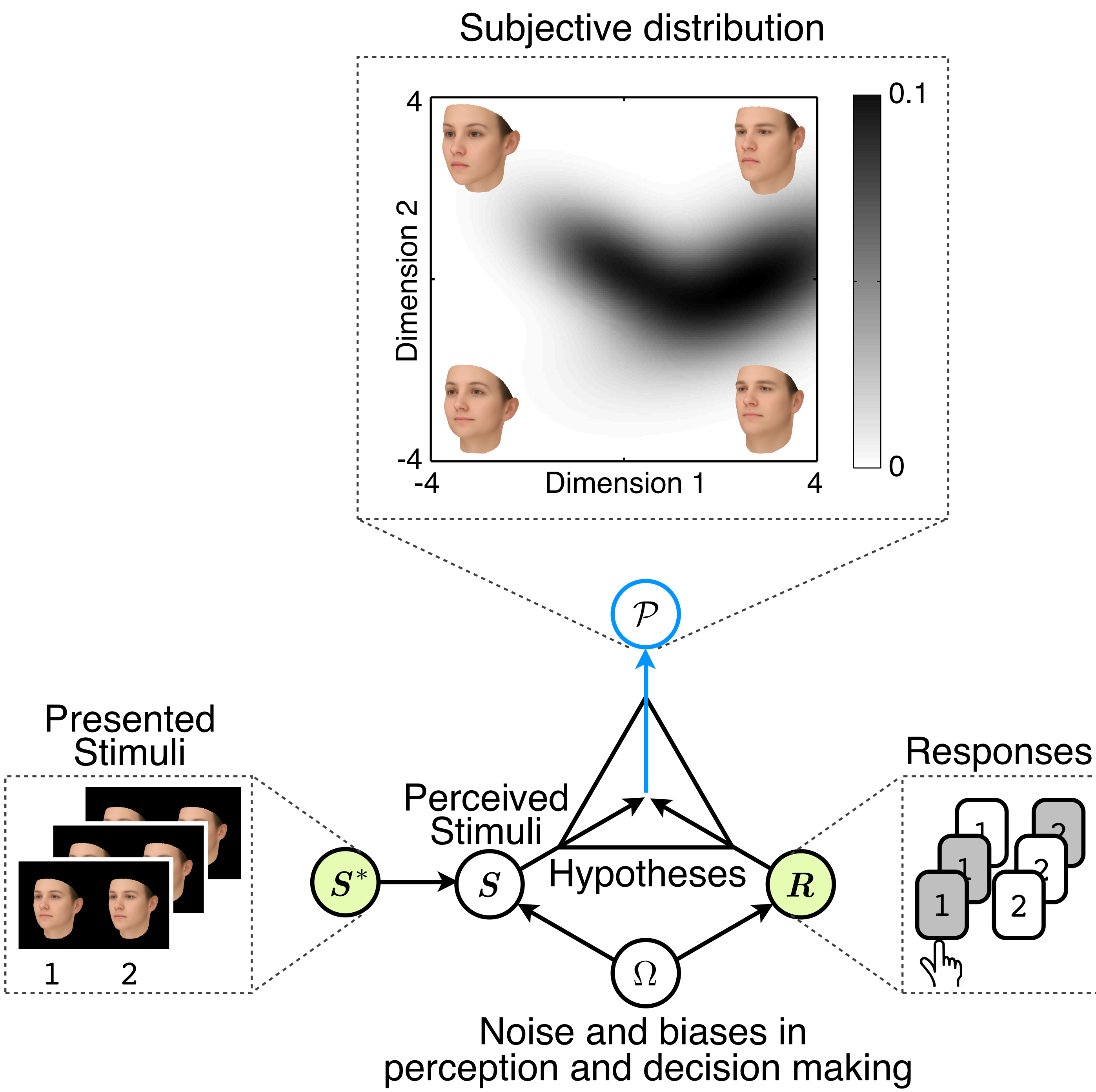


COGNITIVE TOMOGRAPHY

quasi- ideal observer model



cognitive tomography
inverting the ideal observer model

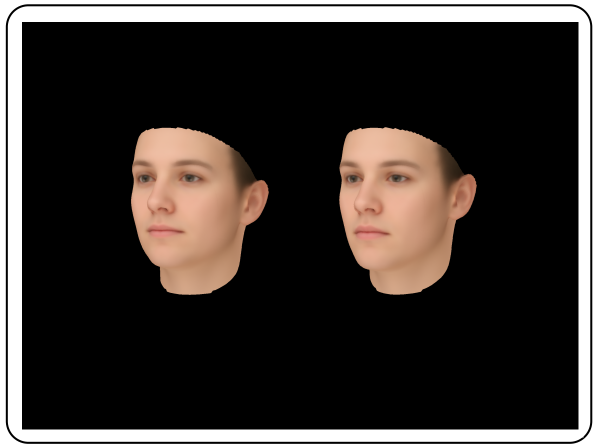


Houlsby et al, Curr Biol 2013

TWO TASKS

TWO TASKS

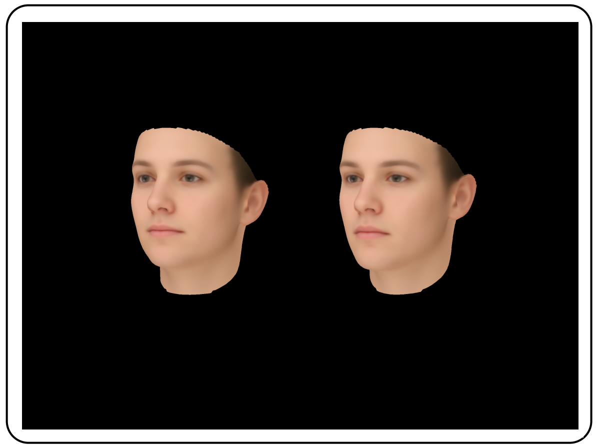
familiarity task
"*which face looks more familiar?*"



1 2

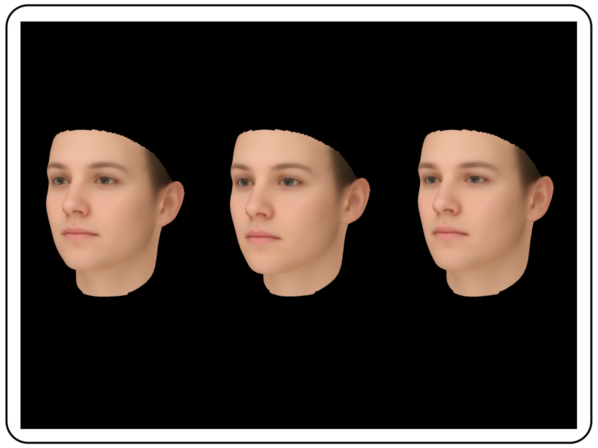
TWO TASKS

familiarity task
"which face looks more familiar?"



1 2

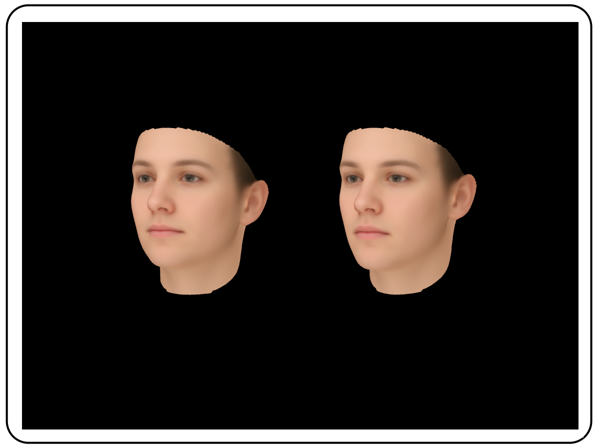
odd-one-out task
"which face is the odd-one-out?"



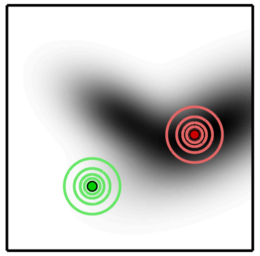
1 2 3

TWO TASKS

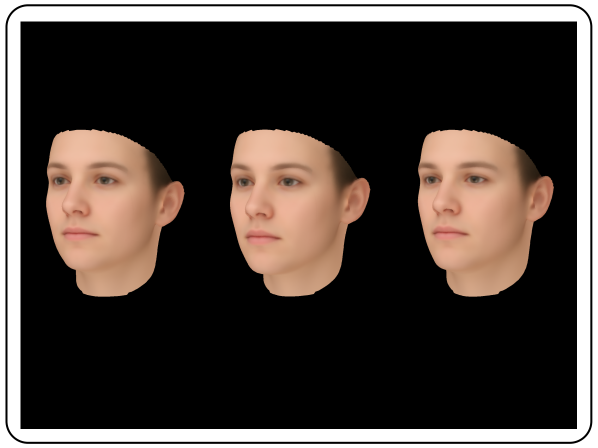
familiarity task
“which face looks more familiar?”



1 2



odd-one-out task
“which face is the odd-one-out?”

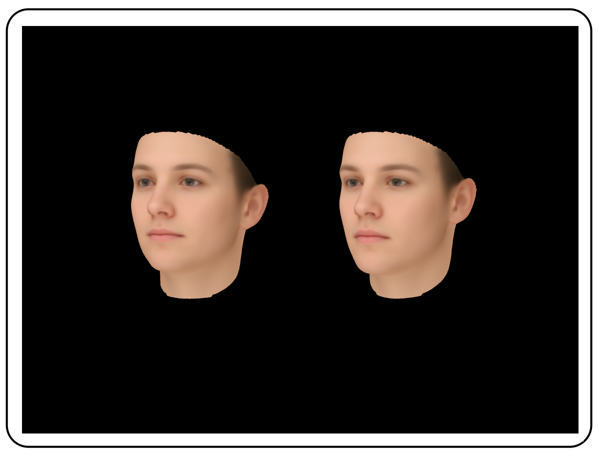


1 2 3

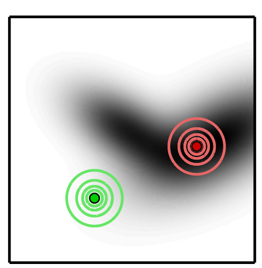
TWO TASKS

familiarity task

"which face looks more familiar?"

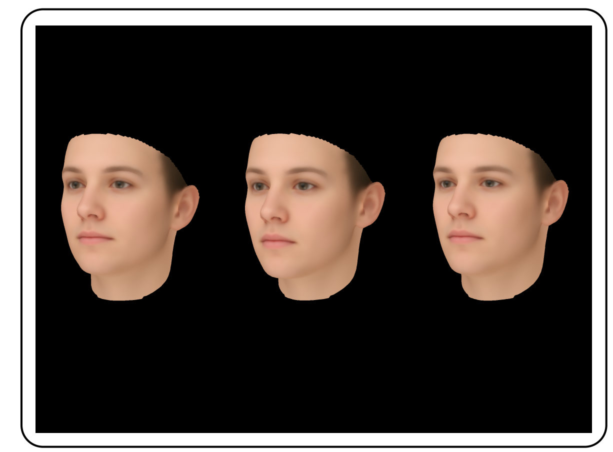


1 2

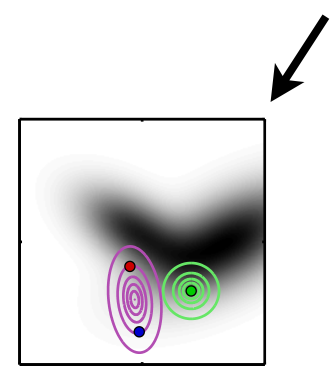


odd-one-out task

"which face is the odd-one-out?"

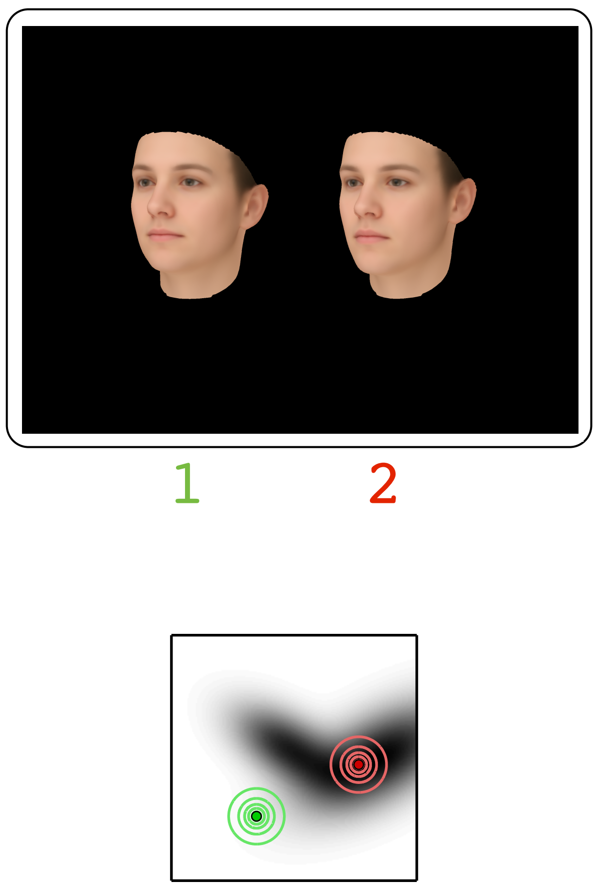


1 2 3

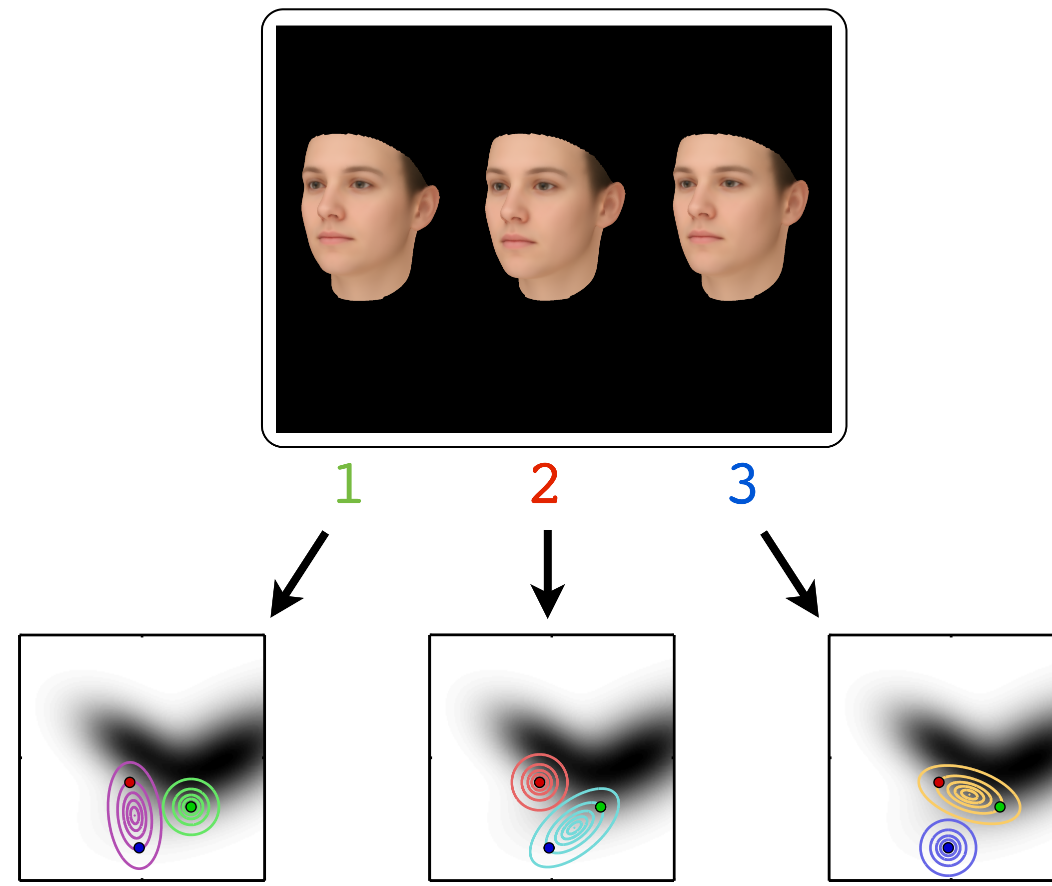


TWO TASKS

familiarity task
"which face looks more familiar?"



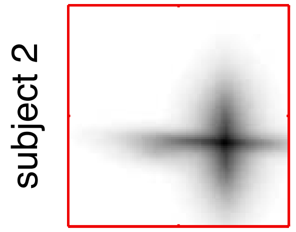
odd-one-out task
"which face is the odd-one-out?"



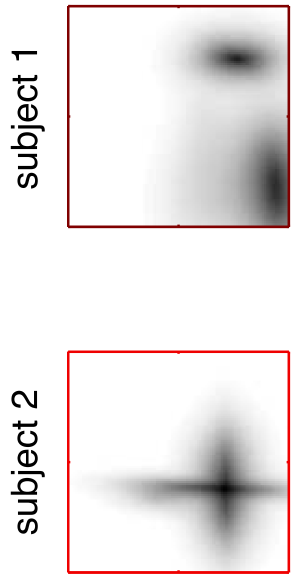
Houlsby et al, Curr Biol 2013

SUBJECTIVE DISTRIBUTIONS

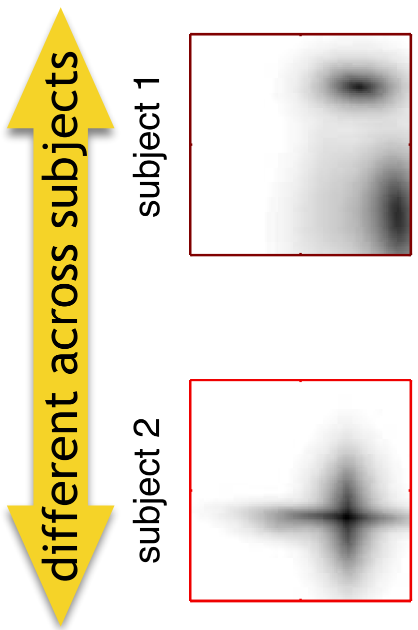
SUBJECTIVE DISTRIBUTIONS



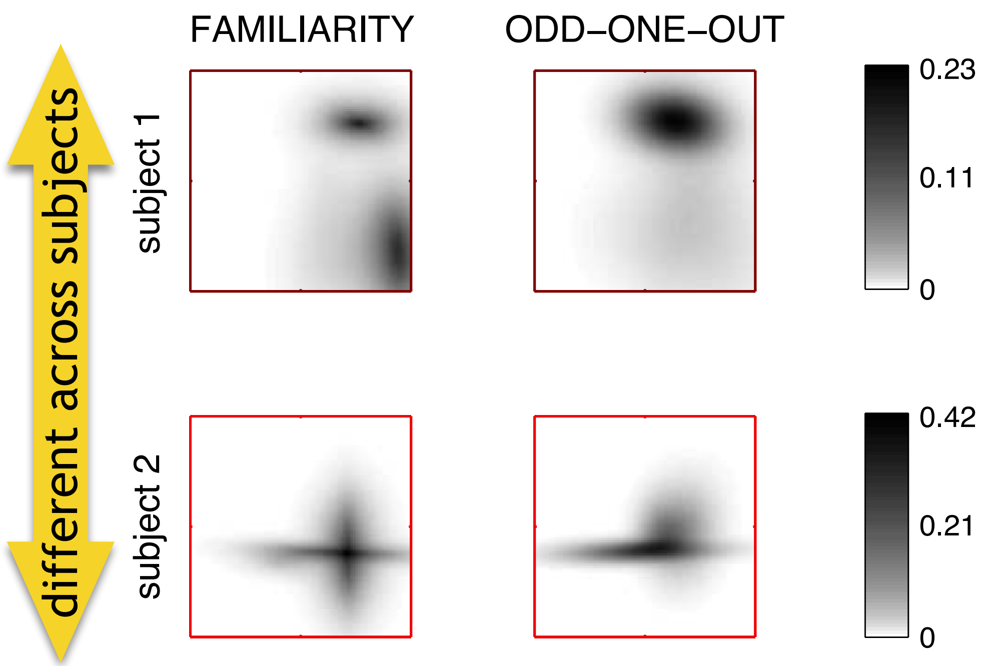
SUBJECTIVE DISTRIBUTIONS



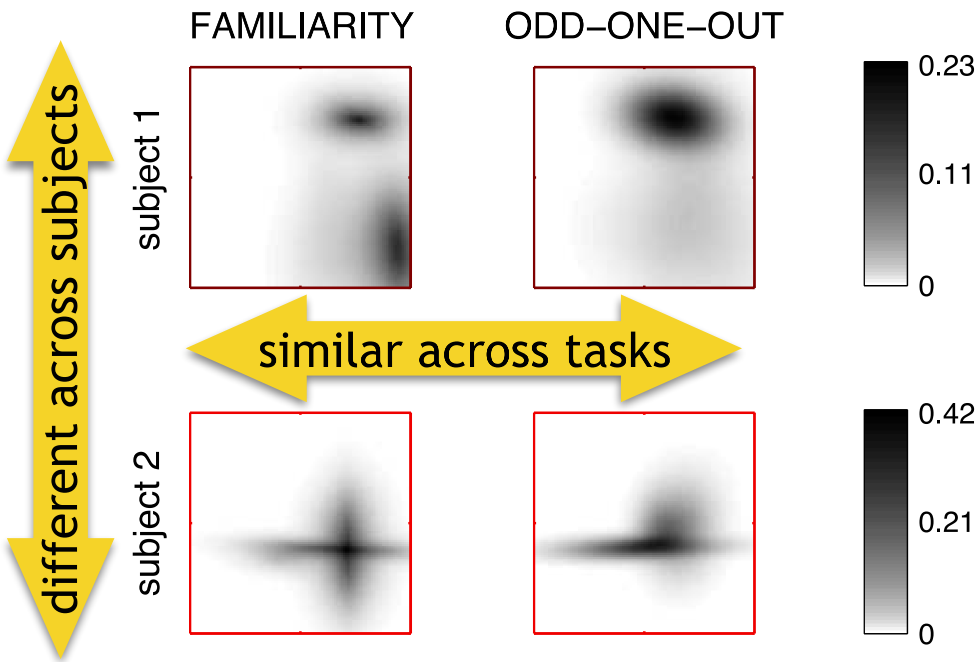
SUBJECTIVE DISTRIBUTIONS



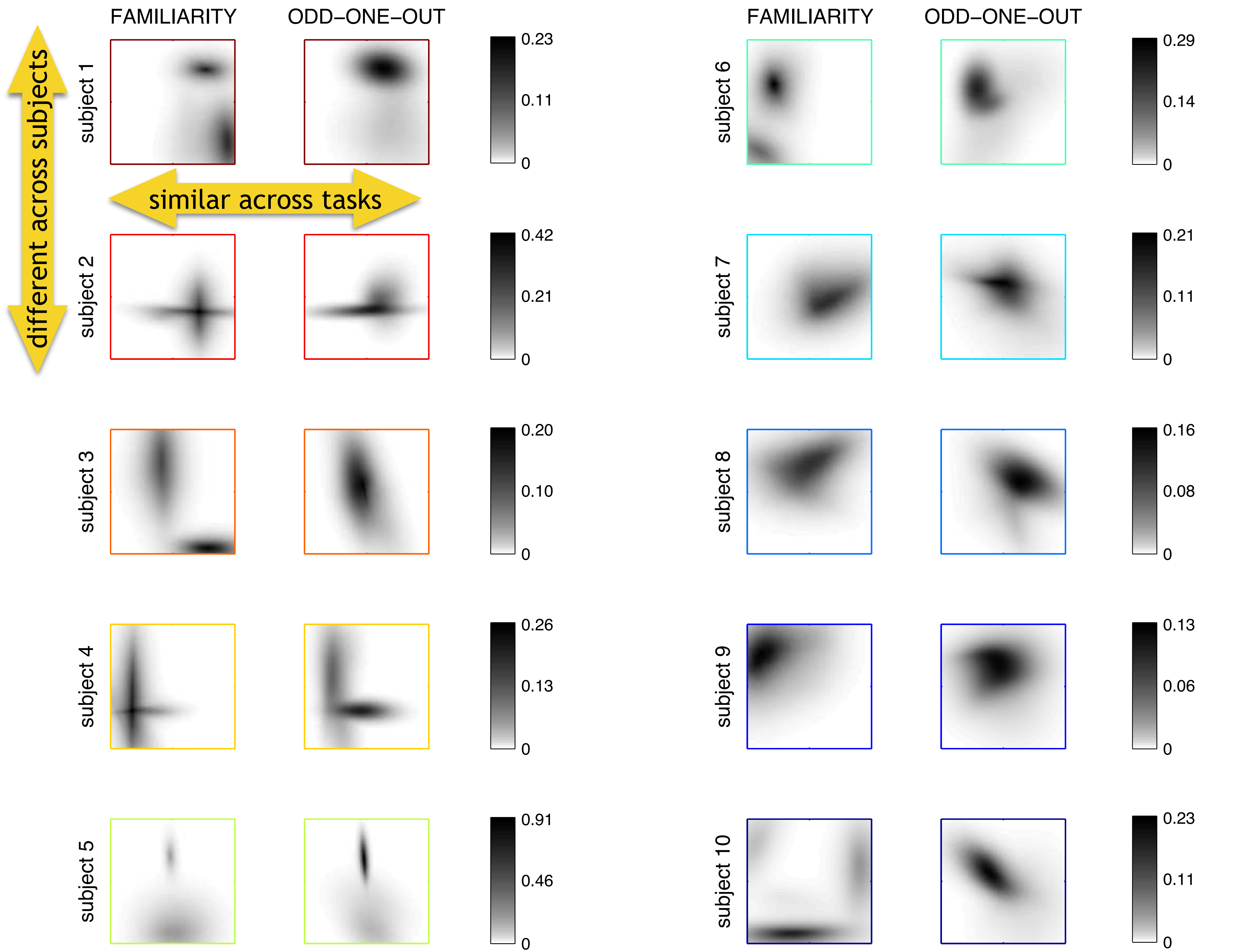
SUBJECTIVE DISTRIBUTIONS



SUBJECTIVE DISTRIBUTIONS



SUBJECTIVE DISTRIBUTIONS

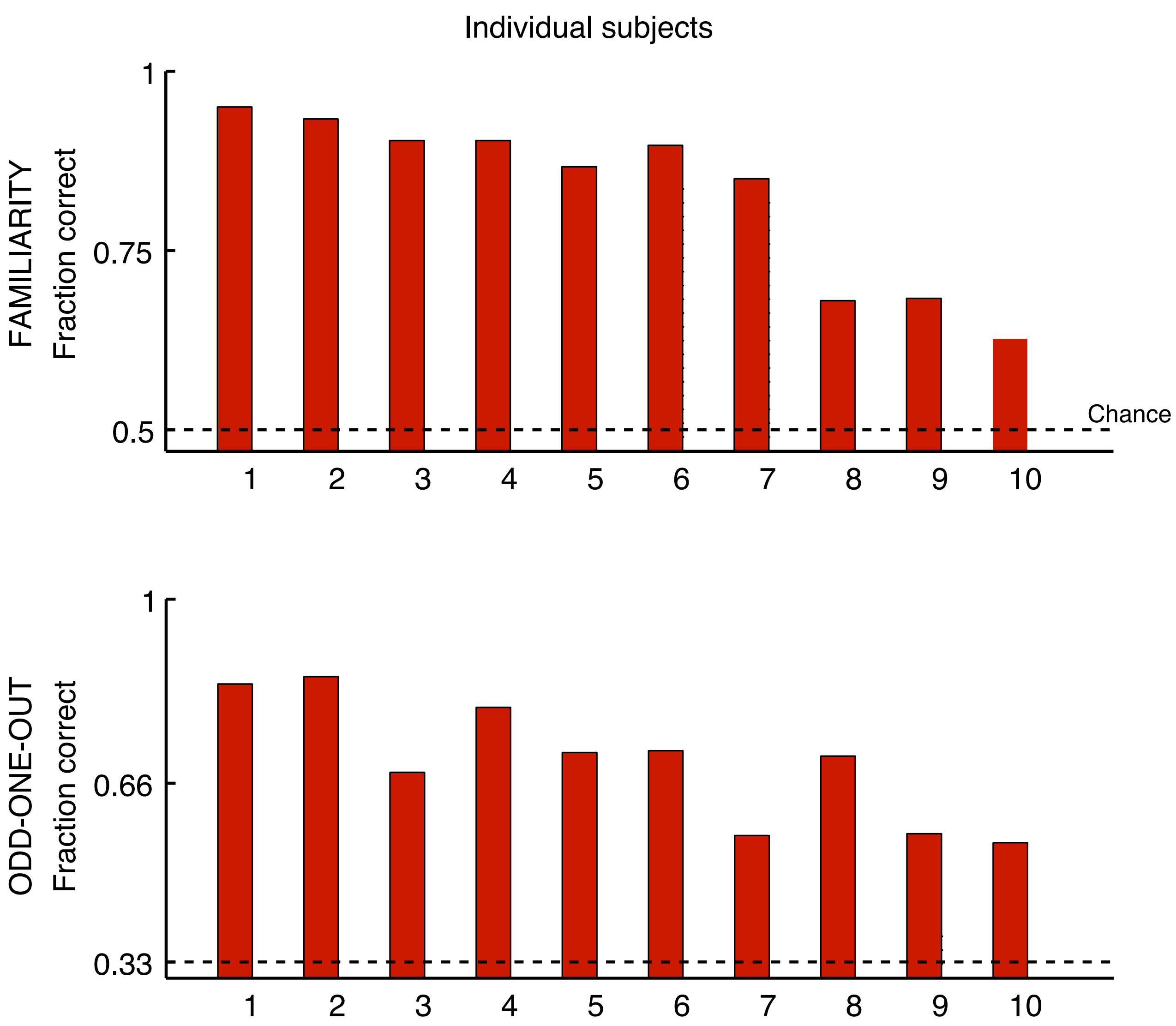


Houlsby et al, Curr Biol 2013

PREDICTING BEHAVIOUR

PREDICTING BEHAVIOUR

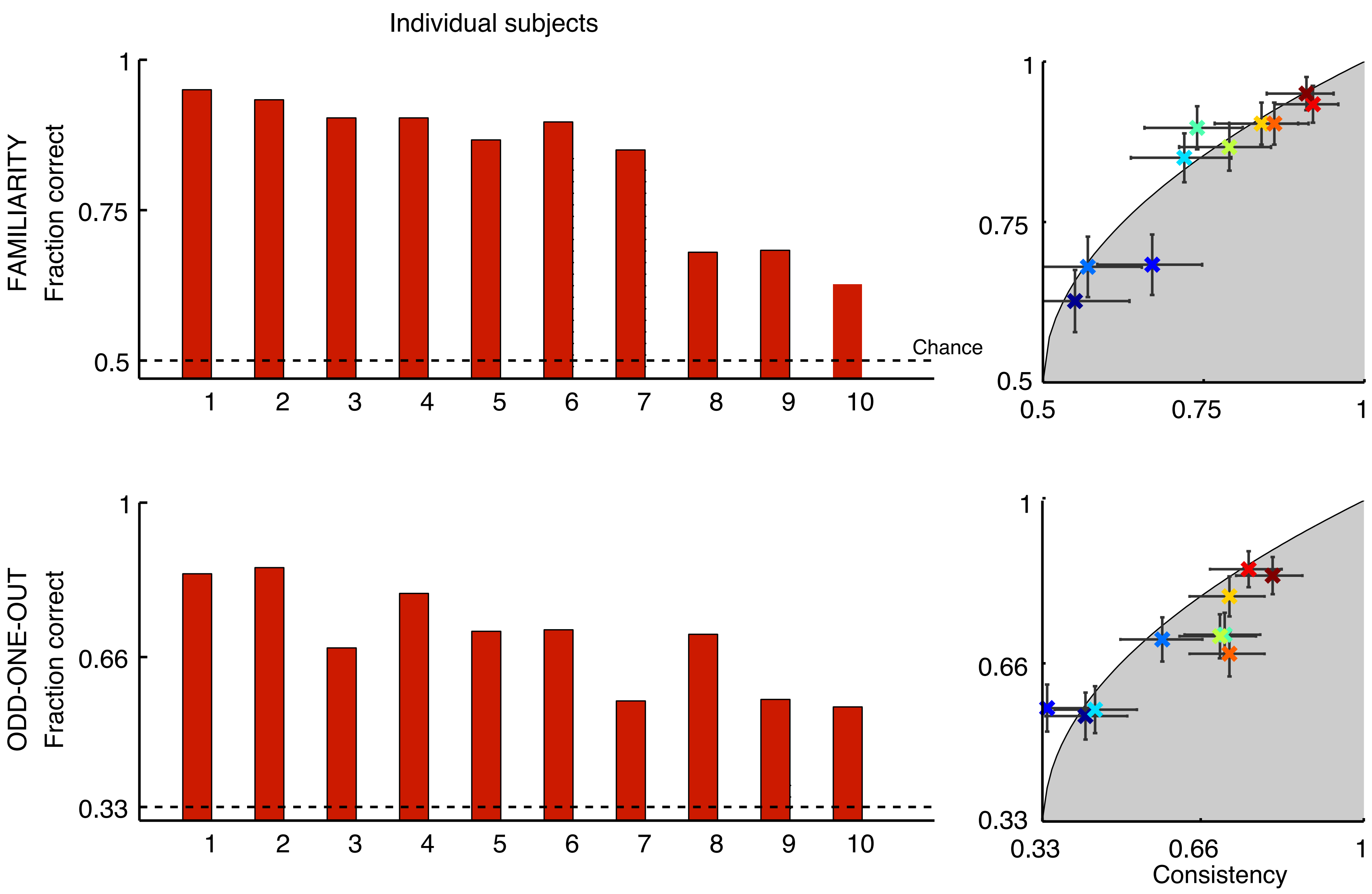
well above chance



PREDICTING BEHAVIOUR

well above chance

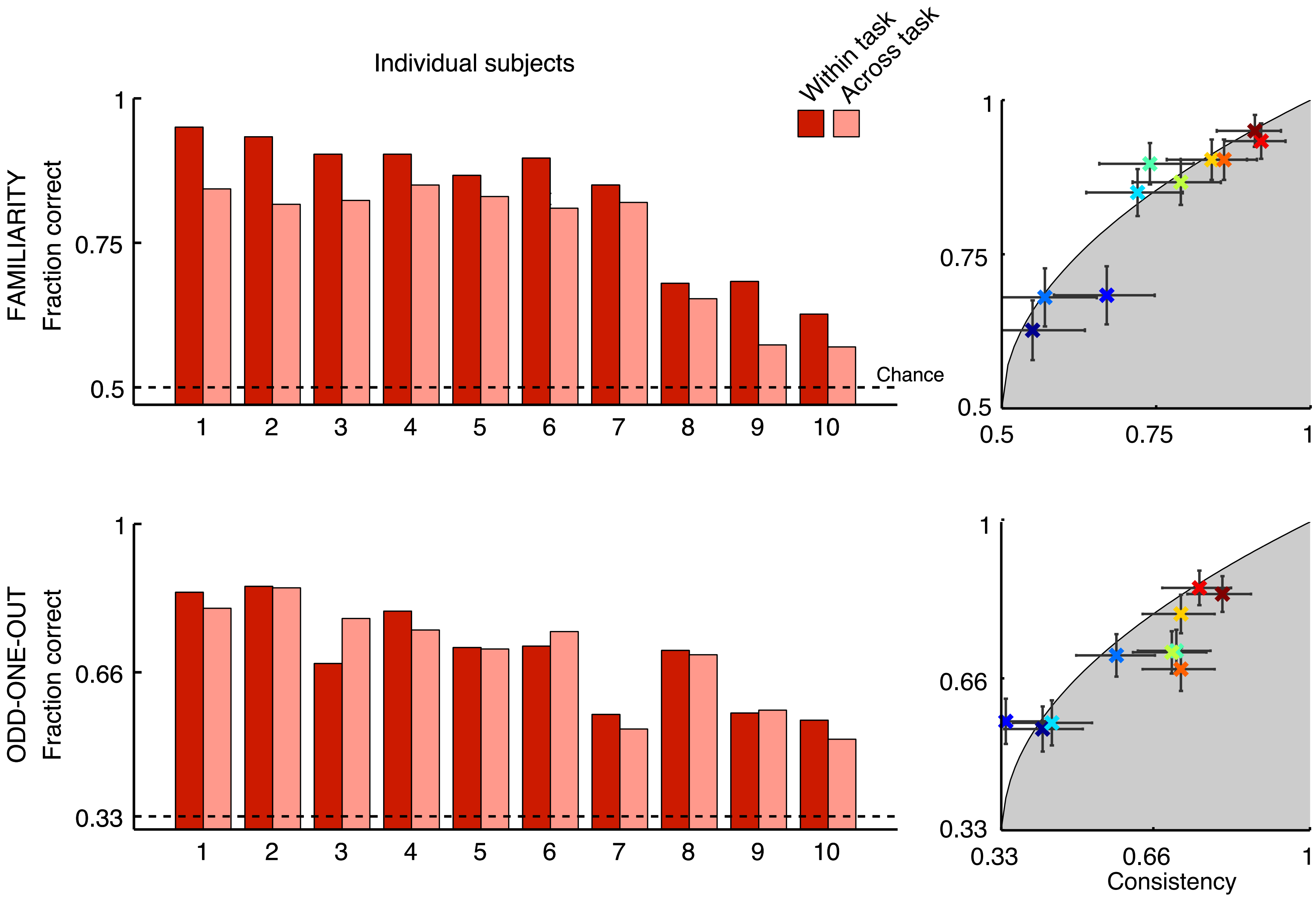
close to theoretical upper bound



PREDICTING BEHAVIOUR

well above chance both within and across tasks

close to theoretical upper bound

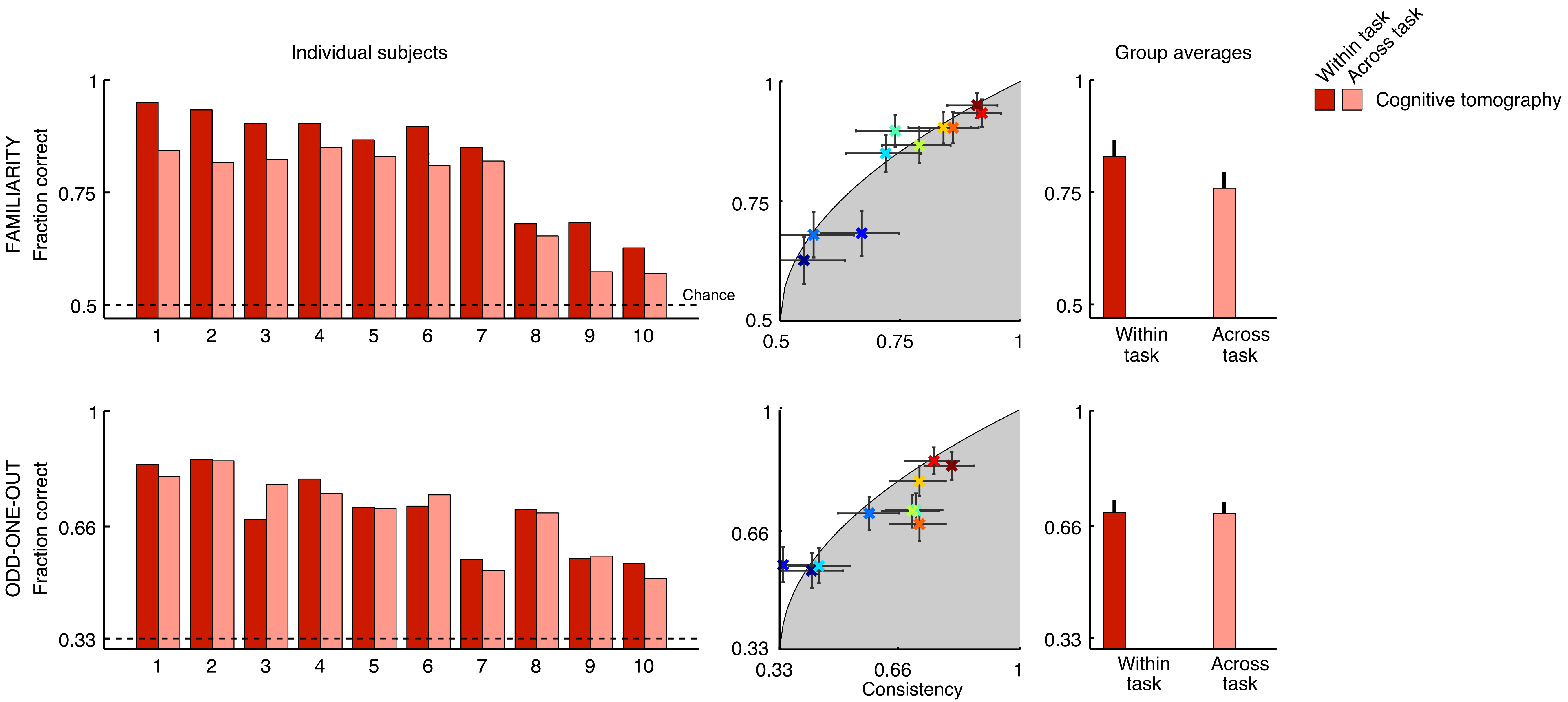


PREDICTING BEHAVIOUR

well above chance both within and across tasks

close to theoretical upper bound

better than controls



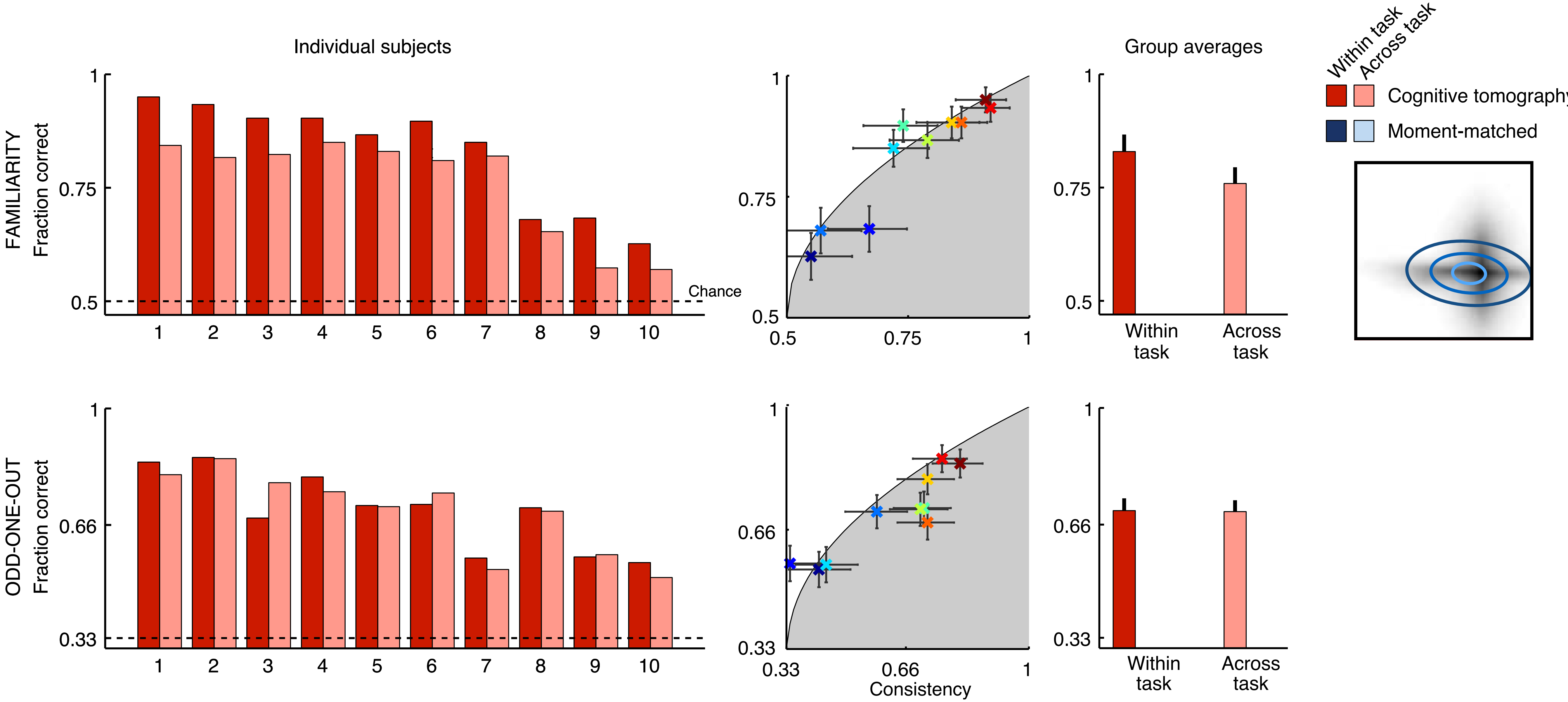
Houlsby et al, Curr Biol 2013

PREDICTING BEHAVIOUR

well above chance both within and across tasks

close to theoretical upper bound

better than controls



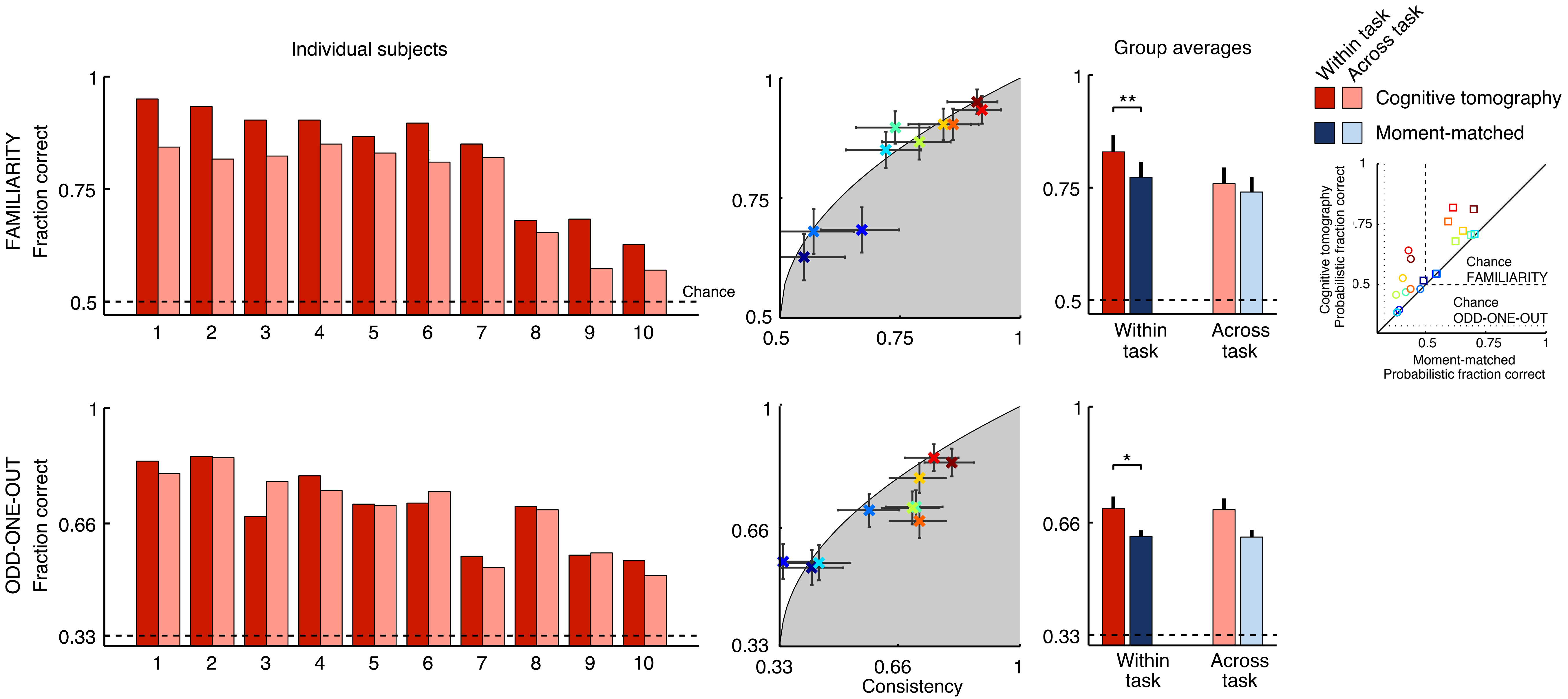
Houlsby et al, Curr Biol 2013

PREDICTING BEHAVIOUR

well above chance both within and across tasks

close to theoretical upper bound

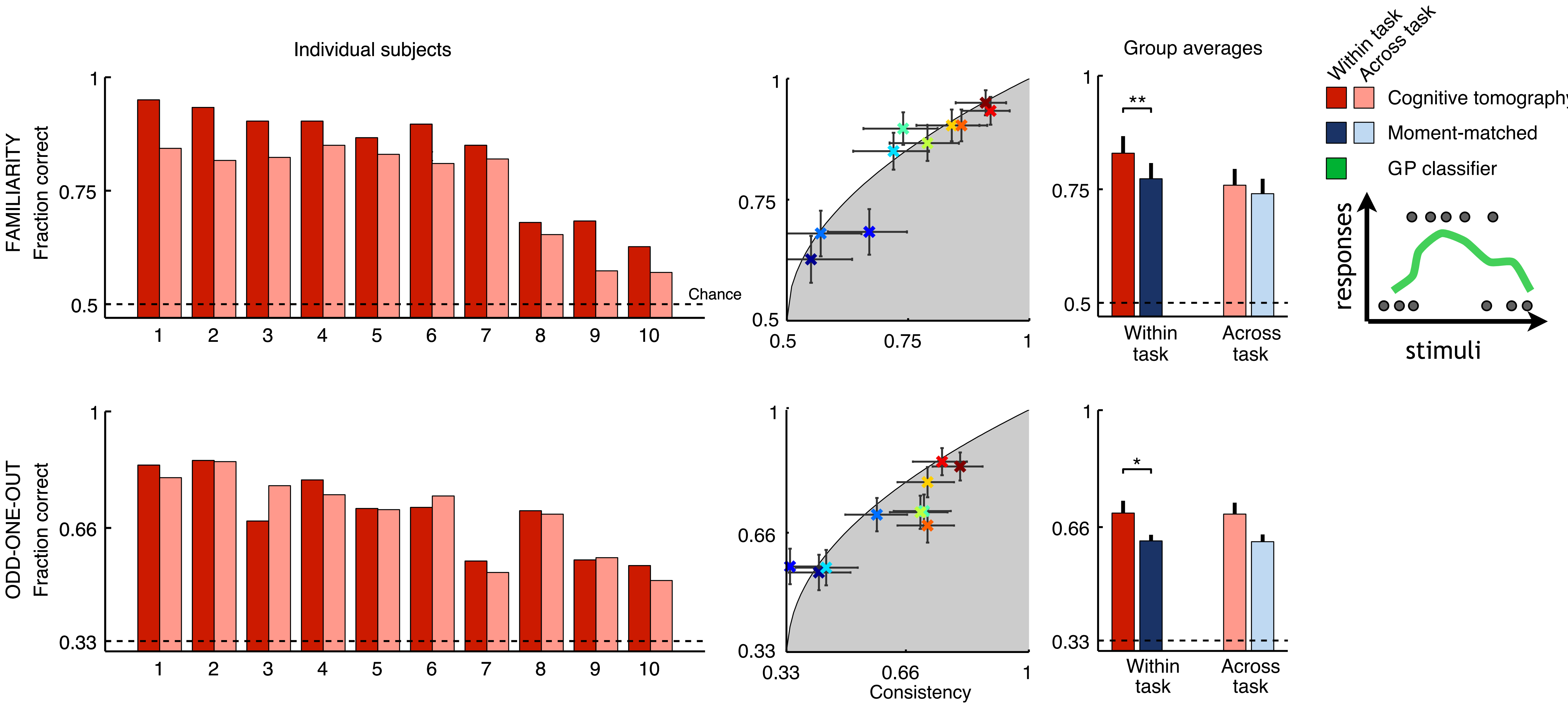
better than controls



Houlsby et al, Curr Biol 2013

PREDICTING BEHAVIOUR

well above chance both within and across tasks | **close to theoretical upper bound** | **better than controls**



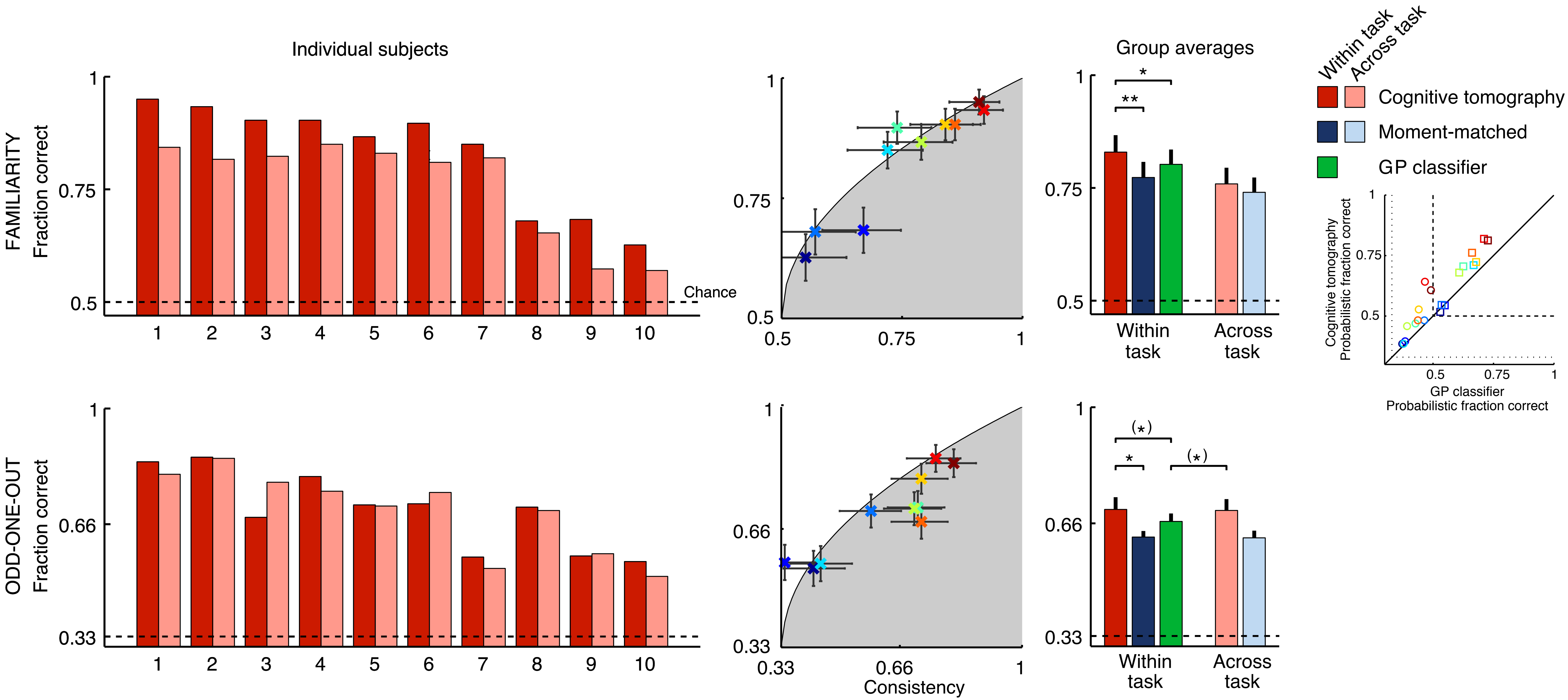
Houlsby et al, Curr Biol 2013

PREDICTING BEHAVIOUR

well above chance both within and across tasks

close to theoretical upper bound

better than controls



Houlsby et al, Curr Biol 2013

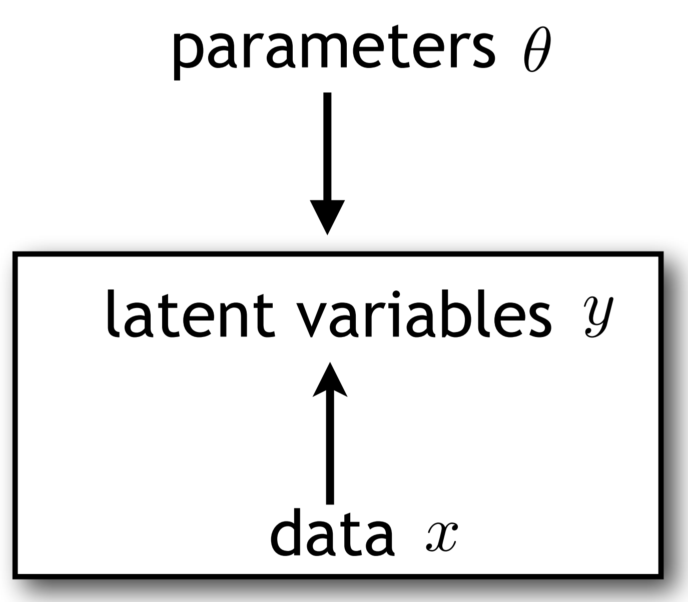
PROBABILISTIC INFERENCE AND LEARNING

latent variables y

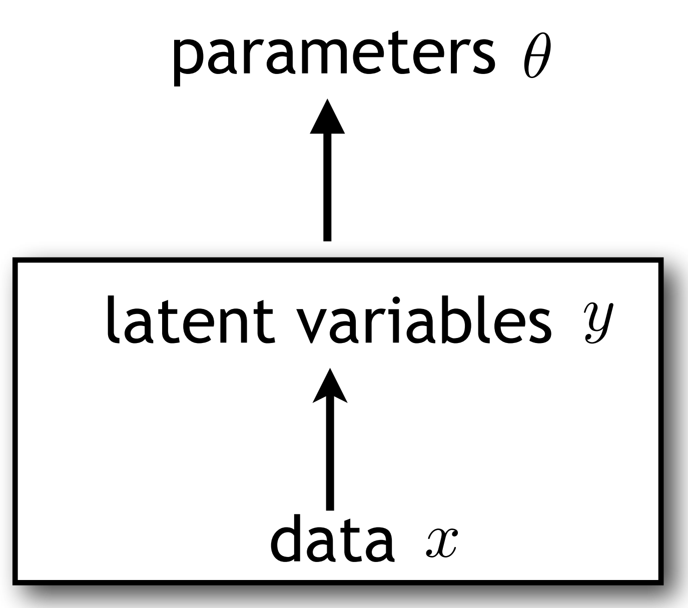
↑

data x

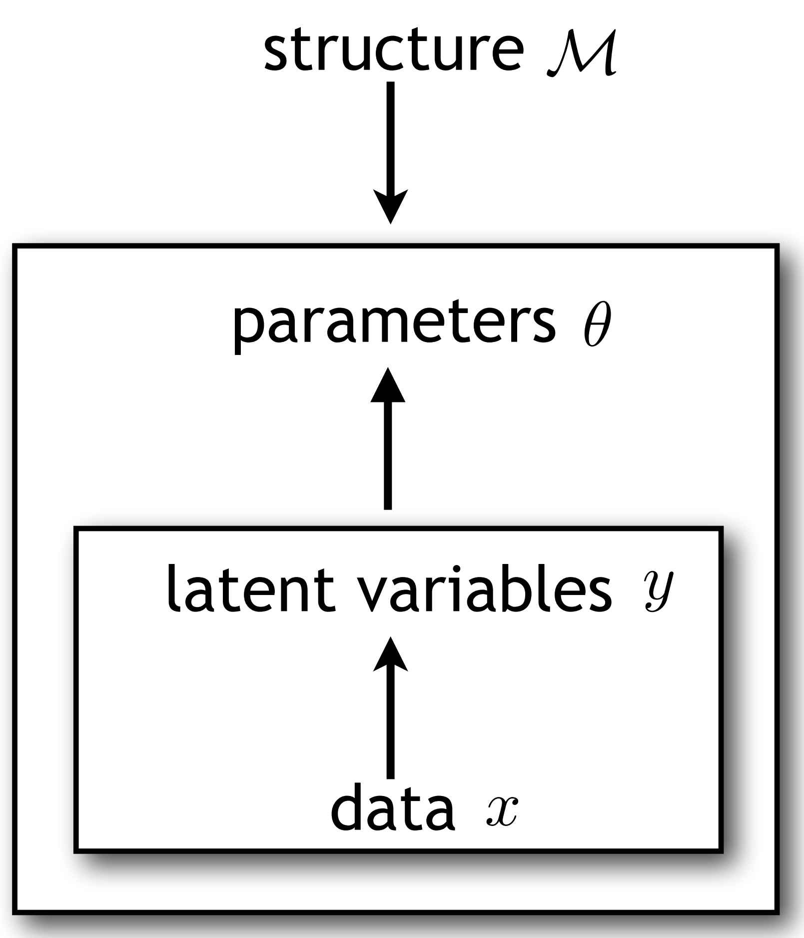
PROBABILISTIC INFERENCE AND LEARNING



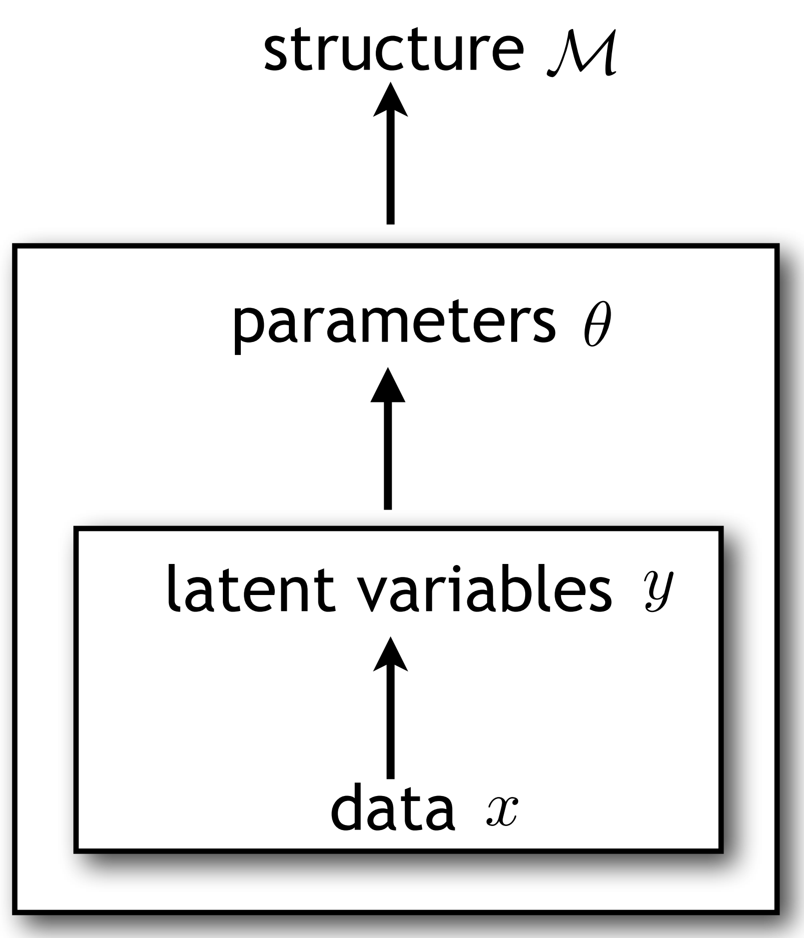
PROBABILISTIC INFERENCE AND LEARNING



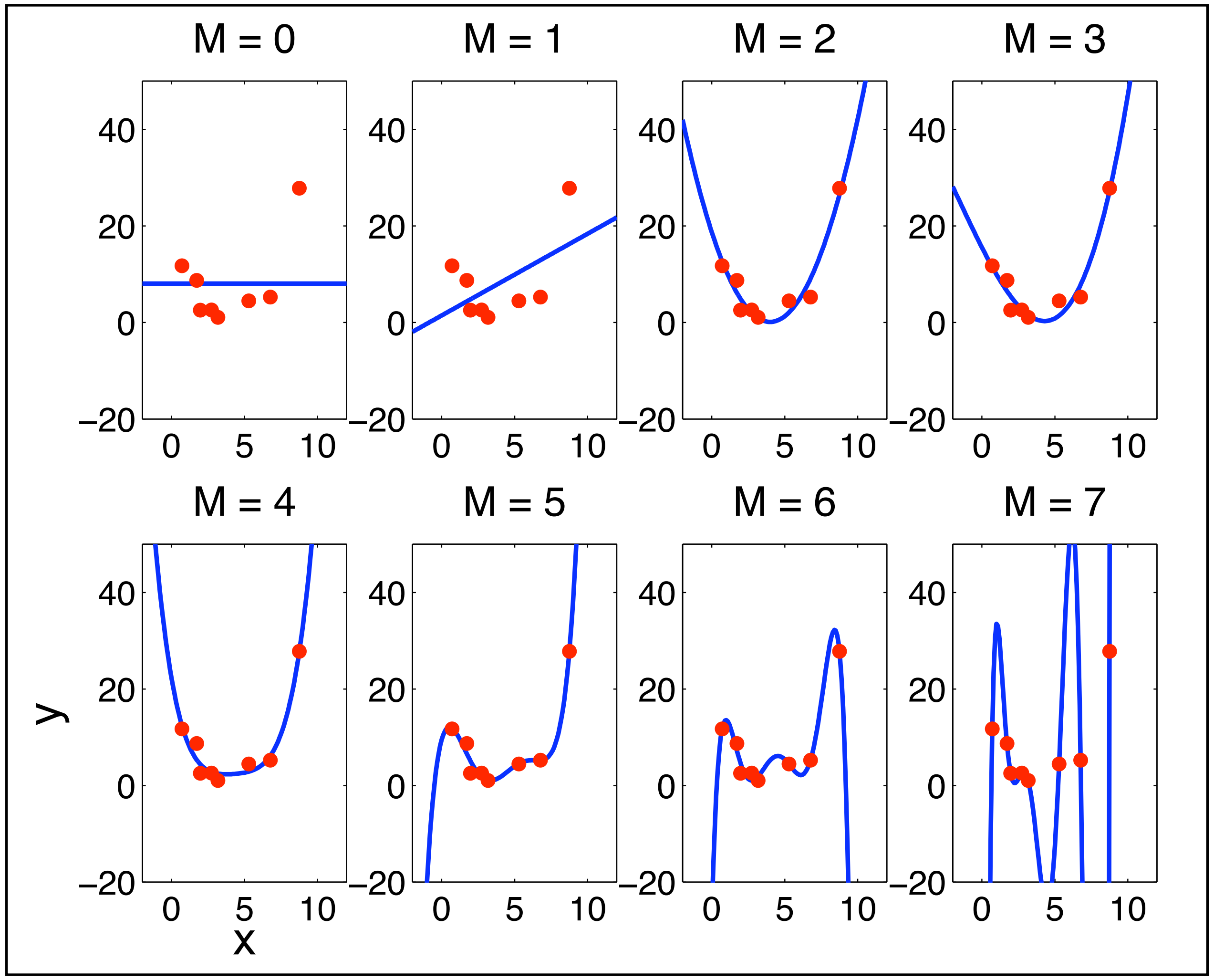
PROBABILISTIC INFERENCE AND LEARNING



PROBABILISTIC INFERENCE AND LEARNING

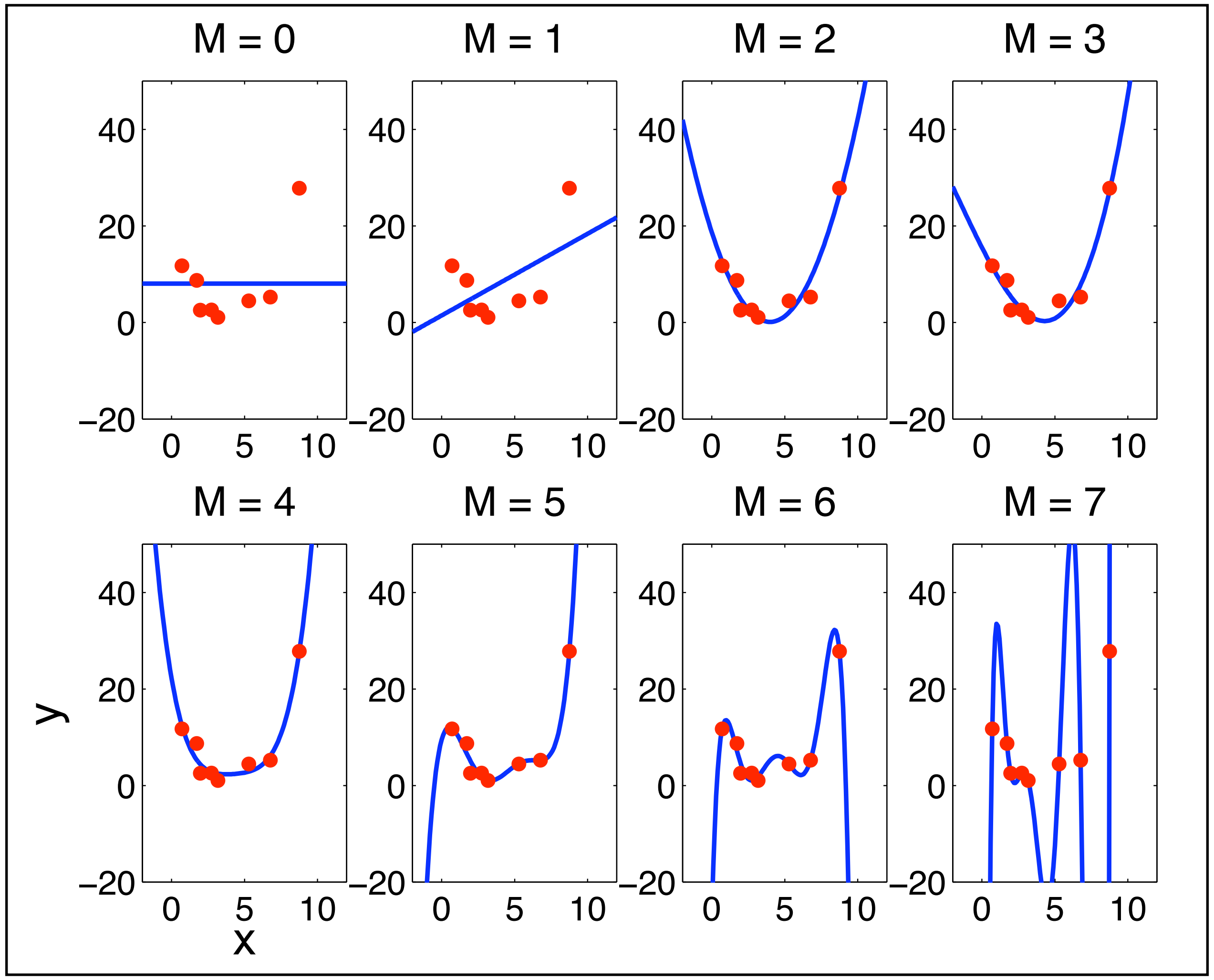


MODEL COMPARISON



courtesy of Zoubin Ghahramani

MODEL COMPARISON



courtesy of Zoubin Ghahramani

- under which model do I get the best fit?

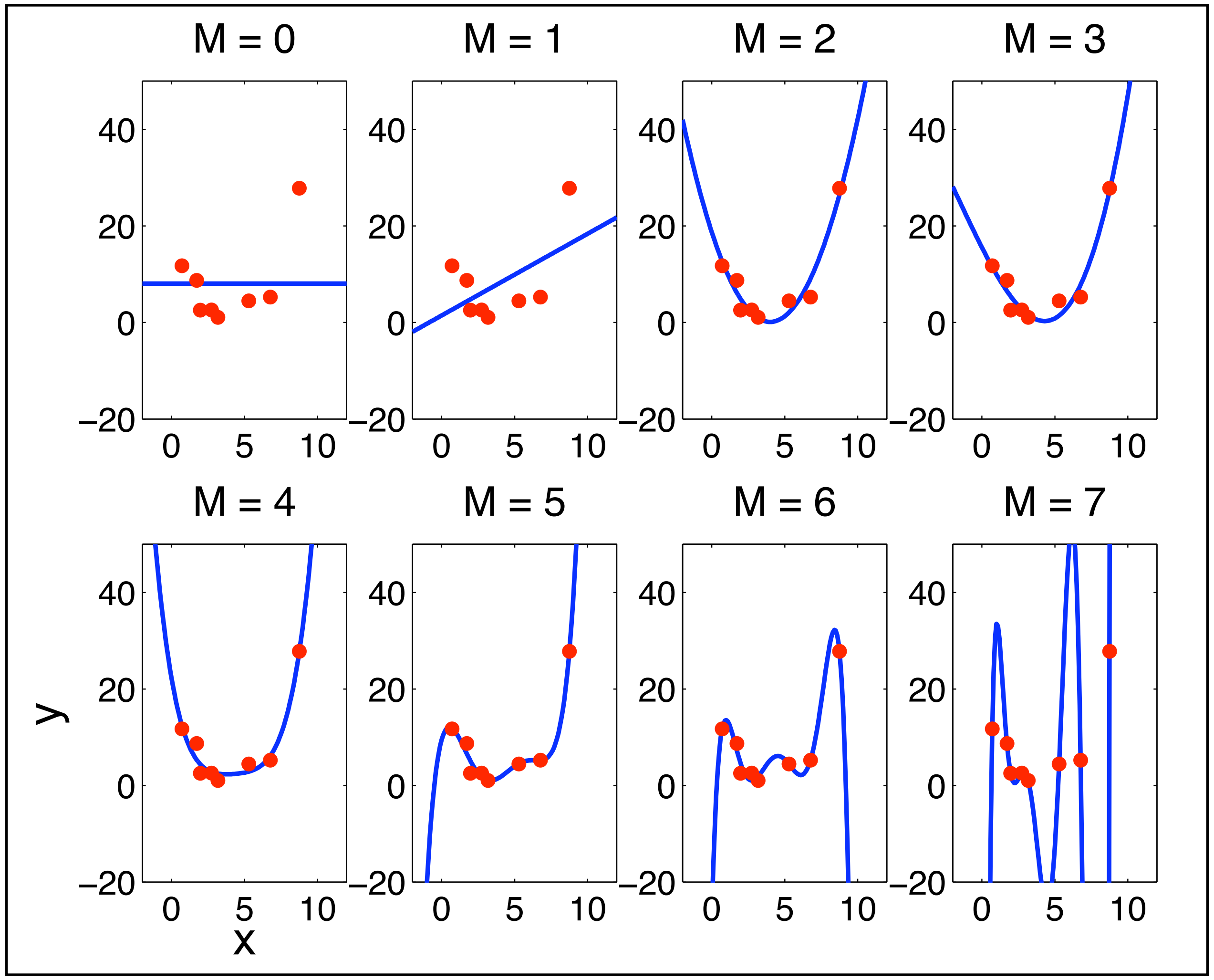
$$\mathrm{P}\left(\mathcal{D}|\hat{\theta}_{\mathrm{ML}}, \mathcal{M}\right)$$

parameters

model structure

what is the likelihood of the model with *the best* parameters?

MODEL COMPARISON



courtesy of Zoubin Ghahramani

- under which model do I get the best fit?

$$\mathrm{P}\left(\mathcal{D}|\hat{\theta}_{\mathrm{ML}}, \mathcal{M}\right)$$

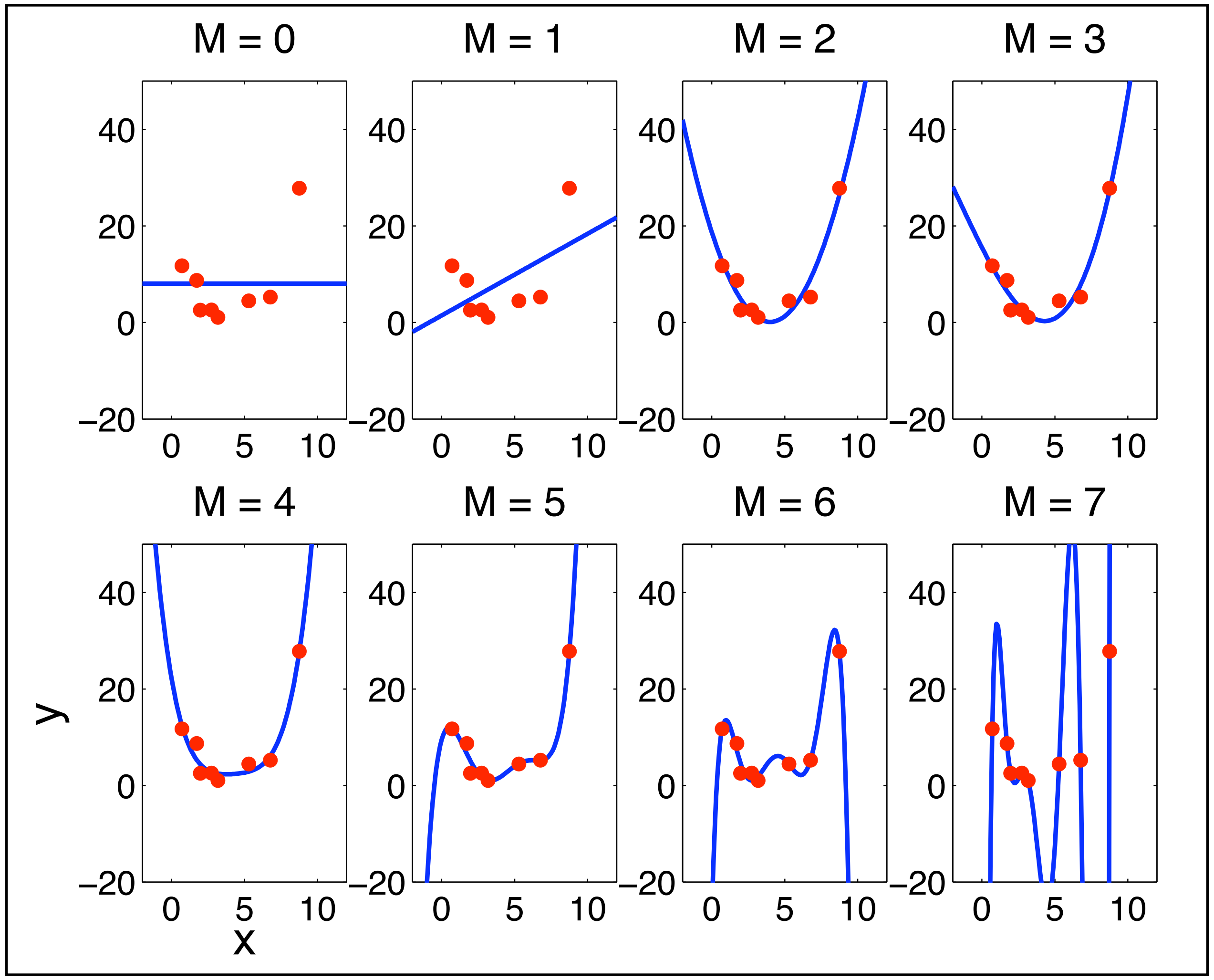
parameters

model structure

what is the likelihood of the model with *the best* parameters?

overfitting!

MODEL COMPARISON



courtesy of Zoubin Ghahramani

- under which model do I get the best fit?

$$\mathrm{P}\left(\mathcal{D}|\hat{\theta}_{\mathrm{ML}}, \mathcal{M}\right)$$

parameters ↑ ← model structure

what is the likelihood of the model with *the best* parameters?

↓

overfitting!

- which model has the highest likelihood?

$$\mathrm{P}(\mathcal{D}|\mathcal{M}) = \sum_{\theta} \mathrm{P}(\mathcal{D}|\theta, \mathcal{M})\, \mathrm{P}(\theta|\mathcal{M})$$

what is the average likelihood of the model with *randomly chosen* parameters?

BAYESIAN MODEL SELECTION

Rev. Bayes

a model defines a probability distribution over data sets

$$P[\mathcal{D}|\mathcal{M}]$$

data coming from the environment

model stored in memory

BAYESIAN MODEL SELECTION

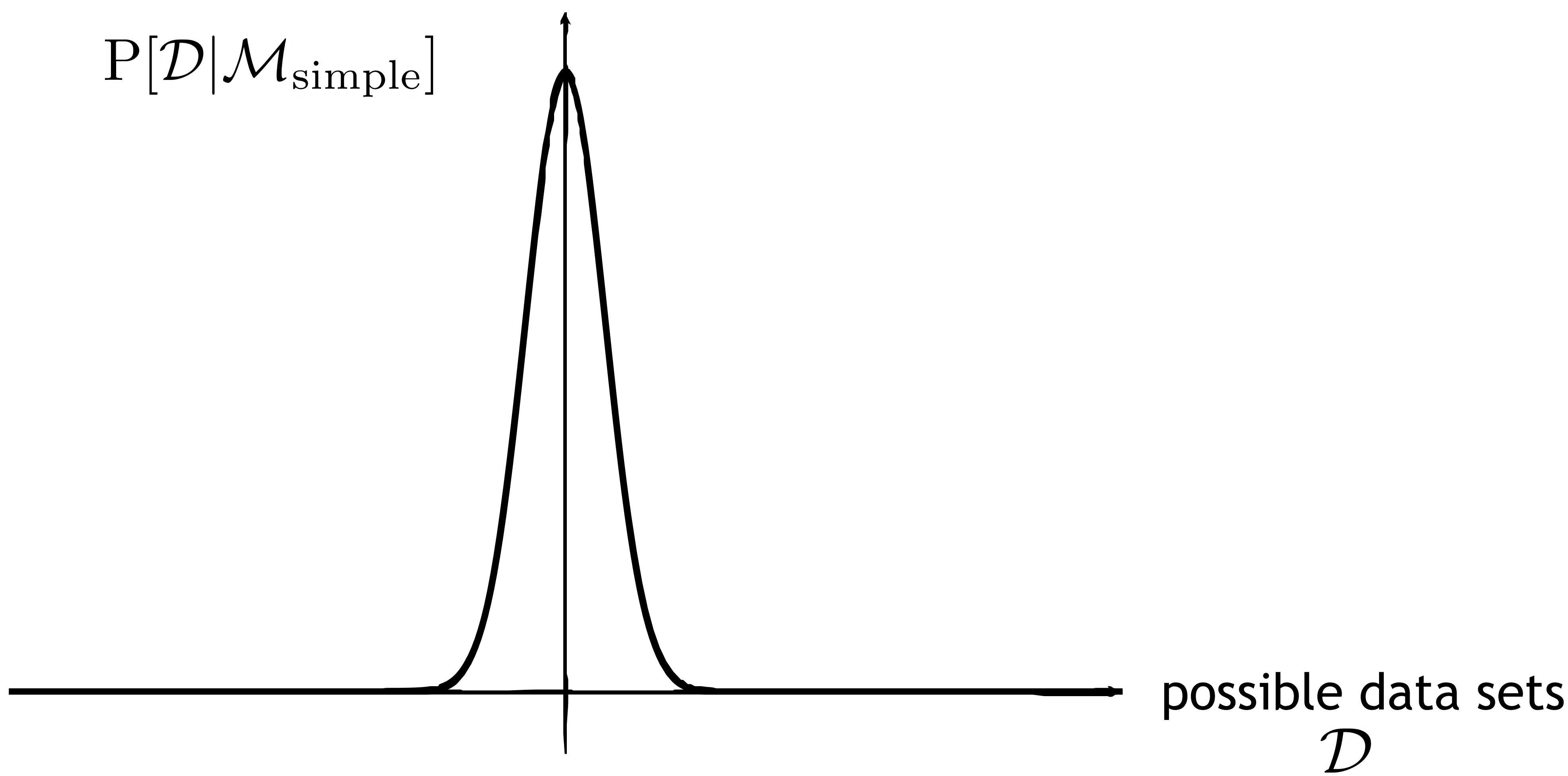
Rev. Bayes

a model defines a probability distribution over data sets

$$\mathrm{P}[\mathcal{D}|\mathcal{M}]$$

data coming from the environment

model stored in memory



BAYESIAN MODEL SELECTION

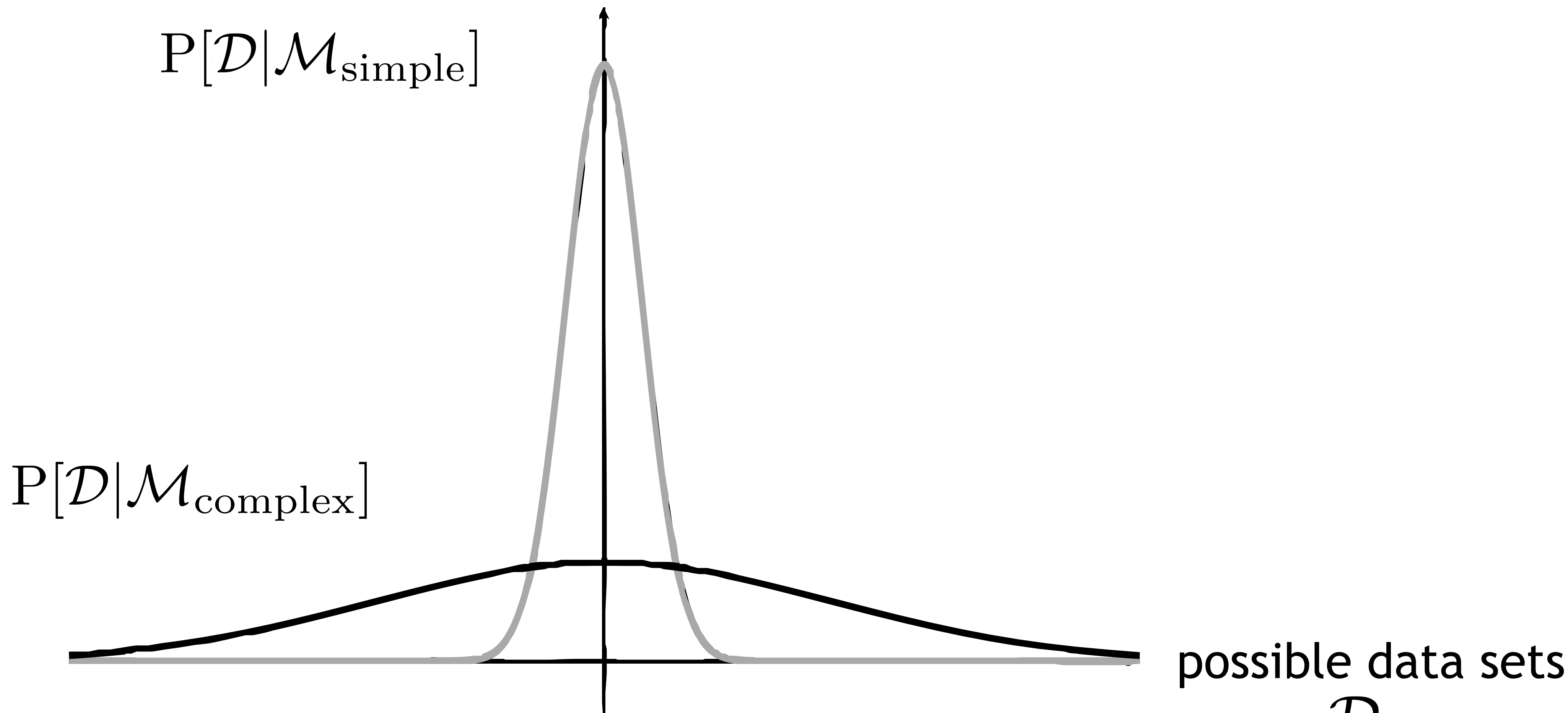
Rev. Bayes

a model defines a probability distribution over data sets

$$\mathrm{P}[\mathcal{D}|\mathcal{M}]$$

data coming from the environment

model stored in memory



BAYESIAN MODEL SELECTION

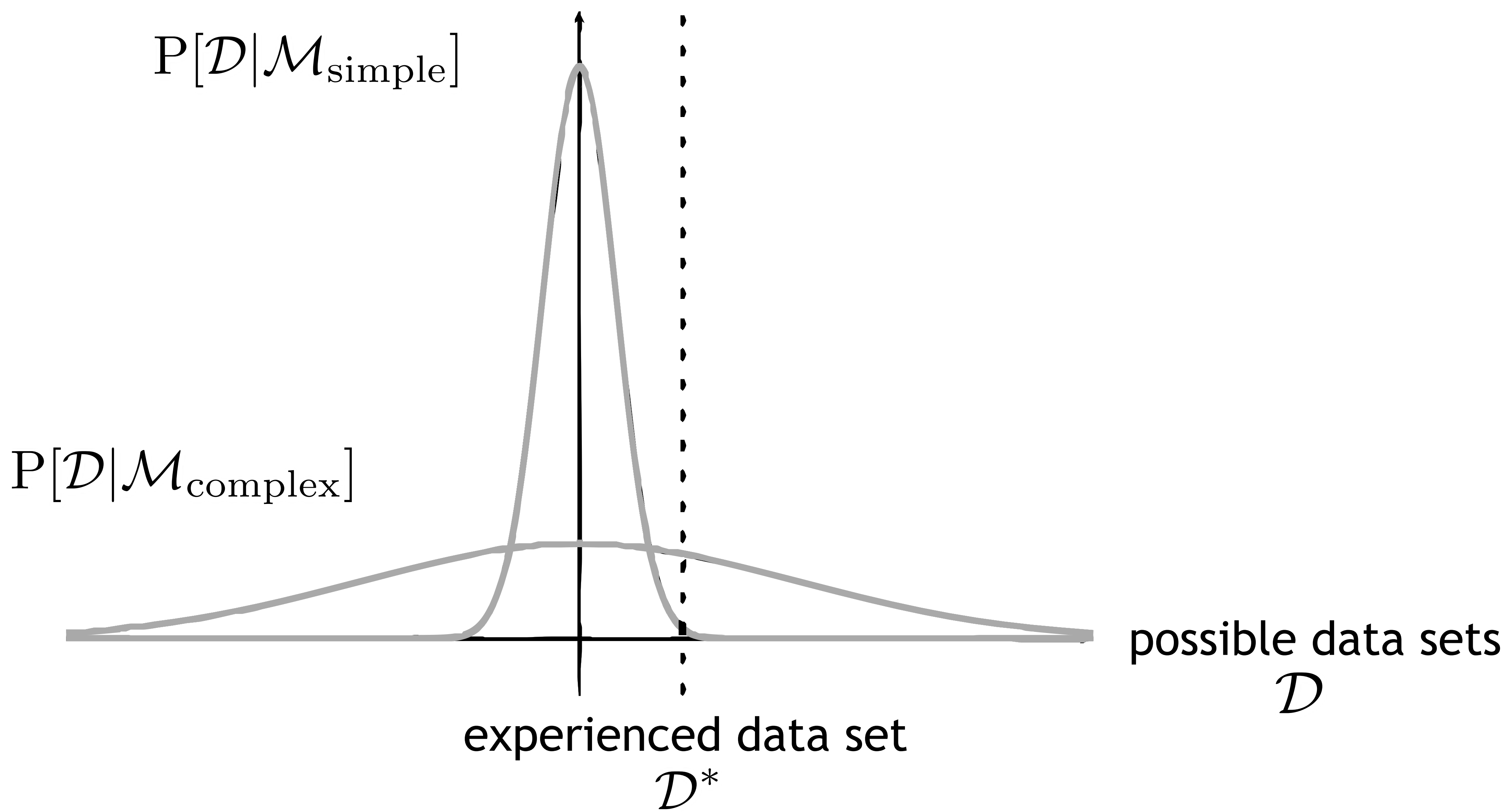
Rev. Bayes

a model defines a probability distribution over data sets

$$\mathrm{P}[\mathcal{D}|\mathcal{M}]$$

data coming from the environment

model stored in memory



BAYESIAN MODEL SELECTION

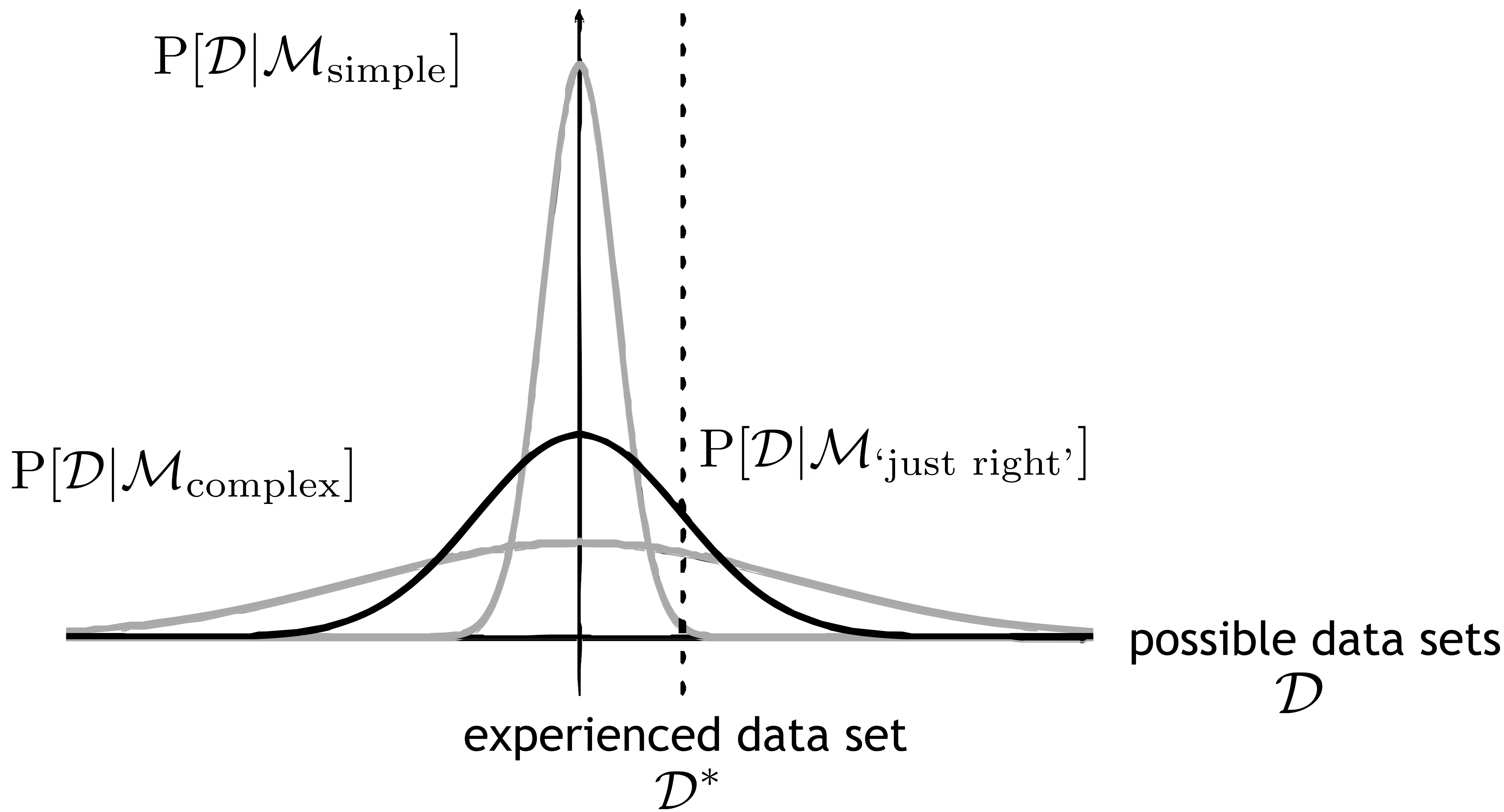
Rev. Bayes

a model defines a probability distribution over data sets

$$\mathrm{P}[\mathcal{D}|\mathcal{M}]$$

data coming from the environment

model stored in memory



BAYESIAN MODEL SELECTION

Rev. Bayes

a model defines a probability distribution over data sets

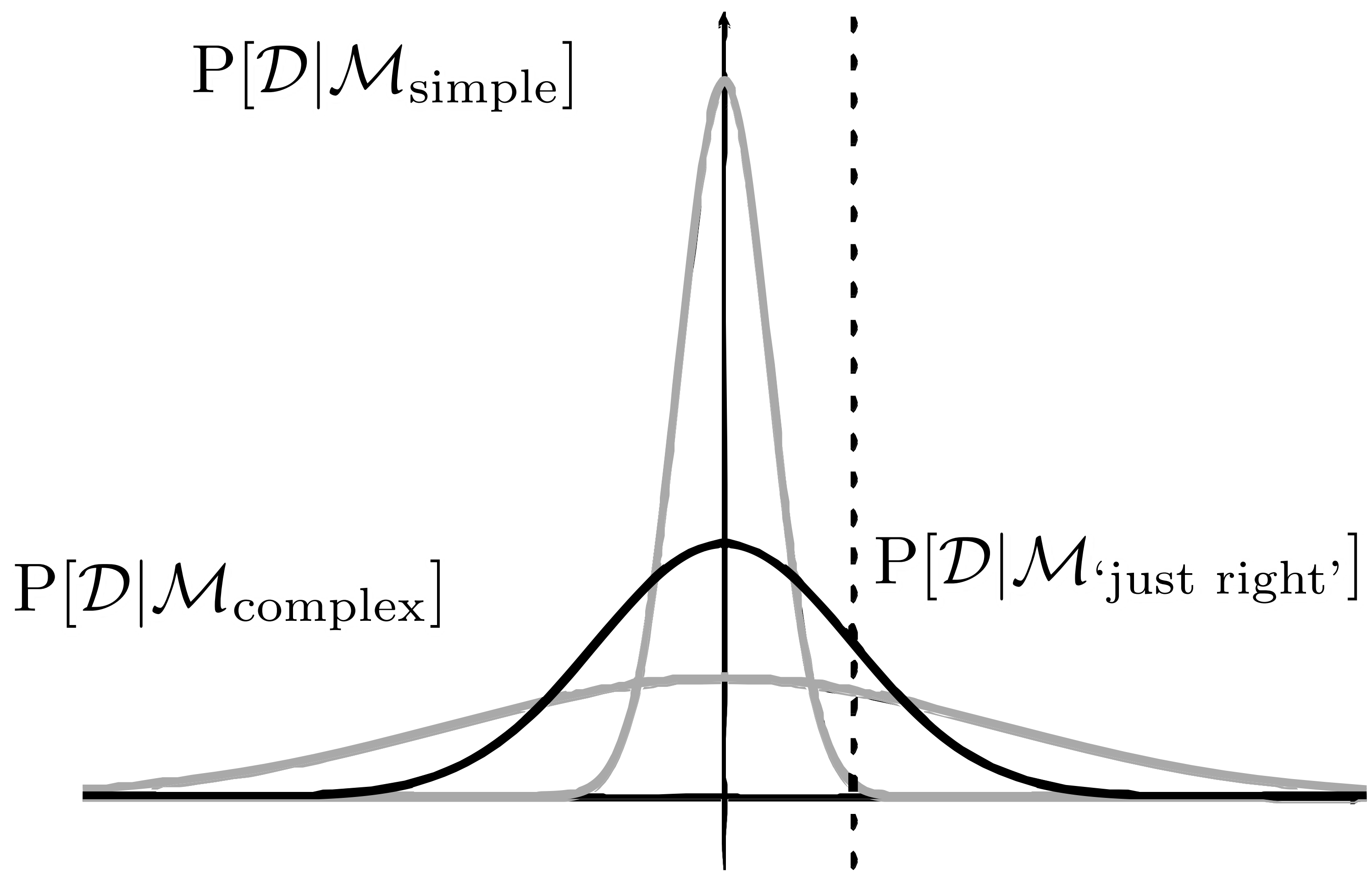
$\mathrm{P}[\mathcal{D}|\mathcal{M}]$

data coming from the environment

model stored in memory

Occam's razor

'entities should not be multiplied beyond necessity'



possible data sets $\mathcal{D}$

experienced data set $\mathcal{D}^*$

VISUAL PATTERN LEARNING

VISUAL PATTERN LEARNING

familiarization

‘pay attention’

VISUAL PATTERN LEARNING

familiarization

'pay attention'

VISUAL PATTERN LEARNING

familiarization

'pay attention'

test

'which one looks more familiar?'

VISUAL PATTERN LEARNING

familiarization

‘pay attention’

test

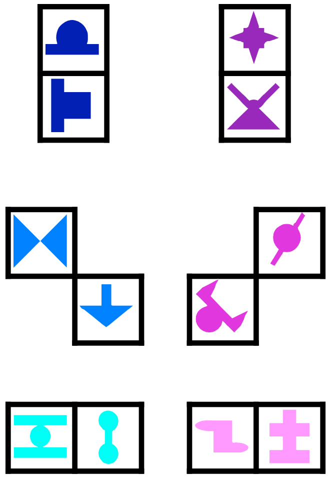
‘which one looks more familiar?’

VISUAL PATTERN LEARNING

familiarization

‘pay attention’

inventory



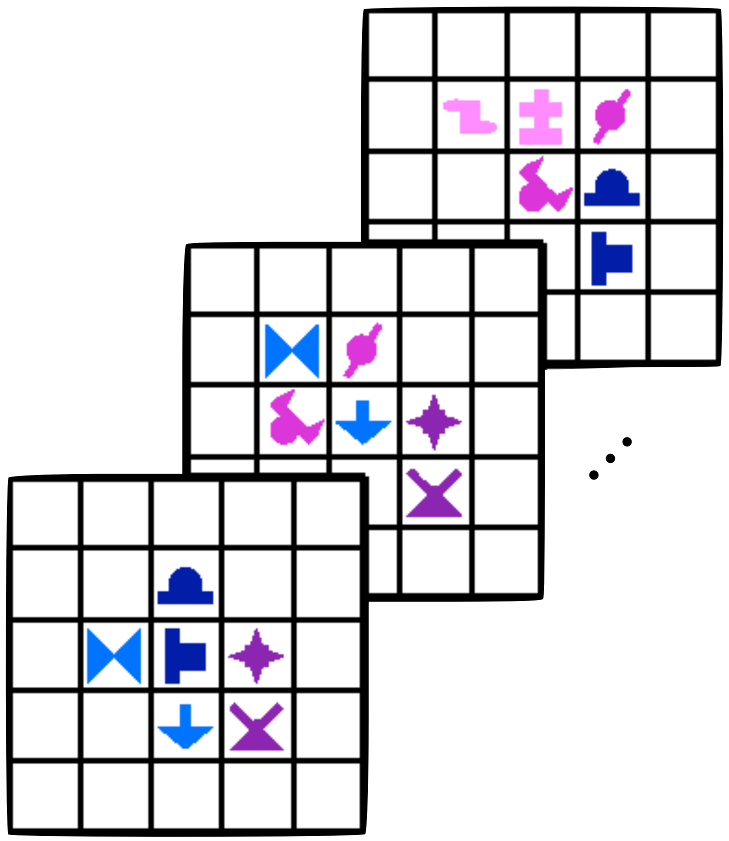
test

‘which one looks more familiar?’

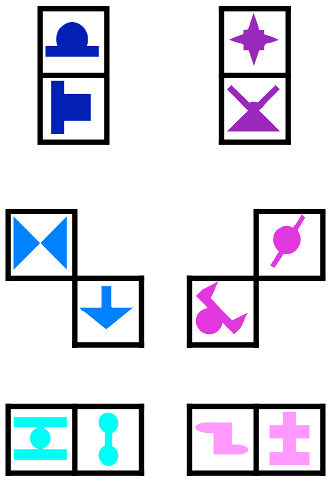
VISUAL PATTERN LEARNING

familiarization

'pay attention'



inventory



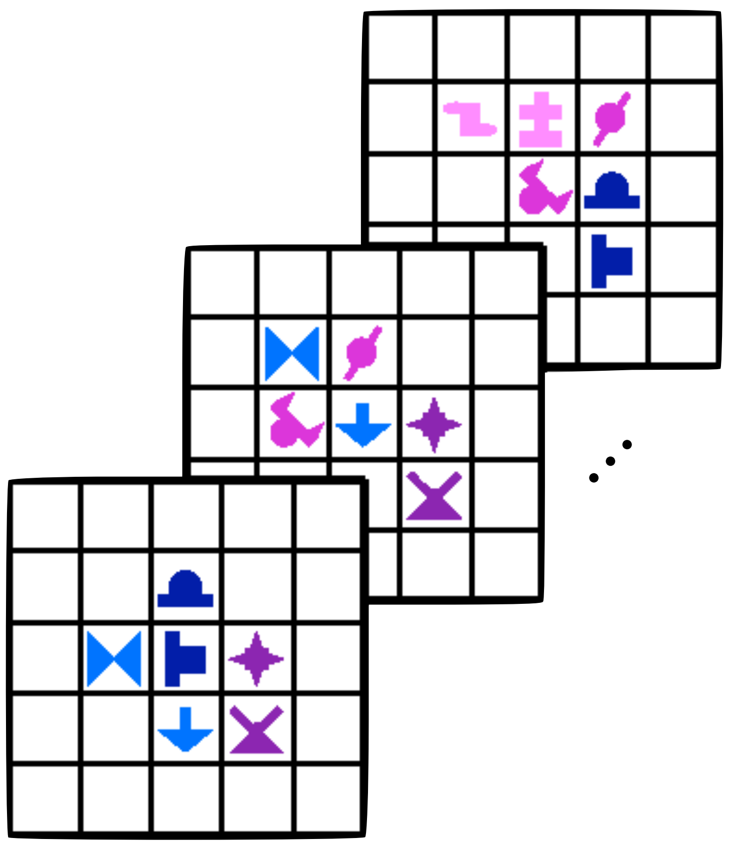
test

'which one looks more familiar?'

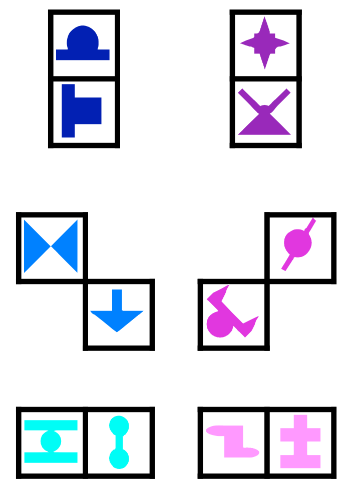
VISUAL PATTERN LEARNING

familiarization

'pay attention'

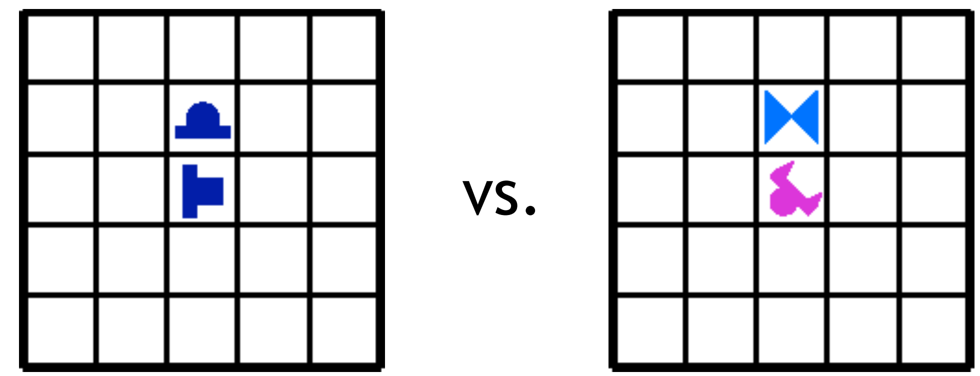


inventory



test

'which one looks more familiar?'

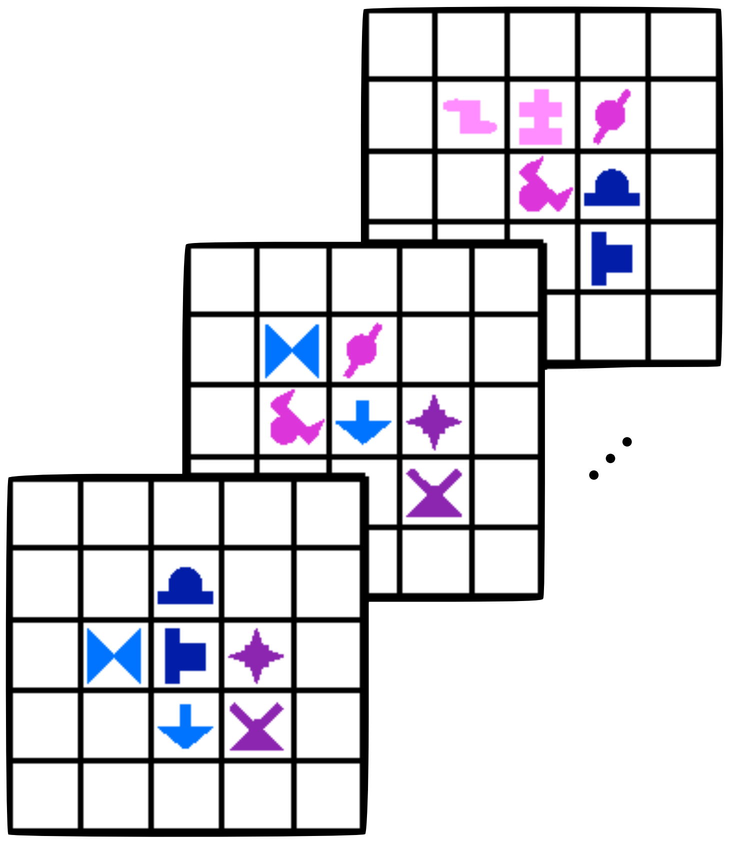


Orbán & al, PNAS 2008

VISUAL PATTERN LEARNING

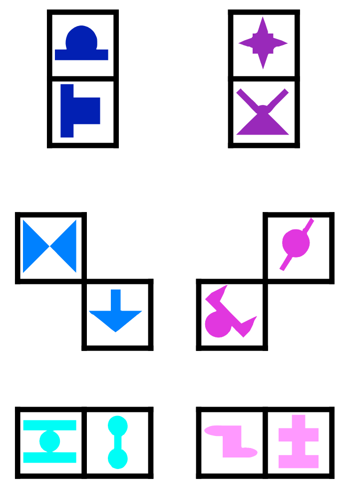
familiarization

'pay attention'



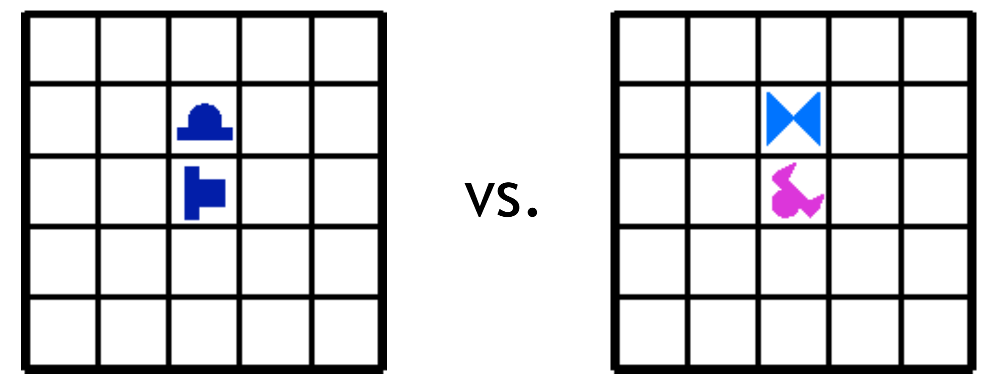
$\mathcal{D} = \text{scene}_1, \ldots \text{scene}_n$

inventory



test

'which one looks more familiar?'

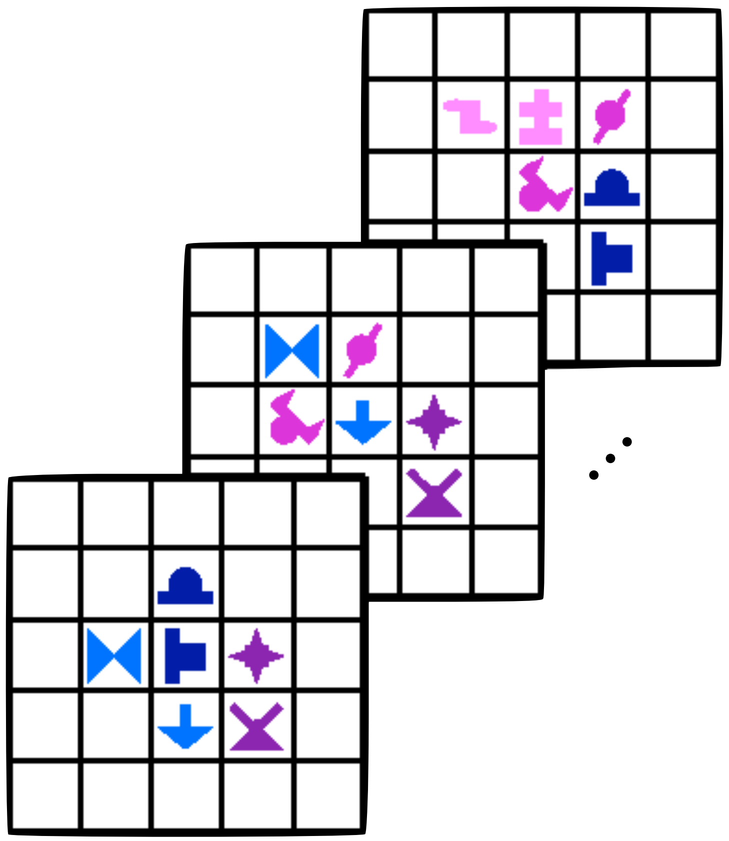


Orbán & al, PNAS 2008

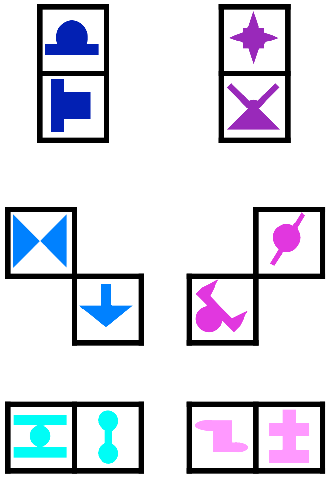
VISUAL PATTERN LEARNING

familiarization

'pay attention'



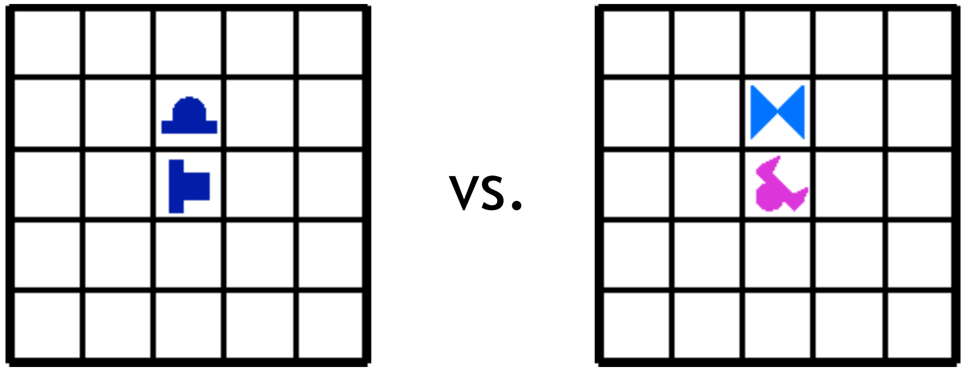
inventory



$\mathcal{M}$

test

'which one looks more familiar?'



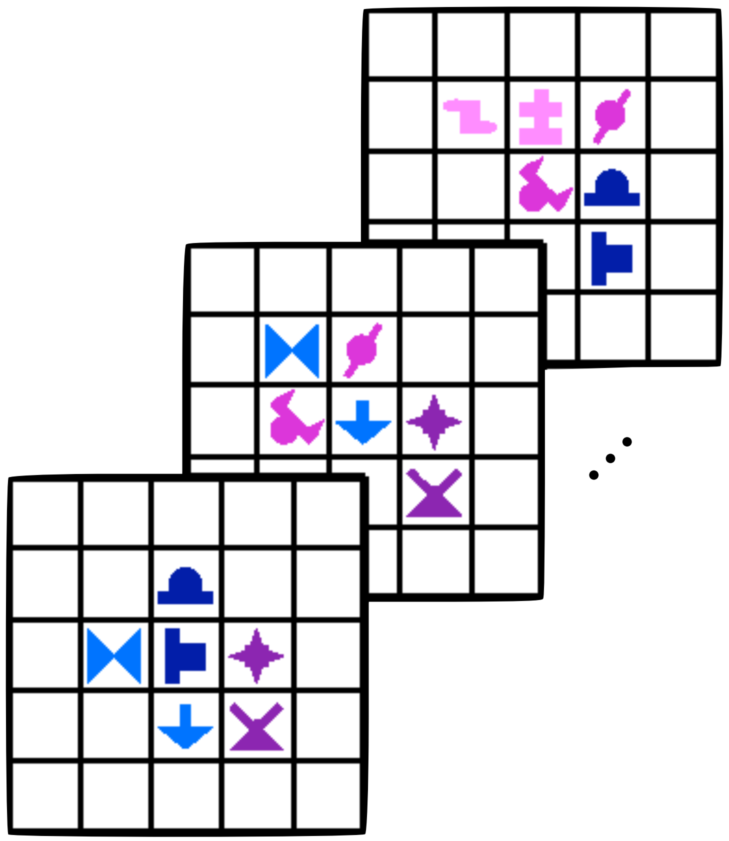
Orbán & al, PNAS 2008

$$\mathcal{D} = \text{scene}_1, \dots \text{scene}_n \longrightarrow \hat{\mathcal{M}} = \underset{\mathcal{M}}{\operatorname{argmax}} \, \mathrm{P}(\mathcal{M}|\mathcal{D})$$

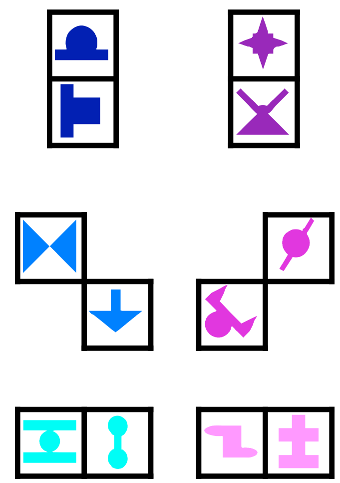
VISUAL PATTERN LEARNING

familiarization

‘pay attention’



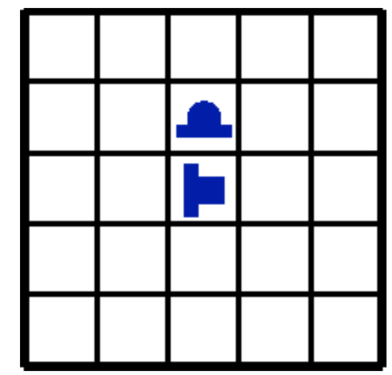
inventory



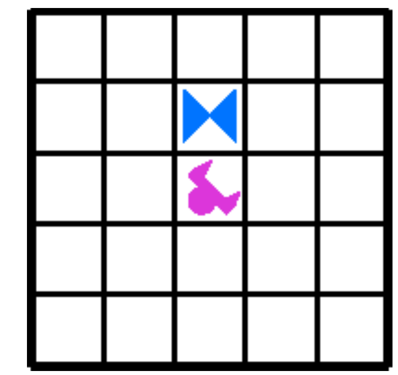
$\mathcal{M}$

test

‘which one looks more familiar?’



vs.



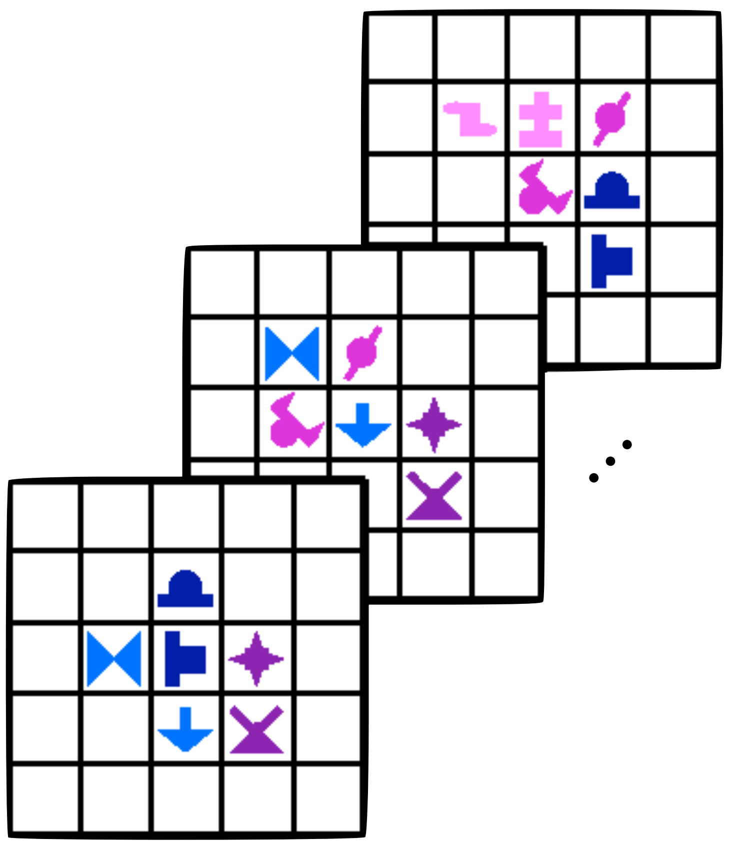
Orbán & al, PNAS 2008

$$\mathcal{D} = \text{scene}_1, \ldots \text{scene}_n \longrightarrow \hat{\mathcal{M}} = \underset{\mathcal{M}}{\text{argmax}}\, \mathrm{P}(\mathcal{M}|\mathcal{D}) \longrightarrow \mathrm{P}\left(\text{scene}_\mathrm{A}|\hat{\mathcal{M}}\right) \text{ vs. } \mathrm{P}\left(\text{scene}_\mathrm{B}|\hat{\mathcal{M}}\right)$$

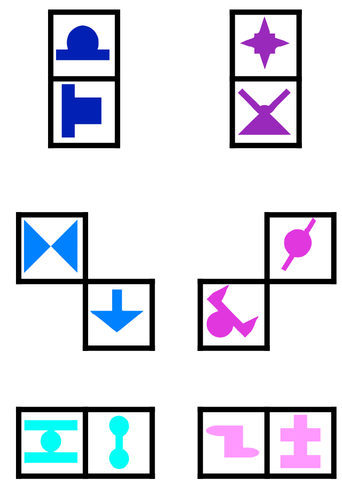
VISUAL PATTERN LEARNING

familiarization

‘pay attention’



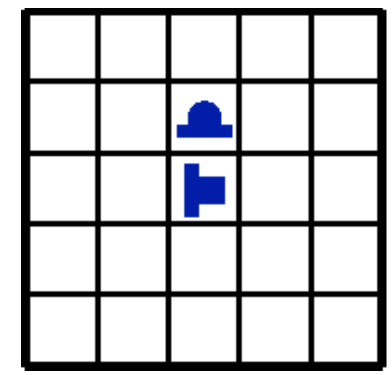
inventory



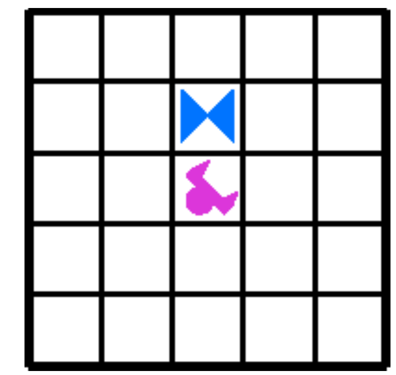
$\mathcal{M}$

test

‘which one looks more familiar?’



vs.



Orbán & al, PNAS 2008

$$\mathcal{D} = \text{scene}_1, \dots \text{scene}_n \rightarrow \hat{\mathcal{M}} = \underset{\mathcal{M}}{\operatorname{argmax}}\, \mathrm{P}(\mathcal{M}|\mathcal{D}) \rightarrow \mathrm{P}\Big(\text{scene}_\mathrm{A}|\hat{\mathcal{M}}\Big) \text{ vs. } \mathrm{P}\Big(\text{scene}_\mathrm{B}|\hat{\mathcal{M}}\Big)$$

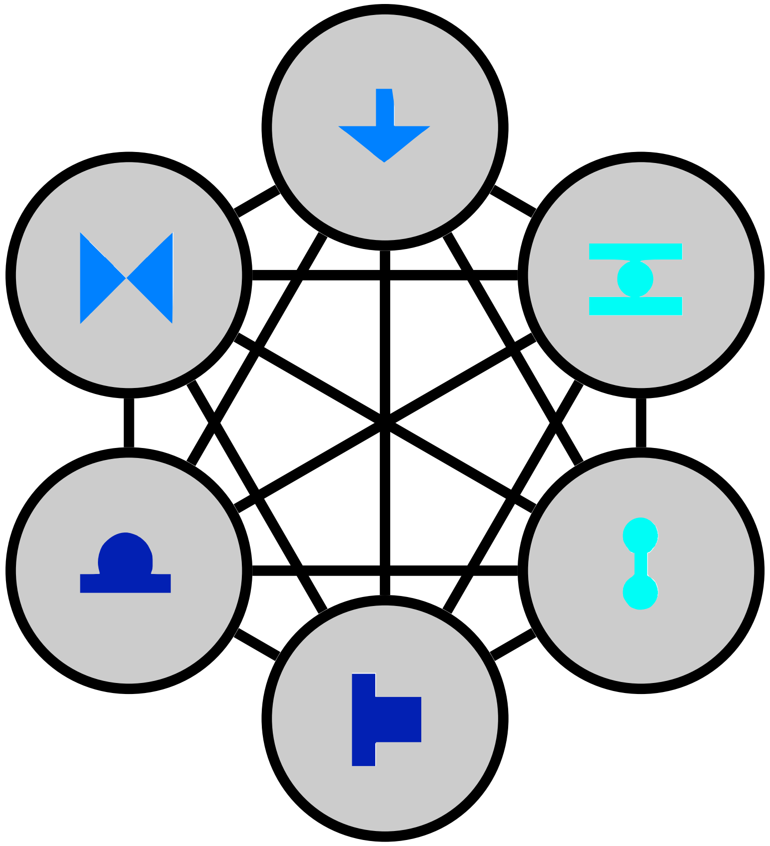
how do humans learn a statistical model of their environment?

- associative learning (fitting 2nd order max-entropy model)
- Bayesian model selection (inferring hidden causal structure)

ALTERNATIVE THEORIES

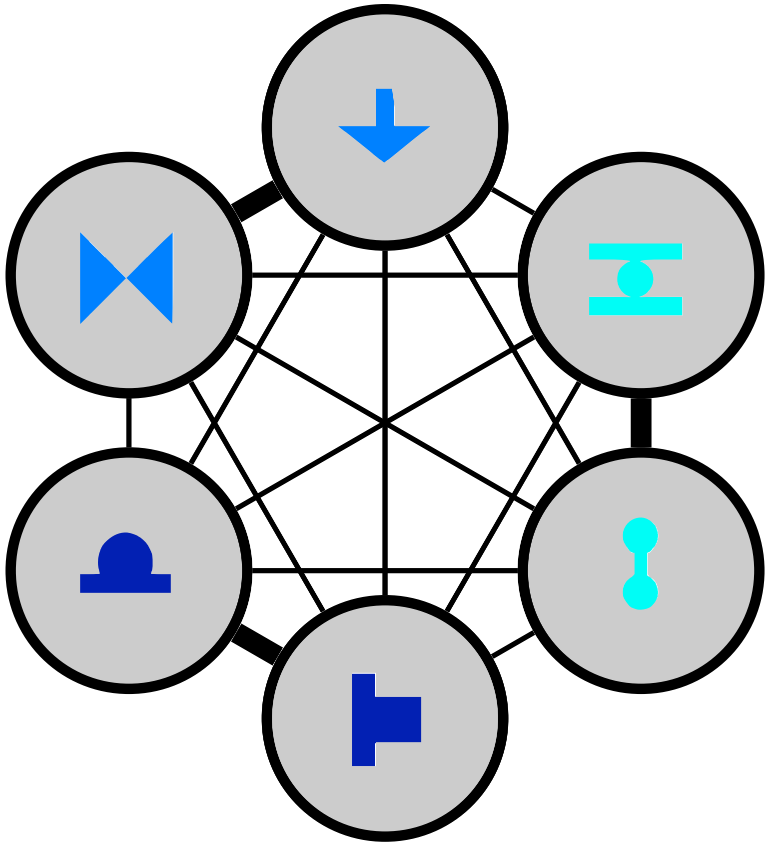
ALTERNATIVE THEORIES

associative learning



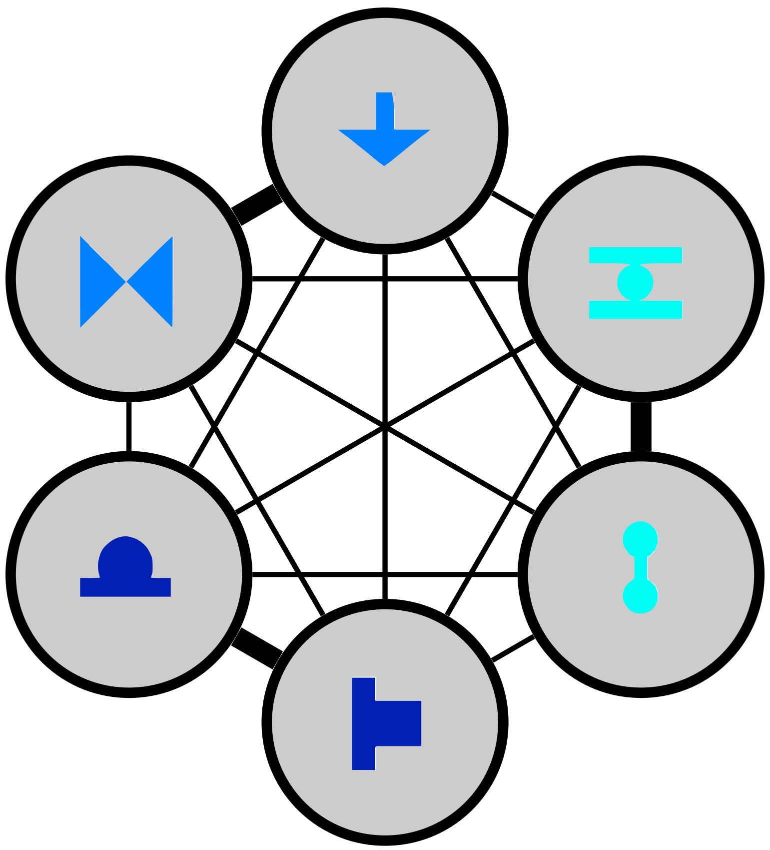
ALTERNATIVE THEORIES

associative learning

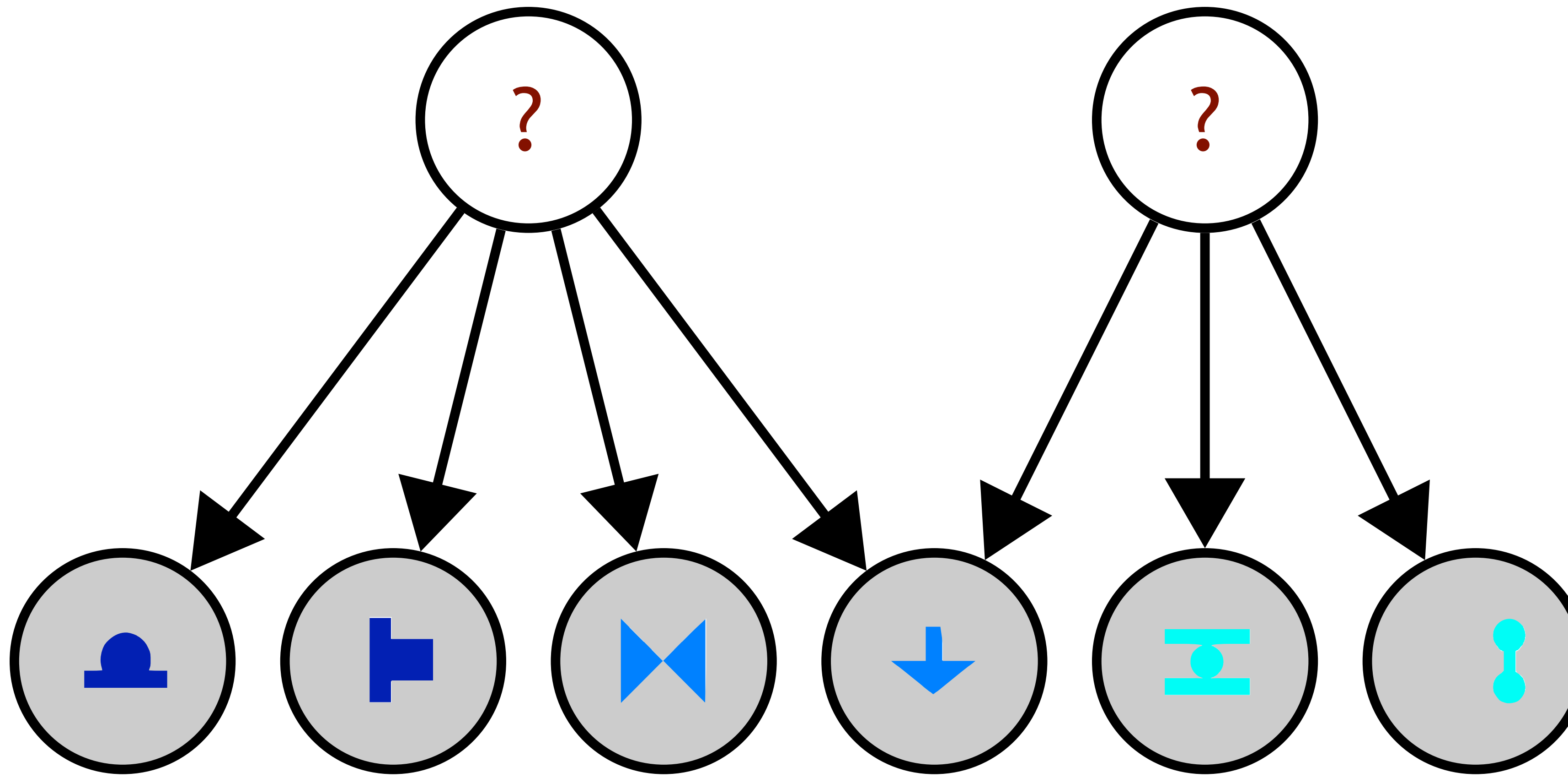


ALTERNATIVE THEORIES

associative learning

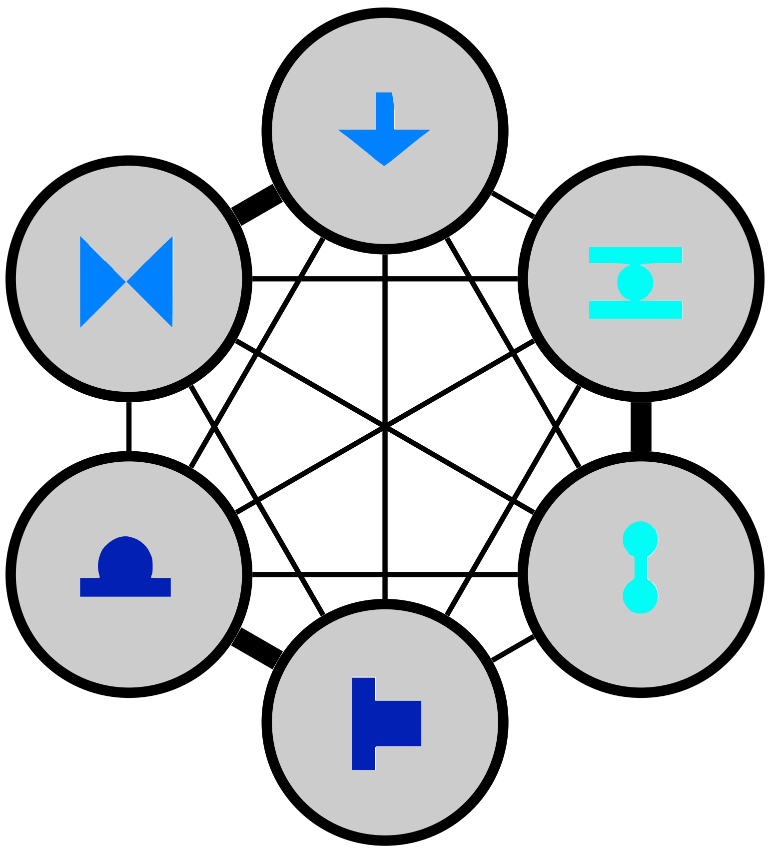


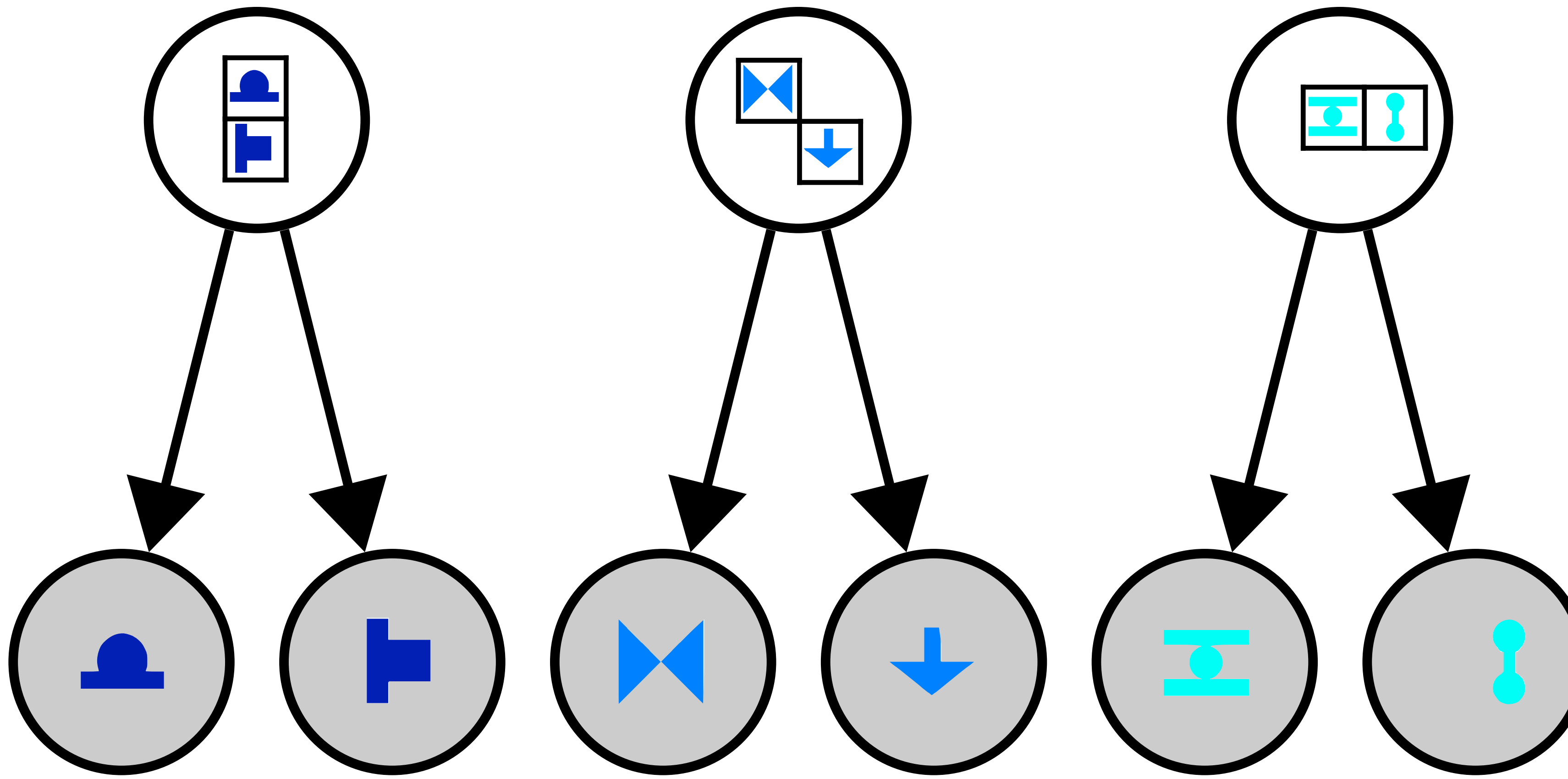
Bayesian learning



ALTERNATIVE THEORIES

associative learning | **Bayesian learning**



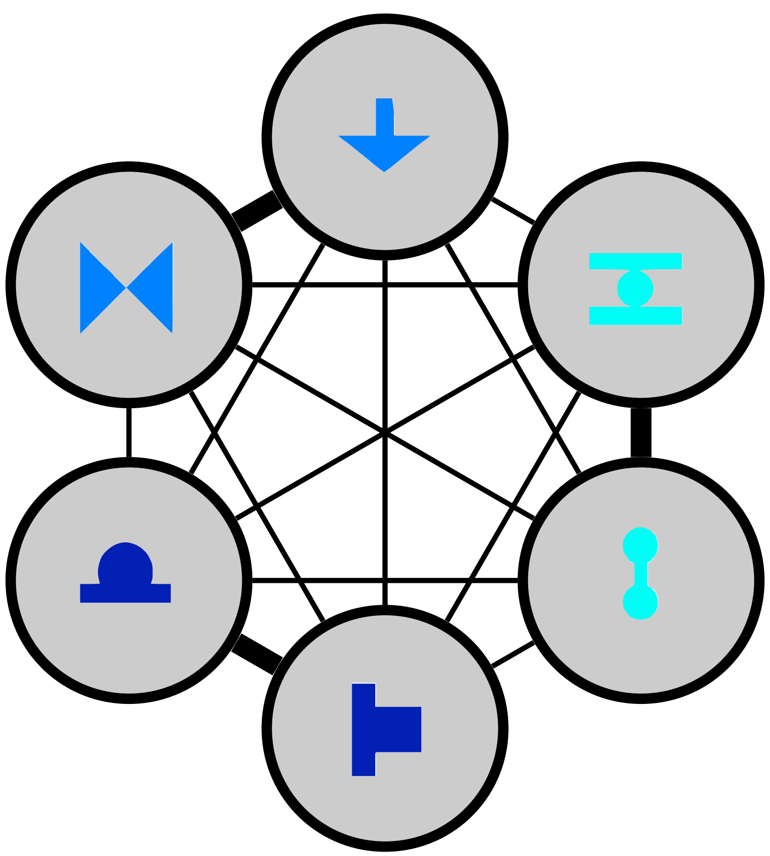


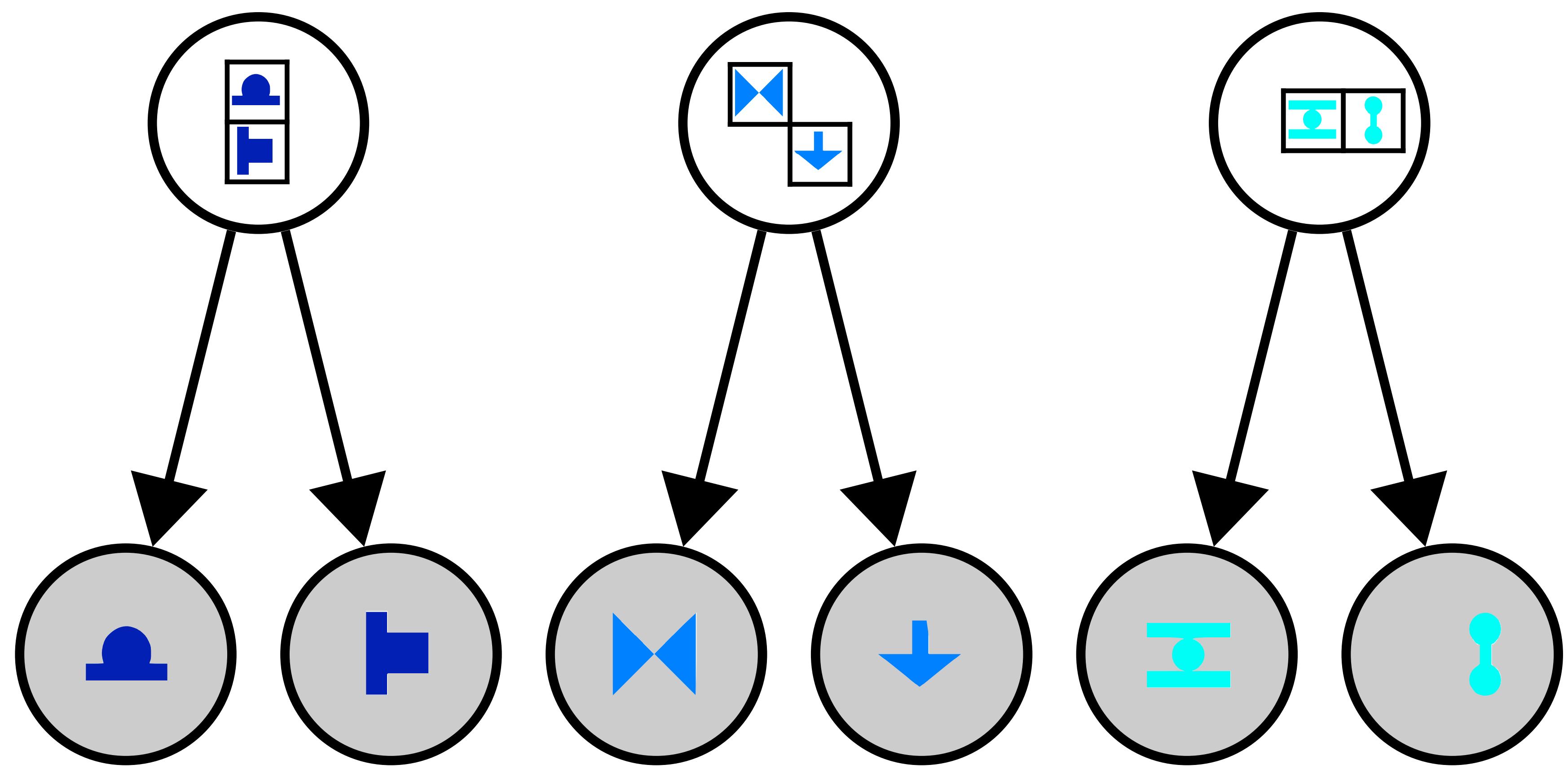
Orbán & al, PNAS 2008

ALTERNATIVE THEORIES

associative learning

Bayesian learning





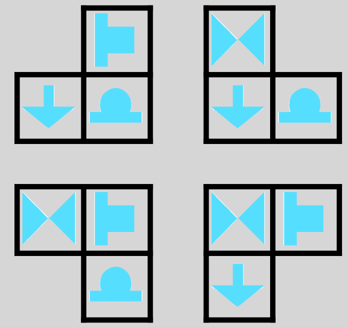
Orbán & al, PNAS 2008

Boltzmann machine
+ Gaussian Markov random field

sigmoid belief network
+ product of (conditional) Gaussian experts

ASSOCIATIVE VS. BAYESIAN LEARNING

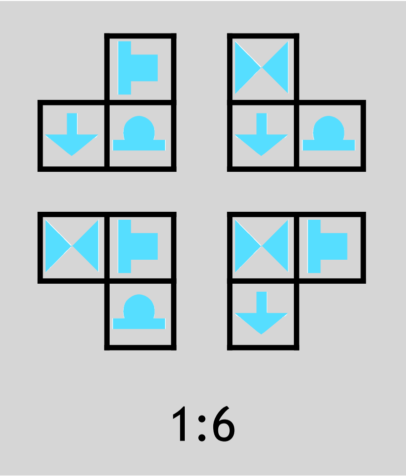
inventory:

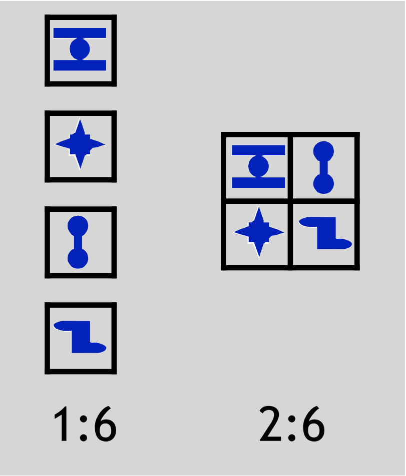


1:6

ASSOCIATIVE VS. BAYESIAN LEARNING

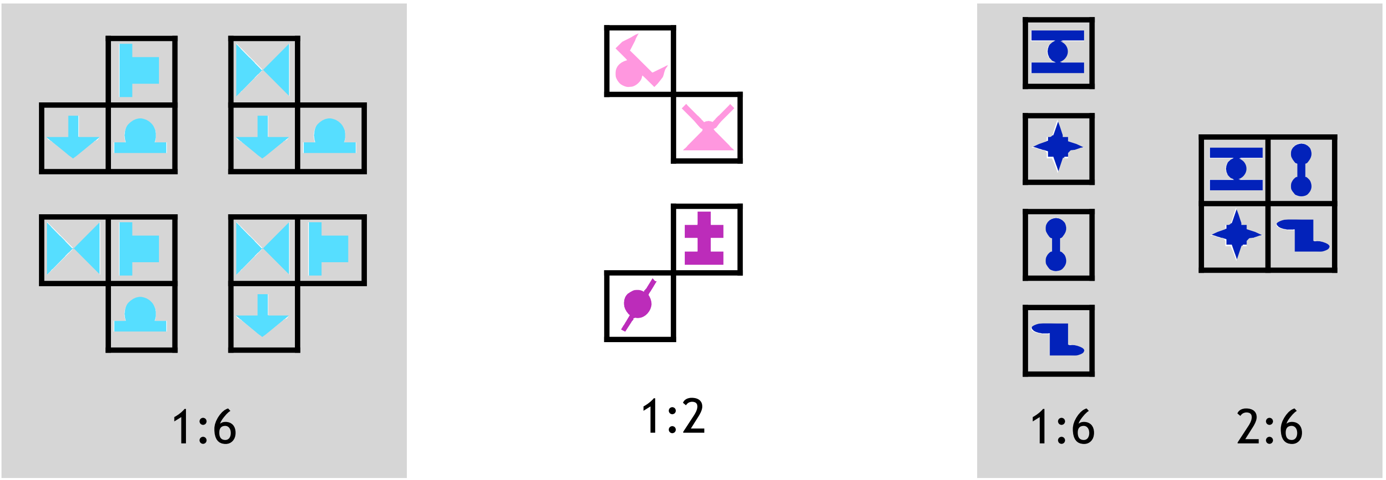
inventory:





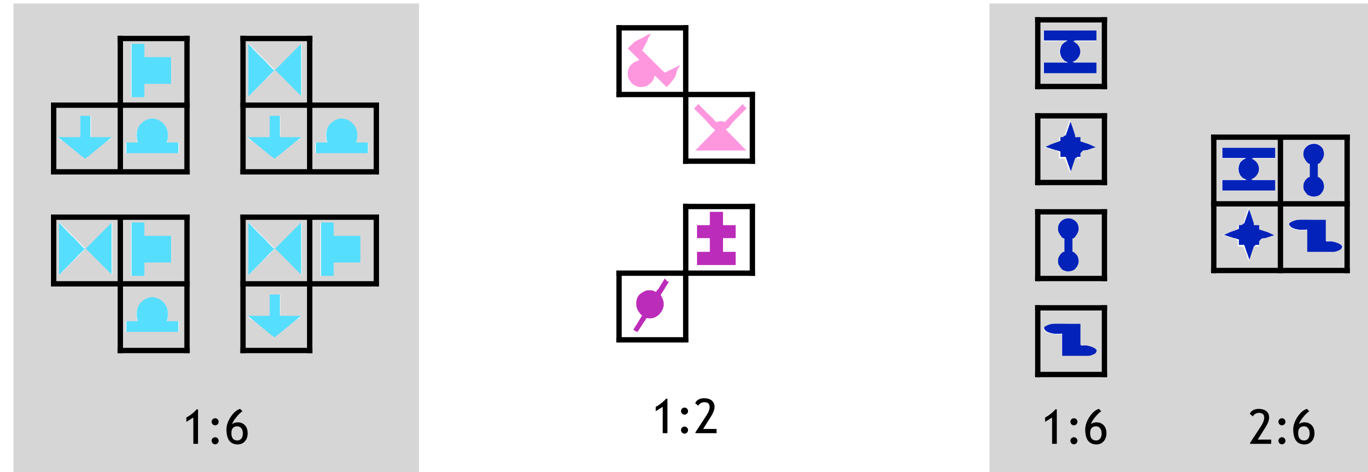
ASSOCIATIVE VS. BAYESIAN LEARNING

inventory:



ASSOCIATIVE VS. BAYESIAN LEARNING

inventory:



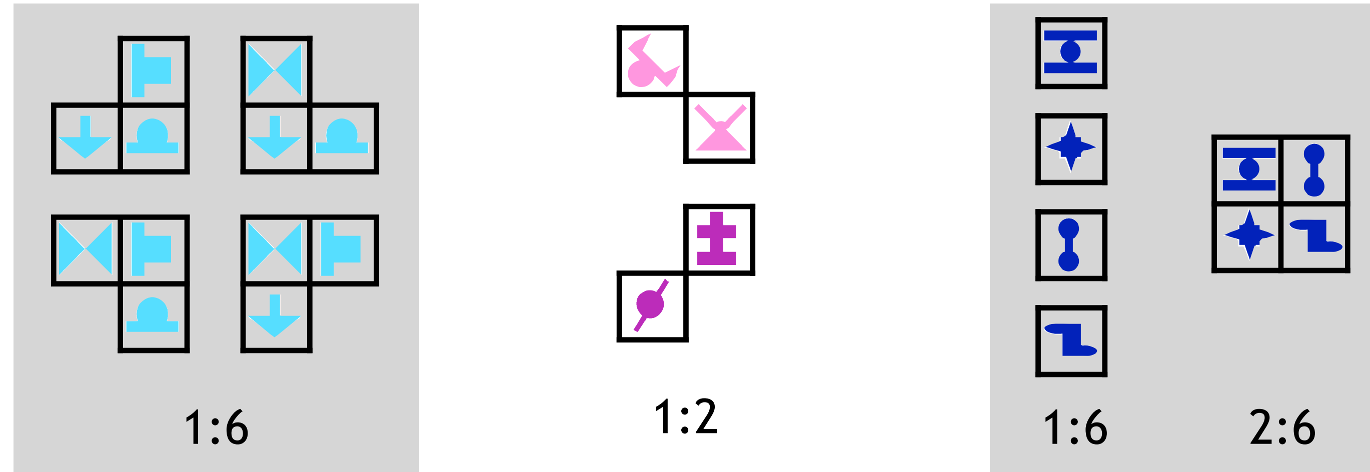
1st order statistic:
shape frequencies

3 × 1/6

1/6 + 2/6

ASSOCIATIVE VS. BAYESIAN LEARNING

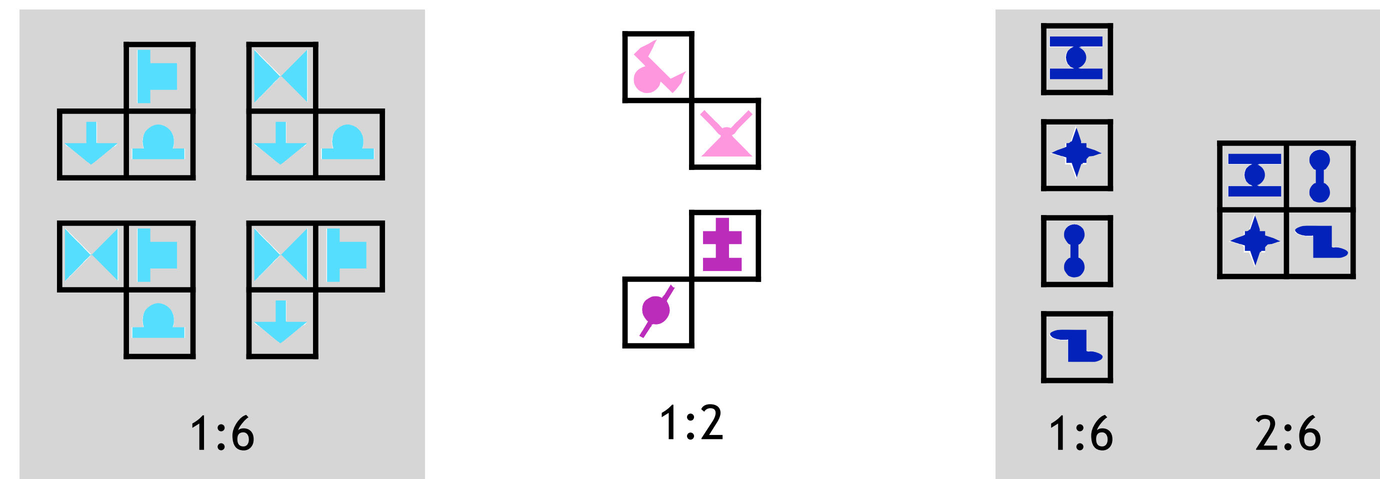
inventory:



1st order statistic: shape frequencies		3 × 1/6		1/6 + 2/6
2nd order statistic: pairwise correlations	both present	2 × 1/6		2/6
	both absent	2/6		2 × 1/6
	one present	2 × 1/6		2 × 1/6

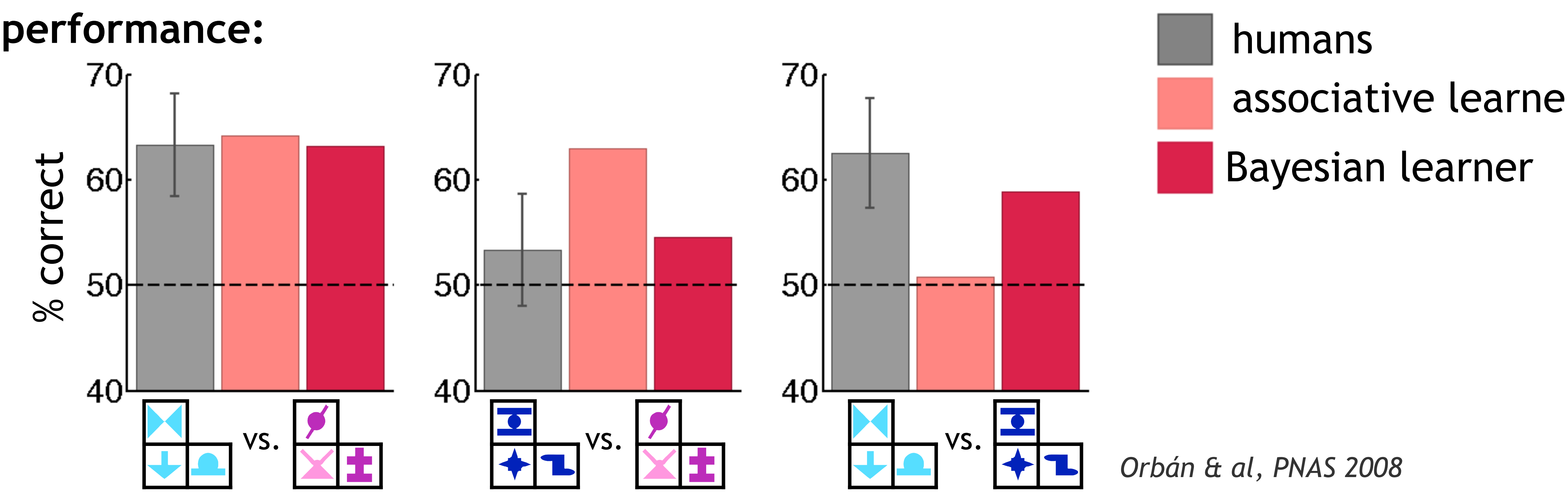
ASSOCIATIVE VS. BAYESIAN LEARNING

inventory:



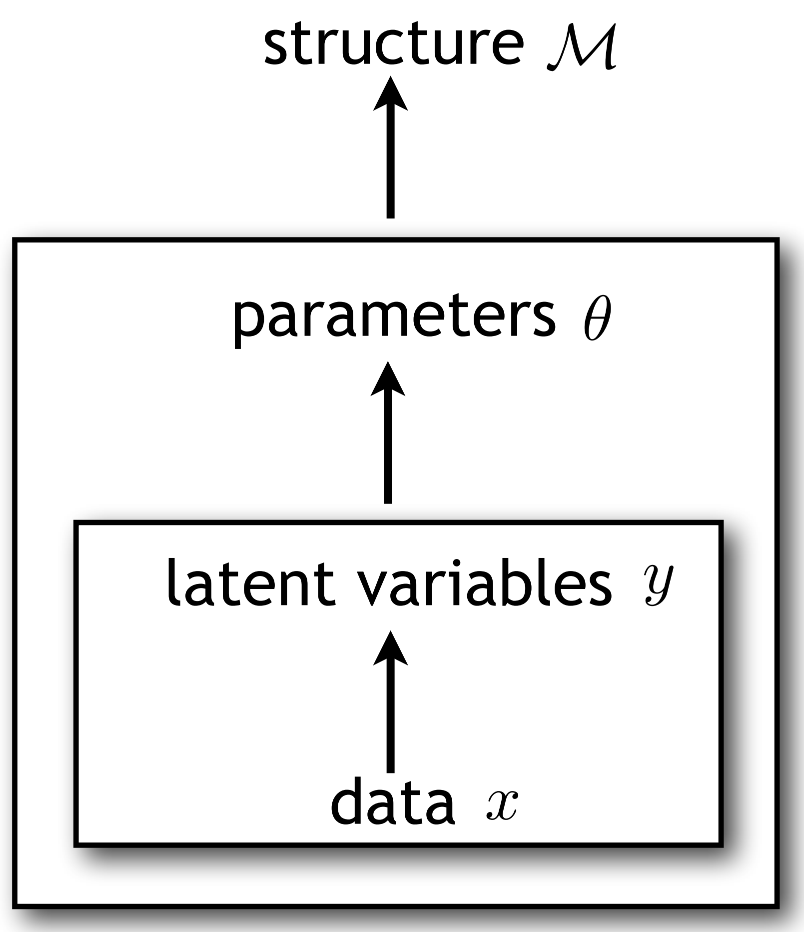
1st order statistic: shape frequencies		$3 \times 1/6$	$1/6 + 2/6$
2nd order statistic: pairwise correlations	both present	$2 \times 1/6$	$2/6$
	both absent	$2/6$	$2 \times 1/6$
	one present	$2 \times 1/6$	$2 \times 1/6$

test performance:

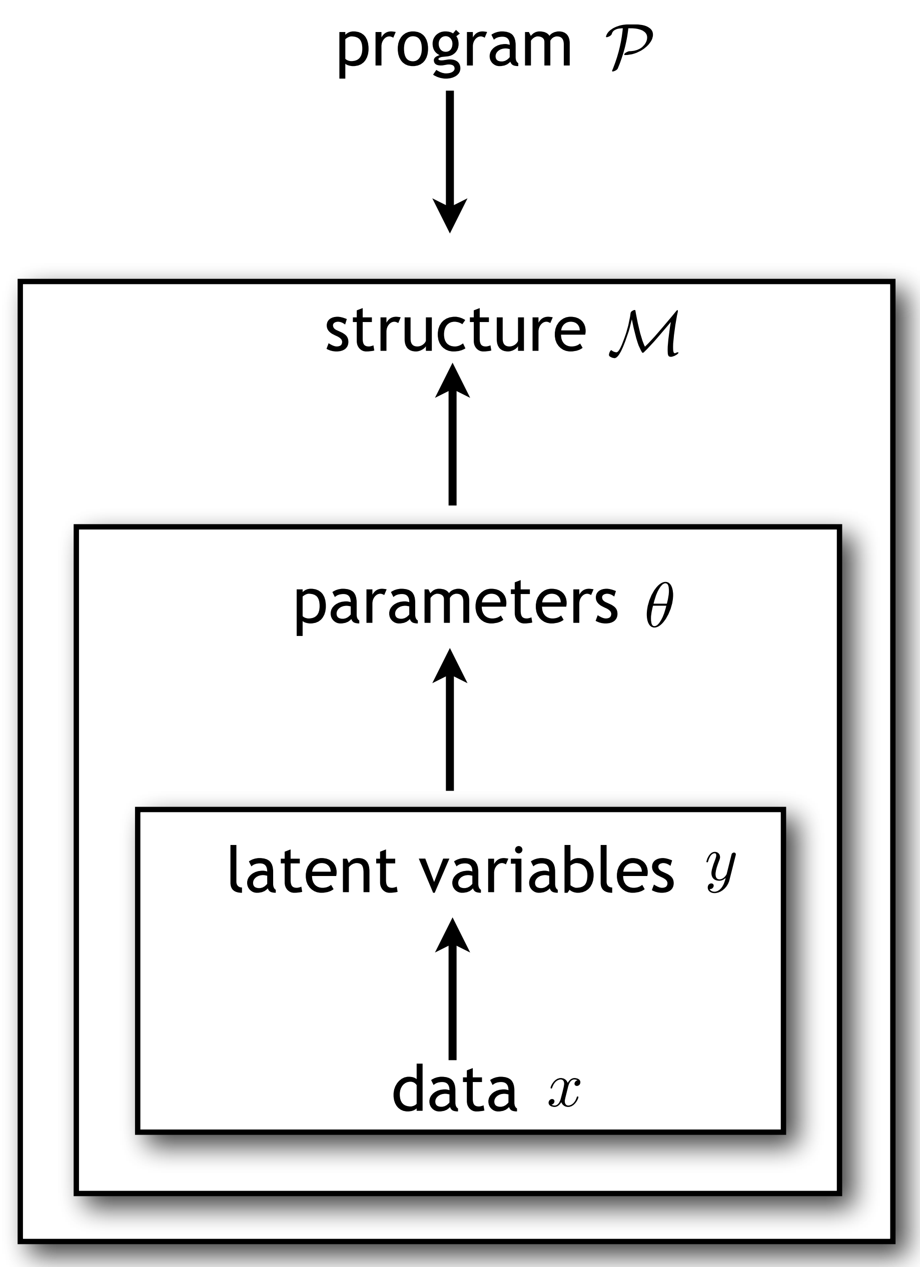


Orbán & al, PNAS 2008

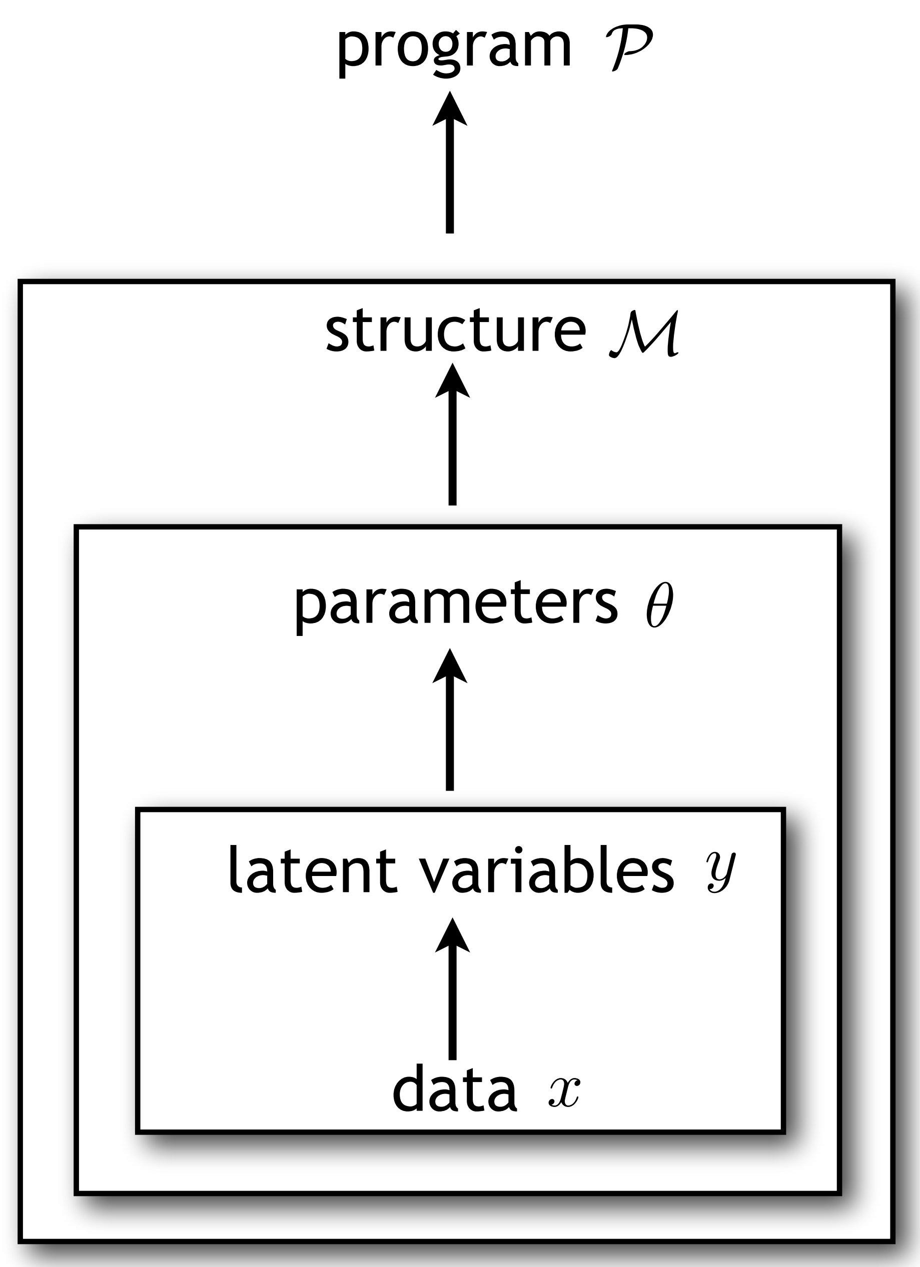
PROBABILISTIC INFERENCE AND LEARNING



PROBABILISTIC INFERENCE AND LEARNING

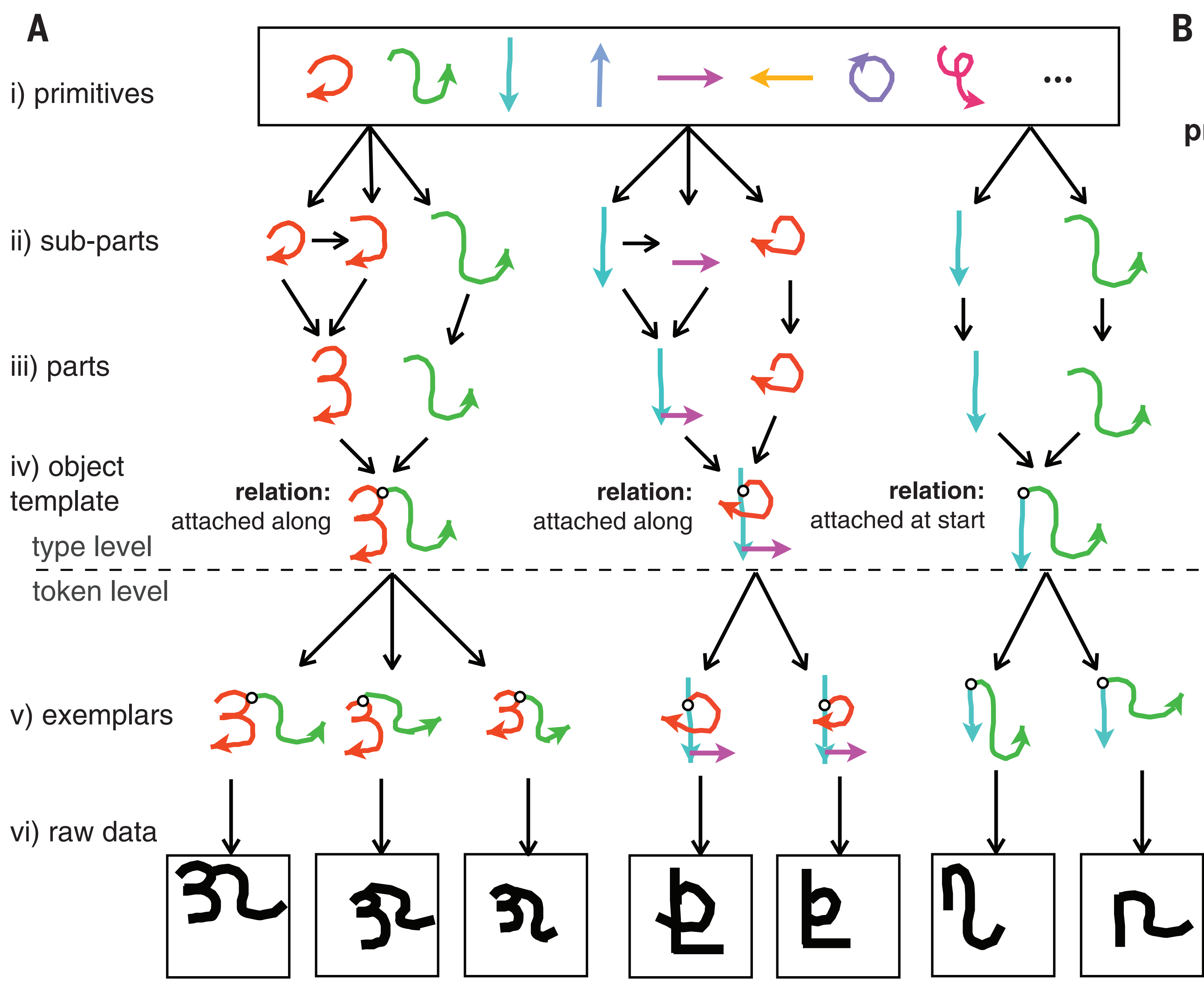


PROBABILISTIC INFERENCE AND LEARNING



CONCEPT LEARNING BY BAYESIAN PROGRAM LEARNING

A



B

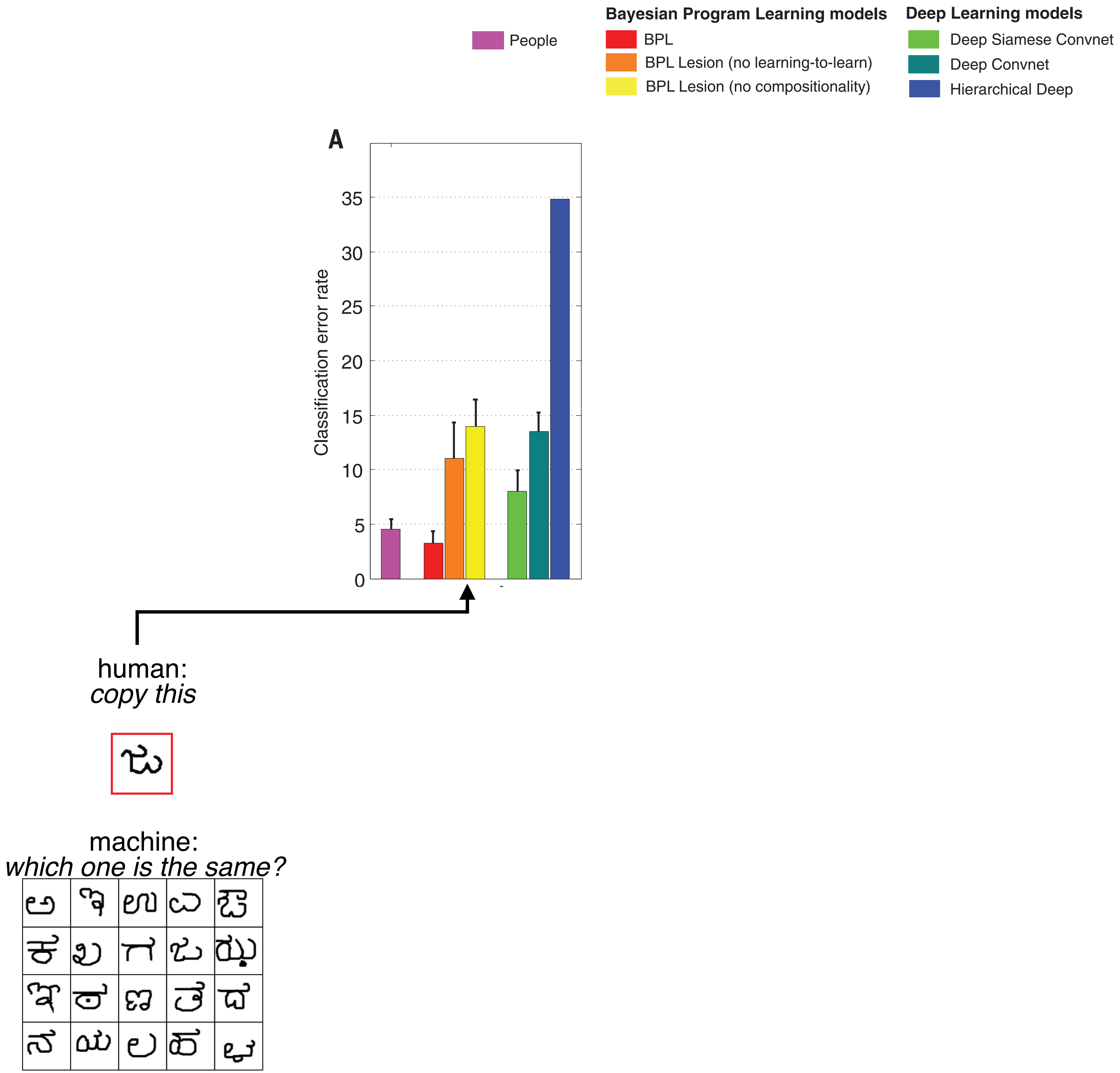
procedure GENERATETYPE
  $\kappa \leftarrow P(\kappa)$ ▷ Sample number of parts
  for i = 1 ... κ **do**
    $n_i \leftarrow P(n_i|\kappa)$ ▷ Sample number of sub-parts
    for j = 1 ... n_i **do**
      $s_{ij} \leftarrow P(s_{ij}|s_{i(j-1)})$ ▷ Sample sub-part sequence
    end for
    $R_i \leftarrow P(R_i|S_1, ..., S_{i-1})$ ▷ Sample relation
  end for
  $\psi \leftarrow \{\kappa, R, S\}$
  return @GENERATETOKEN(ψ) ▷ Return program

procedure GENERATETOKEN(ψ)
  for $i = 1 ... \kappa$ **do**
    $S_i^{(m)} \leftarrow P(S_i^{(m)}|S_i)$ ▷ Add motor variance
    $L_i^{(m)} \leftarrow P(L_i^{(m)}|R_i, T_1^{(m)}, ..., T_{i-1}^{(m)})$ ▷ Sample part's start location
    $T_i^{(m)} \leftarrow f(L_i^{(m)}, S_i^{(m)})$ ▷ Compose a part's trajectory
  end for
  $A^{(m)} \leftarrow P(A^{(m)})$ ▷ Sample affine transform
  $I^{(m)} \leftarrow P(I^{(m)}|T^{(m)}, A^{(m)})$ ▷ Sample image
  return $I^{(m)}$

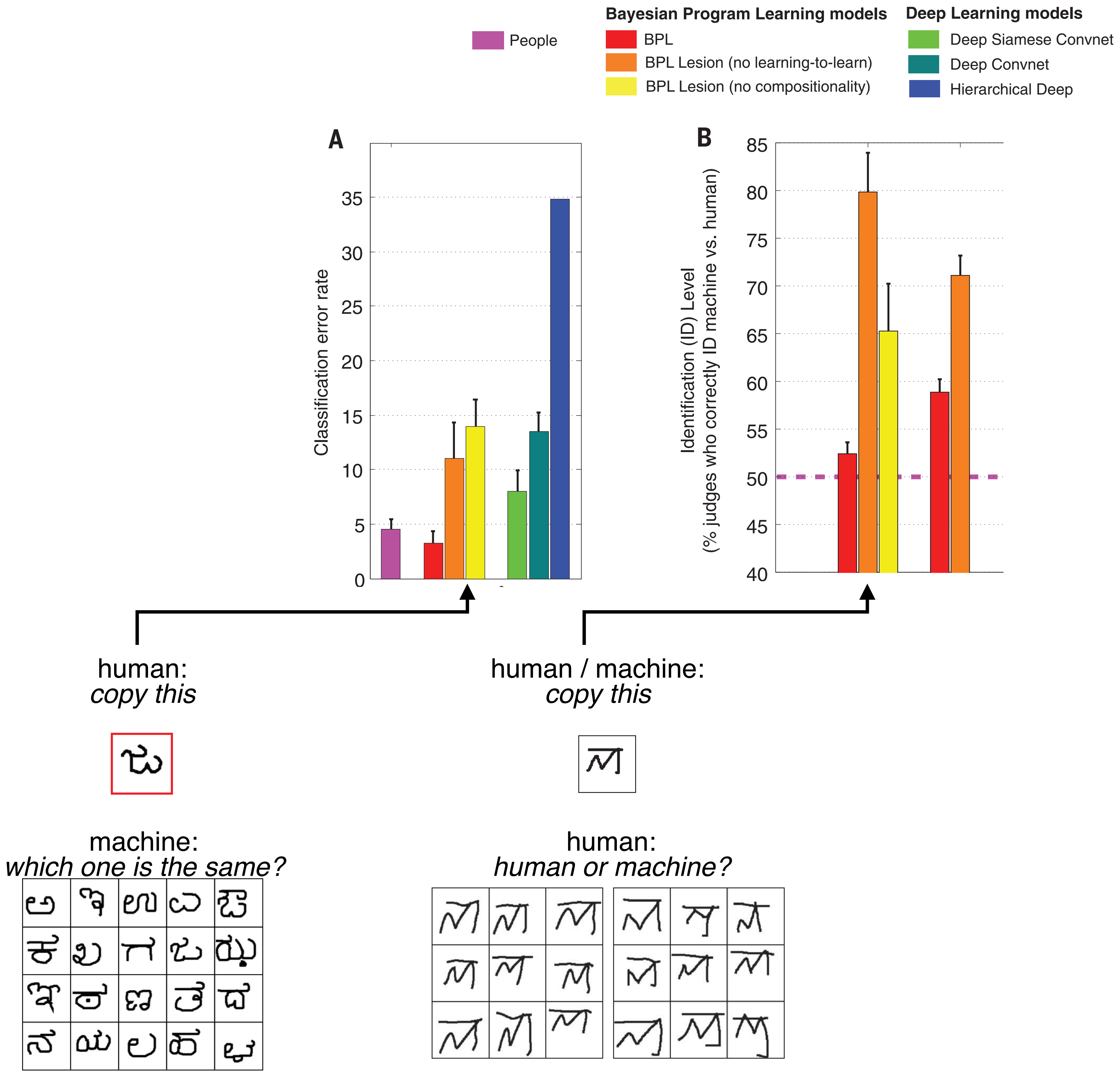
Lake et al, Science 2015

CONCEPT LEARNING
BY BAYESIAN PROGRAM LEARNING

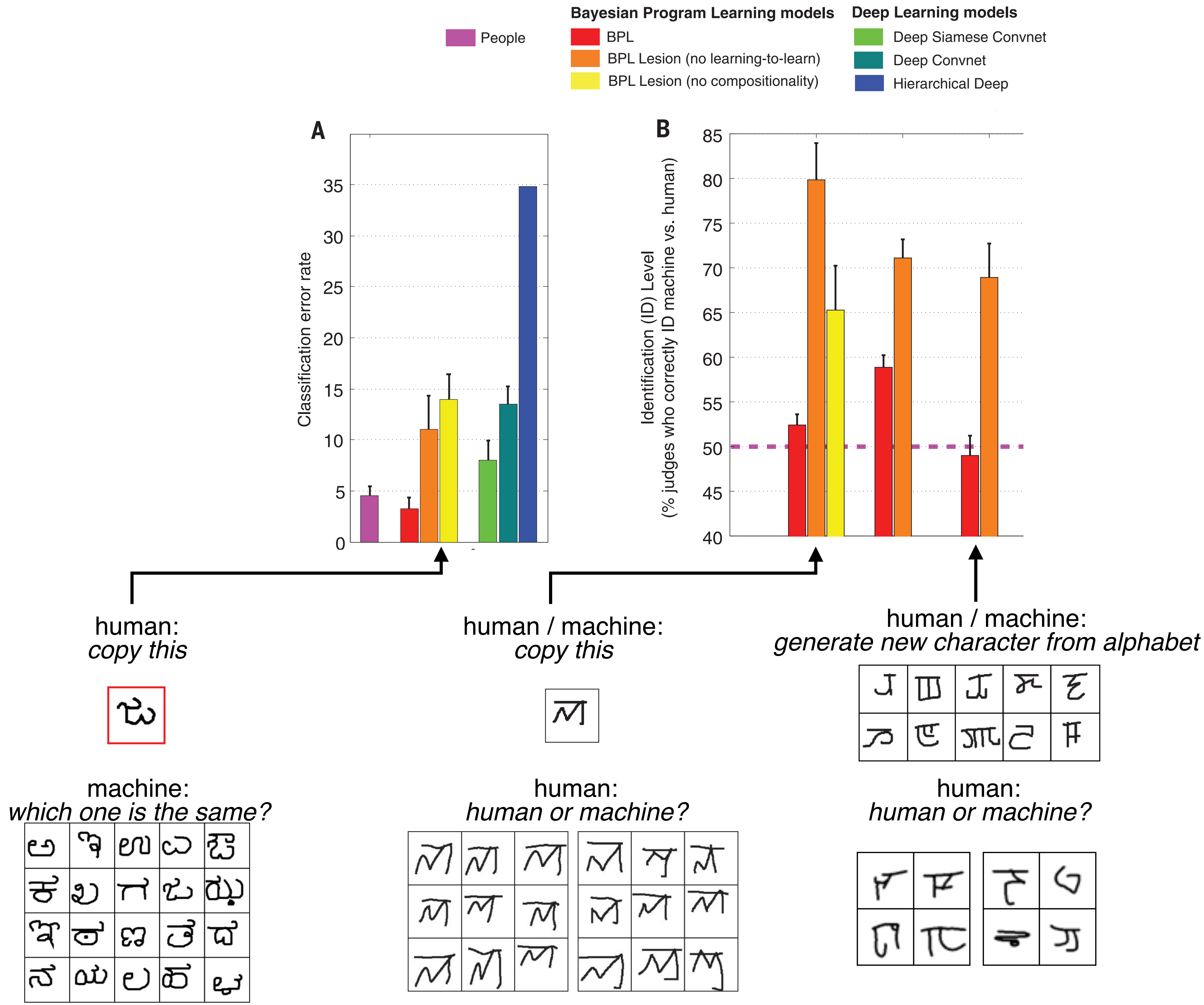
CONCEPT LEARNING
BY BAYESIAN PROGRAM LEARNING



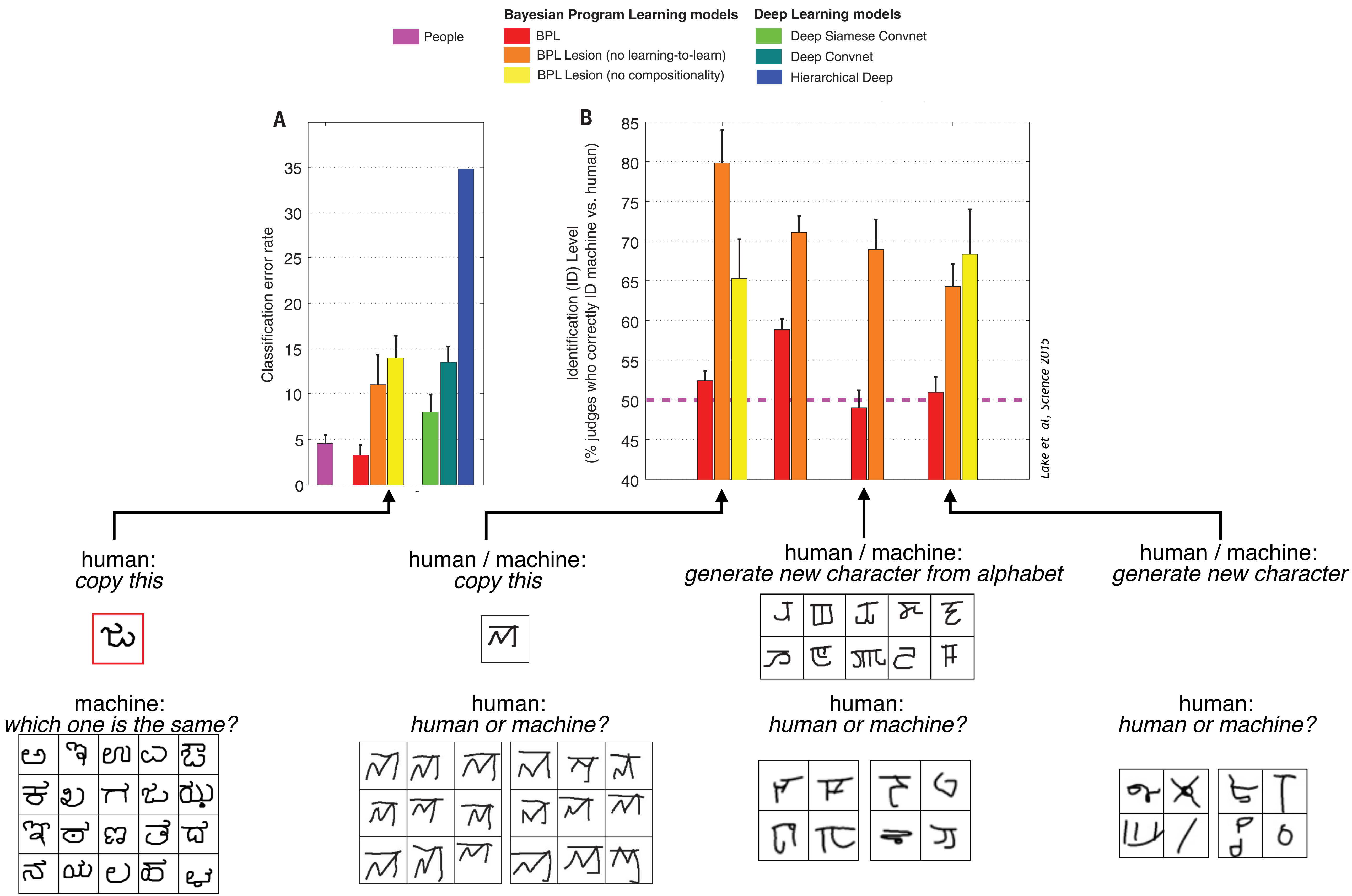
CONCEPT LEARNING BY BAYESIAN PROGRAM LEARNING



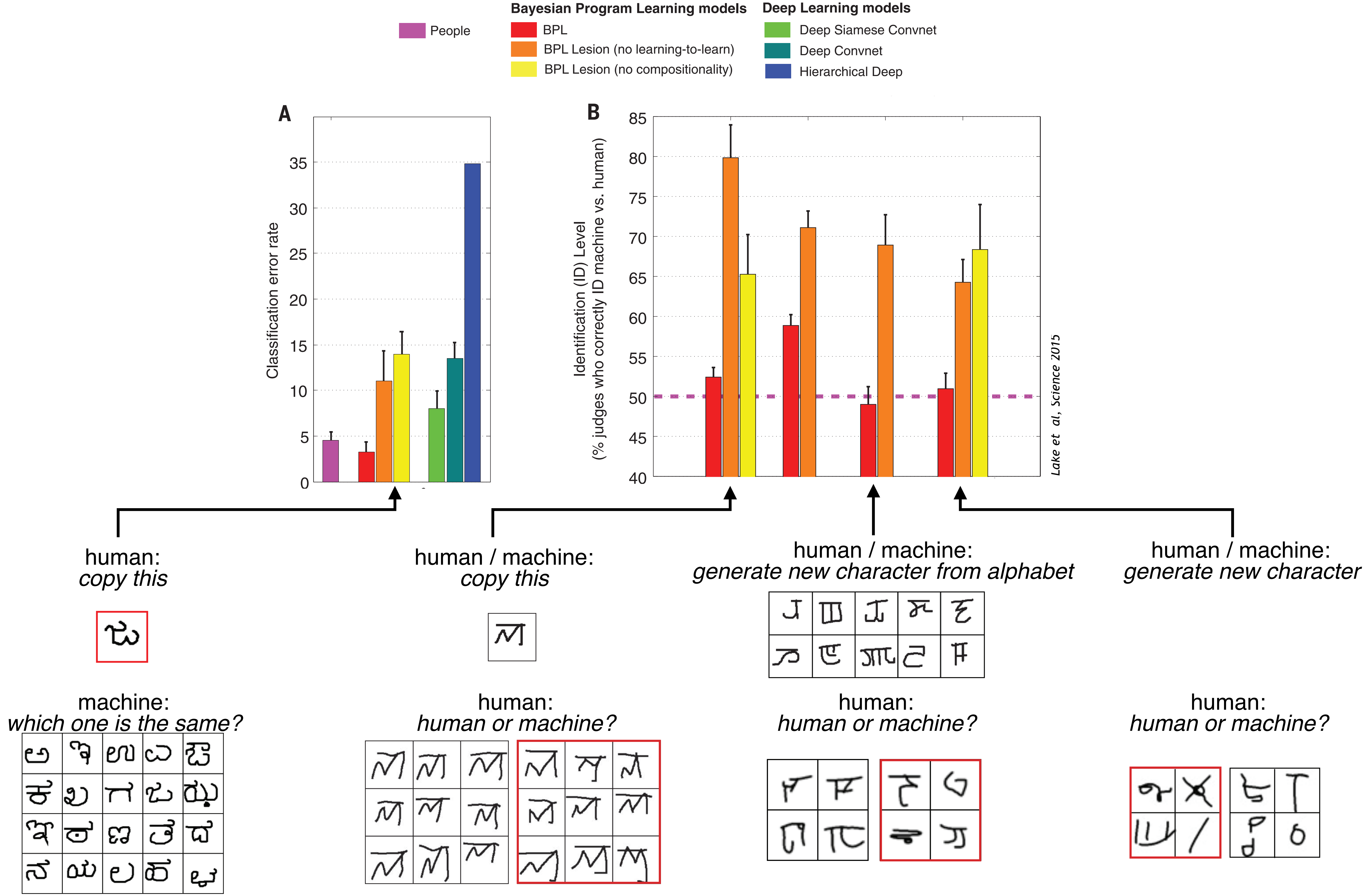
CONCEPT LEARNING
BY BAYESIAN PROGRAM LEARNING



CONCEPT LEARNING BY BAYESIAN PROGRAM LEARNING



CONCEPT LEARNING BY BAYESIAN PROGRAM LEARNING



PROBABILISTIC INFERENCE AND LEARNING

PROBABILISTIC INFERENCE AND LEARNING

data x

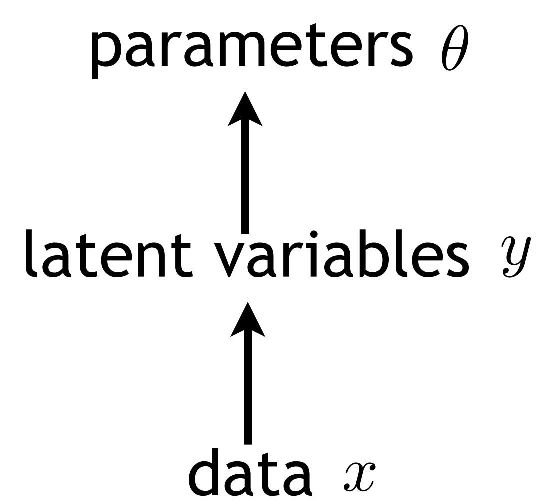
PROBABILISTIC INFERENCE AND LEARNING

latent variables y

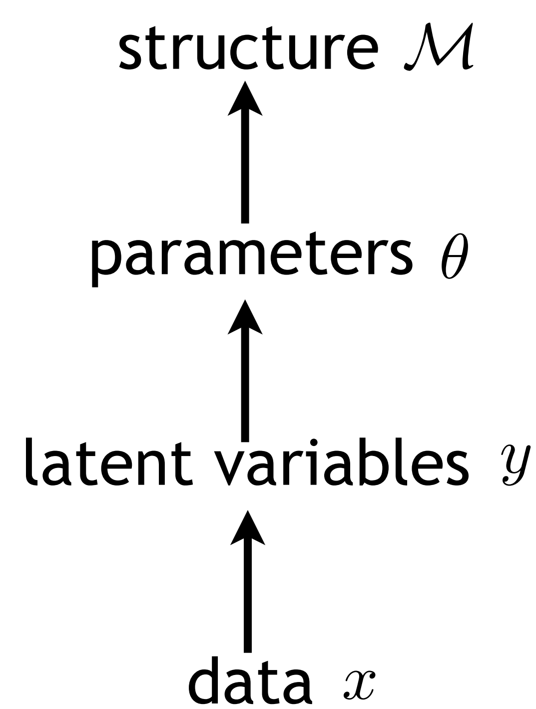
↑

data x

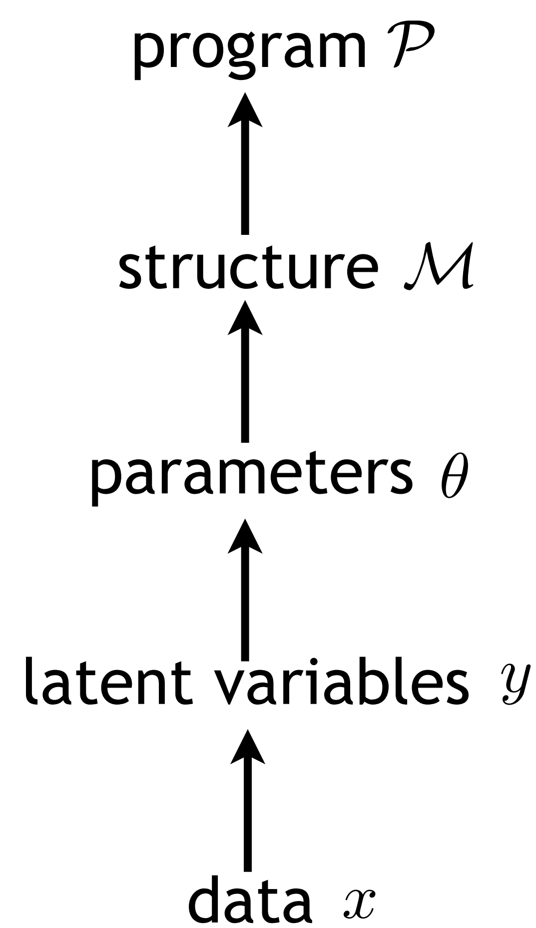
PROBABILISTIC INFERENCE AND LEARNING



PROBABILISTIC INFERENCE AND LEARNING



PROBABILISTIC INFERENCE AND LEARNING



PROBABILISTIC INFERENCE AND LEARNING

cognitive science

program $\mathcal{P}$

↑

structure $\mathcal{M}$

↑

parameters θ

↑

latent variables y

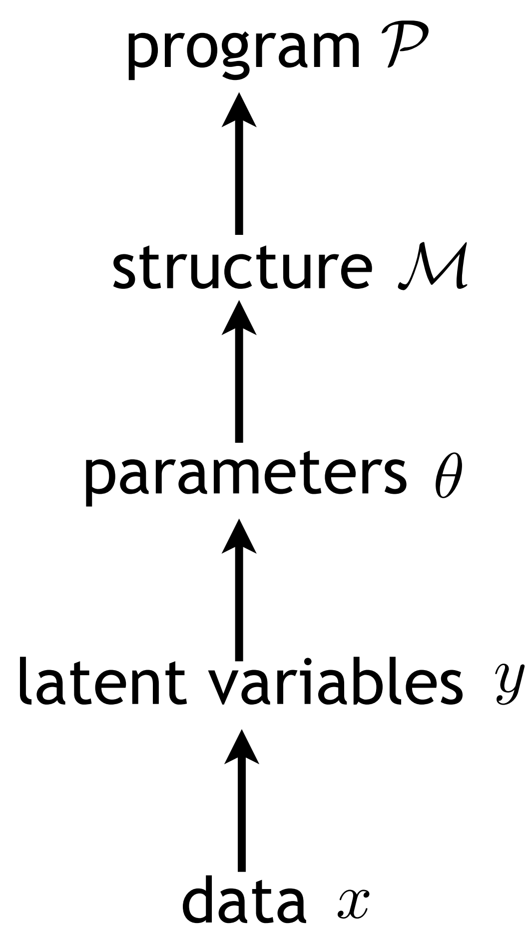
↑

data x

PROBABILISTIC INFERENCE AND LEARNING

cognitive science

theory



PROBABILISTIC INFERENCE AND LEARNING

cognitive science

theory	*experiments*	
		program $\mathcal{P}$
✓		↑
		structure $\mathcal{M}$
✓	✓	↑
		parameters θ
✓	✓	↑
		latent variables y
✓	✓	↑
		data x

PROBABILISTIC INFERENCE AND LEARNING

cognitive science

theory	*experiments*	
		program $\mathcal{P}$
✓	?✓	↑
		structure $\mathcal{M}$
✓	✓	↑
		parameters θ
✓	✓	↑
		latent variables y
✓	✓	↑
		data x

PROBABILISTIC INFERENCE AND LEARNING

cognitive science | **neuroscience**

theory	*experiments*	
		program $\mathcal{P}$
✓	✓?	↑
		structure $\mathcal{M}$
✓	✓	↑
		parameters θ
✓	✓	↑
		latent variables y
✓	✓	↑
		data x

PROBABILISTIC INFERENCE AND LEARNING

cognitive science			neuroscience
theory	*experiments*		*theory*
		program $\mathcal{P}$	
✓	✓?	↑	
		structure $\mathcal{M}$	
✓	✓	↑	
		parameters θ	
✓	✓	↑	
		latent variables y	
✓	✓	↑	✓✓✓
		data x	

PROBABILISTIC INFERENCE AND LEARNING

cognitive science: *theory*	cognitive science: *experiments*		neuroscience: *theory*
		program $\mathcal{P}$	
✓	✓ ?	↑	
		structure $\mathcal{M}$	
✓	✓	↑	
		parameters θ	
✓	✓	↑	✓✓
		latent variables y	
✓	✓	↑	✓✓✓
		data x	

PROBABILISTIC INFERENCE AND LEARNING

cognitive science: *theory*	cognitive science: *experiments*		neuroscience: *theory*
		program $\mathcal{P}$	
✓	✓?	↑	?
		structure $\mathcal{M}$	
✓	✓	↑	?
		parameters θ	
✓	✓	↑	✓✓
		latent variables y	
✓	✓	↑	✓✓✓
		data x	

PROBABILISTIC INFERENCE AND LEARNING

cognitive science: *theory*	cognitive science: *experiments*		neuroscience: *theory*	neuroscience: *experiments*
		program $\mathcal{P}$		
✓	?✓	↑	?	
		structure $\mathcal{M}$		
✓	✓	↑	?	
		parameters θ		
✓	✓	↑	✓✓	
		latent variables y		
✓	✓	↑	✓✓✓	?✓
		data x		

PROBABILISTIC INFERENCE AND LEARNING

cognitive science			neuroscience	
theory	*experiments*		*theory*	*experiments*
		program $\mathcal{P}$		
✓	?✓	↑	?	
		structure $\mathcal{M}$		
✓	✓	↑	?	
		parameters θ		
✓	✓	↑	✓✓	?✓
		latent variables y		
✓	✓	↑	✓✓✓	?✓
		data x		

PROBABILISTIC INFERENCE AND LEARNING

cognitive science			neuroscience	
theory	*experiments*		*theory*	*experiments*
		program $\mathcal{P}$		
✓	✓?	↑	?	?
		structure $\mathcal{M}$		
✓	✓	↑	?	?
		parameters θ		
✓	✓	↑	✓✓	✓?
		latent variables y		
✓	✓	↑	✓✓✓	✓?
		data x		

BAYESIAN INFERENCE IN THE BRAIN BY MCMC SAMPLING

28

BAYESIAN INFERENCE IN THE BRAIN BY MCMC SAMPLING

stimulus

28

BAYESIAN INFERENCE IN THE BRAIN BY MCMC SAMPLING

stimulus

28

BAYESIAN INFERENCE IN THE BRAIN BY MCMC SAMPLING

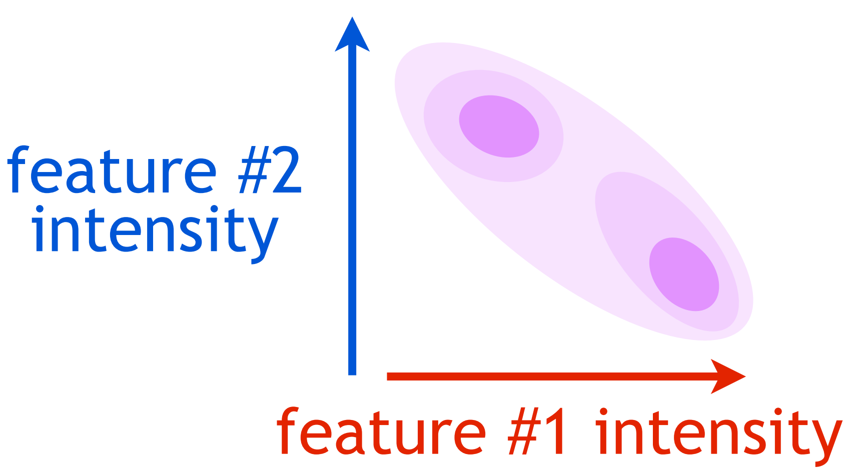
stimulus

BAYESIAN INFERENCE IN THE BRAIN BY MCMC SAMPLING

ideal observer

stimulus

P(feature 1, feature 2 | stimulus)

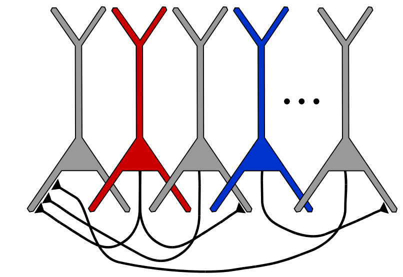


28

BAYESIAN INFERENCE IN THE BRAIN BY MCMC SAMPLING

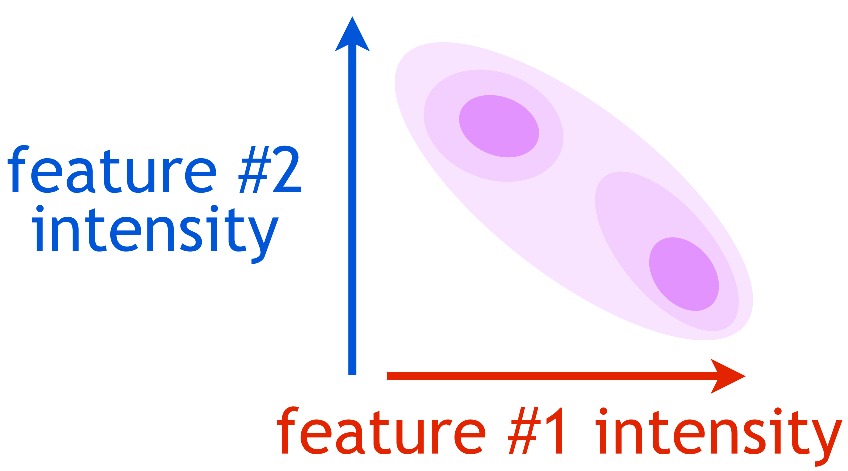
ideal observer

visual cortex



stimulus

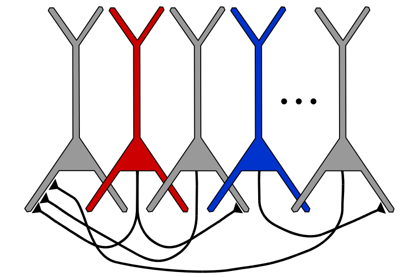
P(feature 1, feature 2 | stimulus)



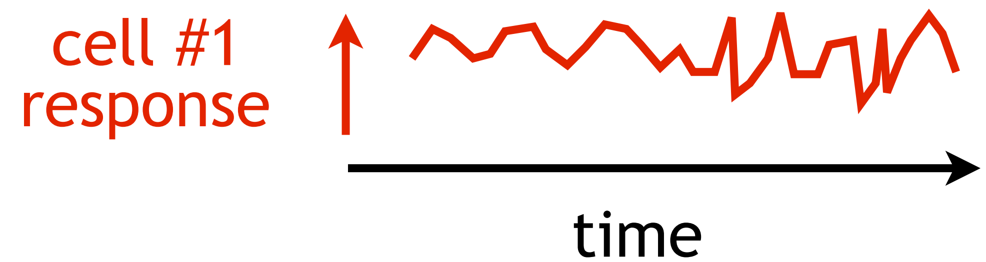
BAYESIAN INFERENCE IN THE BRAIN BY MCMC SAMPLING

ideal observer

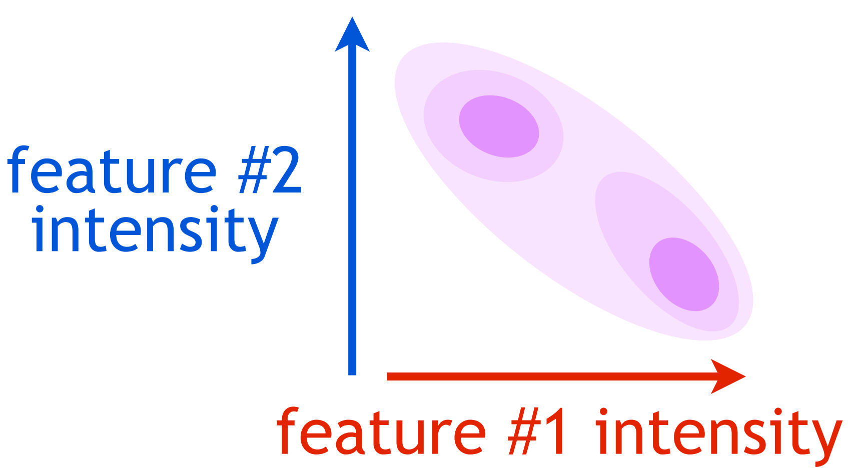
visual cortex



stimulus



P(feature 1, feature 2 | stimulus)

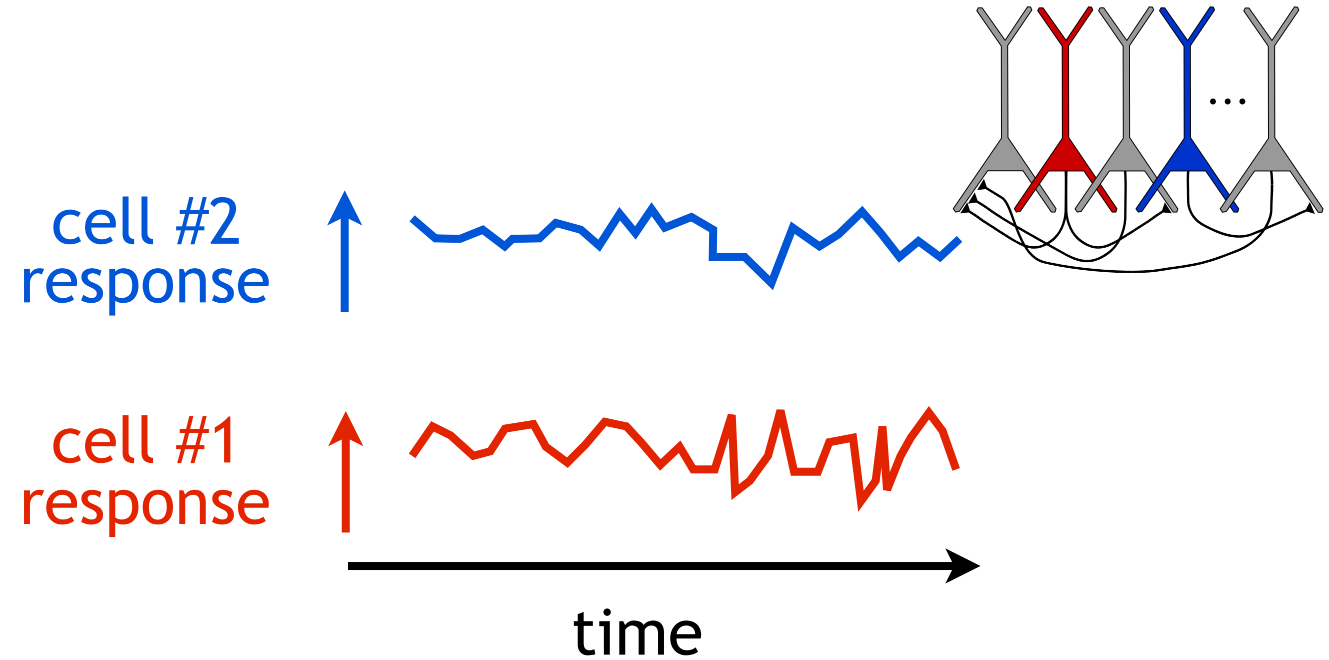


BAYESIAN INFERENCE IN THE BRAIN BY MCMC SAMPLING

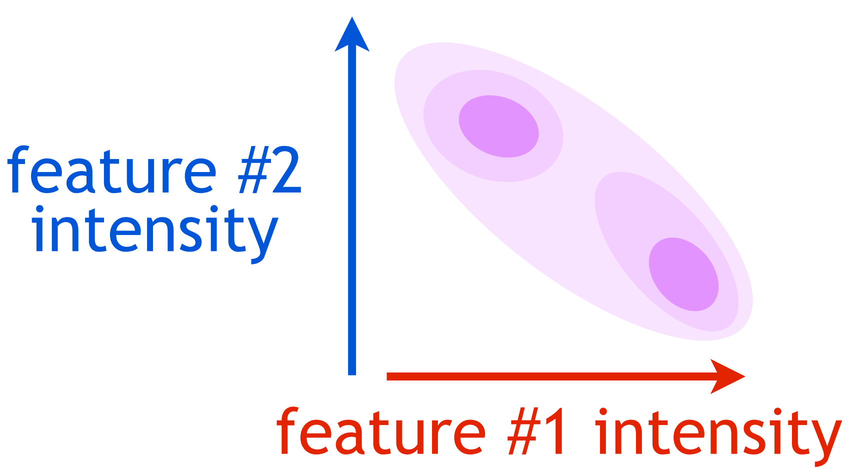
ideal observer

visual cortex

stimulus



P(feature 1, feature 2 | stimulus)



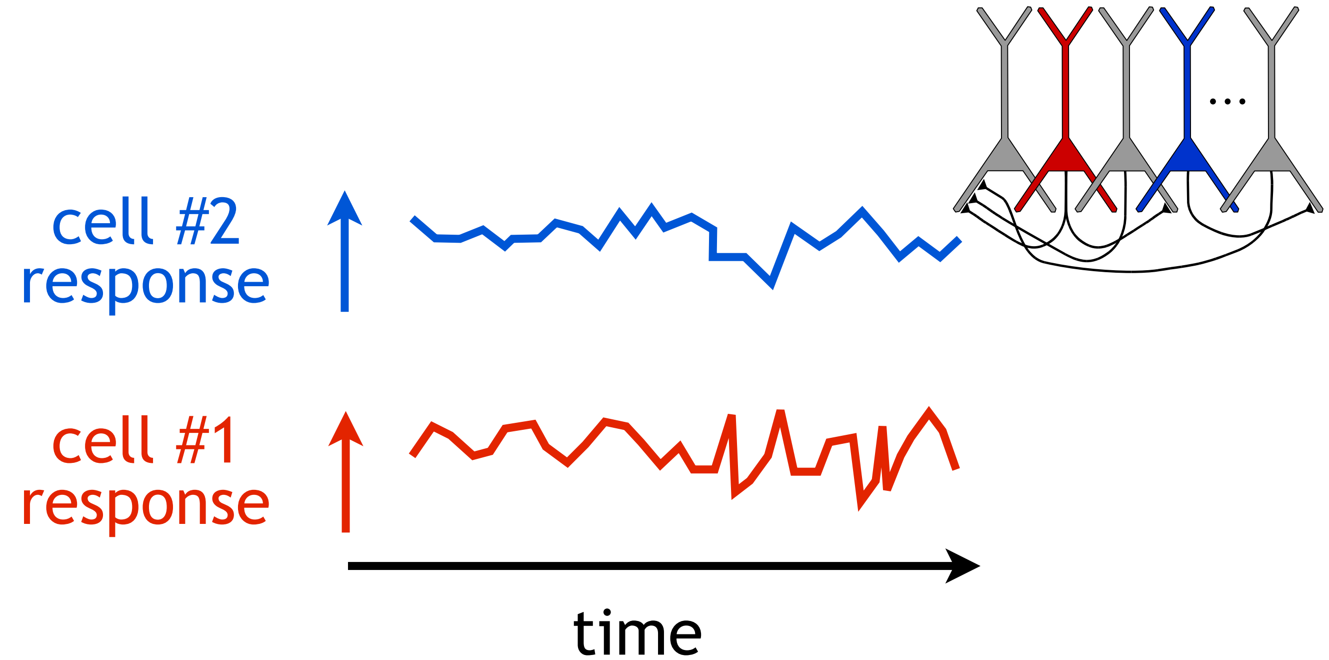
28

BAYESIAN INFERENCE IN THE BRAIN BY MCMC SAMPLING

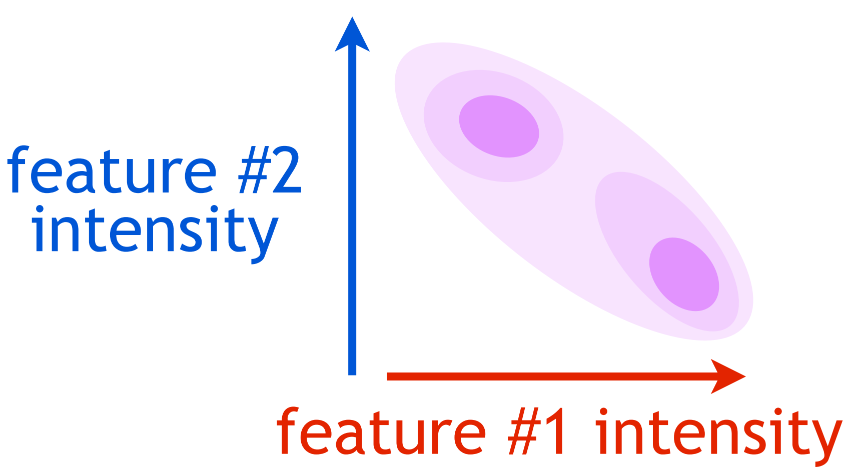
ideal observer

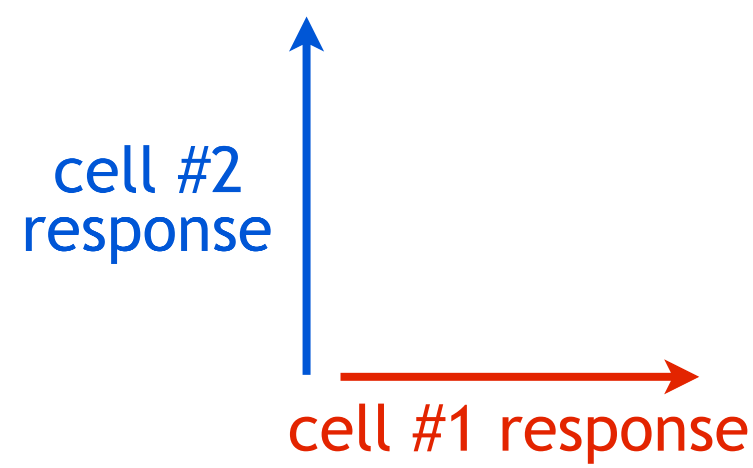
visual cortex

stimulus



P(feature 1, feature 2 | stimulus)

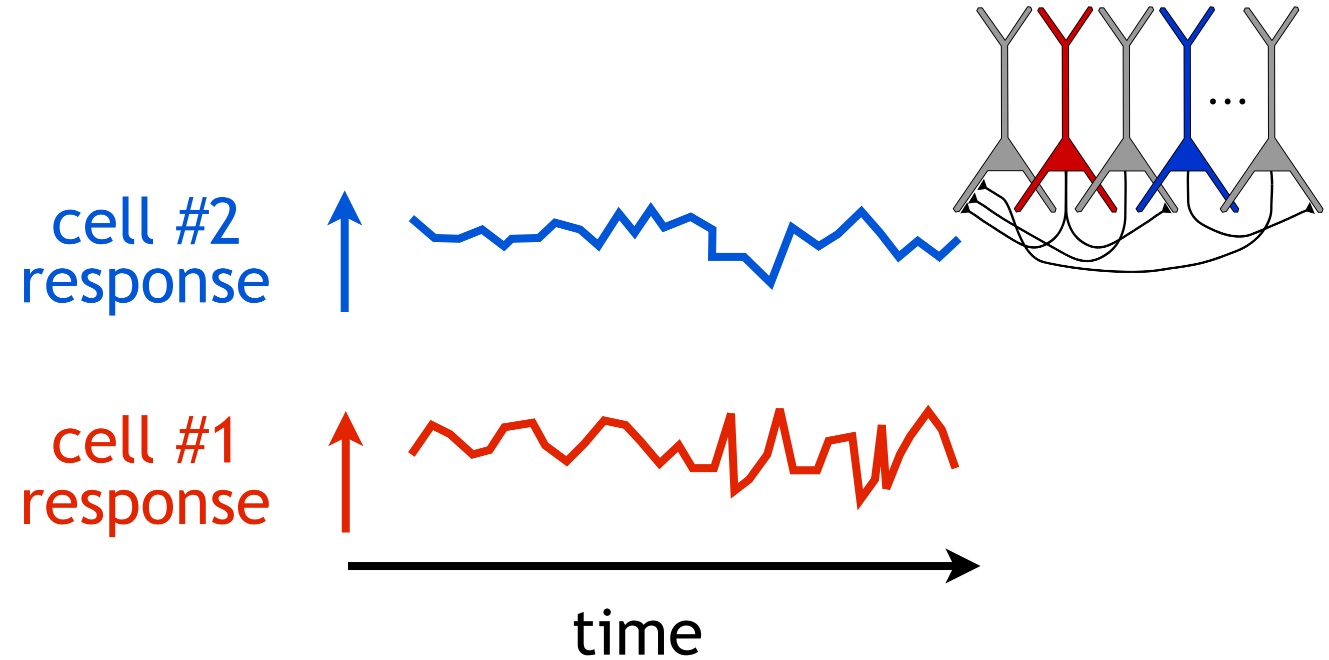




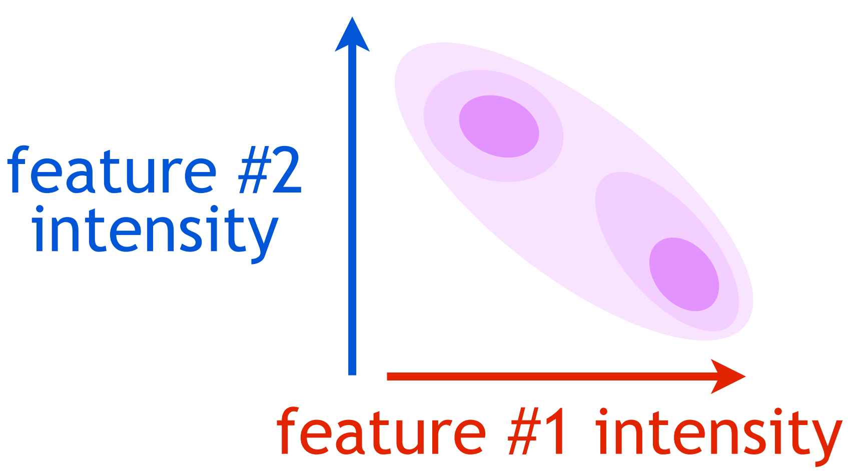
BAYESIAN INFERENCE IN THE BRAIN BY MCMC SAMPLING

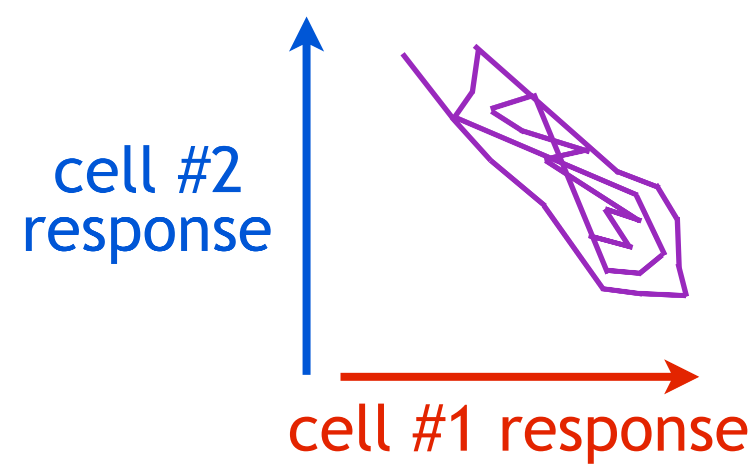
ideal observer | *visual cortex*

stimulus



$P(\text{feature 1}, \text{feature 2} \mid \text{stimulus})$





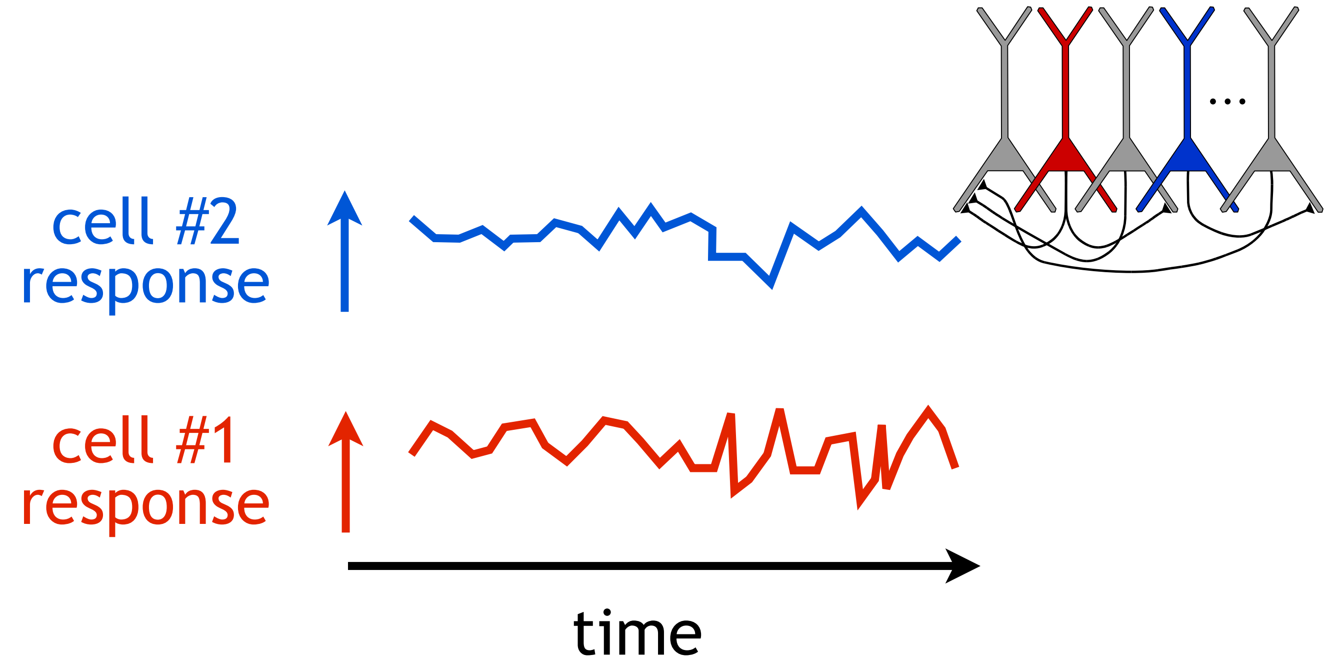
28

BAYESIAN INFERENCE IN THE BRAIN BY MCMC SAMPLING

ideal observer

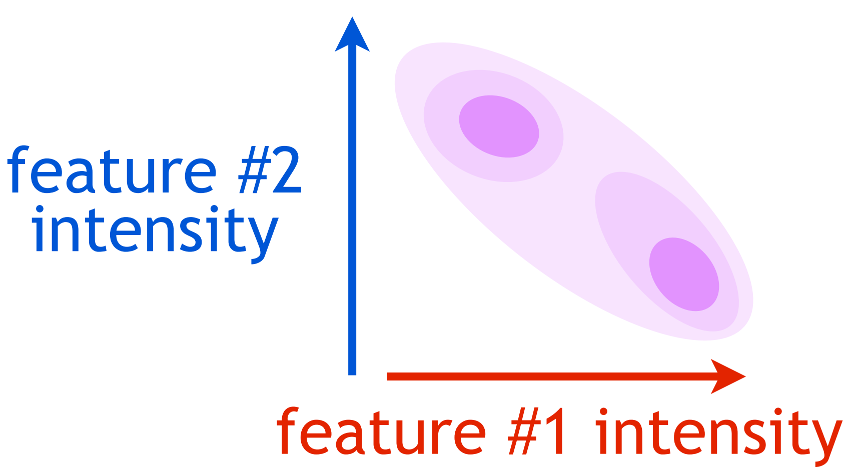
visual cortex

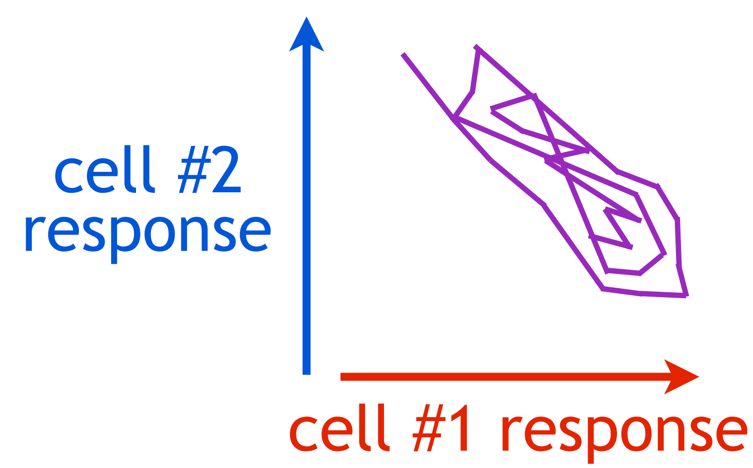
stimulus



P(feature 1, feature 2 | stimulus)

P(response 1, response 2 | stimulus)



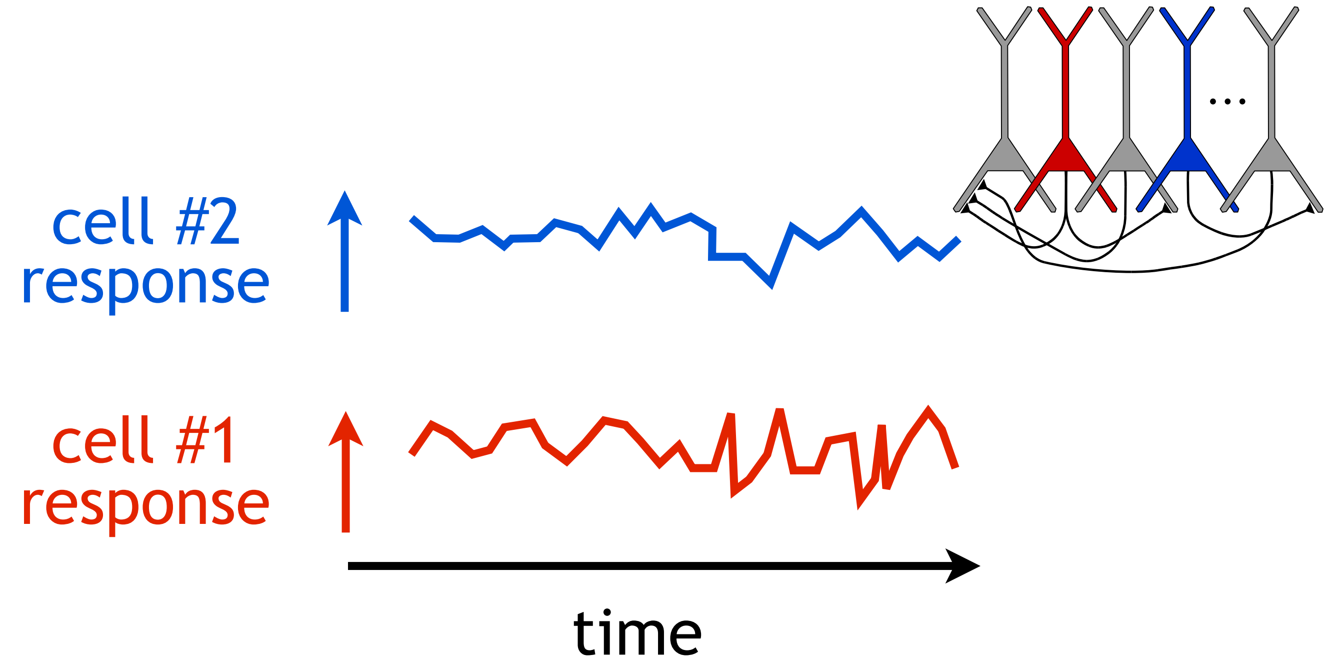


28

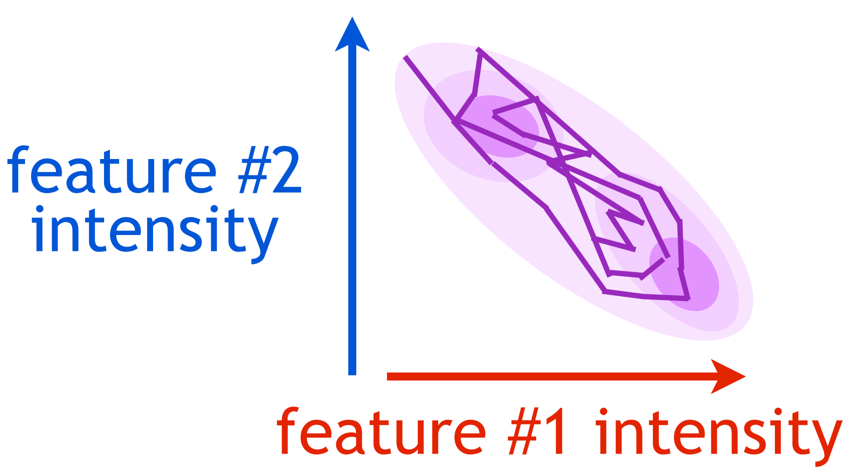
BAYESIAN INFERENCE IN THE BRAIN BY MCMC SAMPLING

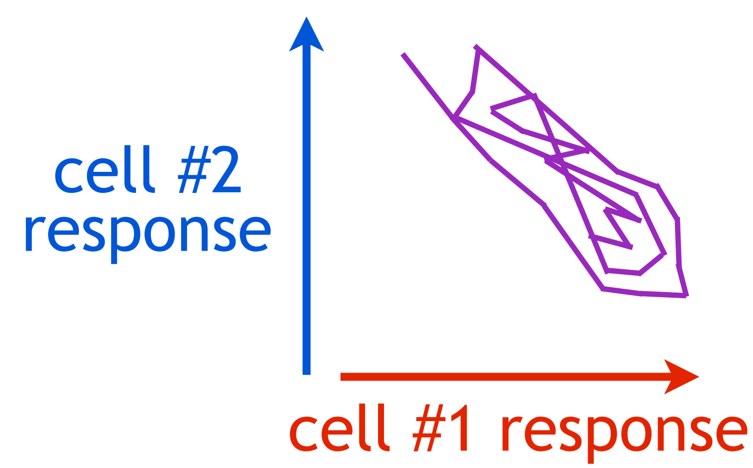
ideal observer | *visual cortex*

stimulus



$$P(\text{feature 1}, \text{feature 2} \mid \text{stimulus}) = P(\text{response 1}, \text{response 2} \mid \text{stimulus})$$





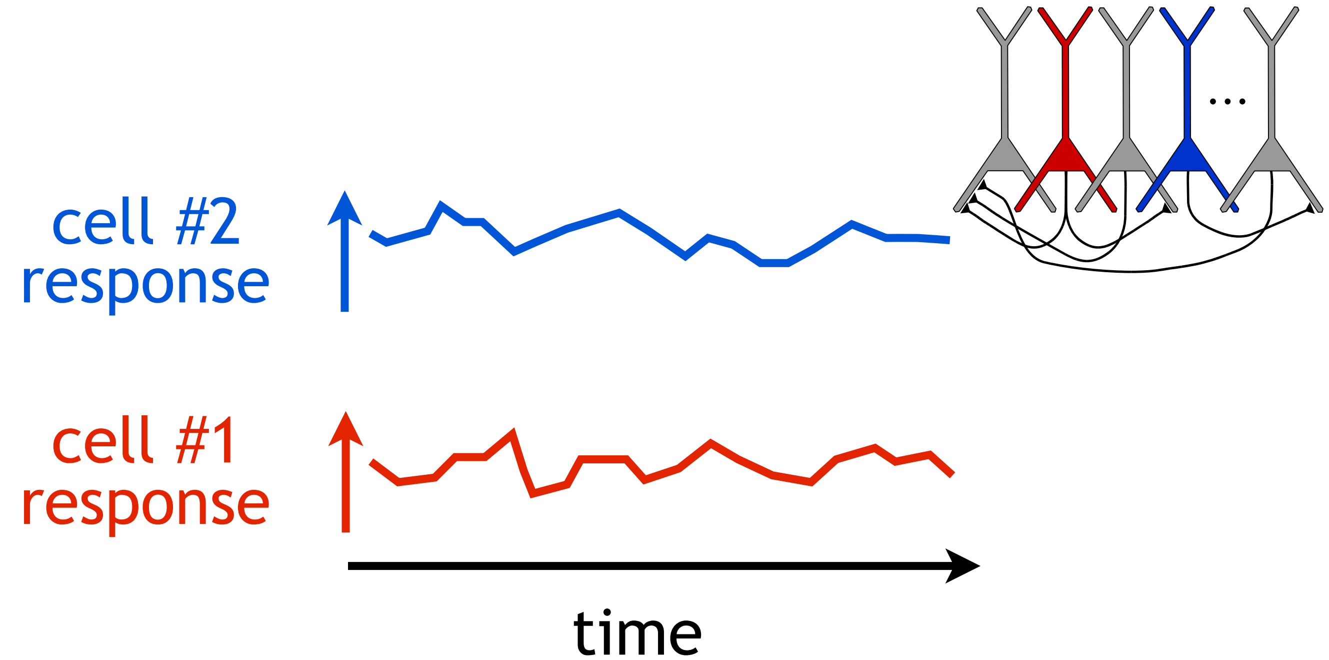
Fiser et al, TICS 2010

see also:
Hinton & Sejnowski, PDP 1986; Hinton et al, Science 1995; Dayan 1999; Hoyer & Hyvarinen, NIPS 2003, Lee & Mumford 2003

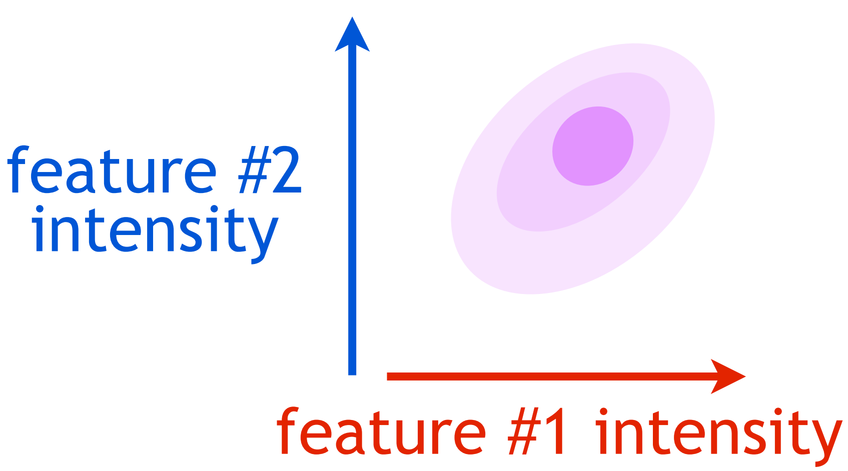
BAYESIAN INFERENCE IN THE BRAIN BY MCMC SAMPLING

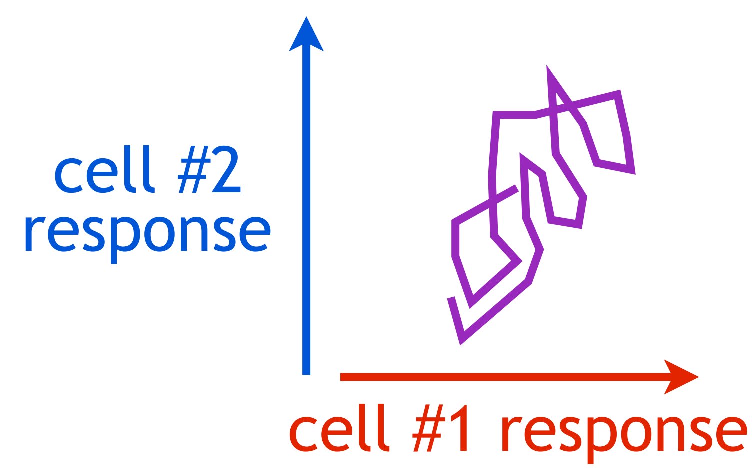
ideal observer *visual cortex*

stimulus



P(feature 1, feature 2 | stimulus) = P(response 1, response 2 | stimulus)





Fiser et al, TICS 2010

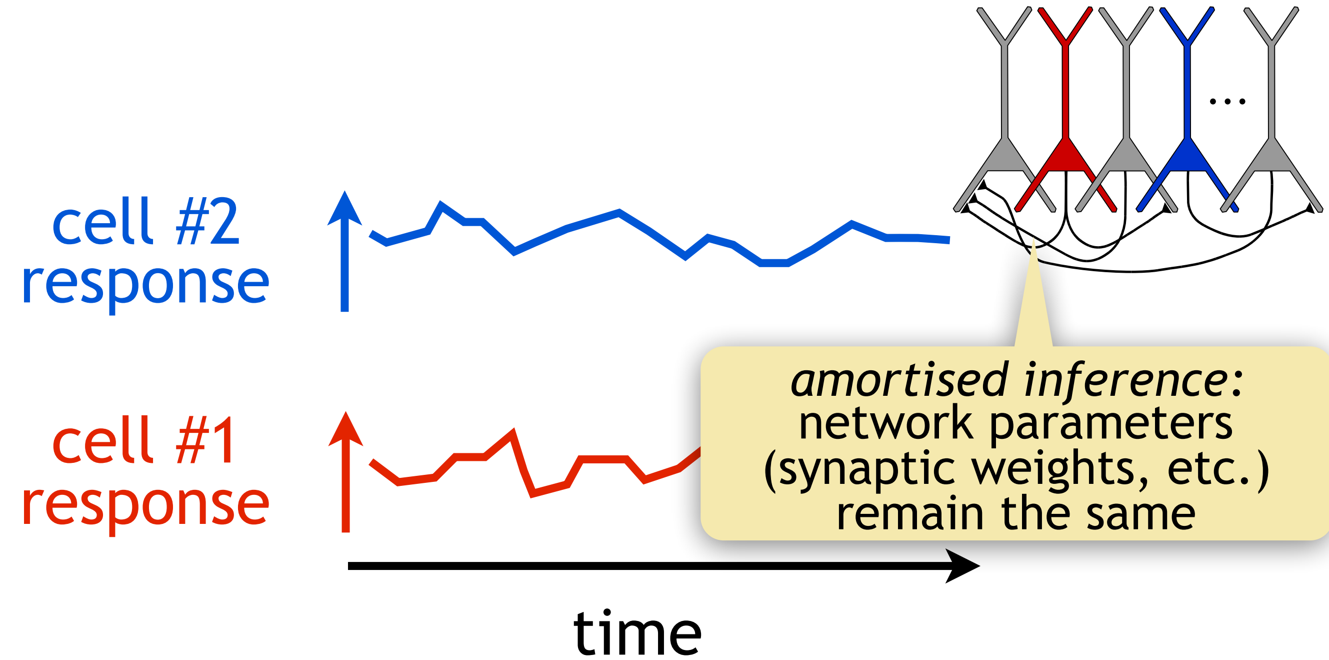
see also:
Hinton & Sejnowski, PDP 1986; Hinton et al, Science 1995; Dayan 1999; Hoyer & Hyvarinen, NIPS 2003, Lee & Mumford 2003

BAYESIAN INFERENCE IN THE BRAIN BY MCMC SAMPLING

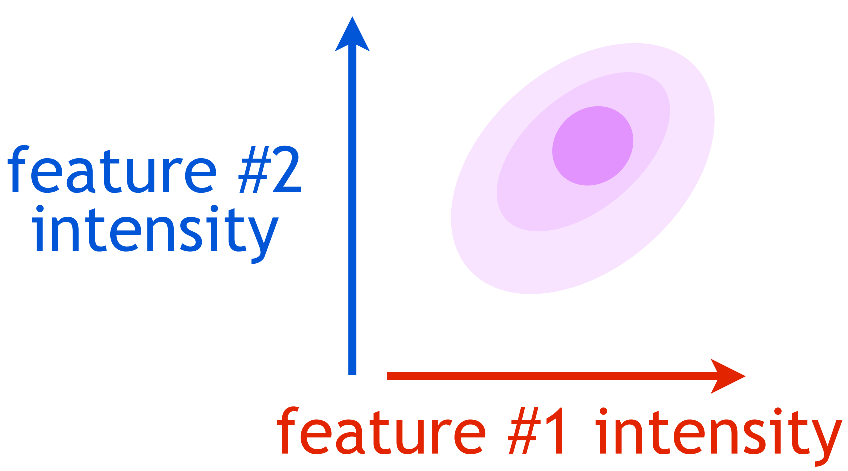
ideal observer

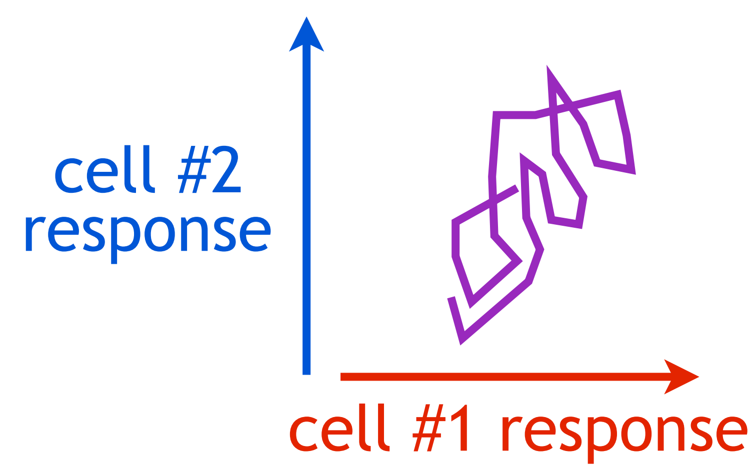
visual cortex

stimulus



$$P(\text{feature 1}, \text{feature 2} \mid \text{stimulus}) = P(\text{response 1}, \text{response 2} \mid \text{stimulus})$$





Fiser et al, TICS 2010

see also:
Hinton & Sejnowski, PDP 1986; Hinton et al, Science 1995; Dayan 1999; Hoyer & Hyvarinen, NIPS 2003, Lee & Mumford 2003

29

NEURAL HALLMARKS OF SAMPLING

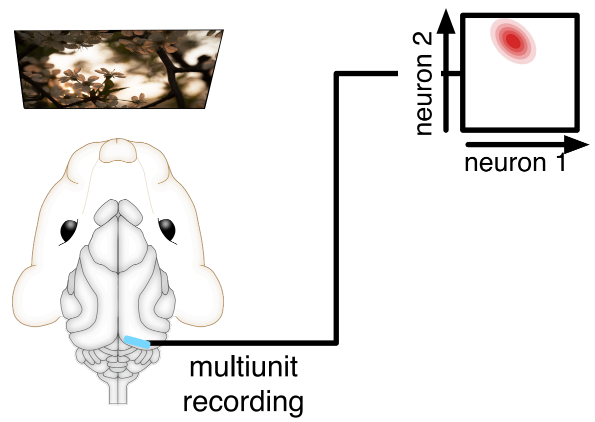

Pietro Berkes


Gergő Orbán


József Fiser

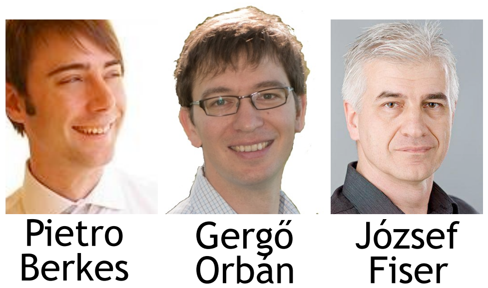
NEURAL HALLMARKS OF SAMPLING

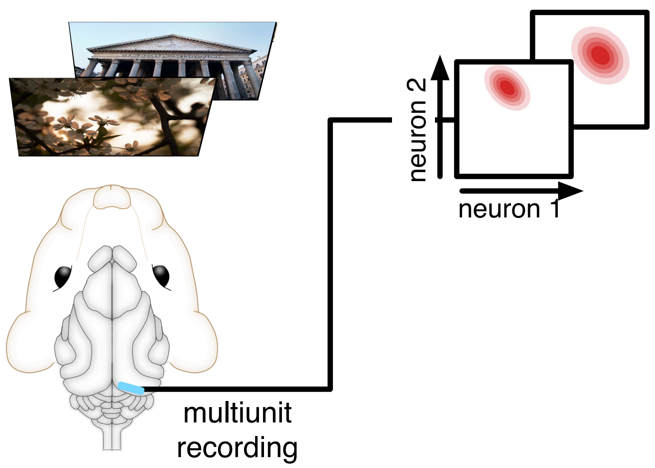

Pietro Berkes
Gergő Orbán
József Fiser



$$P_{\text{model}}(\text{feat}|\text{stim}_1)$$

NEURAL HALLMARKS OF SAMPLING

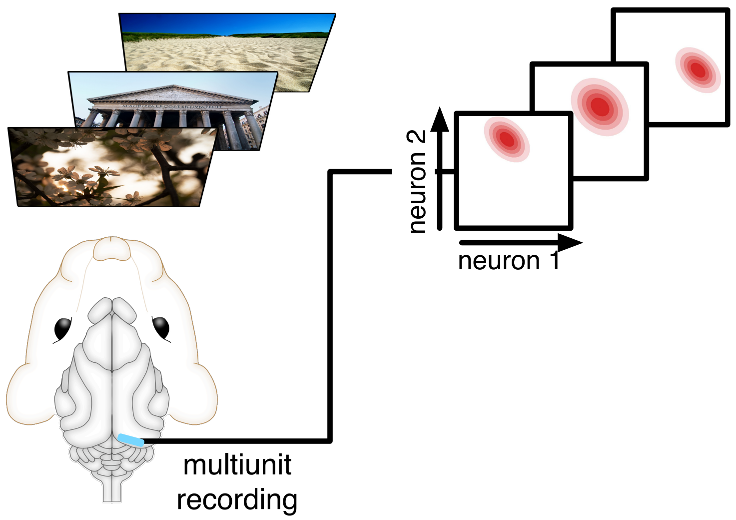




$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}_2)$$
$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}_1)$$

NEURAL HALLMARKS OF SAMPLING

Pietro Berkes　Gergő Orbán　József Fiser



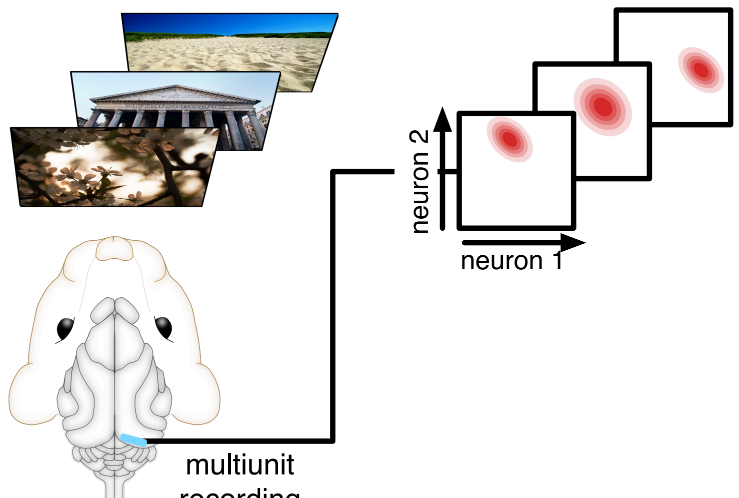
$$P_{\text{model}}(\text{feat}|\text{stim}_3)$$
$$P_{\text{model}}(\text{feat}|\text{stim}_2)$$
$$P_{\text{model}}(\text{feat}|\text{stim}_1)$$

NEURAL HALLMARKS OF SAMPLING

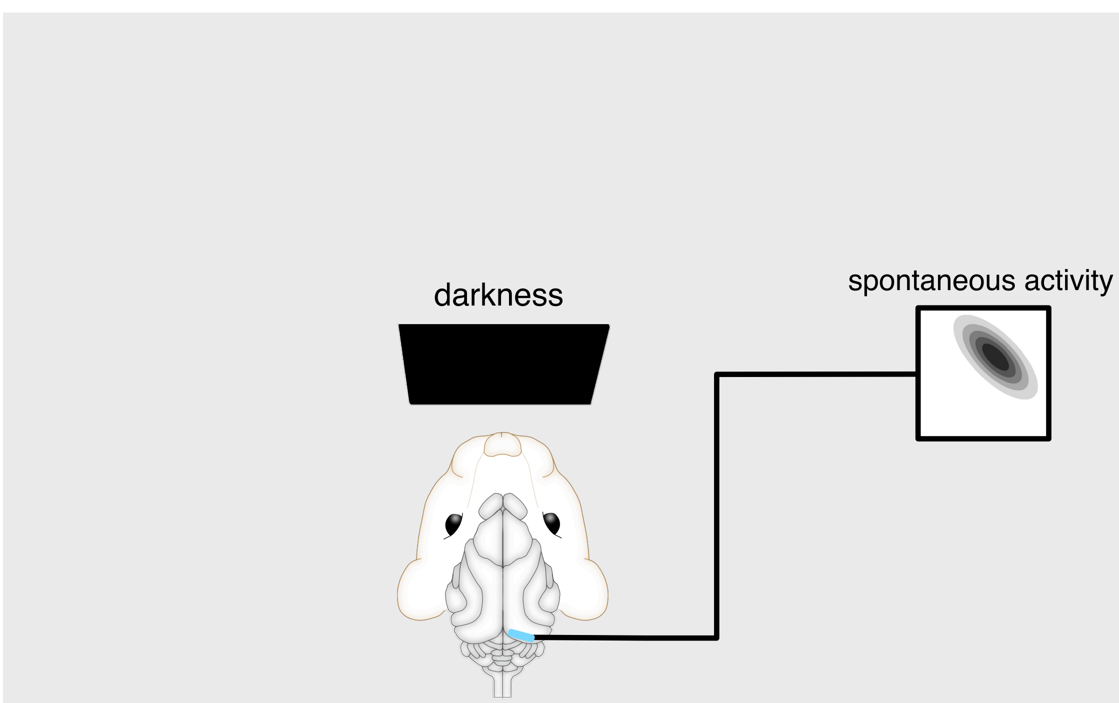

Pietro Berkes


Gergő Orbán


József Fiser



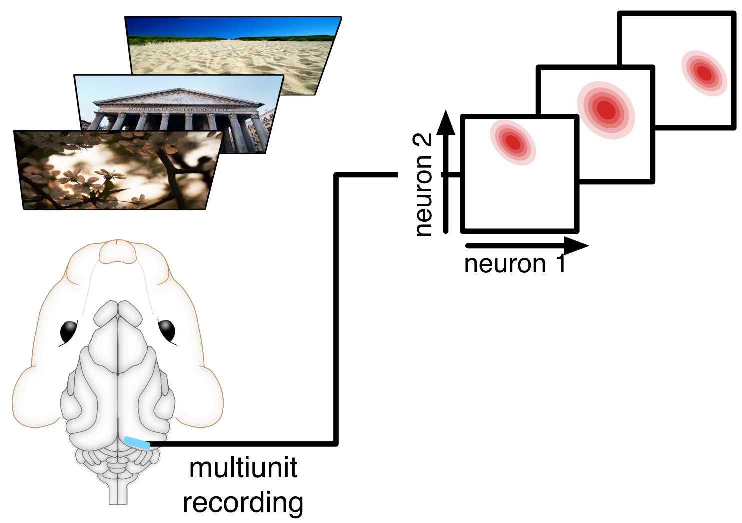
$$P_{\text{model}}(\text{feat}|\text{stim}_3)$$
$$P_{\text{model}}(\text{feat}|\text{stim}_2)$$
$$P_{\text{model}}(\text{feat}|\text{stim}_1)$$



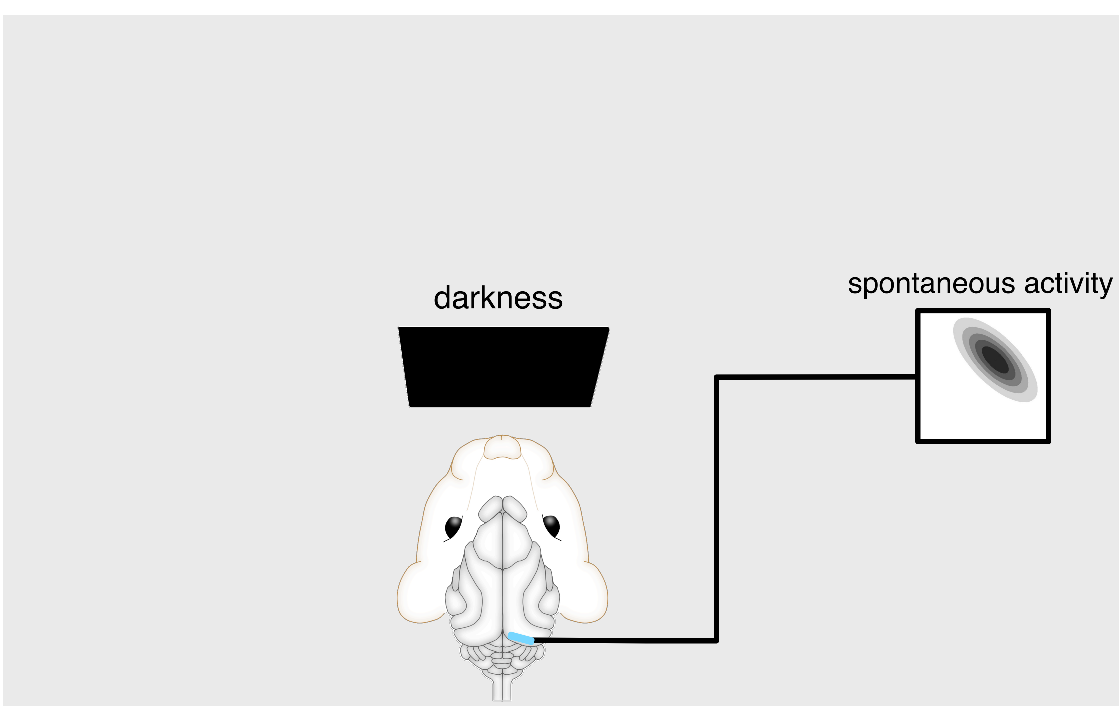
$$P_{\text{model}}(\text{feat}|\text{stim} = 0)$$

Berkes et al, Science 2011

NEURAL HALLMARKS OF SAMPLING



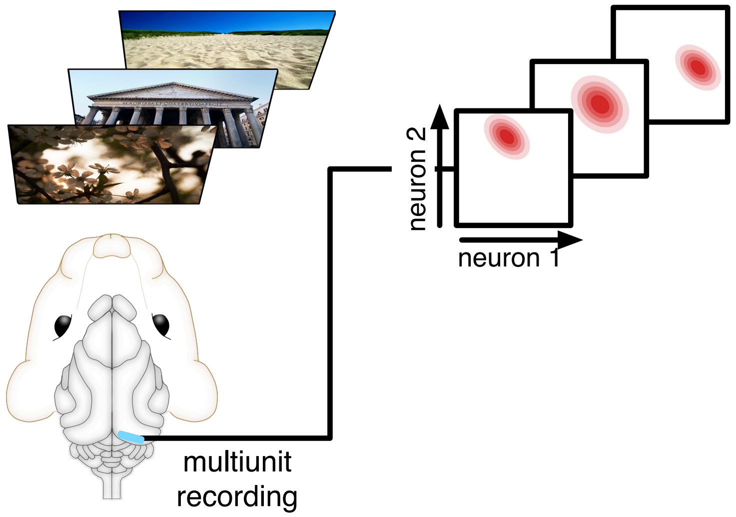
$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}_3)$$
$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}_2)$$
$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}_1)$$



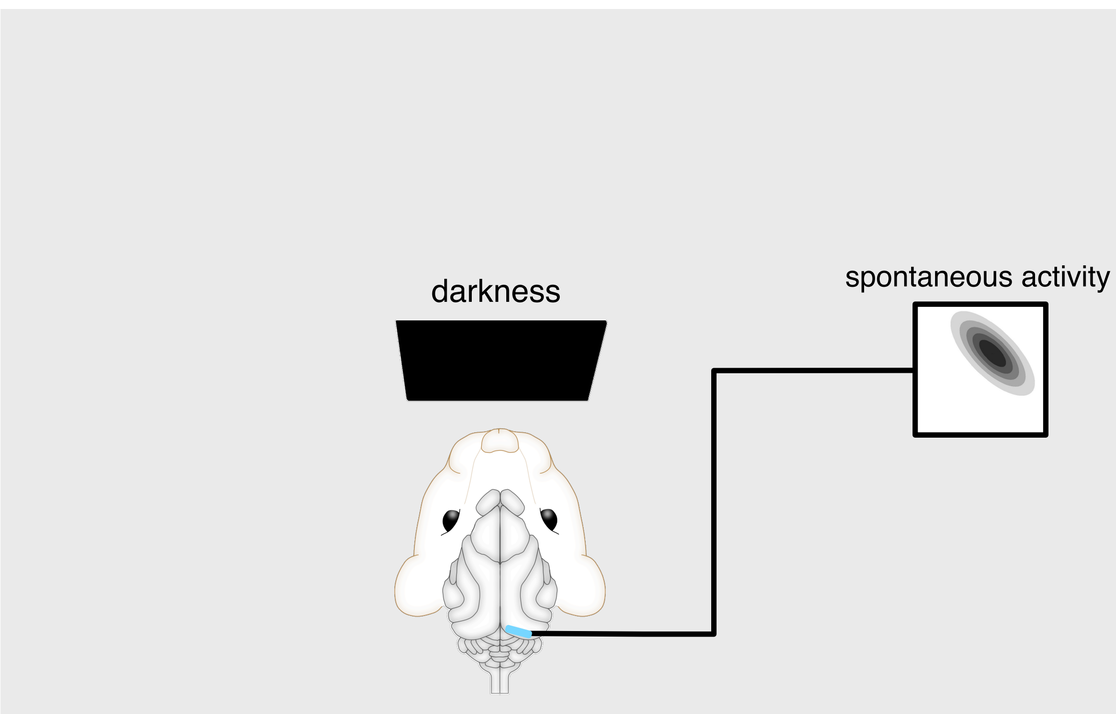
$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}=0) \simeq P_{\mathrm{model}}(\mathsf{feat})$$

Berkes et al, Science 2011

NEURAL HALLMARKS OF SAMPLING



$$P_{\text{model}}(\text{feat}|\text{stim}_3)$$
$$P_{\text{model}}(\text{feat}|\text{stim}_2)$$
$$P_{\text{model}}(\text{feat}|\text{stim}_1)$$

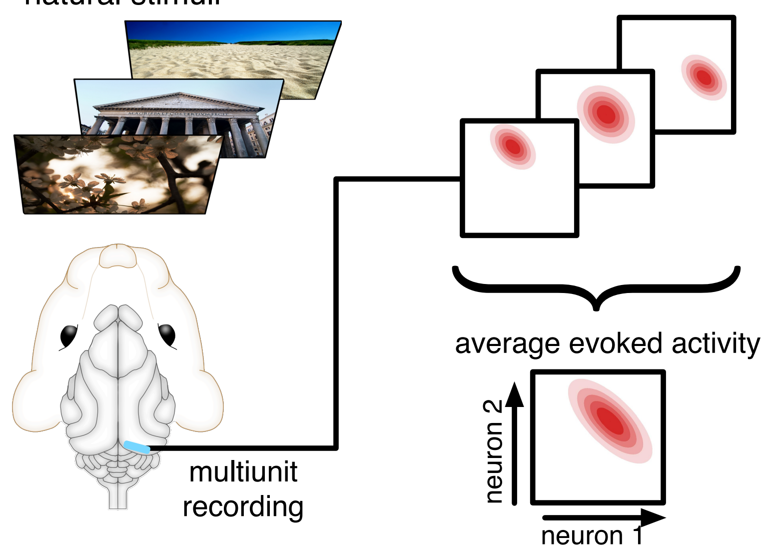


$$P_{\text{model}}(\text{feat}|\text{stim}=0) \simeq P_{\text{model}}(\text{feat})$$
$$\simeq \int P_{\text{model}}(\text{feat}|\text{stim})\, P_{\text{model}}(\text{stim}) \quad d\text{stim}$$

Berkes et al, Science 2011

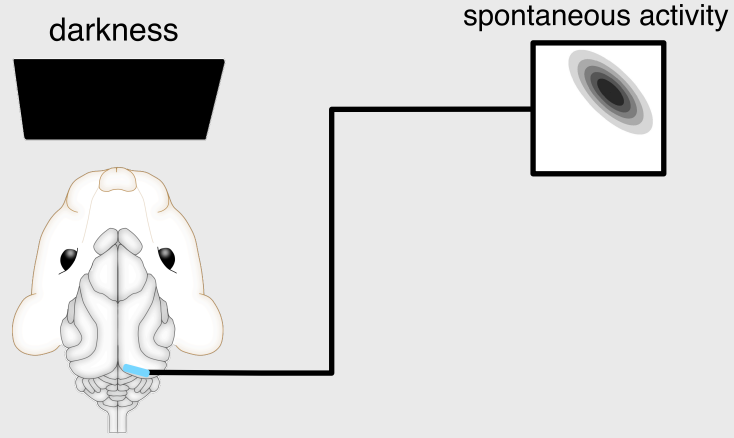
NEURAL HALLMARKS OF SAMPLING


Pietro Berkes
Gergő Orbán
József Fiser



$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}_3)$$
$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}_2)$$
$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}_1)$$

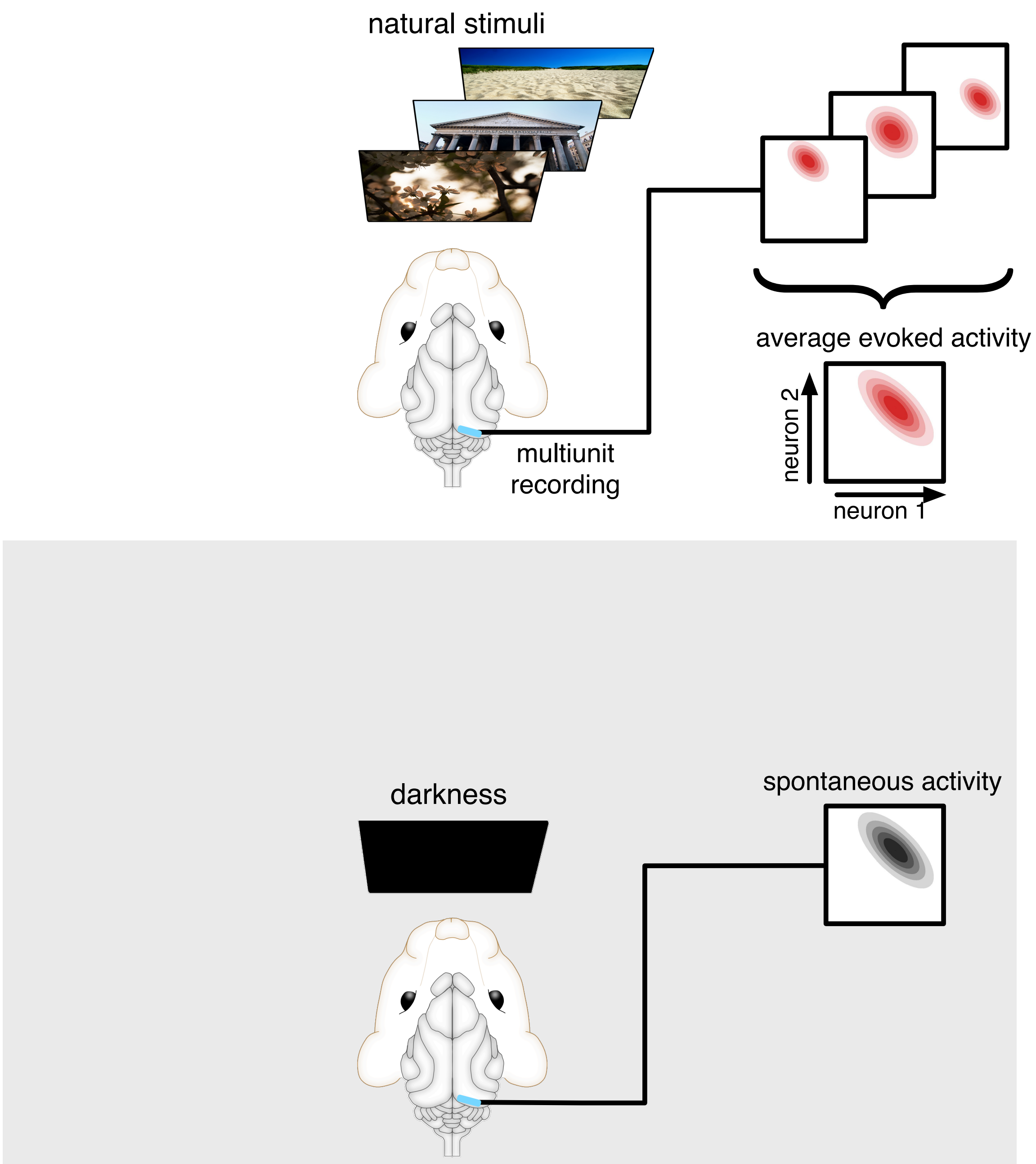
$$\int P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim})\, P_{\mathrm{natural}}(\mathsf{stim})\, d\mathsf{stim}$$



$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}=0) \simeq P_{\mathrm{model}}(\mathsf{feat})$$
$$\simeq \int P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim})\, P_{\mathrm{model}}(\mathsf{stim})\quad d\mathsf{stim}$$

Berkes et al, Science 2011

NEURAL HALLMARKS OF SAMPLING



$$P_{\mathrm{model}}(\text{feat}|\text{stim}_3)$$
$$P_{\mathrm{model}}(\text{feat}|\text{stim}_2)$$
$$P_{\mathrm{model}}(\text{feat}|\text{stim}_1)$$

$$\int P_{\mathrm{model}}(\text{feat}|\text{stim})\, P_{\mathrm{natural}}(\text{stim})\, d\text{stim}$$

$$P_{\mathrm{model}}(\text{feat}|\text{stim}=0) \simeq P_{\mathrm{model}}(\text{feat})$$

$$\simeq \int P_{\mathrm{model}}(\text{feat}|\text{stim})\, P_{\mathrm{model}}(\text{stim})\; d\text{stim}$$

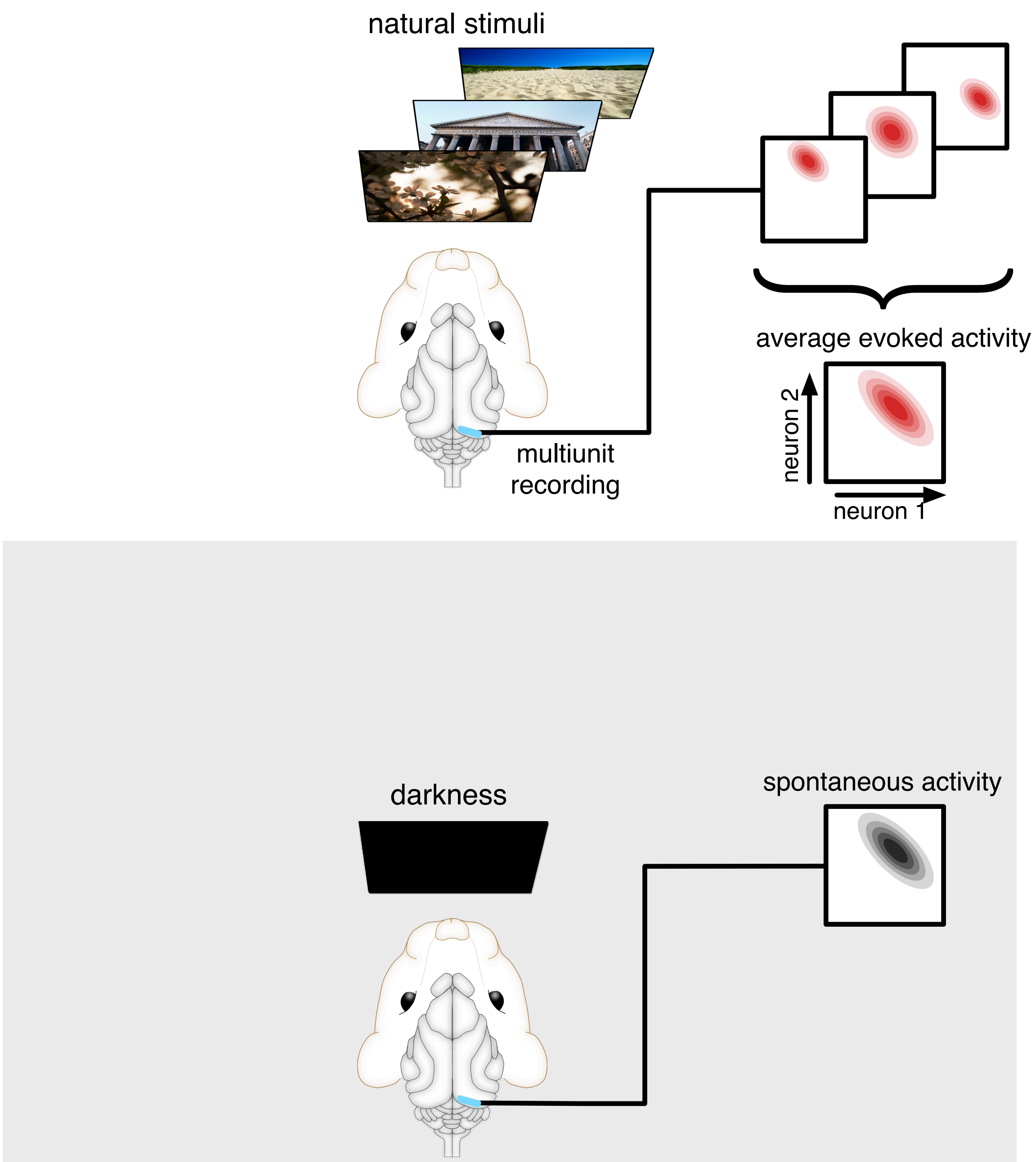
Berkes et al, Science 2011

NEURAL HALLMARKS OF SAMPLING

+ STATISTICALLY OPTIMAL ADAPTATION

$$\mathrm{P}_{\mathrm{model}}(\text{stim}) = \mathrm{P}_{\mathrm{natural}}(\text{stim})$$

Pietro Berkes
Gergő Orbán
József Fiser



$$\mathrm{P}_{\mathrm{model}}(\text{feat}|\text{stim}_3)$$
$$\mathrm{P}_{\mathrm{model}}(\text{feat}|\text{stim}_2)$$
$$\mathrm{P}_{\mathrm{model}}(\text{feat}|\text{stim}_1)$$

$$\int \mathrm{P}_{\mathrm{model}}(\text{feat}|\text{stim})\, \mathrm{P}_{\mathrm{natural}}(\text{stim})\, d\text{stim}$$

$$\mathrm{P}_{\mathrm{model}}(\text{feat}|\text{stim}=0) \simeq \mathrm{P}_{\mathrm{model}}(\text{feat})$$

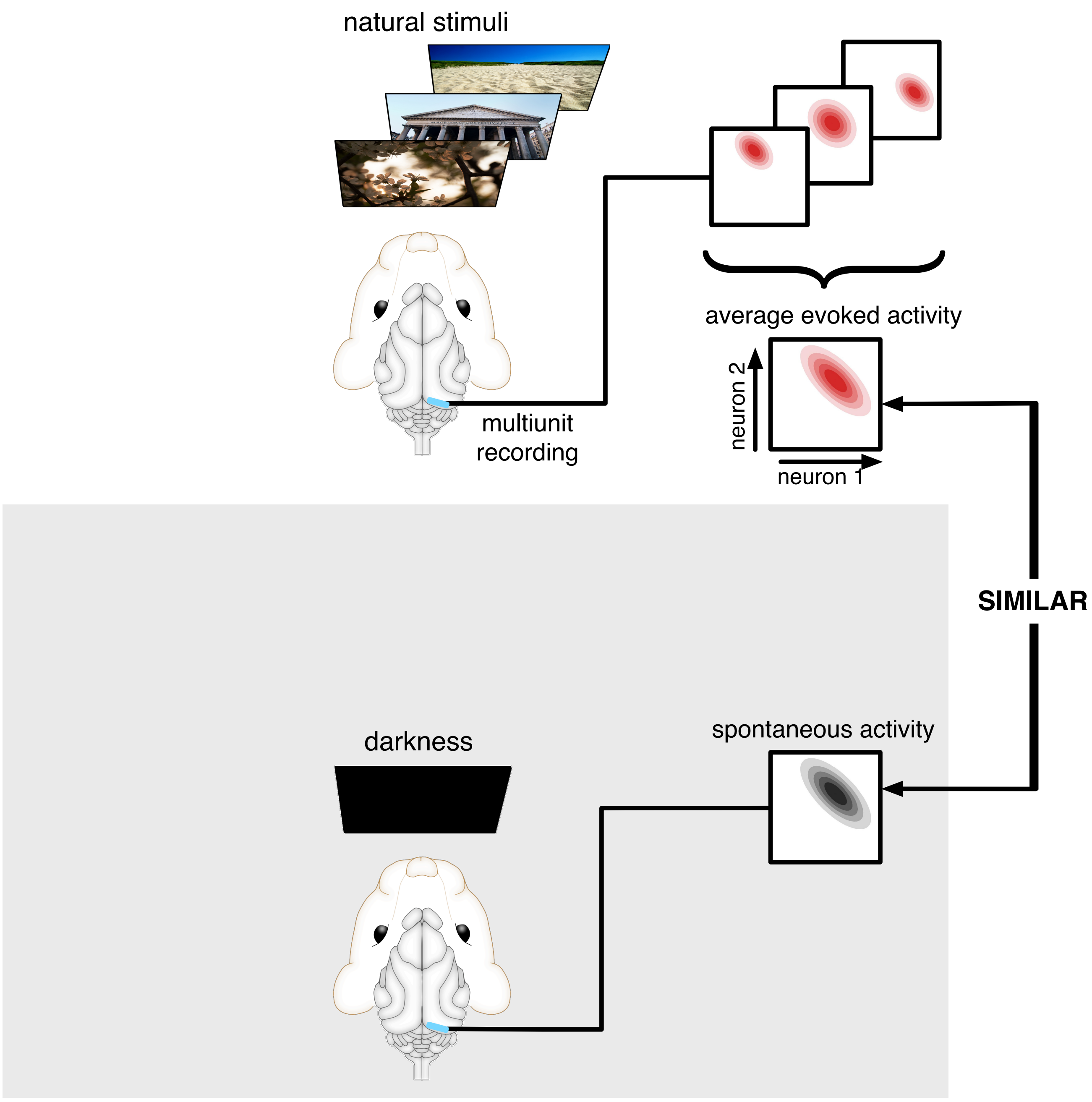
$$\simeq \int \mathrm{P}_{\mathrm{model}}(\text{feat}|\text{stim})\, \mathrm{P}_{\mathrm{model}}(\text{stim})\, d\text{stim}$$

Berkes et al, Science 2011

NEURAL HALLMARKS OF SAMPLING

+ STATISTICALLY OPTIMAL ADAPTATION

$$P_{\text{model}}(\text{stim}) = P_{\text{natural}}(\text{stim})$$



$$P_{\text{model}}(\text{feat}|\text{stim}_3)$$
$$P_{\text{model}}(\text{feat}|\text{stim}_2)$$
$$P_{\text{model}}(\text{feat}|\text{stim}_1)$$

$$\int P_{\text{model}}(\text{feat}|\text{stim})\, P_{\text{natural}}(\text{stim})\, d\text{stim}$$

$$P_{\text{model}}(\text{feat}|\text{stim}=0) \simeq P_{\text{model}}(\text{feat})$$

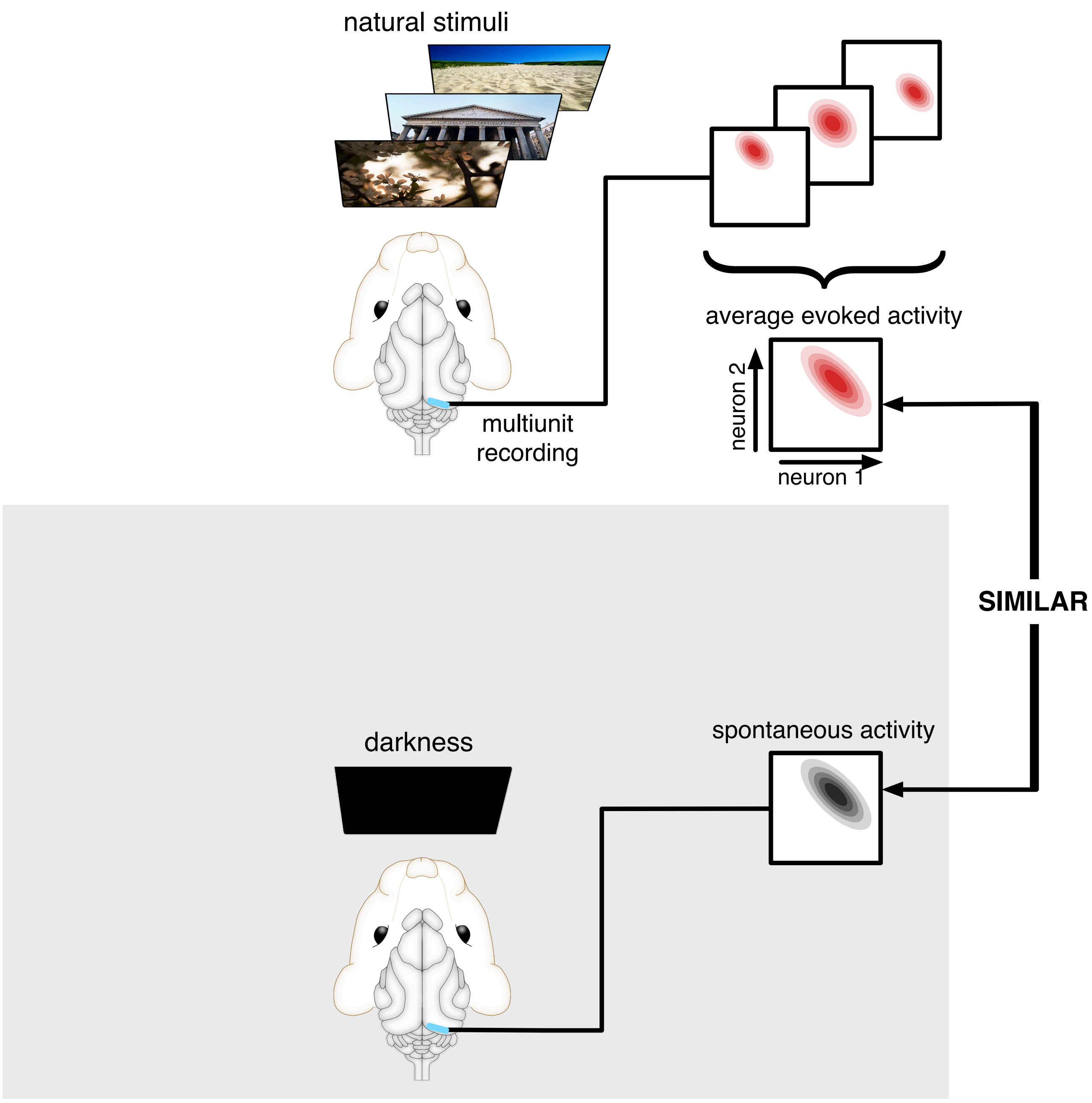
$$\simeq \int P_{\text{model}}(\text{feat}|\text{stim})\, P_{\text{model}}(\text{stim})\, d\text{stim}$$

Berkes et al, Science 2011

NEURAL HALLMARKS OF SAMPLING

+ STATISTICALLY OPTIMAL ADAPTATION

$$\mathrm{P}_{\mathrm{model}}(\text{stim}) = \mathrm{P}_{\mathrm{natural}}(\text{stim})$$



$$\mathrm{P}_{\mathrm{model}}(\text{feat}|\text{stim}_3)$$
$$\mathrm{P}_{\mathrm{model}}(\text{feat}|\text{stim}_2)$$
$$\mathrm{P}_{\mathrm{model}}(\text{feat}|\text{stim}_1)$$

$$\int \mathrm{P}_{\mathrm{model}}(\text{feat}|\text{stim})\, \mathrm{P}_{\mathrm{natural}}(\text{stim})\, d\text{stim}$$

average inference

=

prior expectations

$$\mathrm{P}_{\mathrm{model}}(\text{feat}|\text{stim}=0) \simeq \mathrm{P}_{\mathrm{model}}(\text{feat})$$

$$\simeq \int \mathrm{P}_{\mathrm{model}}(\text{feat}|\text{stim})\, \mathrm{P}_{\mathrm{model}}(\text{stim})\, d\text{stim}$$

Berkes et al, Science 2011

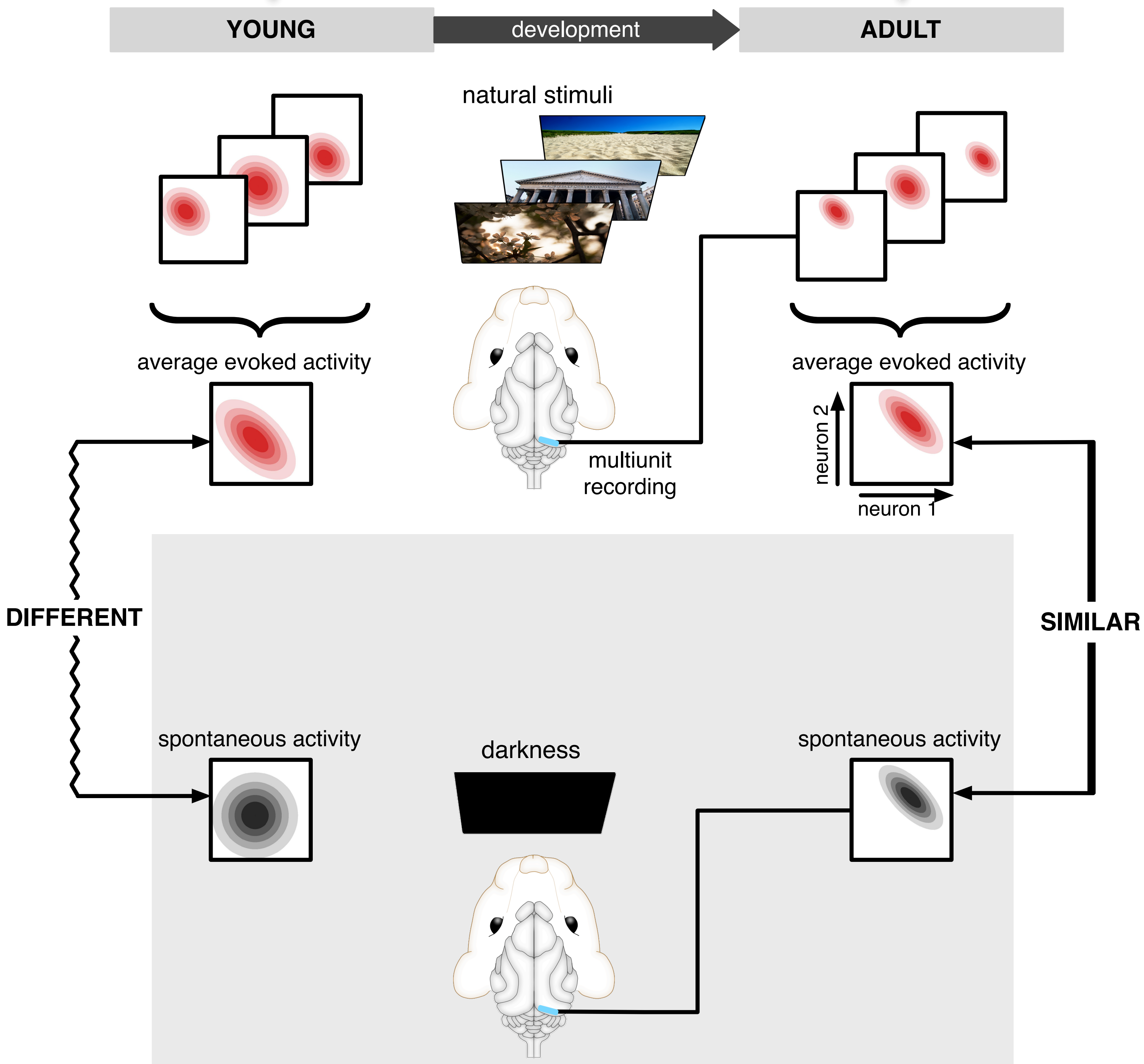
NEURAL HALLMARKS OF SAMPLING

+ STATISTICALLY OPTIMAL ADAPTATION

Pietro Berkes | Gergő Orbán | József Fiser

$\mathrm{P}_{\mathrm{model}}(\mathrm{stim}) \neq \mathrm{P}_{\mathrm{natural}}(\mathrm{stim})$

$\mathrm{P}_{\mathrm{model}}(\mathrm{stim}) = \mathrm{P}_{\mathrm{natural}}(\mathrm{stim})$



$$\mathrm{P}_{\mathrm{model}}(\mathrm{feat}|\mathrm{stim}_3)$$
$$\mathrm{P}_{\mathrm{model}}(\mathrm{feat}|\mathrm{stim}_2)$$
$$\mathrm{P}_{\mathrm{model}}(\mathrm{feat}|\mathrm{stim}_1)$$

$$\int \mathrm{P}_{\mathrm{model}}(\mathrm{feat}|\mathrm{stim})\,\mathrm{P}_{\mathrm{natural}}(\mathrm{stim})\,d\mathrm{stim}$$

average inference

=

prior expectations

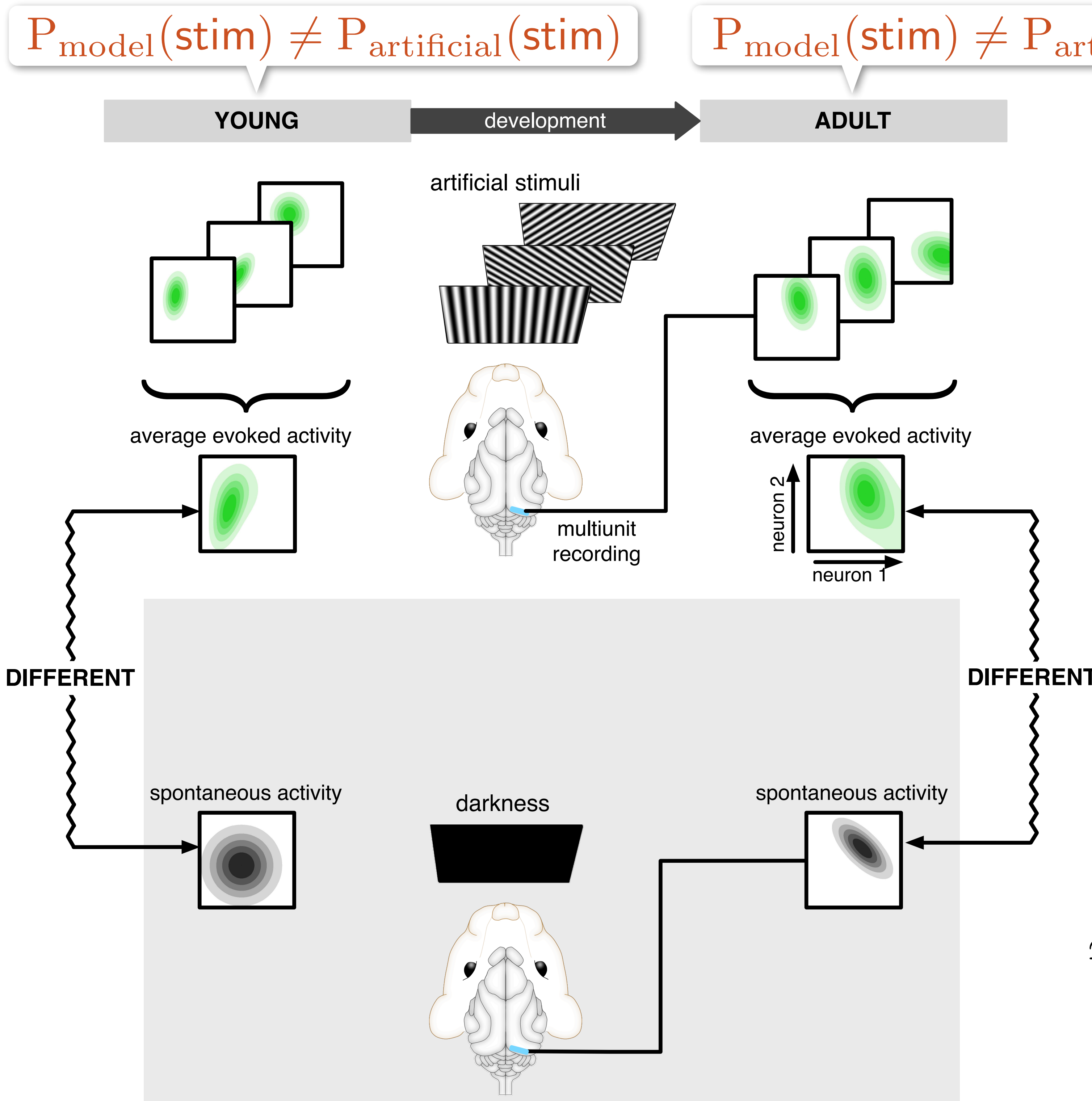
$$\mathrm{P}_{\mathrm{model}}(\mathrm{feat}|\mathrm{stim}=0) \simeq \mathrm{P}_{\mathrm{model}}(\mathrm{feat})$$

$$\simeq \int \mathrm{P}_{\mathrm{model}}(\mathrm{feat}|\mathrm{stim})\,\mathrm{P}_{\mathrm{model}}(\mathrm{stim})\,d\mathrm{stim}$$

Berkes et al, Science 2011

NEURAL HALLMARKS OF SAMPLING

+ STATISTICALLY OPTIMAL ADAPTATION



$$P_{model}(feat|stim_3)$$
$$P_{model}(feat|stim_2)$$
$$P_{model}(feat|stim_1)$$

$$\int P_{model}(feat|stim)\, P_{artificial}(stim)\, dstim$$

average inference

≠

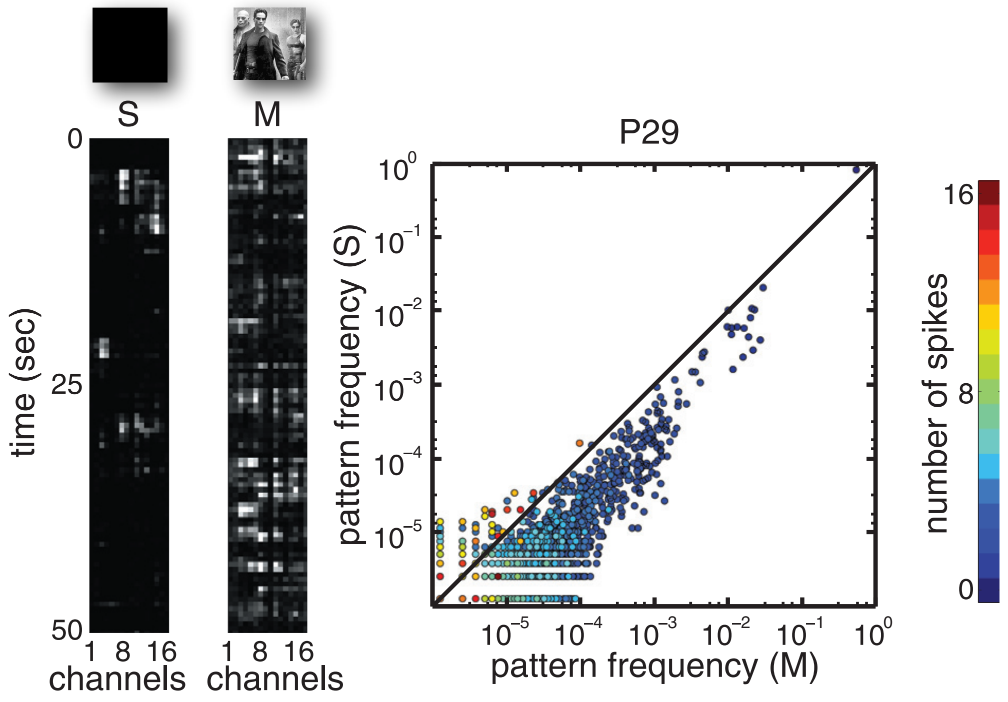
prior expectations

$$P_{model}(feat|stim = 0) \simeq P_{model}(feat)$$

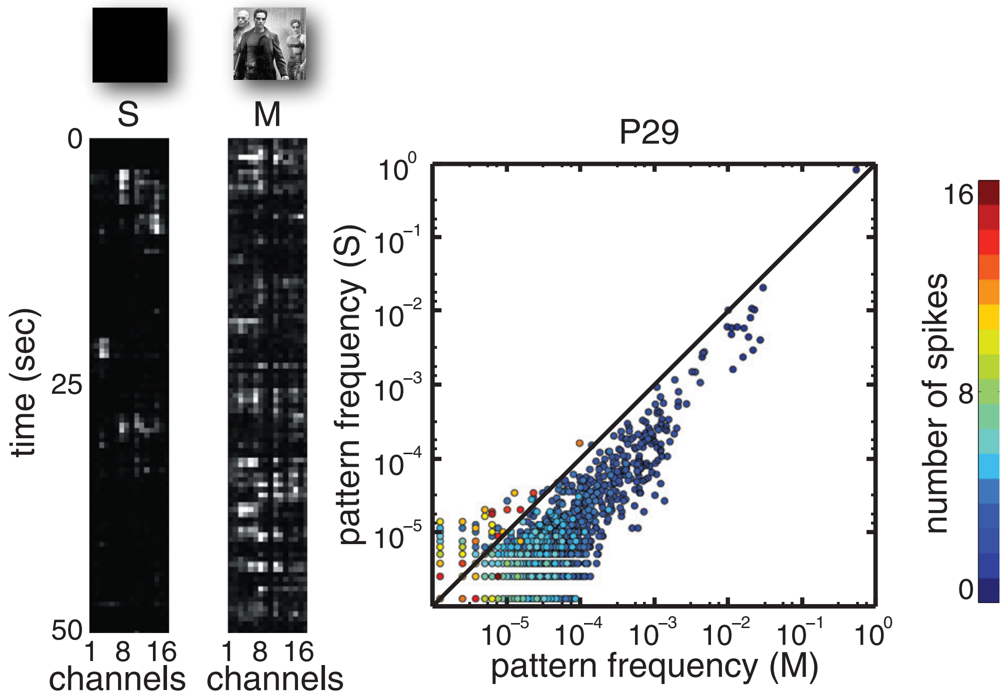
$$\simeq \int P_{model}(feat|stim)\, P_{model}(stim)\, dstim$$

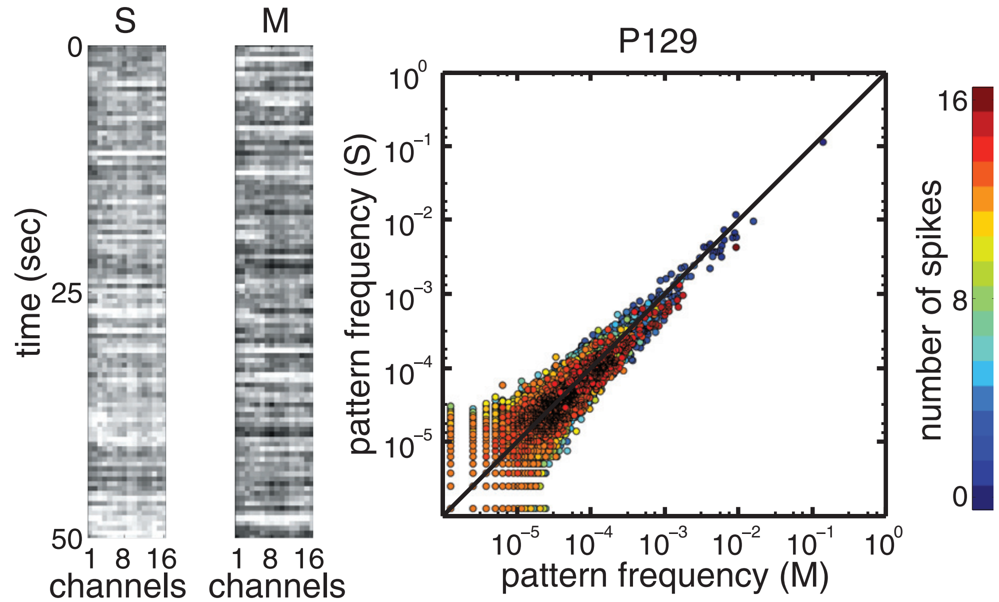
Berkes et al, Science 2011

DEVELOPMENTAL CHANGES

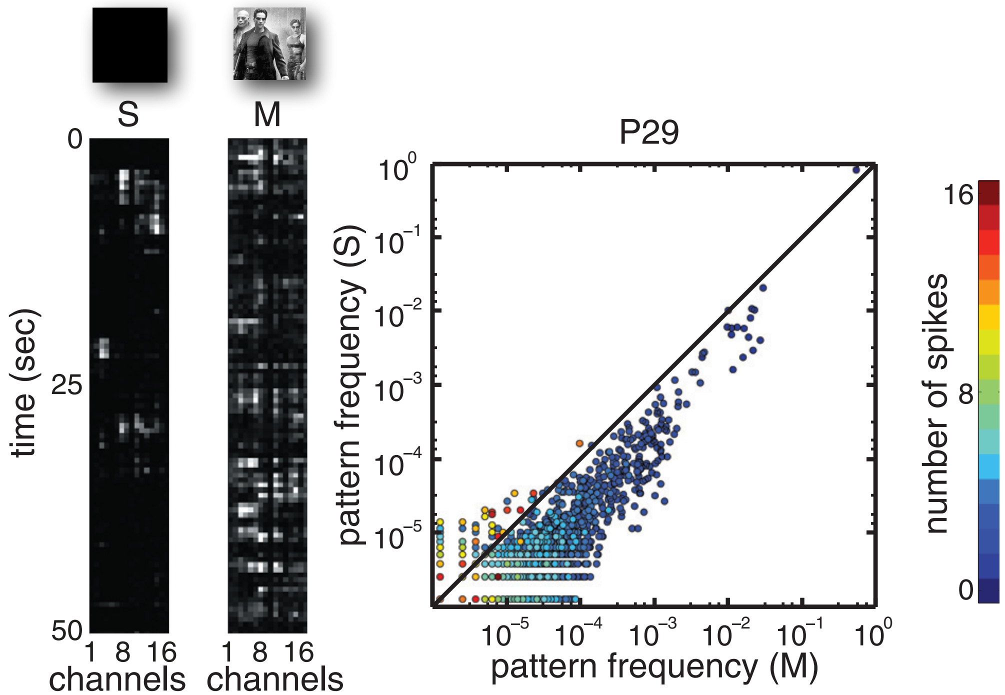


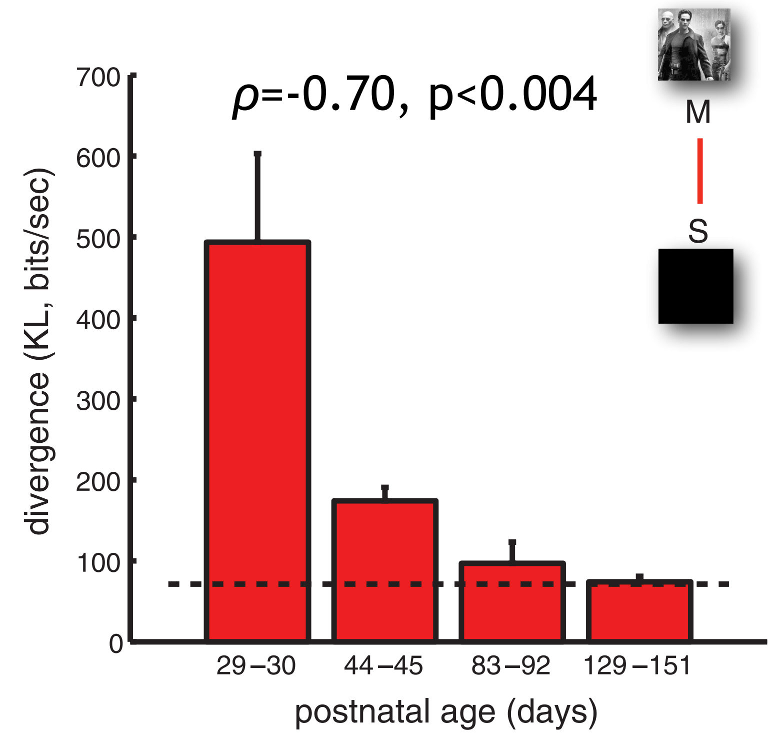
DEVELOPMENTAL CHANGES

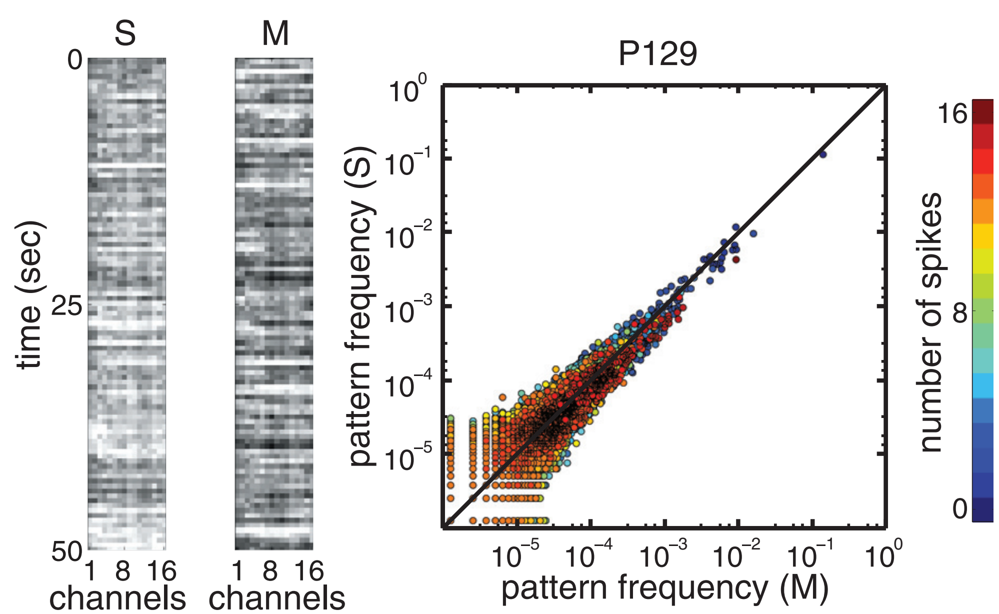




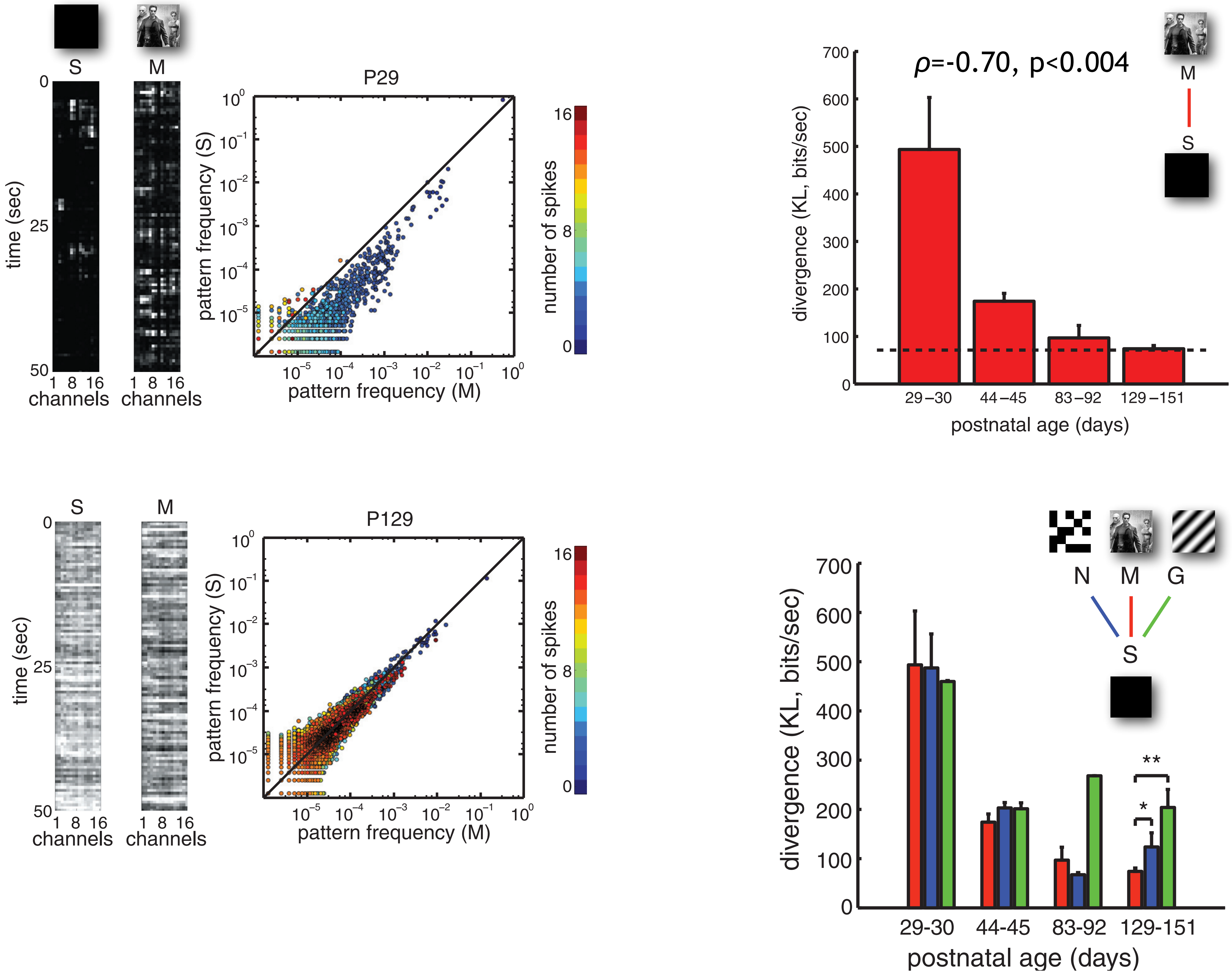
DEVELOPMENTAL CHANGES





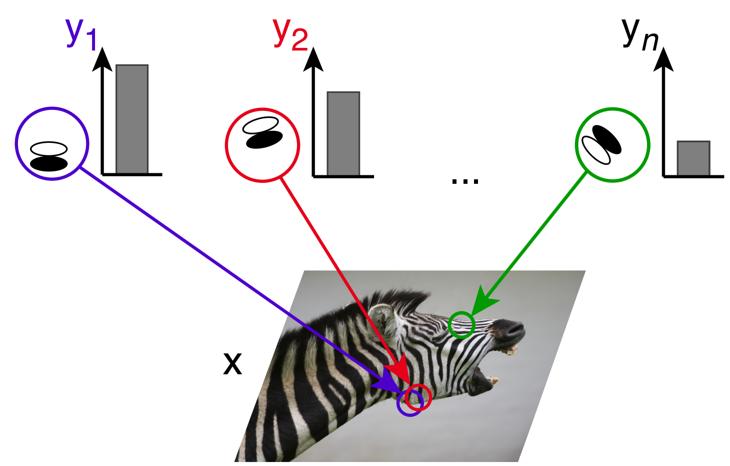


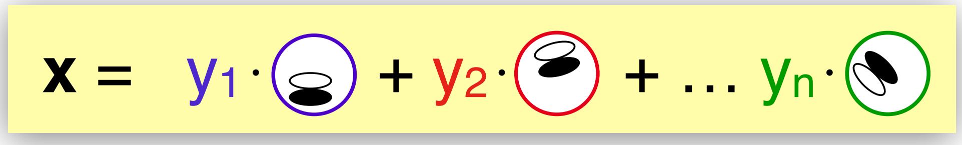
DEVELOPMENTAL CHANGES



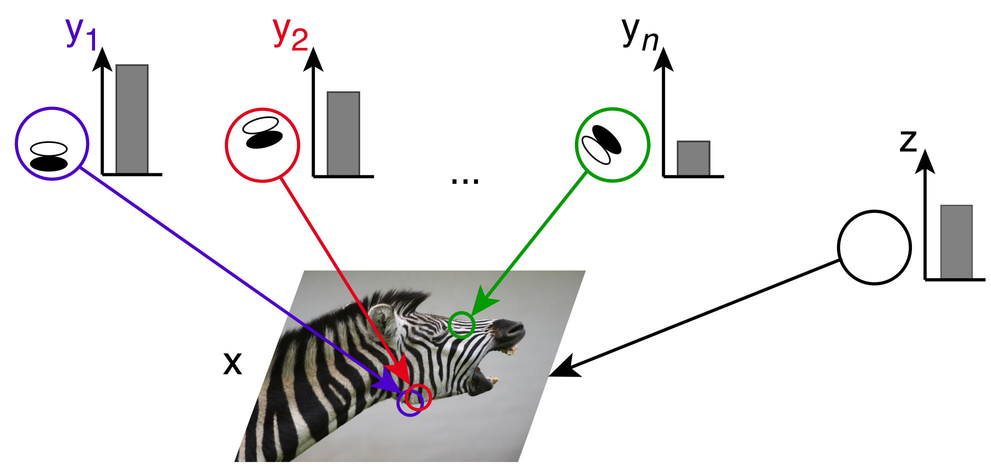
Berkes et al, Science 2011

AN IDEAL OBSERVER MODEL: GAUSSIAN SCALE MIXTURE



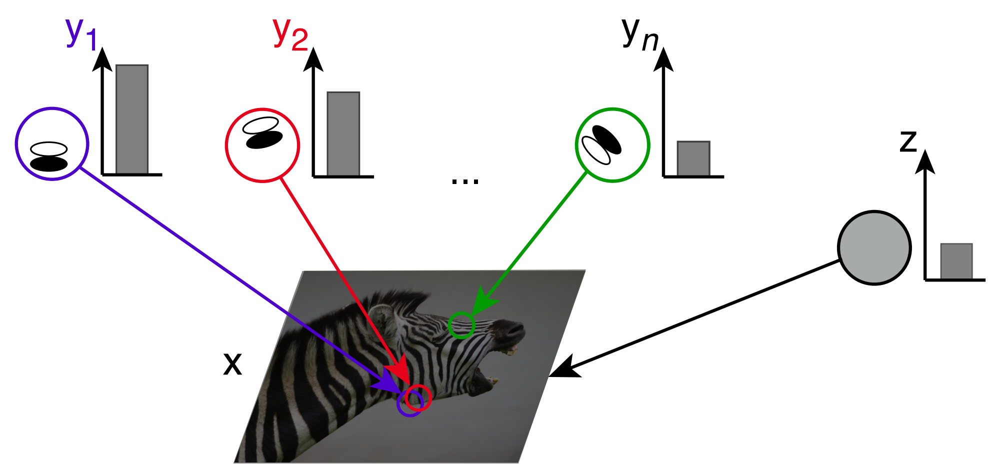


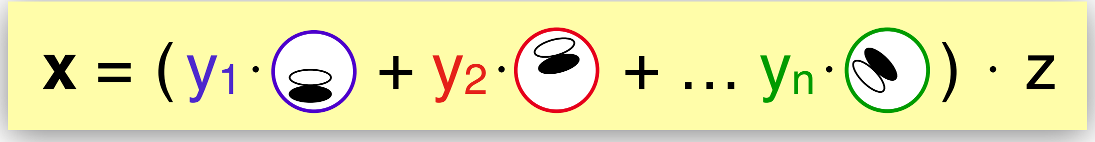
AN IDEAL OBSERVER MODEL: GAUSSIAN SCALE MIXTURE



32

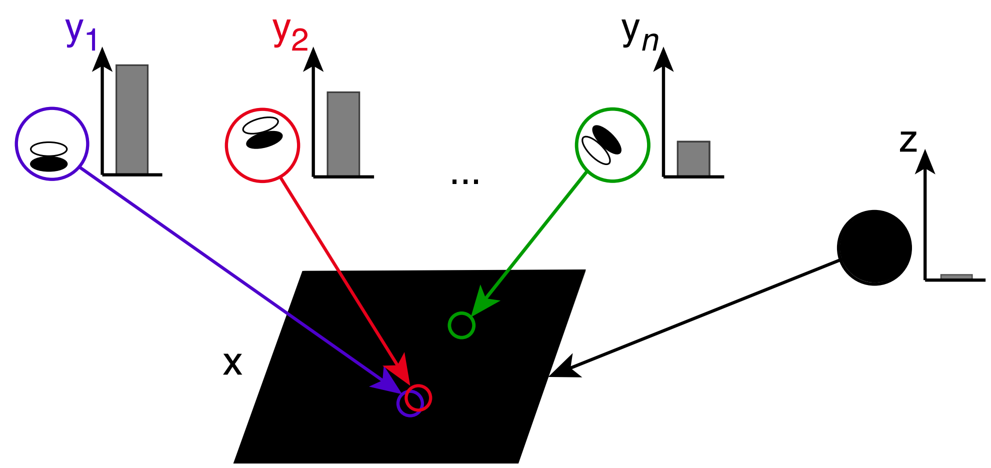
AN IDEAL OBSERVER MODEL: GAUSSIAN SCALE MIXTURE





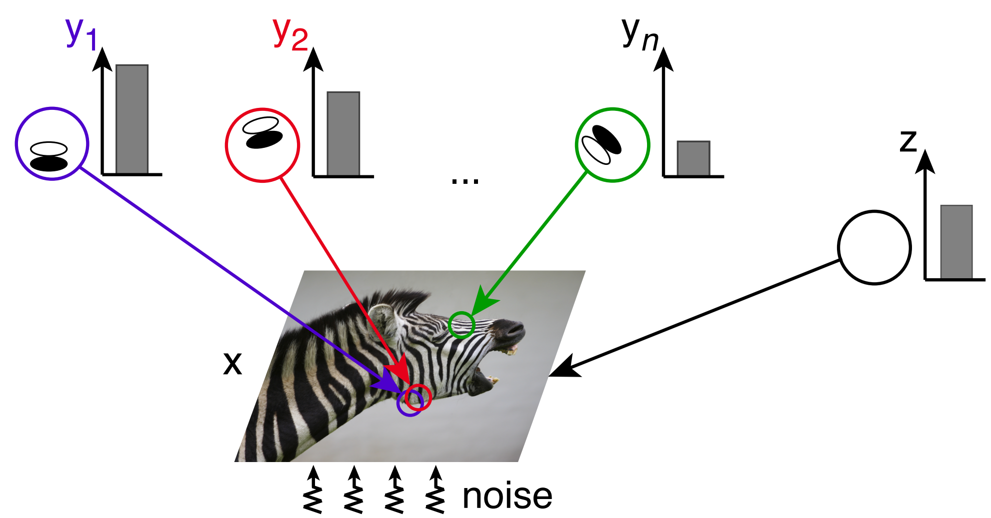
32

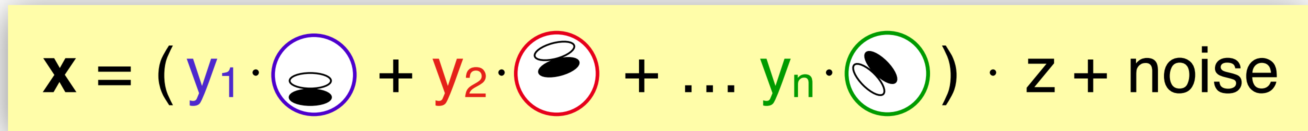
AN IDEAL OBSERVER MODEL: GAUSSIAN SCALE MIXTURE



32

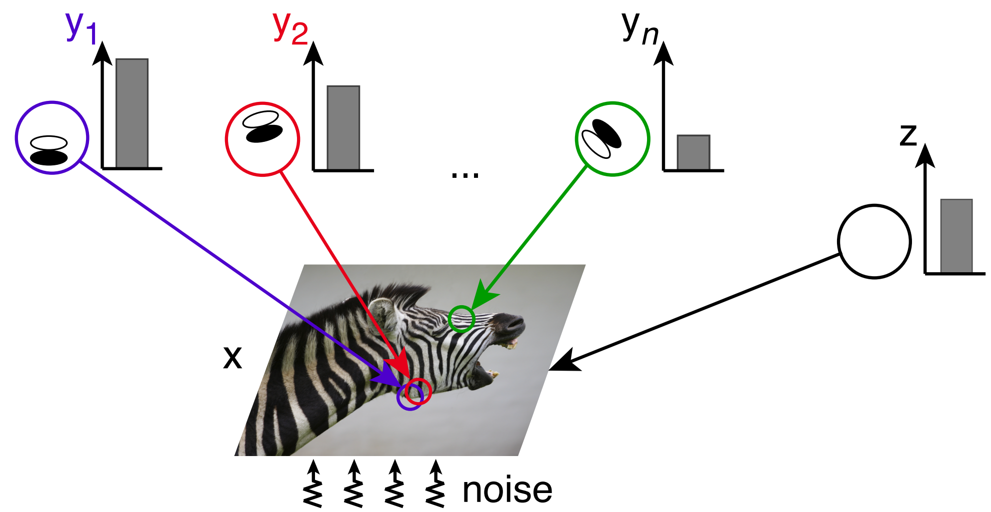
AN IDEAL OBSERVER MODEL: GAUSSIAN SCALE MIXTURE

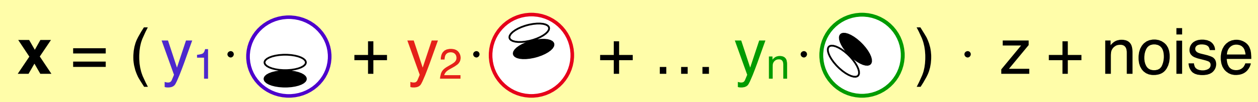




32

AN IDEAL OBSERVER MODEL: GAUSSIAN SCALE MIXTURE

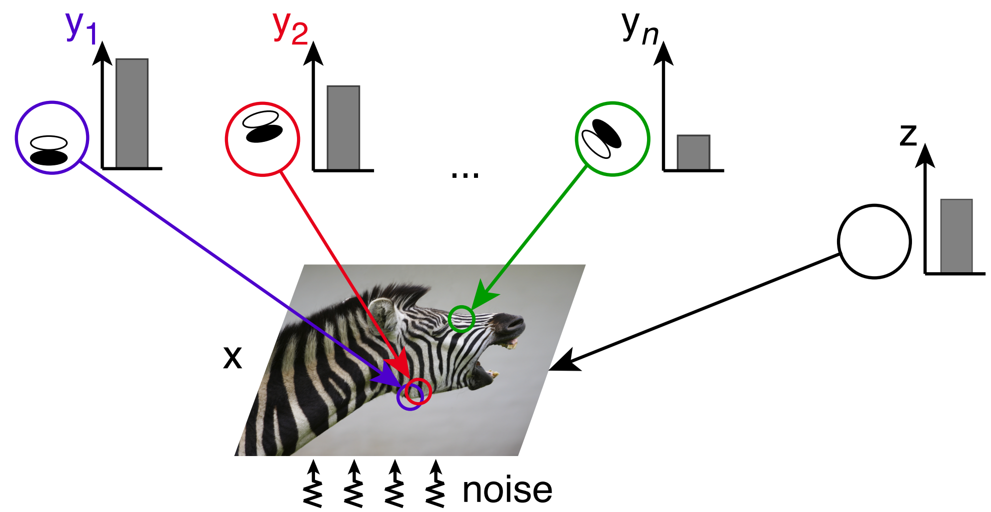




$\mathbf{y} \sim \mathcal{N}(\mathbf{0},\mathbf{C})$, $z \sim \Gamma(\ldots)$

32

AN IDEAL OBSERVER MODEL: GAUSSIAN SCALE MIXTURE

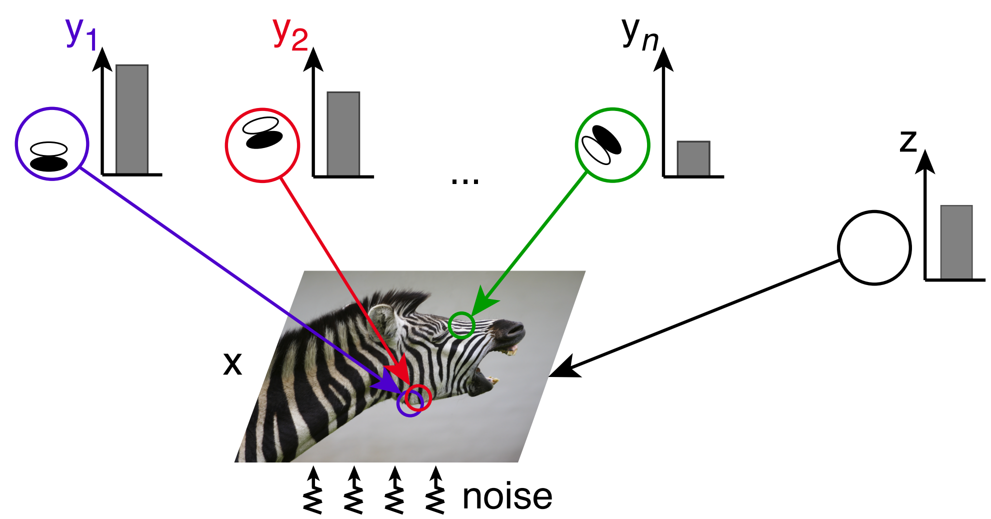


$$\mathbf{x} = (\, y_1 \cdot \bigcirc + y_2 \cdot \bigcirc + \ldots\, y_n \cdot \bigcirc \,) \cdot z + \text{noise}$$

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0},\mathbf{C}),\ z \sim \Gamma(\ldots)$$

Wainwright & Simoncelli 2000 → natural image statistics

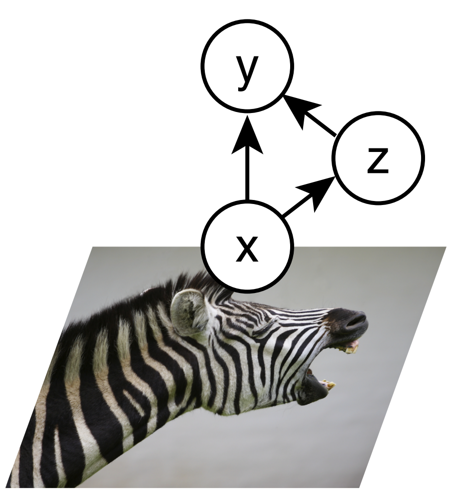
AN IDEAL OBSERVER MODEL: GAUSSIAN SCALE MIXTURE



$\mathbf{x}$ = (y_1 · ◯ + y_2 · ◯ + … y_n · ◯) · z + noise

$\mathbf{y} \sim \mathcal{N}(\mathbf{0},\mathbf{C})$, $z \sim \Gamma(\ldots)$

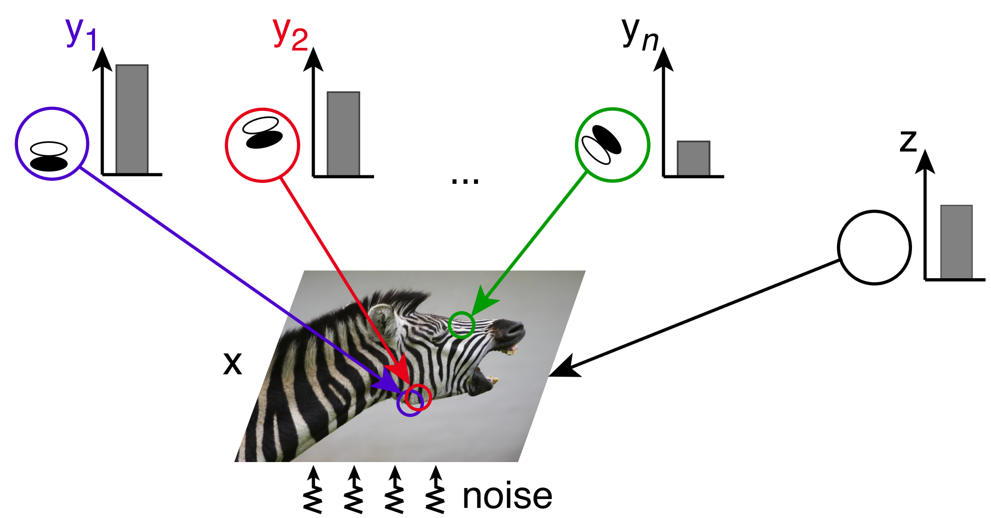
Wainwright & Simoncelli 2000 → natural image statistics



Schwartz & Simoncelli 2001
Coen-Cagli et al 2015 → trial-averaged neural responses

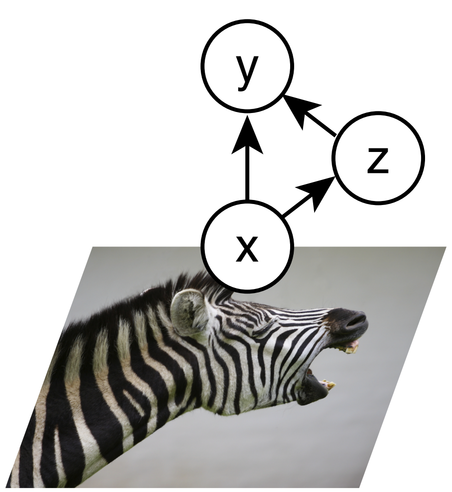
Schwartz & Dayan, J Vis 2009
Coen-Cagli et al, PLoS CB 2012 → psychophysics

AN IDEAL OBSERVER MODEL: GAUSSIAN SCALE MIXTURE



$$\mathbf{x} = (y_1 \cdot \bigcirc + y_2 \cdot \bigcirc + \ldots\ y_n \cdot \bigcirc) \cdot z + \text{noise}$$

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0},\mathbf{C}),\ z \sim \Gamma(\ldots)$$



Wainwright & Simoncelli 2000 → natural image statistics

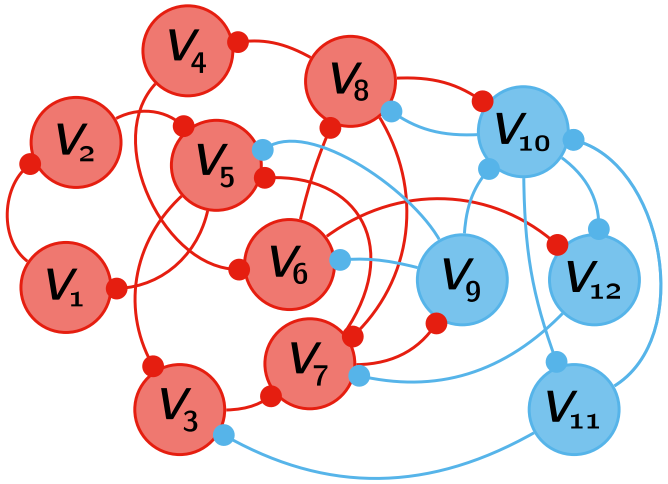
Schwartz & Simoncelli 2001
Coen-Cagli et al 2015 → trial-averaged neural responses

Schwartz & Dayan, J Vis 2009
Coen-Cagli et al, PLoS CB 2012 → psychophysics

Orbán et al, Neuron 2016 → neural *variability*

THE STABILIZED SUPRALINEAR NETWORK

recurrently coupled E/I network



$$\tau_i \frac{\mathrm{d}V_i}{\mathrm{d}t} = -V_i + V_{\mathrm{rest}} + \sum_j W_{ij}\, r_j(t) + h_i(t)$$

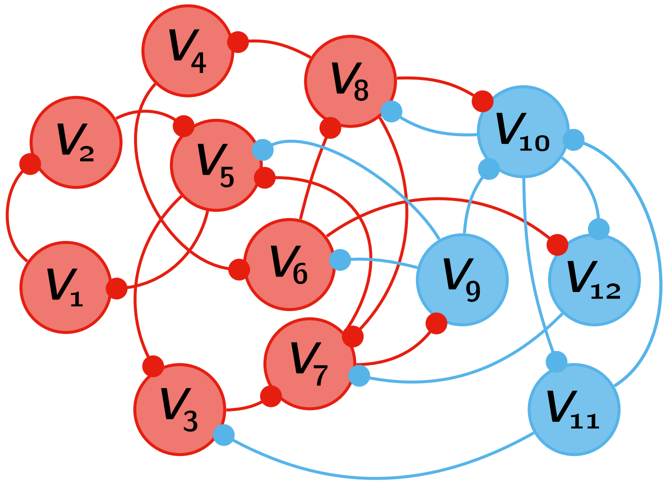
Ahmadian et al, Neural Comput 2013
Rubin et al, Neuron 2015

THE STABILIZED SUPRALINEAR NETWORK


Yashar Ahmadian


Ken Miller

recurrently coupled E/I network



low-pass filtered membrane potential

$$\tau_i \frac{\mathrm{d}V_i}{\mathrm{d}t} = -V_i + V_{\mathrm{rest}} + \sum_j W_{ij}\, r_j(t) + h_i(t)$$

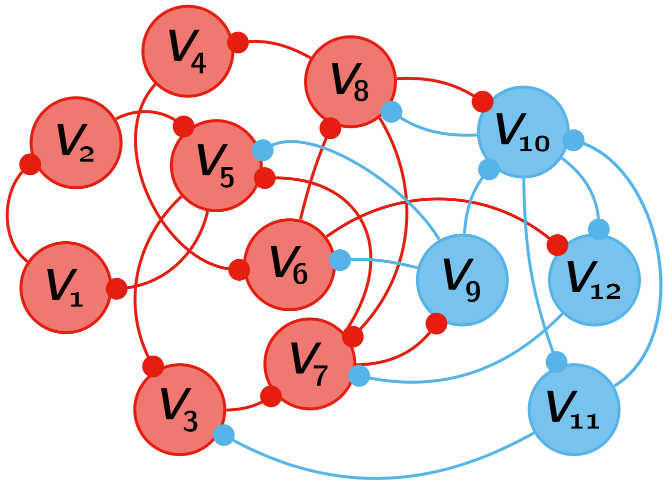
Ahmadian et al, Neural Comput 2013
Rubin et al, Neuron 2015

THE STABILIZED SUPRALINEAR NETWORK

Yashar Ahmadian

Ken Miller

recurrently coupled E/I network



low-pass filtered membrane potential

$$\tau_i \frac{\mathrm{d}V_i}{\mathrm{d}t} = -V_i + V_{\mathrm{rest}} + \sum_j W_{ij}\, r_j(t) + h_i(t)$$

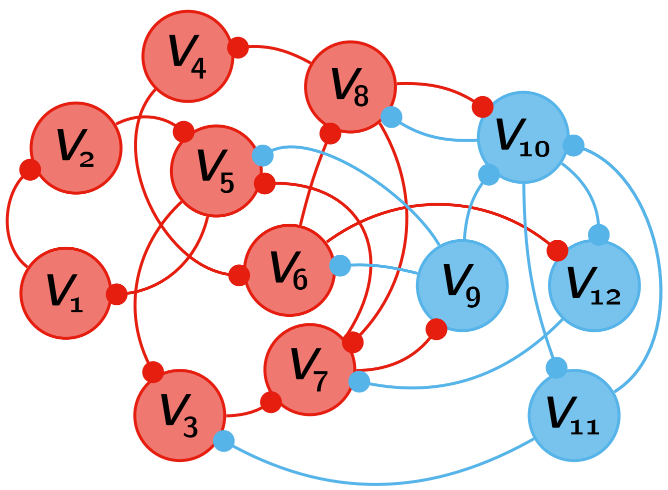
membrane time constant

Ahmadian et al, Neural Comput 2013
Rubin et al, Neuron 2015

THE STABILIZED SUPRALINEAR NETWORK

Yashar Ahmadian   Ken Miller

recurrently coupled E/I network



low-pass filtered membrane potential

$$\tau_i \frac{\mathrm{d}V_i}{\mathrm{d}t} = -V_i + V_{\mathrm{rest}} + \sum_j W_{ij}\, r_j(t) + h_i(t)$$

membrane time constant

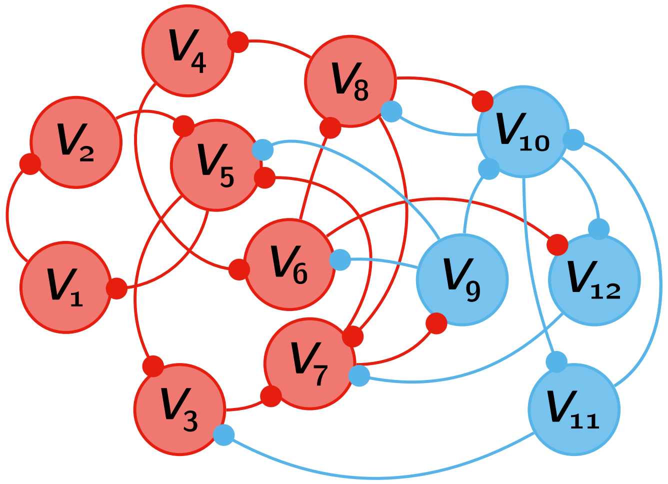
resting potential

Ahmadian et al, Neural Comput 2013
Rubin et al, Neuron 2015

THE STABILIZED SUPRALINEAR NETWORK


Yashar Ahmadian   Ken Miller

recurrently coupled E/I network



low-pass filtered membrane potential

$$\tau_i \frac{\mathrm{d}V_i}{\mathrm{d}t} = -V_i + V_{\mathrm{rest}} + \sum_j W_{ij}\, r_j(t) + h_i(t)$$

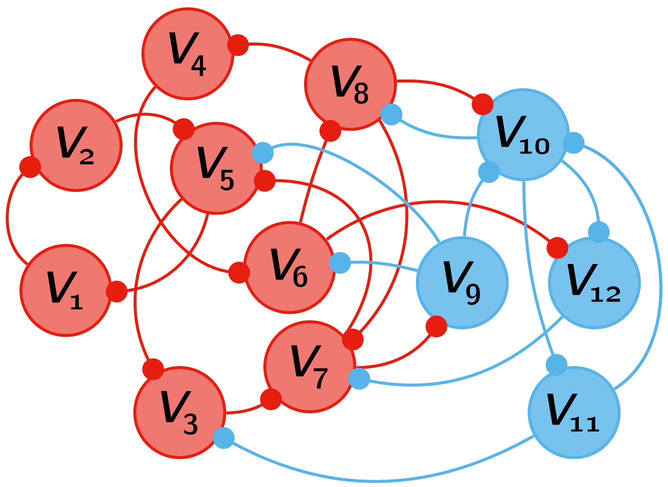
membrane time constant | resting potential | synaptic weights

Ahmadian et al, Neural Comput 2013
Rubin et al, Neuron 2015

THE STABILIZED SUPRALINEAR NETWORK


Yashar Ahmadian Ken Miller

recurrently coupled E/I network



$$\tau_i \frac{\mathrm{d}V_i}{\mathrm{d}t} = -V_i + V_{\mathrm{rest}} + \sum_j W_{ij}\, r_j(t) + h_i(t)$$

low-pass filtered membrane potential

firing rates

membrane time constant

resting potential

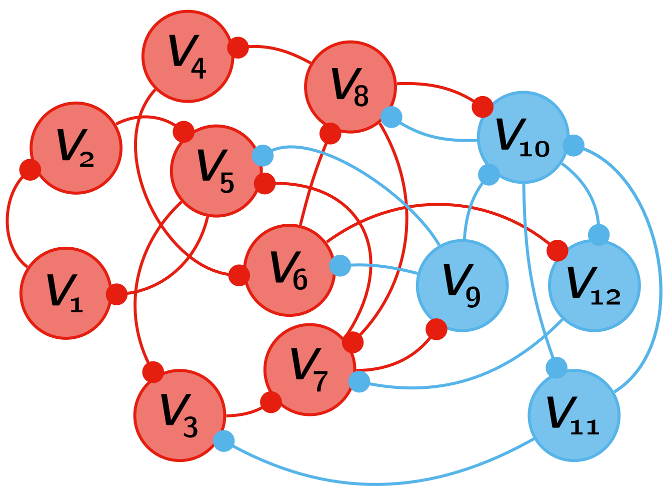
synaptic weights

Ahmadian et al, Neural Comput 2013
Rubin et al, Neuron 2015

THE STABILIZED SUPRALINEAR NETWORK


Yashar Ahmadian   Ken Miller

recurrently coupled E/I network



low-pass filtered membrane potential

firing rates

external input

$$\tau_i \frac{\mathrm{d}V_i}{\mathrm{d}t} = -V_i + V_{\mathrm{rest}} + \sum_j W_{ij}\, r_j(t) + h_i(t)$$

membrane time constant

resting potential

synaptic weights

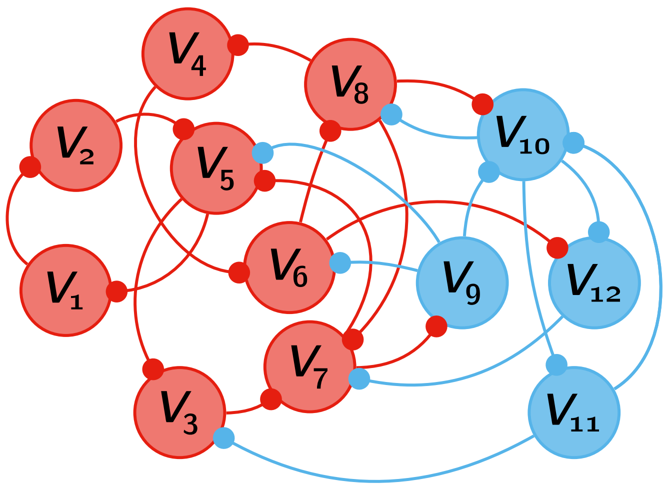
Ahmadian et al, Neural Comput 2013
Rubin et al, Neuron 2015

THE STABILIZED SUPRALINEAR NETWORK

Yashar Ahmadian   Ken Miller

recurrently coupled E/I network



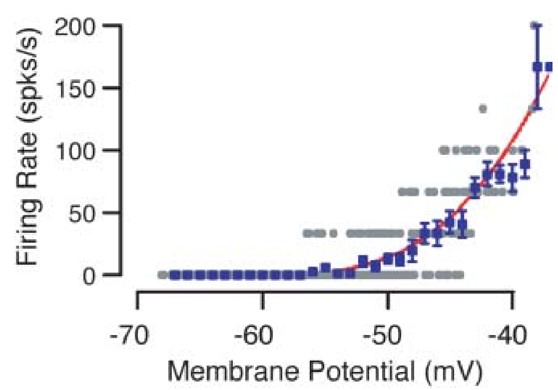
$$\tau_i \frac{\mathrm{d}V_i}{\mathrm{d}t} = -V_i + V_{\mathrm{rest}} + \sum_j W_{ij}\, r_j(t) + h_i(t)$$

low-pass filtered membrane potential ($\frac{\mathrm{d}V_i}{\mathrm{d}t}$)
firing rates ($r_j(t)$)
external input ($h_i(t)$)
membrane time constant (τ_i)
resting potential (V_{rest})
synaptic weights (W_{ij})

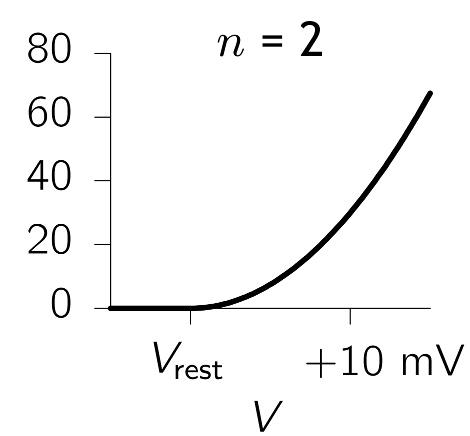
Ahmadian et al, Neural Comput 2013
Rubin et al, Neuron 2015

with expansive firing rate nonlinearities

$$r_i(t) = k \left\lfloor V_i(t) - V_{\mathrm{rest}} \right\rfloor_+^n$$

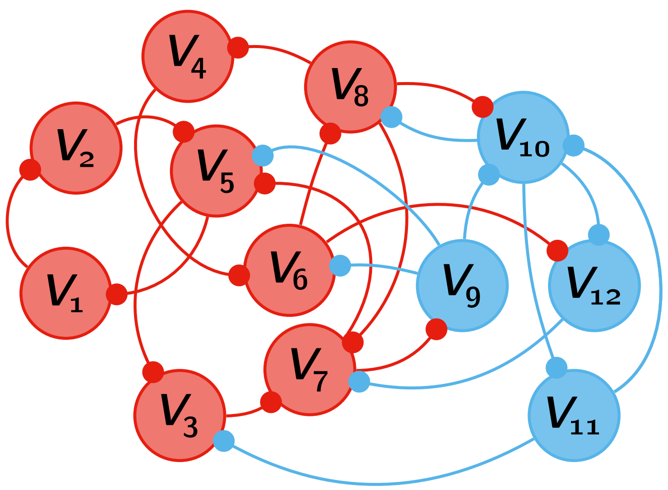


Priebe & Ferster, Neuron 2005



THE STOCHASTIC STABILIZED SUPRALINEAR NETWORK

Guillaume Hennequin   Yashar Ahmadian   Ken Miller



recurrently coupled E/I network

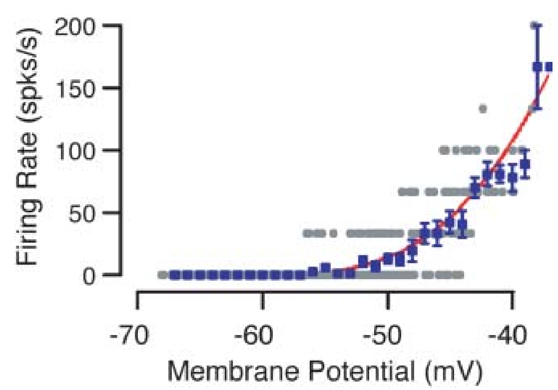
$$\tau_i \frac{\mathrm{d}V_i}{\mathrm{d}t} = -V_i + V_{\mathrm{rest}} + \sum_j W_{ij}\, r_j(t) + h_i(t) + \mathrm{noise}$$

- low-pass filtered membrane potential (V_i)
- firing rates ($r_j(t)$)
- external input ($h_i(t)$)
- membrane time constant (τ_i)
- resting potential (V_{rest})
- synaptic weights (W_{ij})

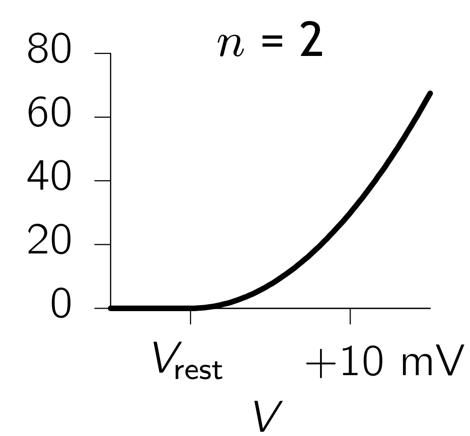
Ahmadian et al, Neural Comput 2013
Rubin et al, Neuron 2015
Hennequin et al, accepted

with expansive firing rate nonlinearities

$$r_i(t) = k \left\lfloor V_i(t) - V_{\mathrm{rest}} \right\rfloor_+^n$$



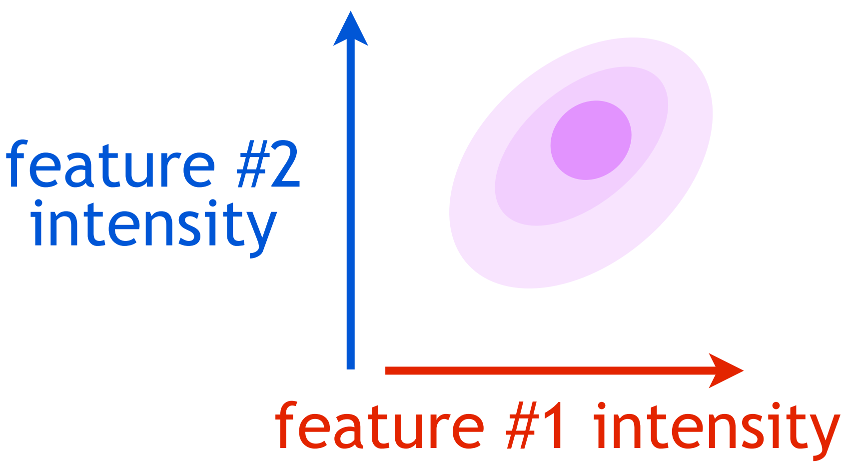
Priebe & Ferster, Neuron 2005

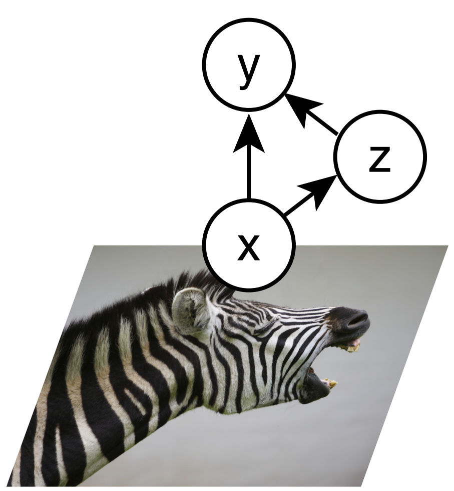


BRINGING THE TWO WORLDS TOGETHER

ideal observer: GSM

$$P(y_1, y_2 \mid x)$$





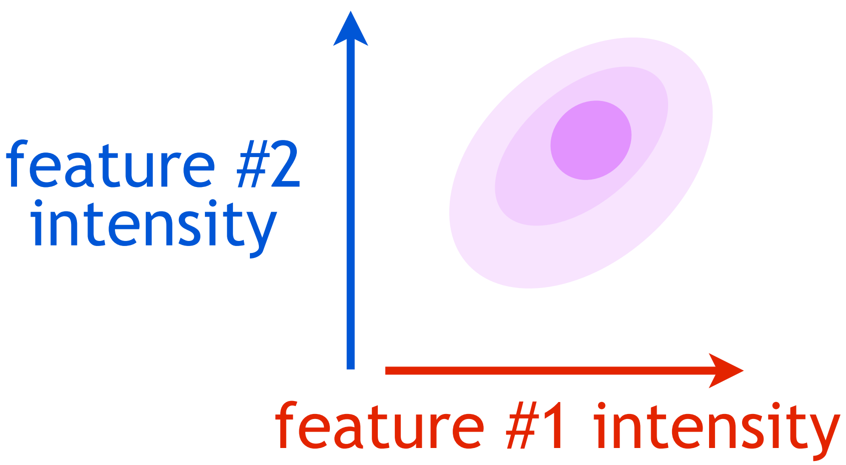
BRINGING THE TWO WORLDS TOGETHER

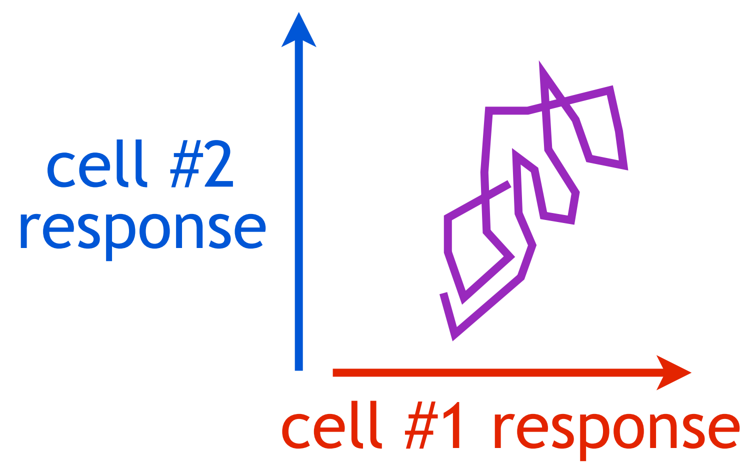
ideal observer: GSM

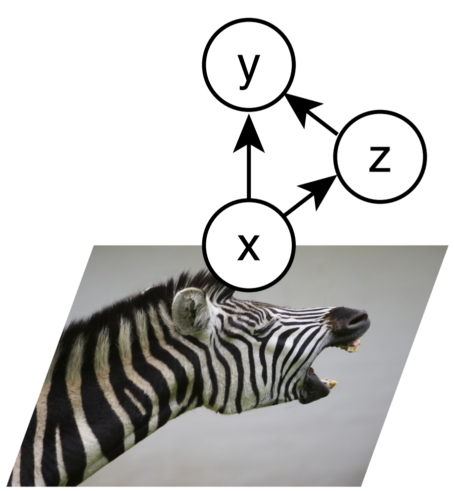
visual cortex: neural network

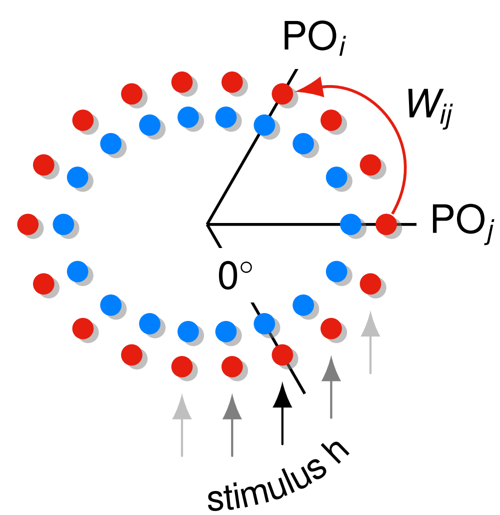
$$P(y_1, y_2 \mid x)$$

$$P(r_1, r_2 \mid x; \text{network parameters})$$









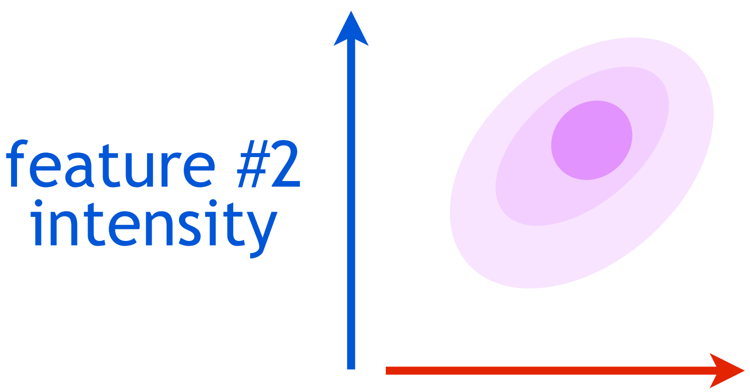
BRINGING THE TWO WORLDS TOGETHER

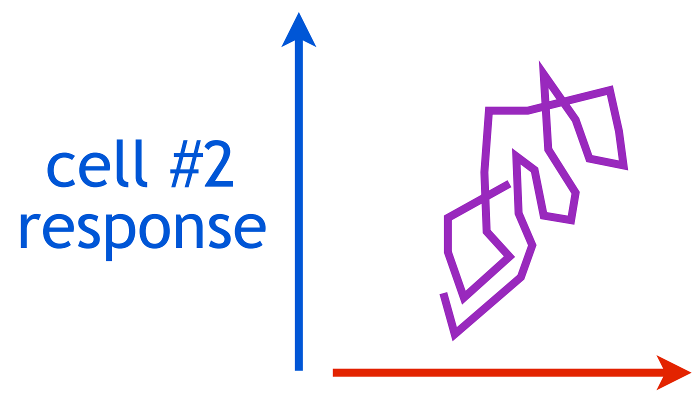
ideal observer: GSM

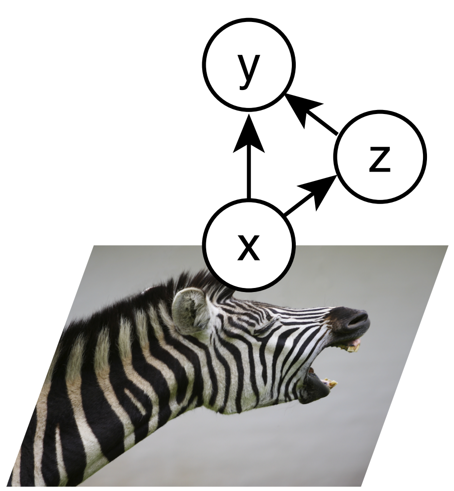
visual cortex: neural network

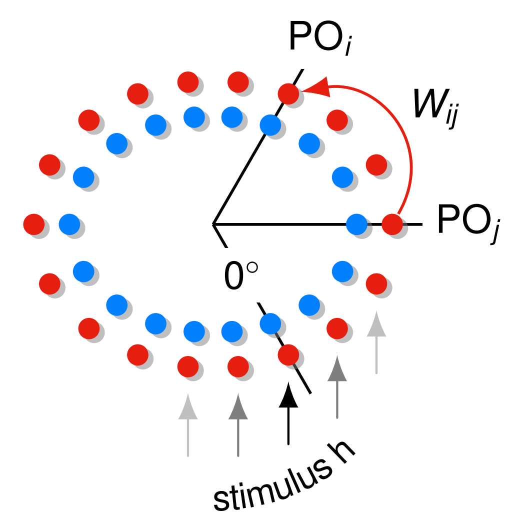
Rodrigo Echeveste   Guillaume Hennequin

$P(y_1, y_2 \mid x)$ — train → $P(r_1, r_2 \mid x; \text{network parameters})$









BRINGING THE TWO WORLDS TOGETHER

ideal observer: GSM
- full distribution (2nd order)

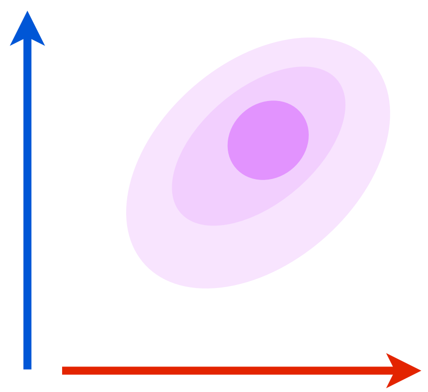
visual cortex: neural network
- recurrent
- E-I
- expansive nonlinearity


Rodrigo Echeveste


Guillaume Hennequin

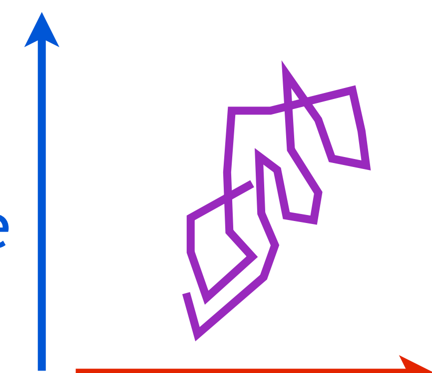
$P(y_1, y_2 \mid x)$ → train → $P(r_1, r_2 \mid x; \text{network parameters})$

feature #2 intensity

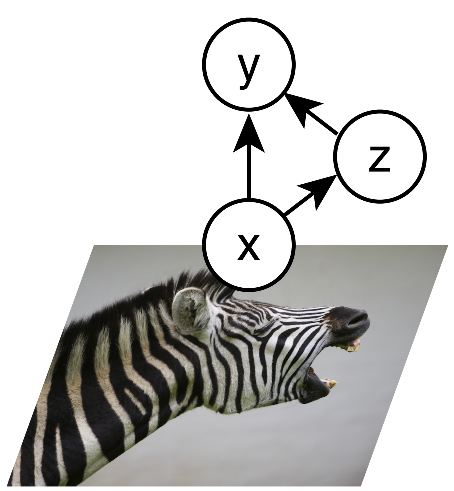


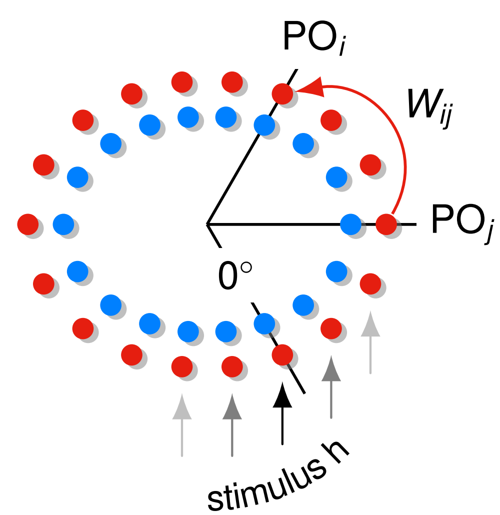
feature #1 intensity

cell #2 response



cell #1 response





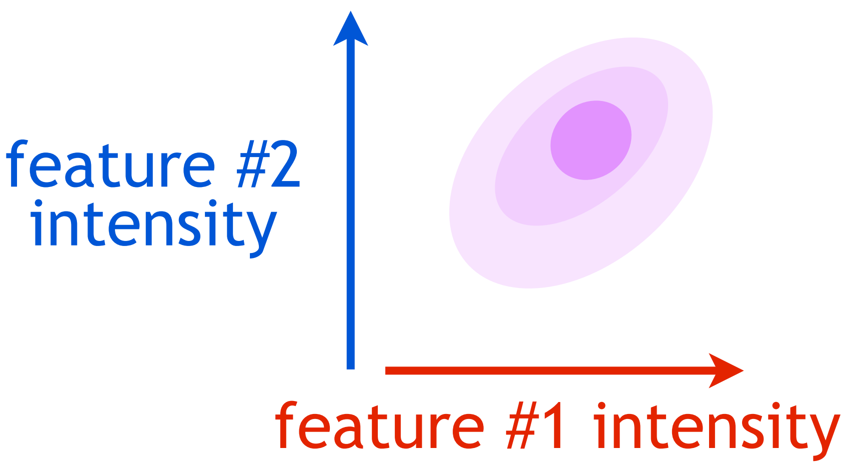
BRINGING THE TWO WORLDS TOGETHER

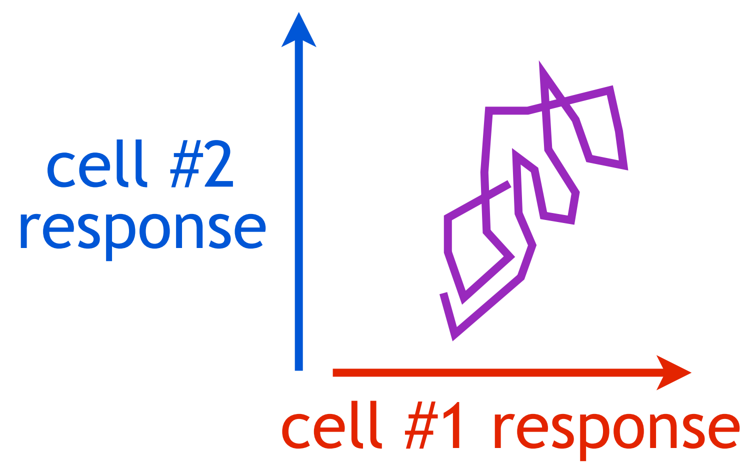
ideal observer: GSM
- full distribution (2nd order)

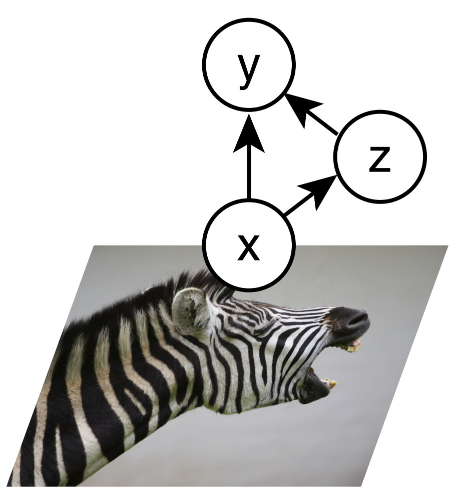
visual cortex: neural network
- recurrent
- E-I
- expansive nonlinearity

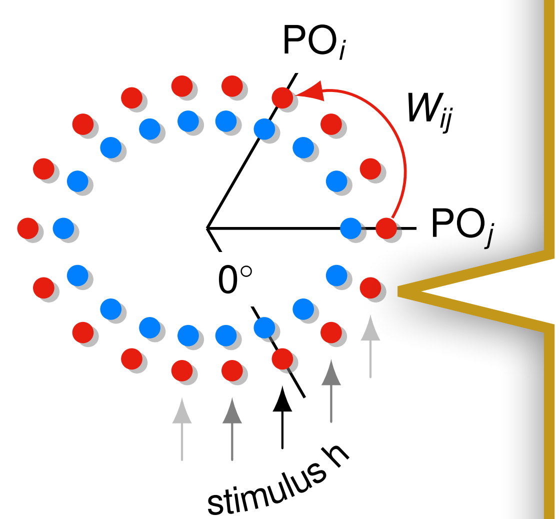

Rodrigo Echeveste   Guillaume Hennequin

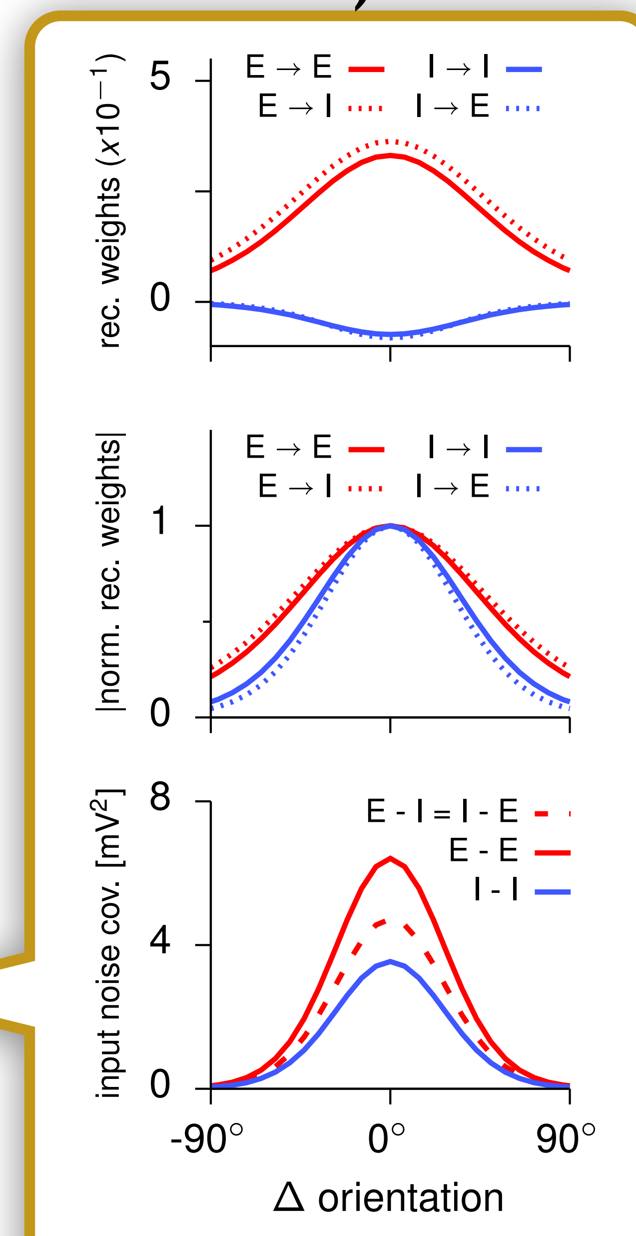
$P(y_1, y_2 \mid x)$ train $P(r_1, r_2 \mid x;\ \text{network parameters})$





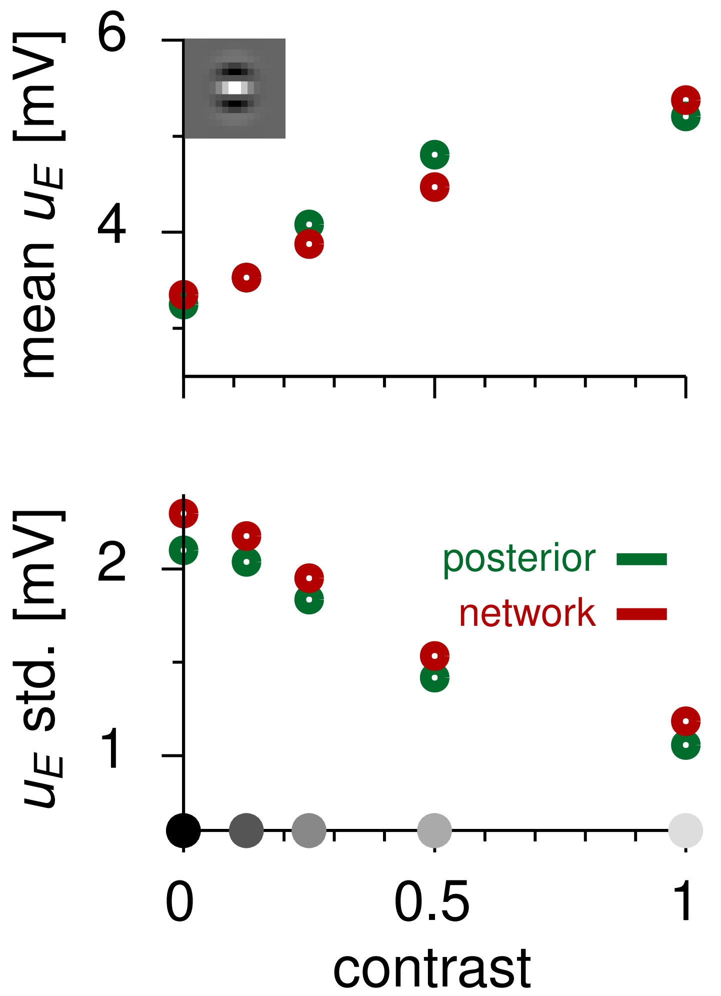




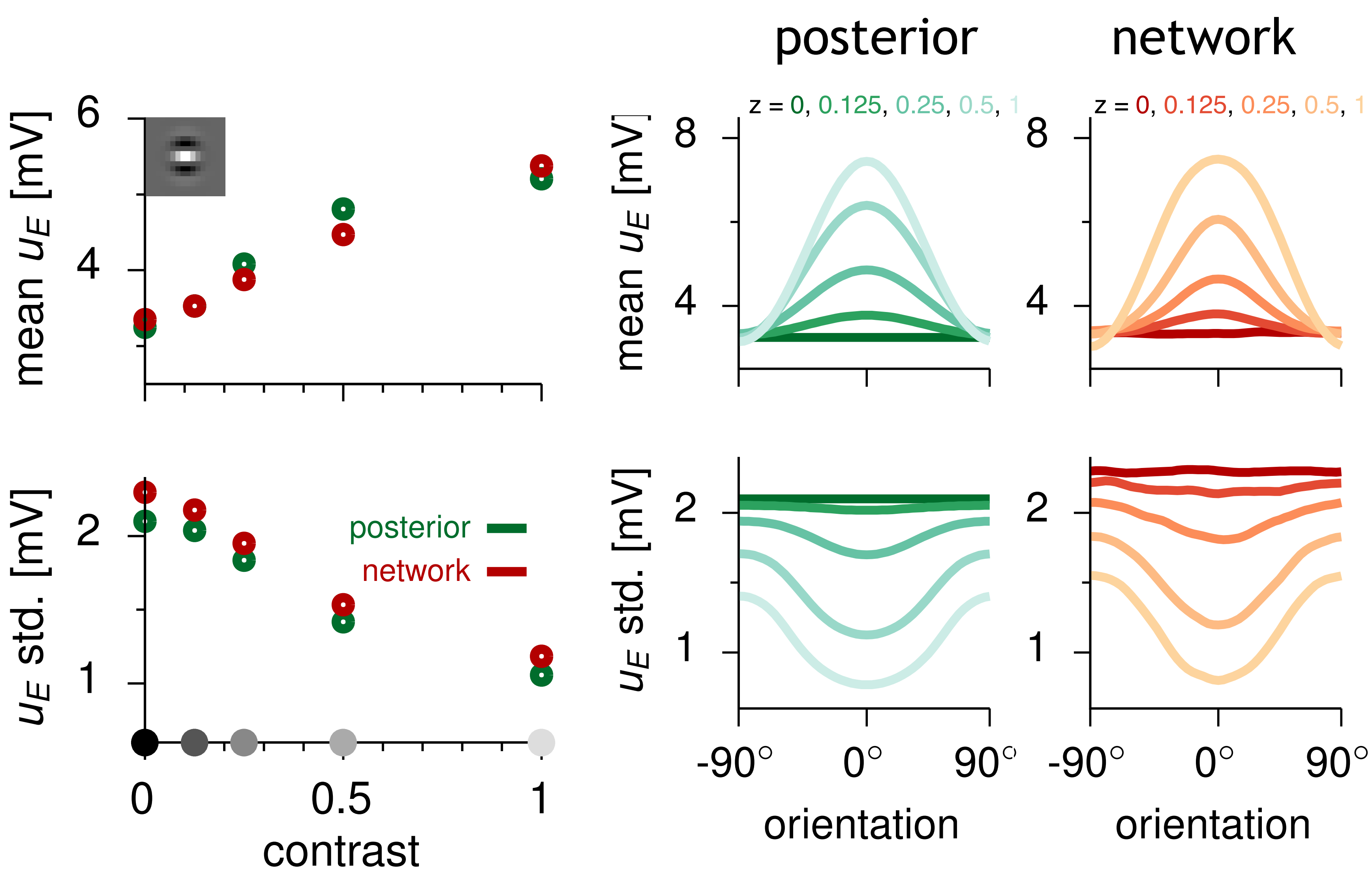


Echeveste et al, in prep

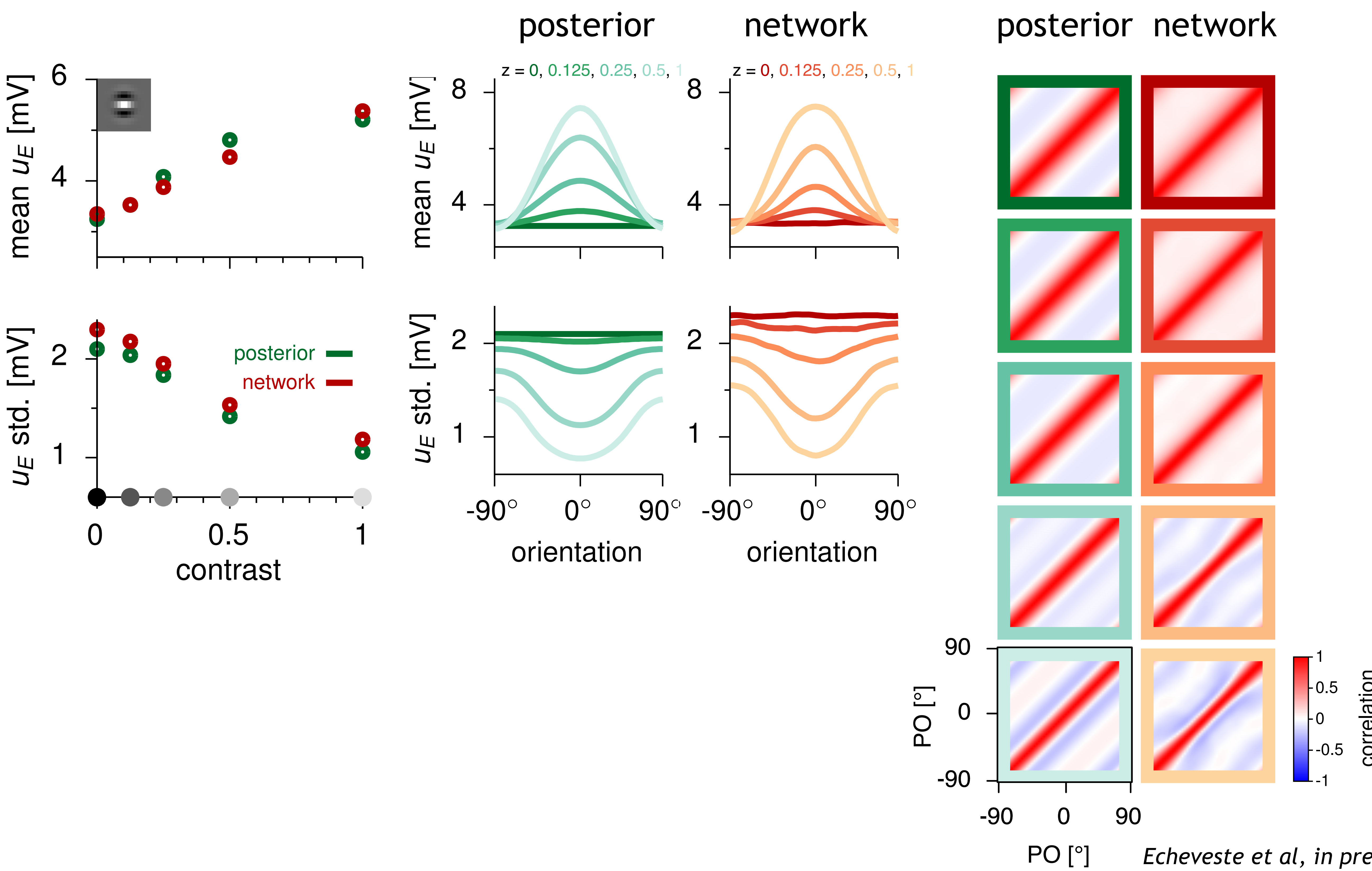
GSM POSTERIOR VS. NETWORK: TRAINING



GSM POSTERIOR VS. NETWORK: TRAINING

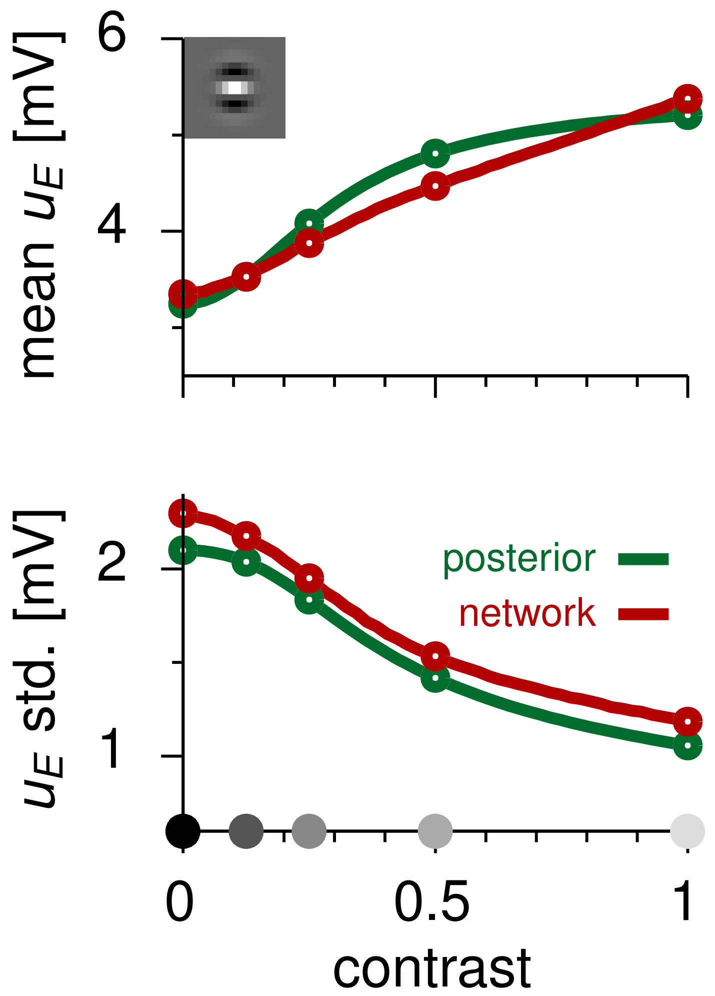


GSM POSTERIOR VS. NETWORK: TRAINING

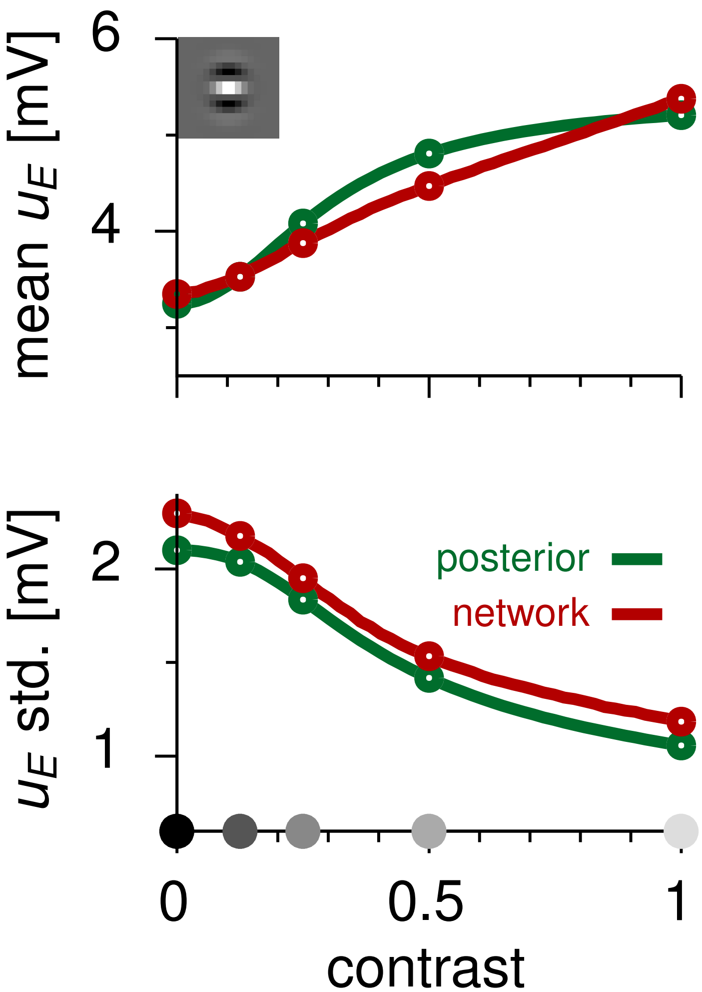


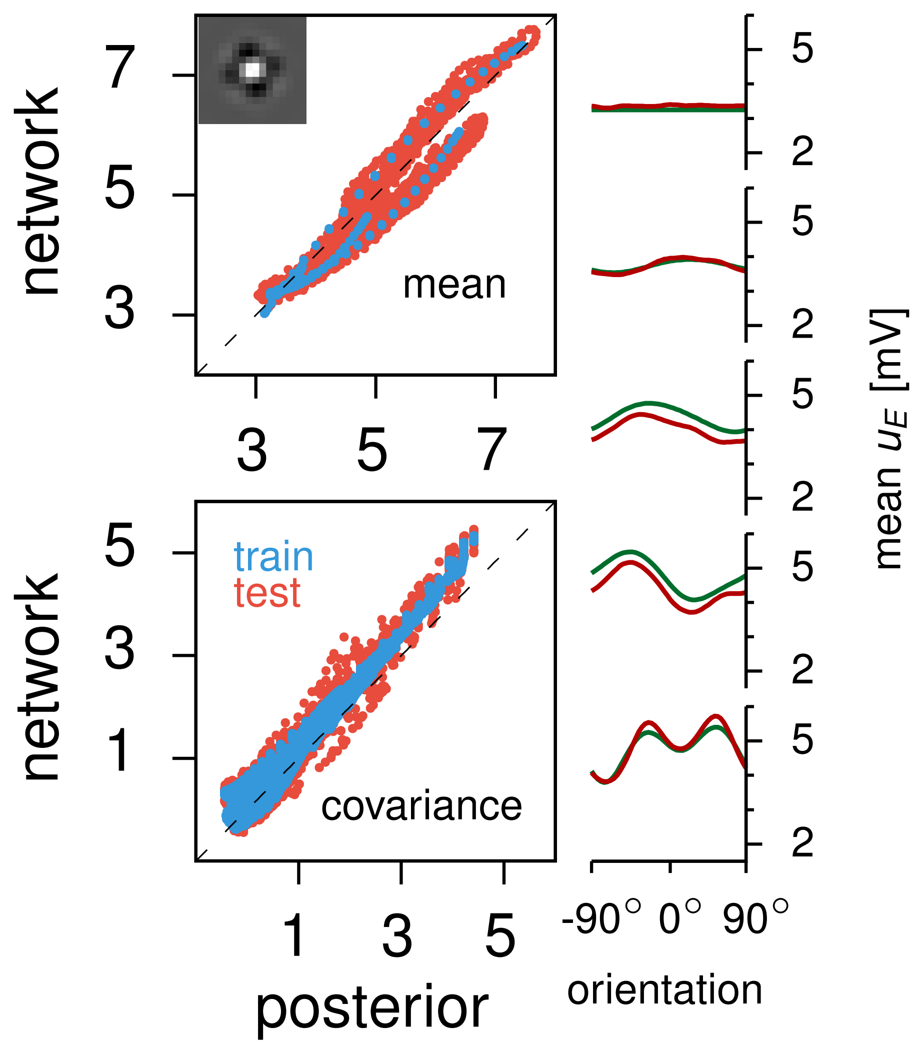
Echeveste et al, in prep

GSM POSTERIOR VS. NETWORK: GENERALISATION



GSM POSTERIOR VS. NETWORK: GENERALISATION

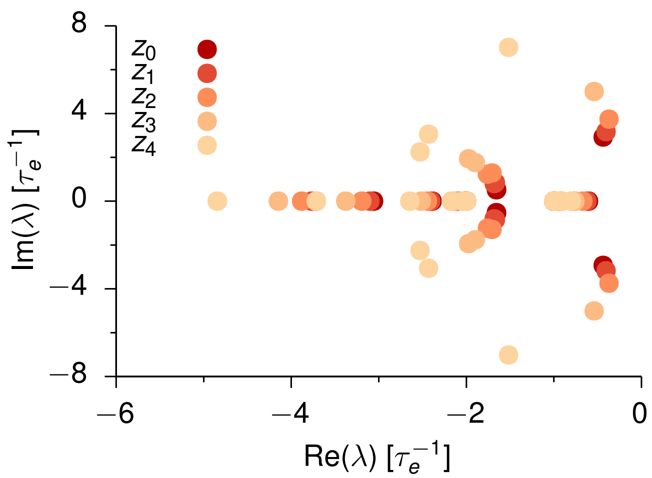




Echeveste et al, in prep

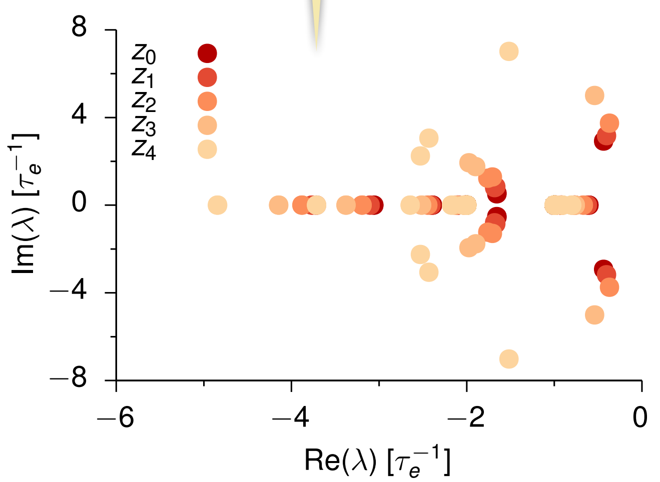
THE SAMPLING-OPTIMIZED NETWORK IS STABLE, FAST & UNBALANCED

THE SAMPLING-OPTIMIZED NETWORK IS STABLE, FAST & UNBALANCED



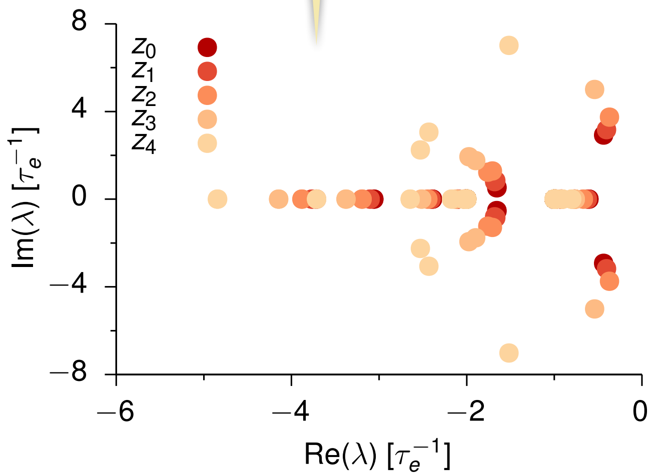
THE SAMPLING-OPTIMIZED NETWORK IS STABLE, FAST & UNBALANCED

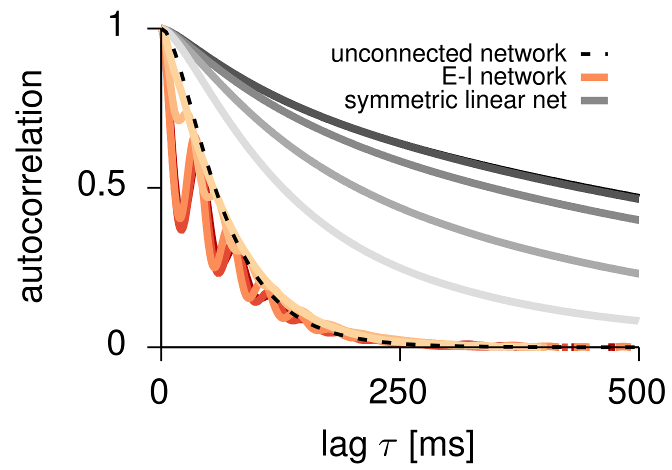
always stable



THE SAMPLING-OPTIMIZED NETWORK IS STABLE, FAST & UNBALANCED

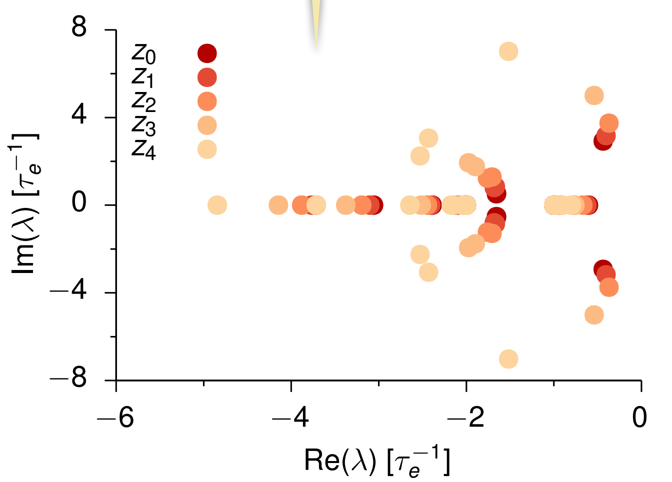
always stable





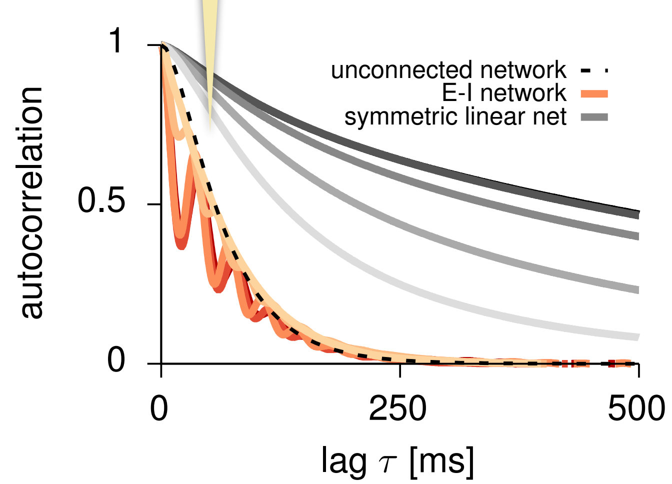
THE SAMPLING-OPTIMIZED NETWORK IS STABLE, FAST & UNBALANCED

always stable



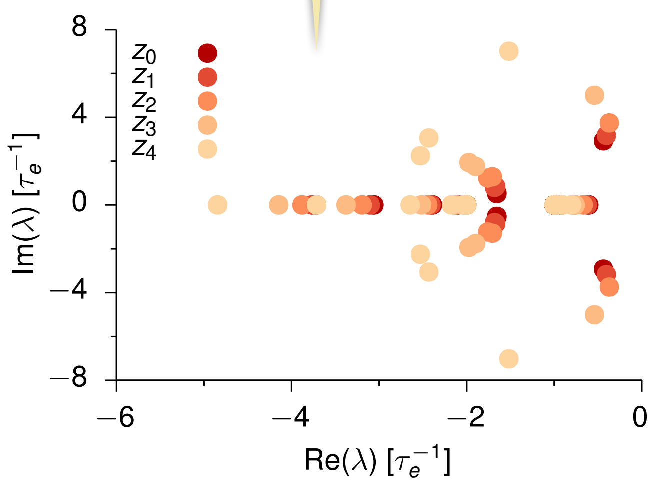
near-optimally fast

Hennequin et al, 2014



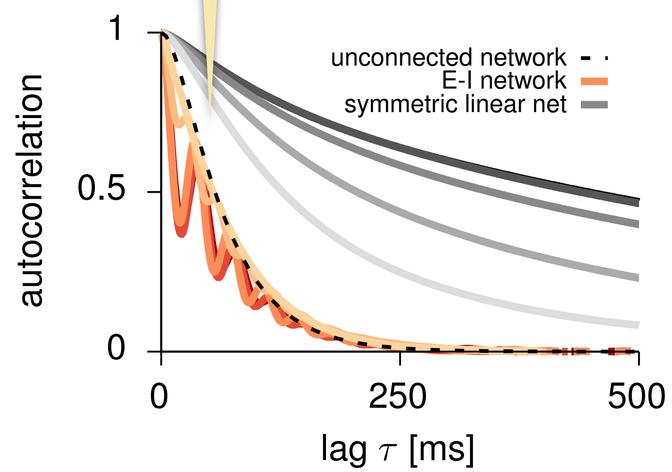
THE SAMPLING-OPTIMIZED NETWORK IS STABLE, FAST & UNBALANCED

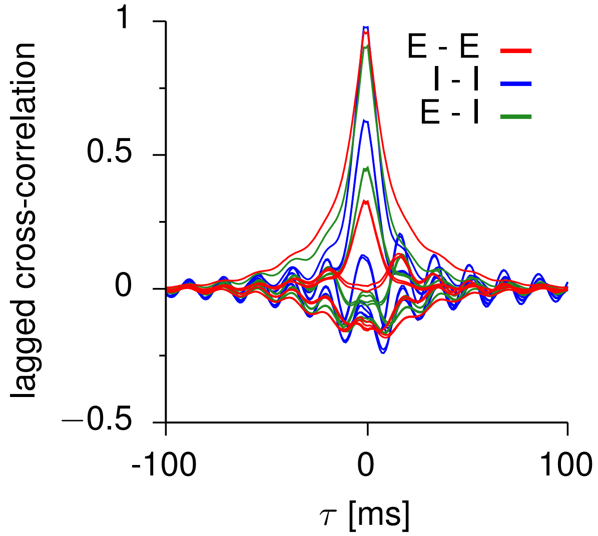
always stable



near-optimally fast

Hennequin et al, 2014



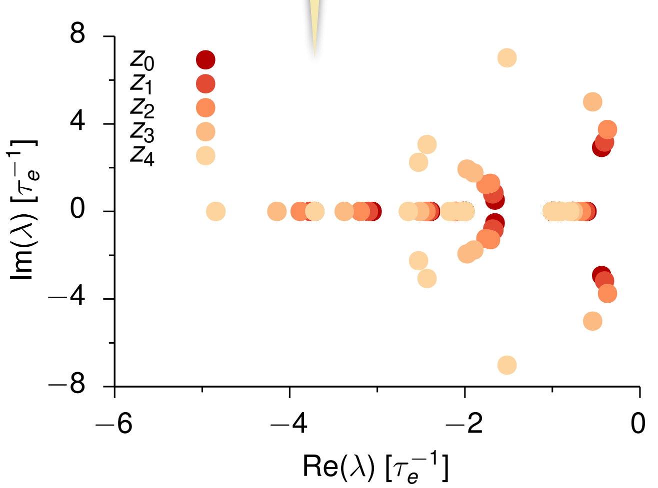


Echeveste et al, in prep

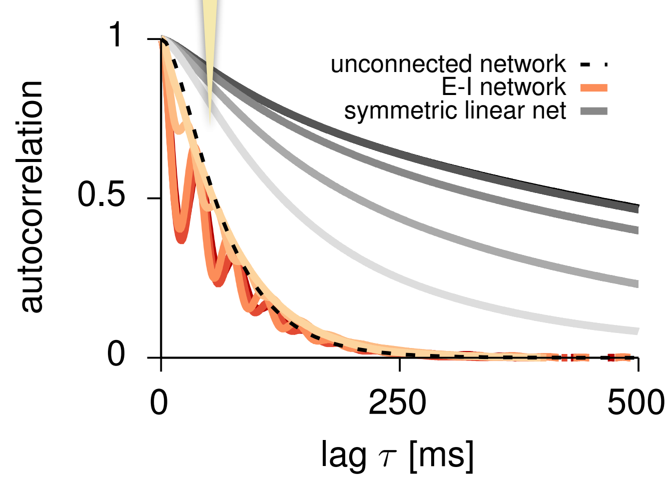
THE SAMPLING-OPTIMIZED NETWORK IS STABLE, FAST & UNBALANCED

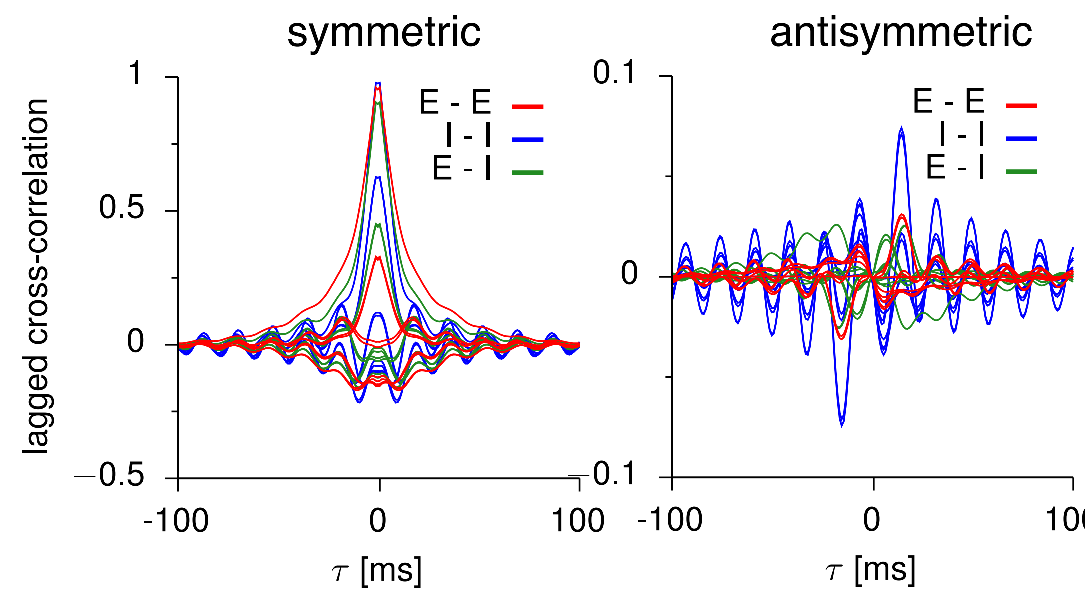
always stable

near-optimally fast



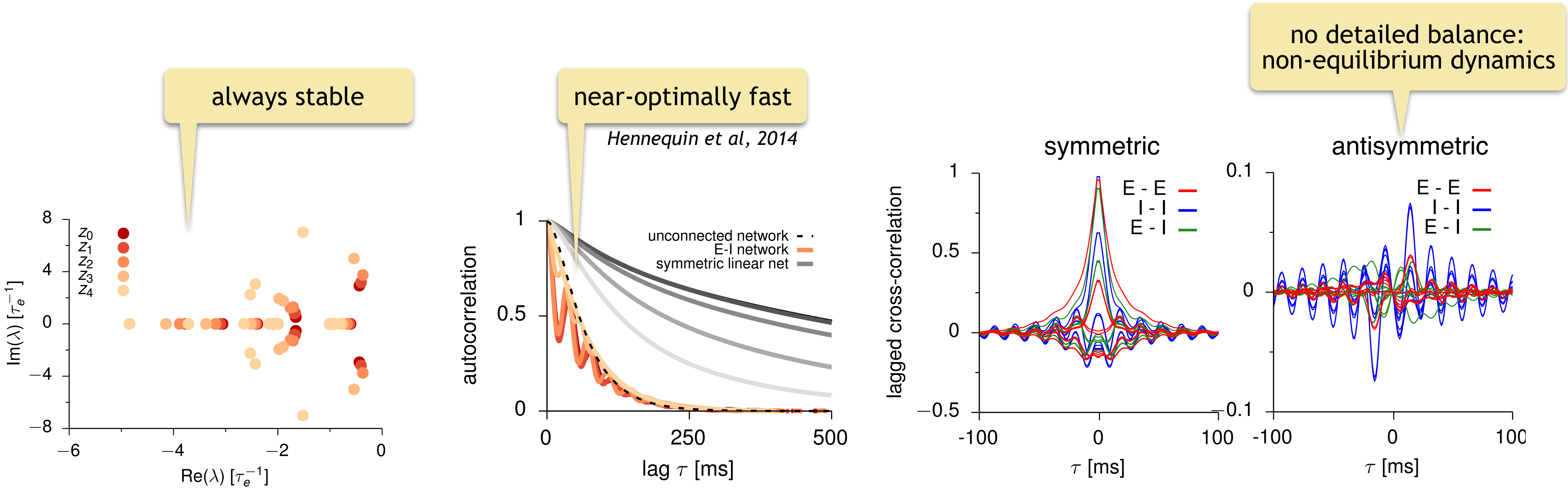
Hennequin et al, 2014





Echeveste et al, in prep

THE SAMPLING-OPTIMIZED NETWORK IS STABLE, FAST & UNBALANCED



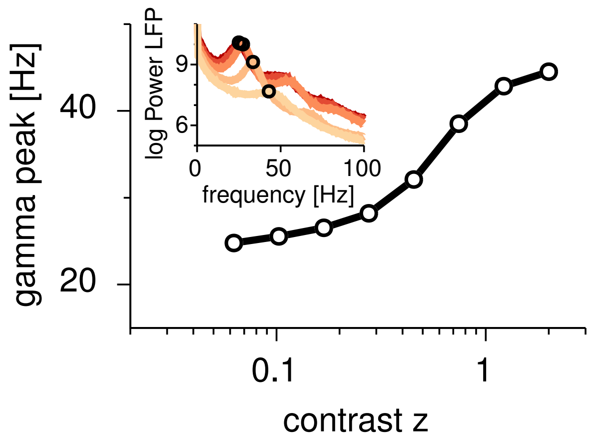
Echeveste et al, in prep

THE SAMPLING-OPTIMIZED NETWORK IS IN A REALISTIC DYNAMICAL REGIME

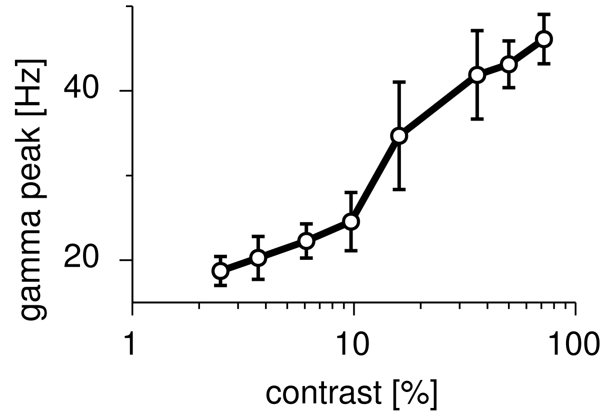
THE SAMPLING-OPTIMIZED NETWORK IS IN A REALISTIC DYNAMICAL REGIME

contrast-dependent
oscillations

MODEL



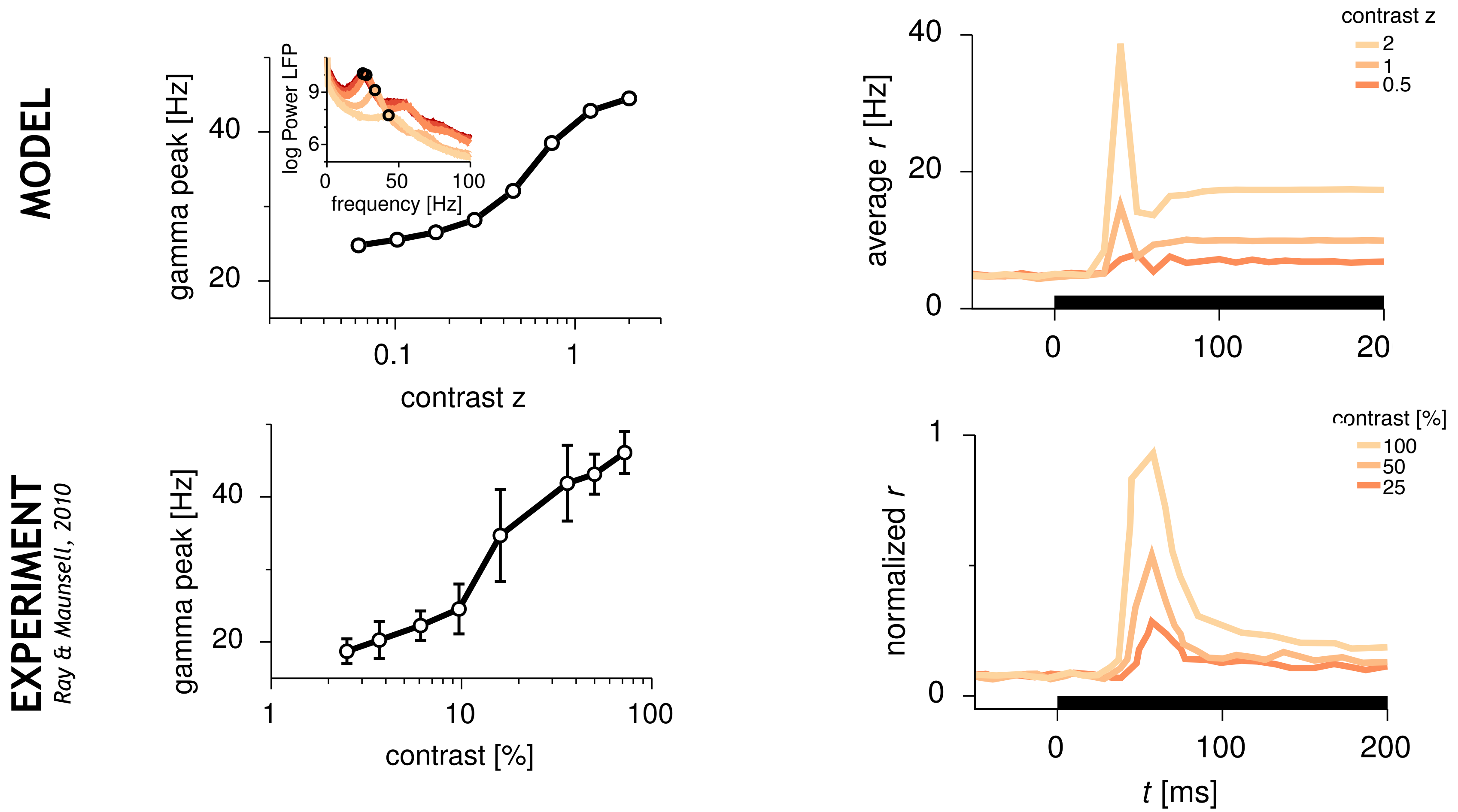
EXPERIMENT
Ray & Maunsell, 2010



THE SAMPLING-OPTIMIZED NETWORK IS IN A REALISTIC DYNAMICAL REGIME

contrast-dependent oscillations

contrast-dependent stimulus-onset transients

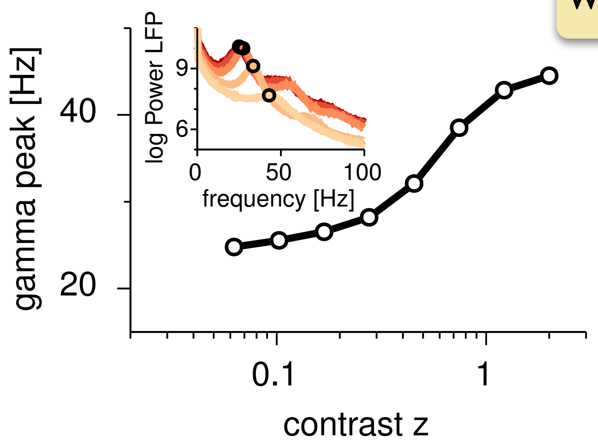


THE SAMPLING-OPTIMIZED NETWORK IS IN A REALISTIC DYNAMICAL REGIME

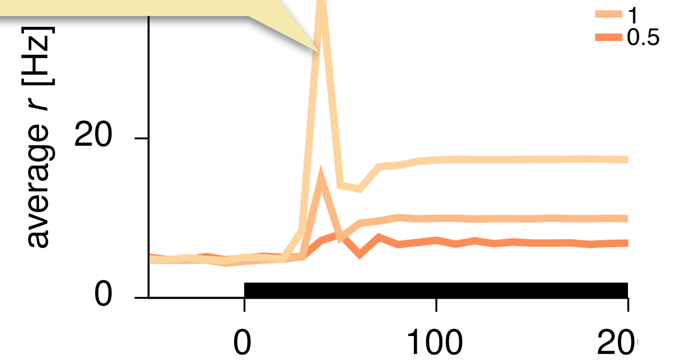
contrast-dependent oscillations

contrast-dependent stimulus-onset transients

MODEL

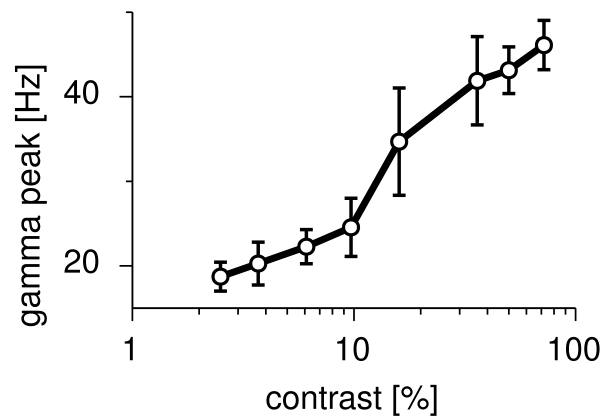


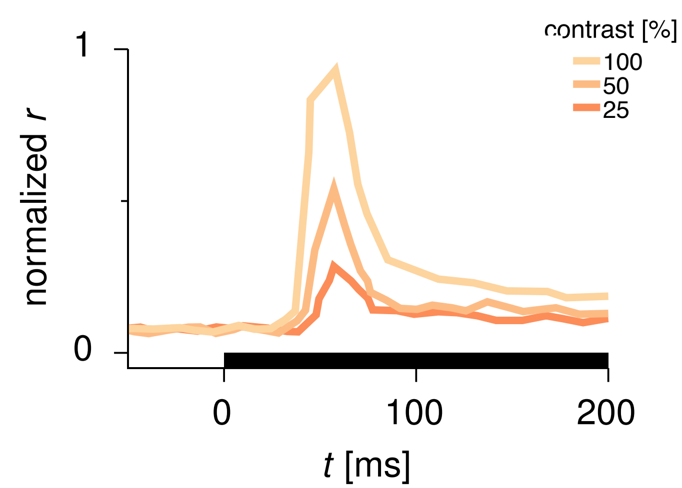
without predictive coding



EXPERIMENT

Ray & Maunsell, 2010





THE SAMPLING-OPTIMIZED NETWORK IS IN A REALISTIC DYNAMICAL REGIME

contrast-dependent oscillations

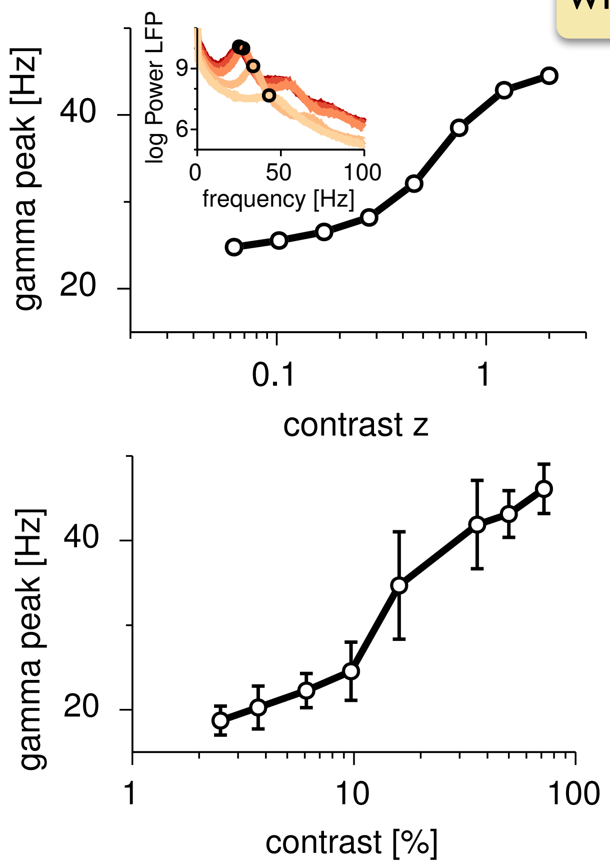
contrast-dependent stimulus-onset transients

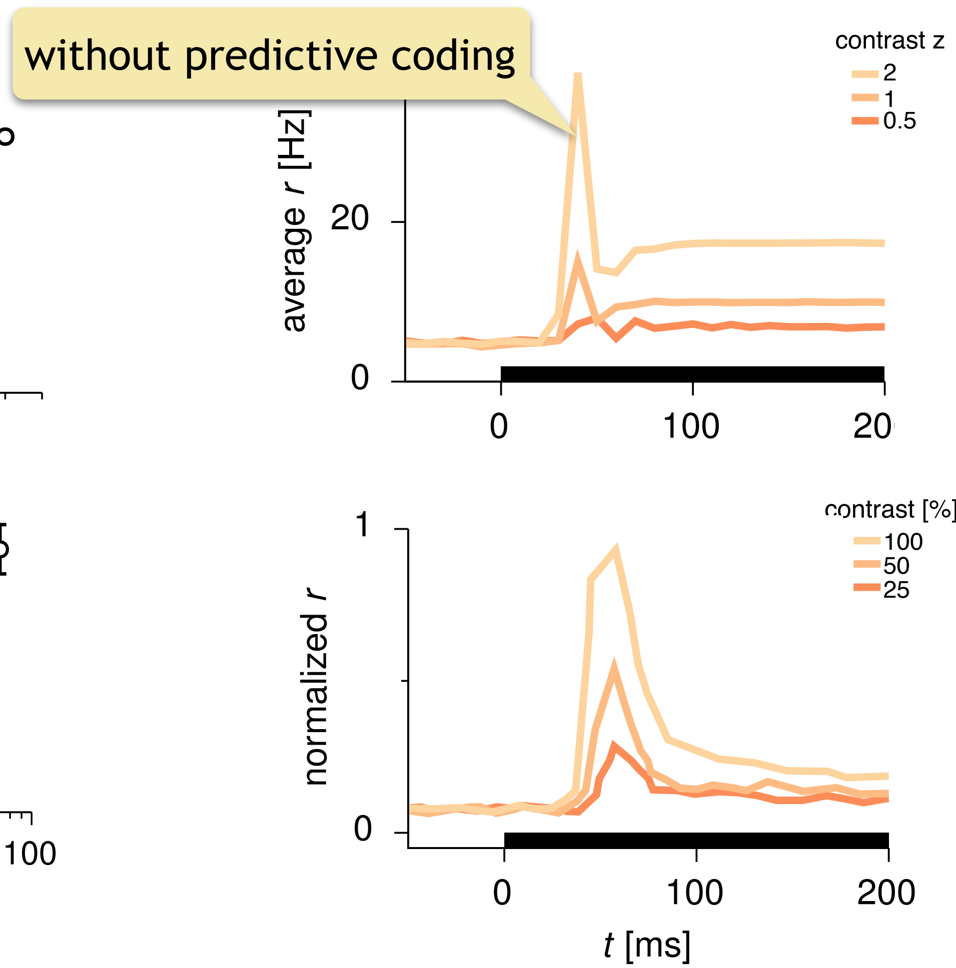
inhibition dominated

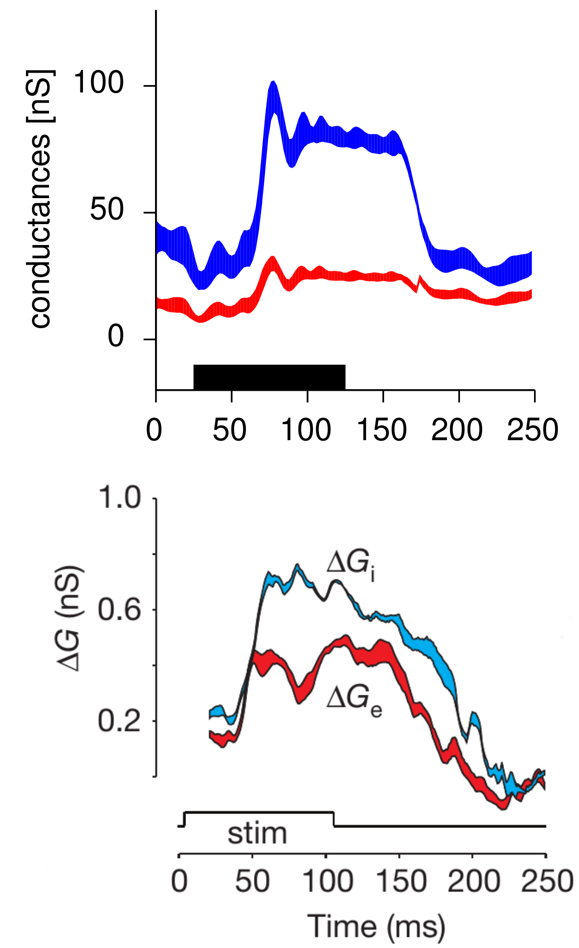
MODEL

EXPERIMENT

Ray & Maunsell, 2010
Haider et al, 2013







Echeveste et al, in prep

USING VARIABILITY IS CRITICAL

ideal observer: GSM
- full distribution (2nd order)

train →

visual cortex: neural network
- recurrent
- E-I
- expansive nonlinearity

USING VARIABILITY IS CRITICAL

ideal observer: GSM

- just the means (1st order)

train

visual cortex: neural network

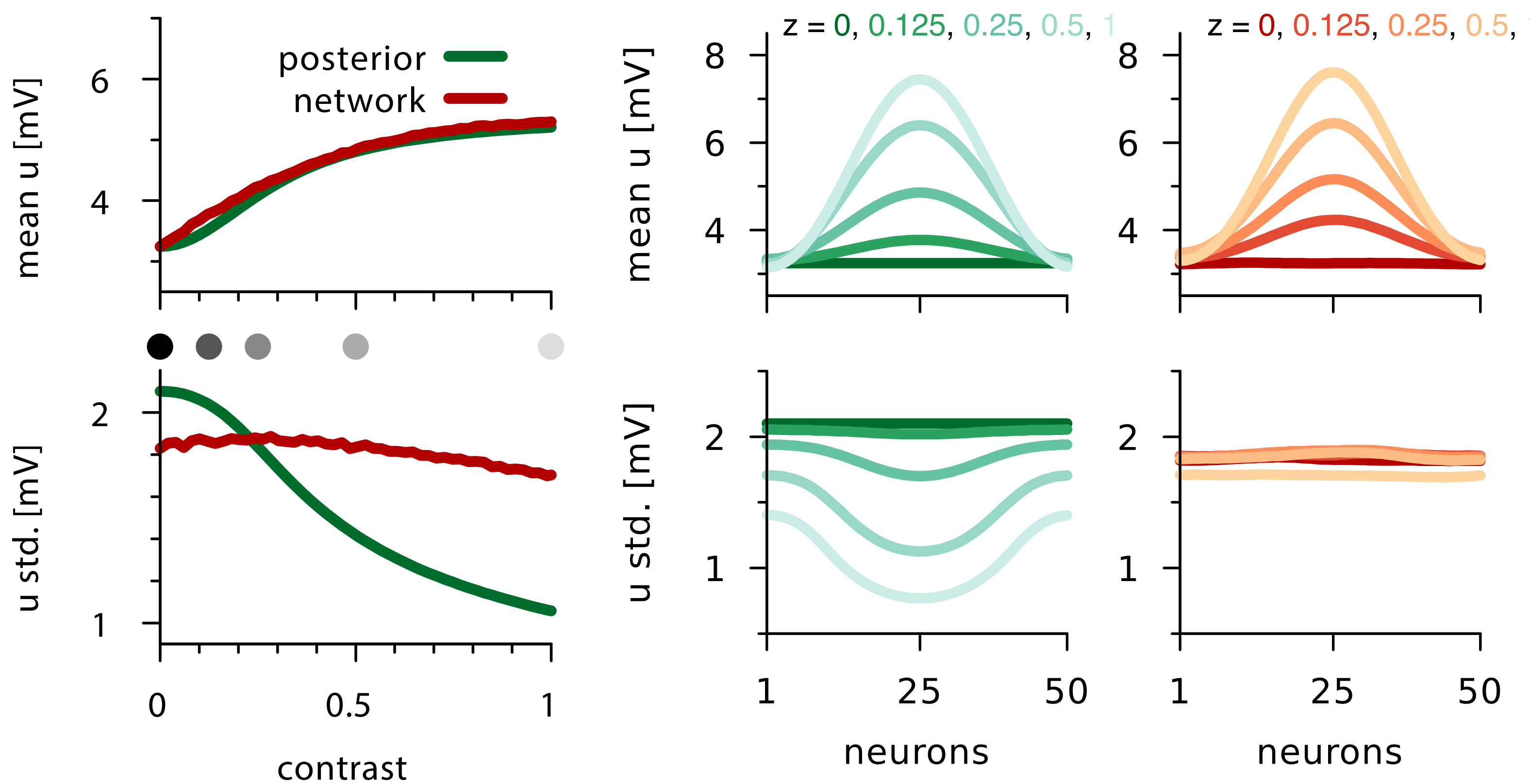
- recurrent
- E-I
- expansive nonlinearity

USING VARIABILITY IS CRITICAL

ideal observer: GSM
- just the means (1st order)

train

visual cortex: neural network
- recurrent
- E-I
- expansive nonlinearity

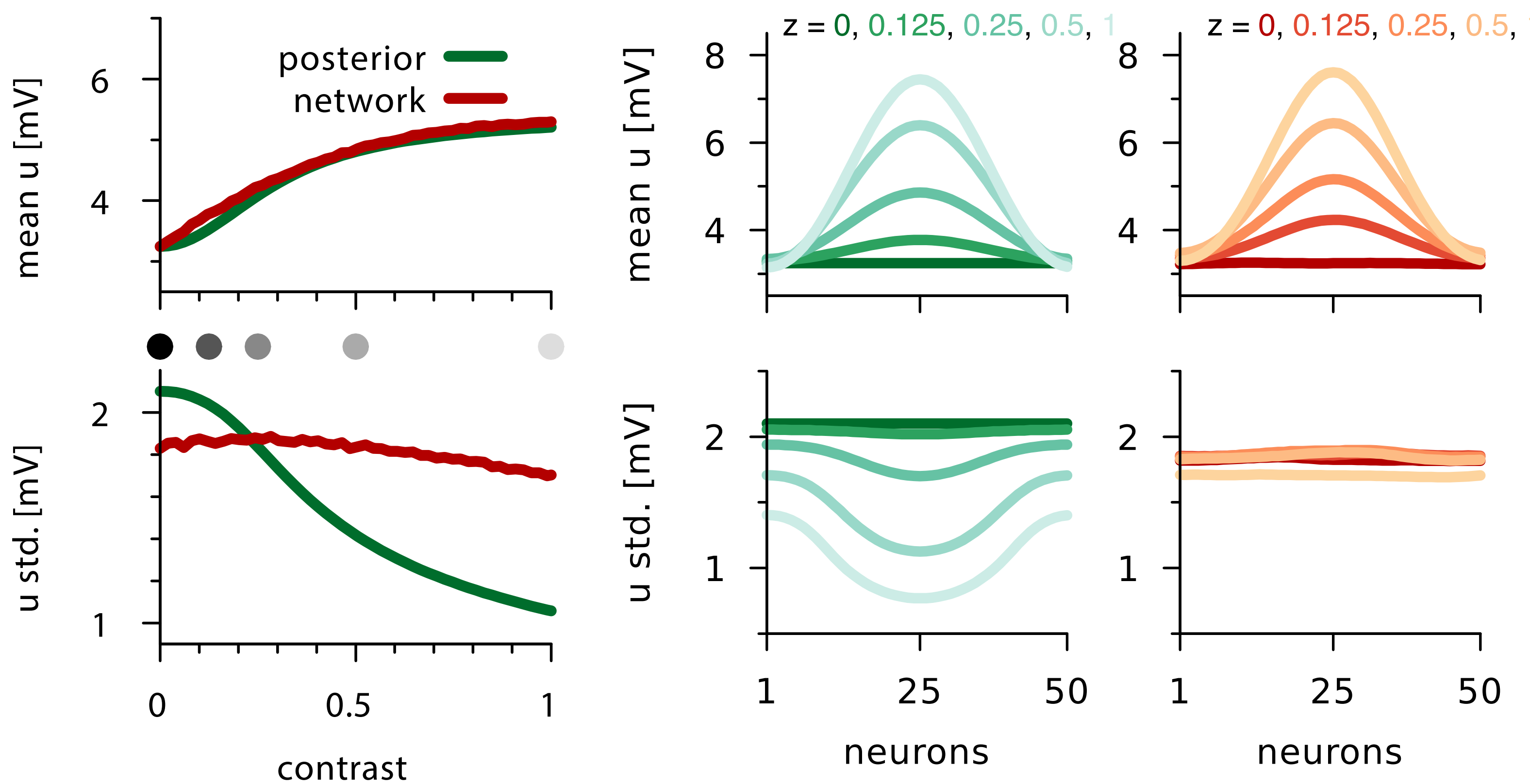


USING VARIABILITY IS CRITICAL

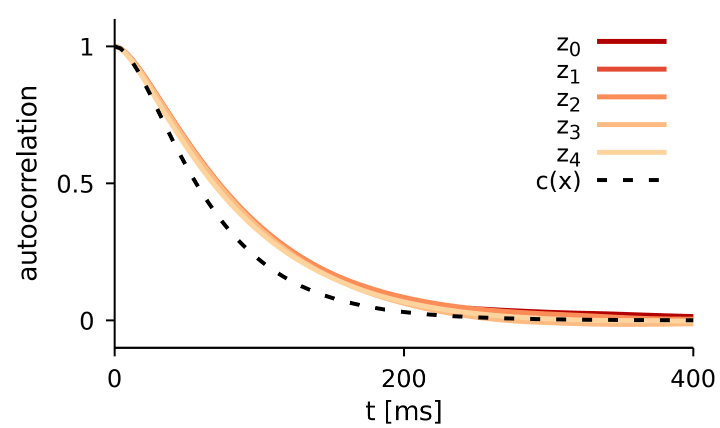
ideal observer: GSM → train → *visual cortex: neural network*

- just the means (1st order)

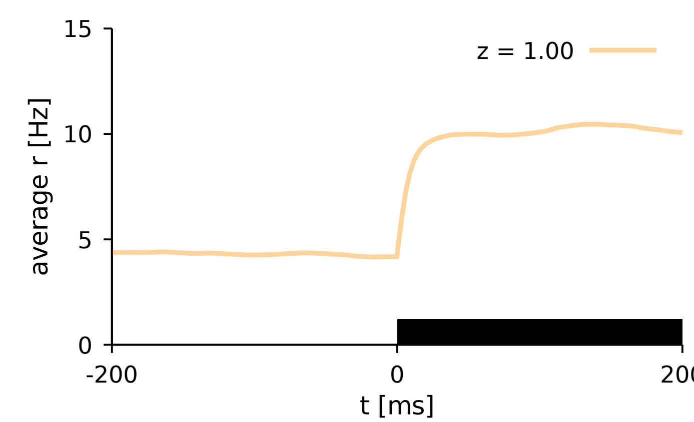
- recurrent
- E-I
- expansive nonlinearity



no fastness, no gamma



no transients



PROBABILISTIC INFERENCE AND LEARNING

cognitive science: *theory*	cognitive science: *experiments*		neuroscience: *theory*	neuroscience: *experiments*
		program $\mathcal{P}$		
✓	✓?	↑	?	?
		structure $\mathcal{M}$		
✓	✓	↑	?	?
		parameters θ		
✓	✓	↑	✓✓	✓?
		latent variables y		
✓	✓	↑	✓✓✓	✓?
		data x		

POSTDOC POSITION AVAILABLE

estimate humans' complex, high dimensional, dynamically changing internal models from behaviour

collaborate with world-leading experimental cognitive scientists

CEU

erc