A BAYESIAN APPROACH TO INTERNAL MODELS

MÁTÉ LENGYEL

Computational and Biological Learning Lab
Department of Engineering
University of Cambridge

Department of Cognitive Science
Central European University

INTERNAL MODELS

INTERNAL MODELS

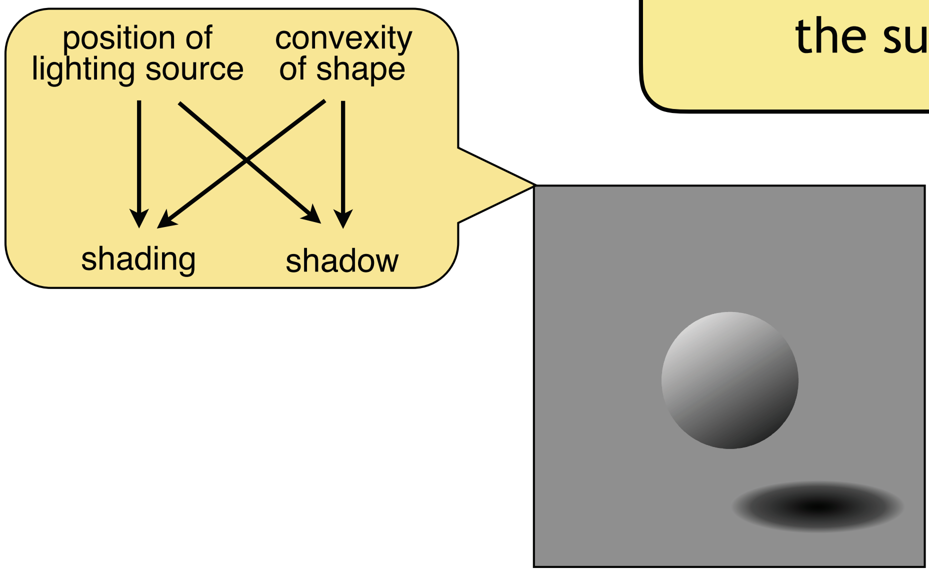
internal model:

a mental representation of the surrounding world

INTERNAL MODELS

internal model:

a mental representation of the surrounding world

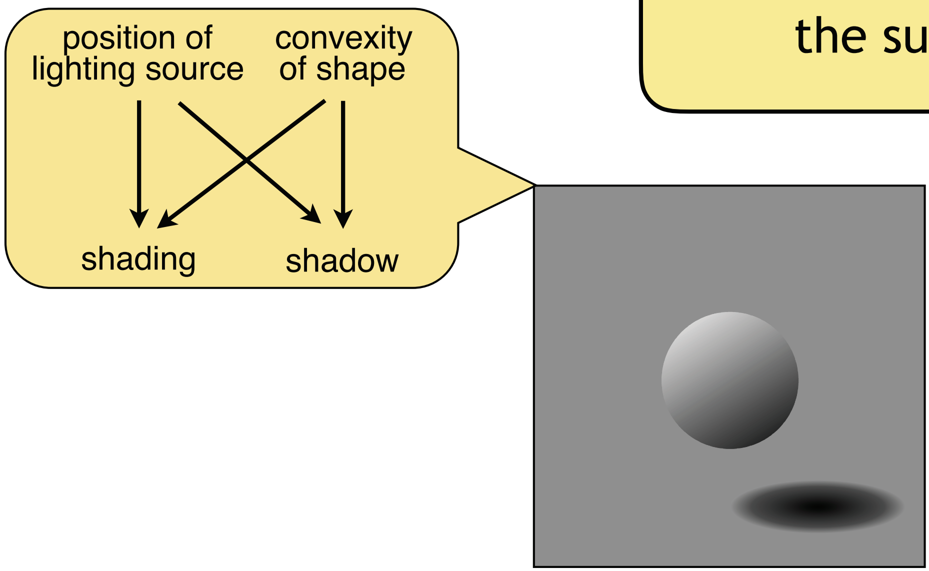


perception

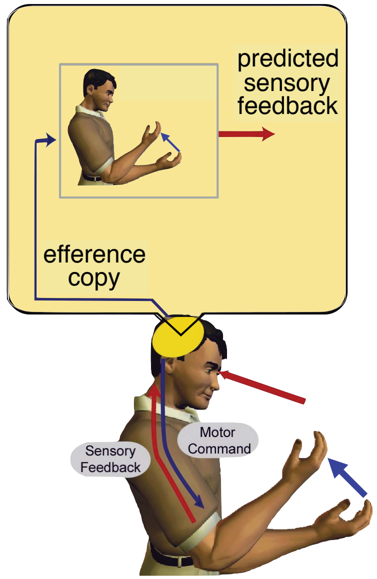
INTERNAL MODELS

internal model:

a mental representation of the surrounding world



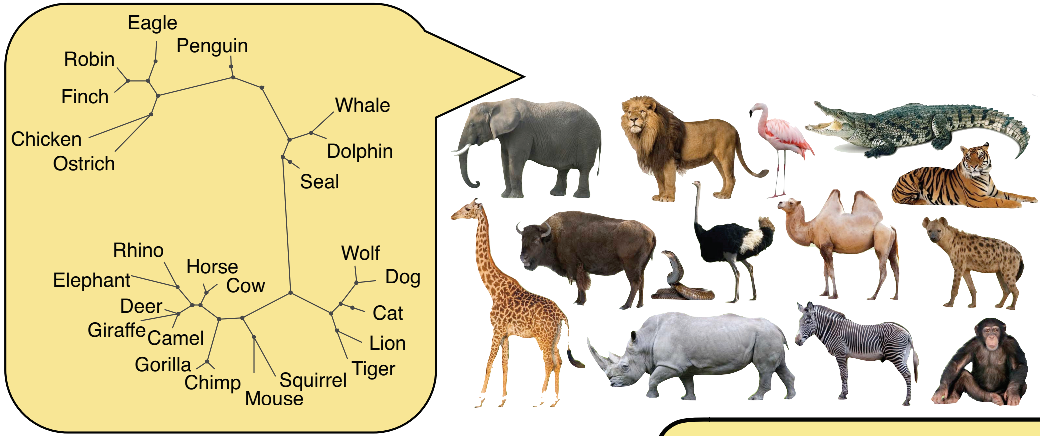
perception



action

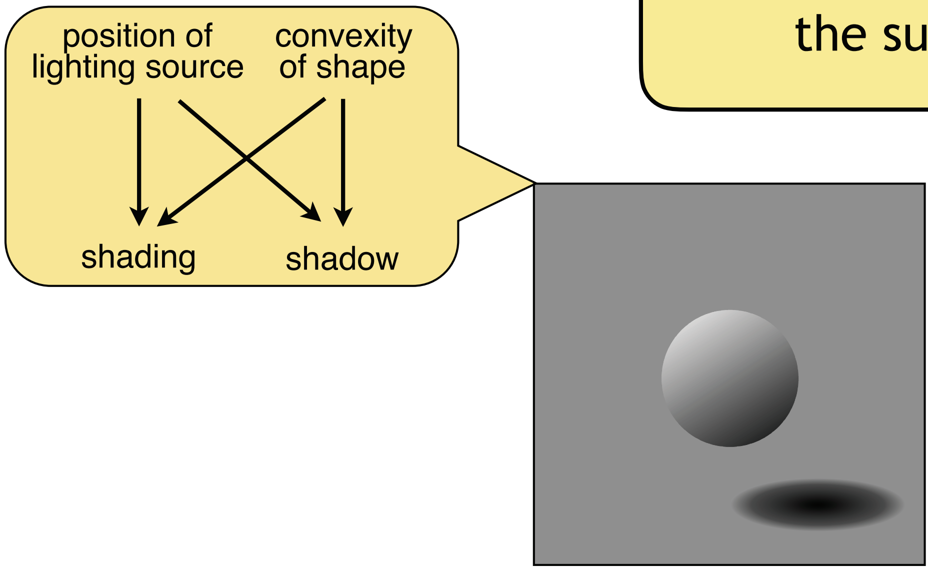
INTERNAL MODELS

semantic knowledge

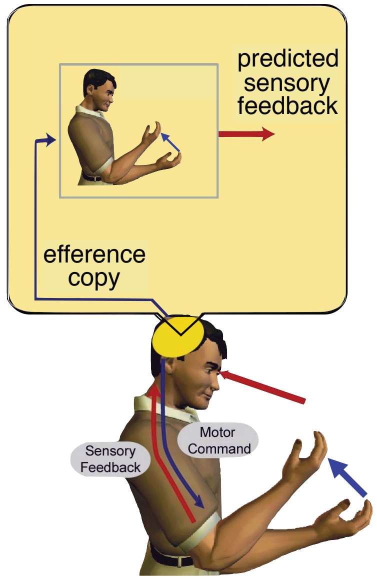


internal model:

a mental representation of the surrounding world



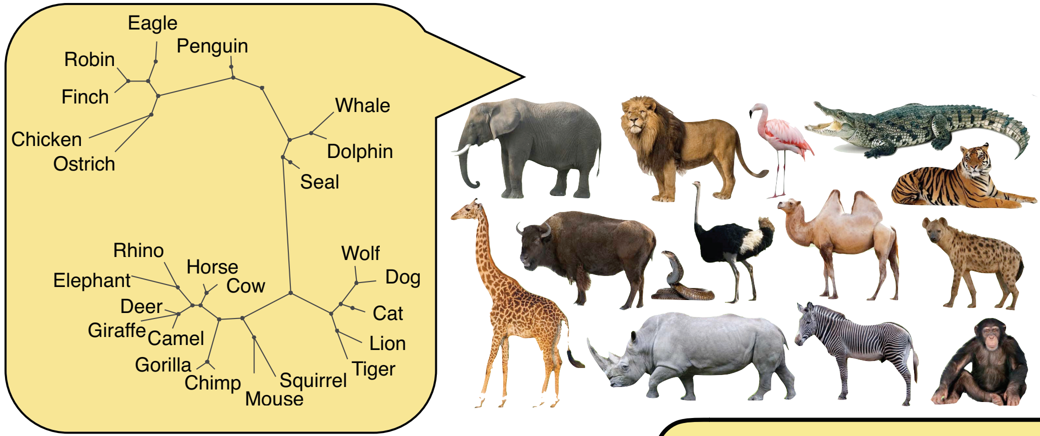
perception



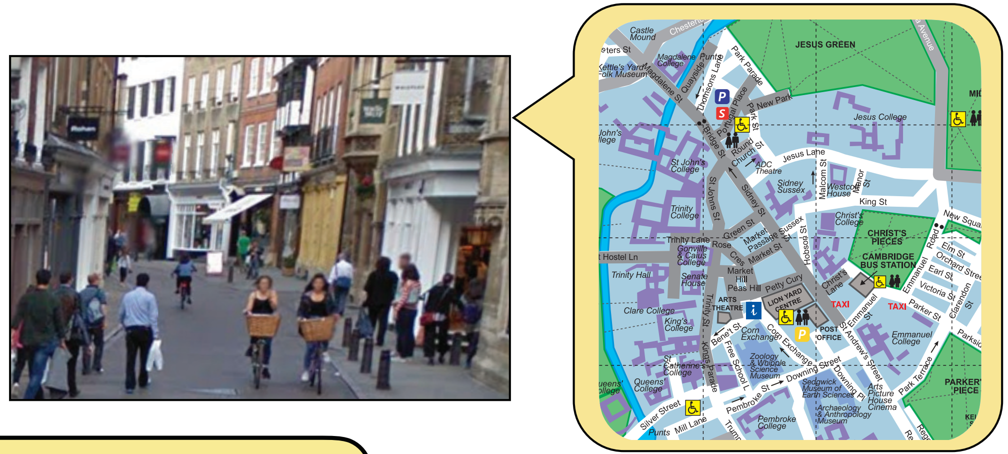
action

INTERNAL MODELS

semantic knowledge

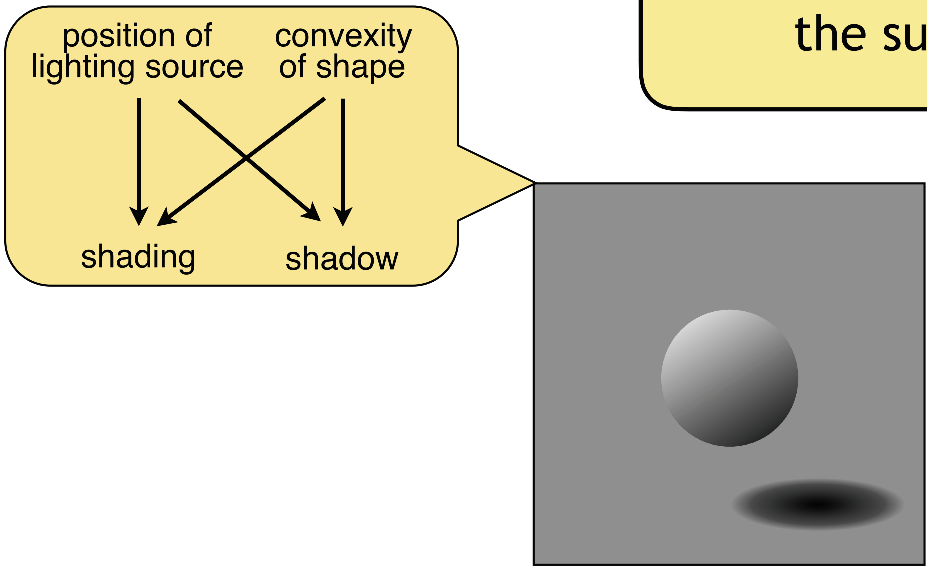


navigation

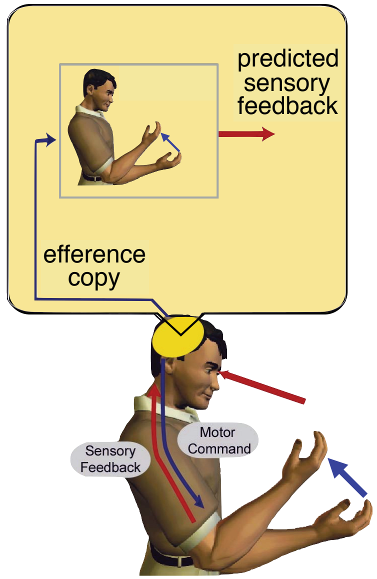


internal model:

a mental representation of
the surrounding world



perception



action

PROBABILISTIC INTERNAL MODELS

PROBABILISTIC INTERNAL MODELS

internal model

$\mathrm{P}_{\mathrm{model}}(\mathbf{y})$

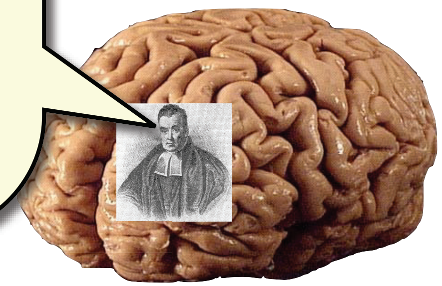
$\mathbf{y}$

objects, features

$\mathrm{P}_{\mathrm{model}}(\mathbf{x}|\mathbf{y})$

$\mathbf{x}$

sensory stimuli



PROBABILISTIC INTERNAL MODELS

internal model

$\mathrm{P}_{\mathrm{model}}(\mathbf{y})$

$\mathbf{y}$

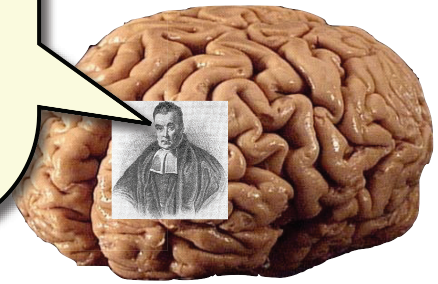
objects, features

$\mathrm{P}_{\mathrm{model}}(\mathbf{y}|\mathbf{x})$

percept

$\mathbf{x}$

sensory stimuli



PROBABILISTIC INTERNAL MODELS

$P_{\text{true}}(\mathbf{y})$

$\mathbf{y}$

objects, features

$P_{\text{true}}(\mathbf{x}|\mathbf{y})$

$\mathbf{x}$

sensory stimuli

internal model

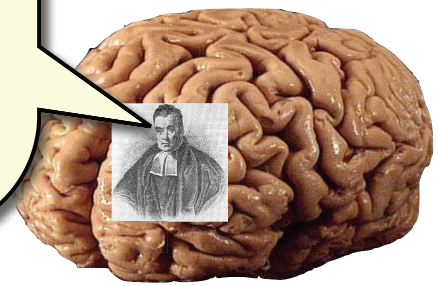
$P_{\text{model}}(\mathbf{y})$

$\mathbf{y}$

objects, features

$P_{\text{model}}(\mathbf{y}|\mathbf{x})$

percept



PROBABILISTIC INTERNAL MODELS

$\mathrm{P}_{\mathrm{true}}(\mathbf{y})$

$\mathbf{y}$

objects, features

$\mathrm{P}_{\mathrm{true}}(\mathbf{x}|\mathbf{y})$

$\mathbf{x}$

sensory stimuli

internal model

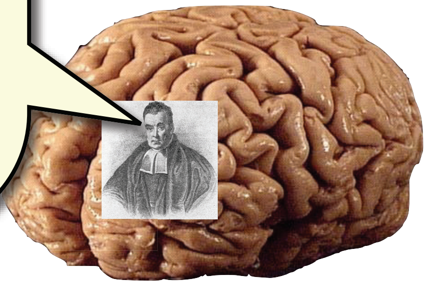
$\mathrm{P}_{\mathrm{model}}(\mathbf{y})$

$\mathbf{y}$

objects, features

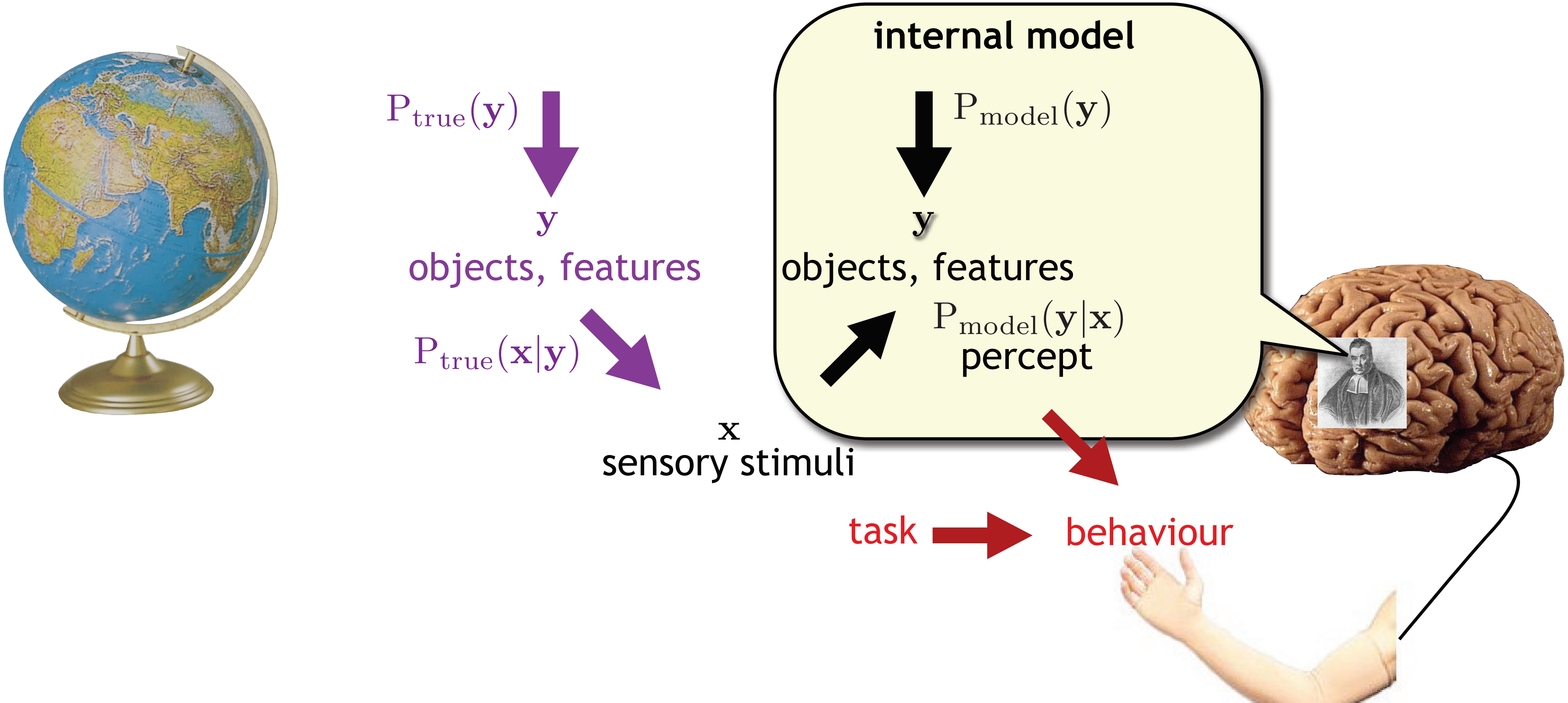
$\mathrm{P}_{\mathrm{model}}(\mathbf{y}|\mathbf{x})$

percept



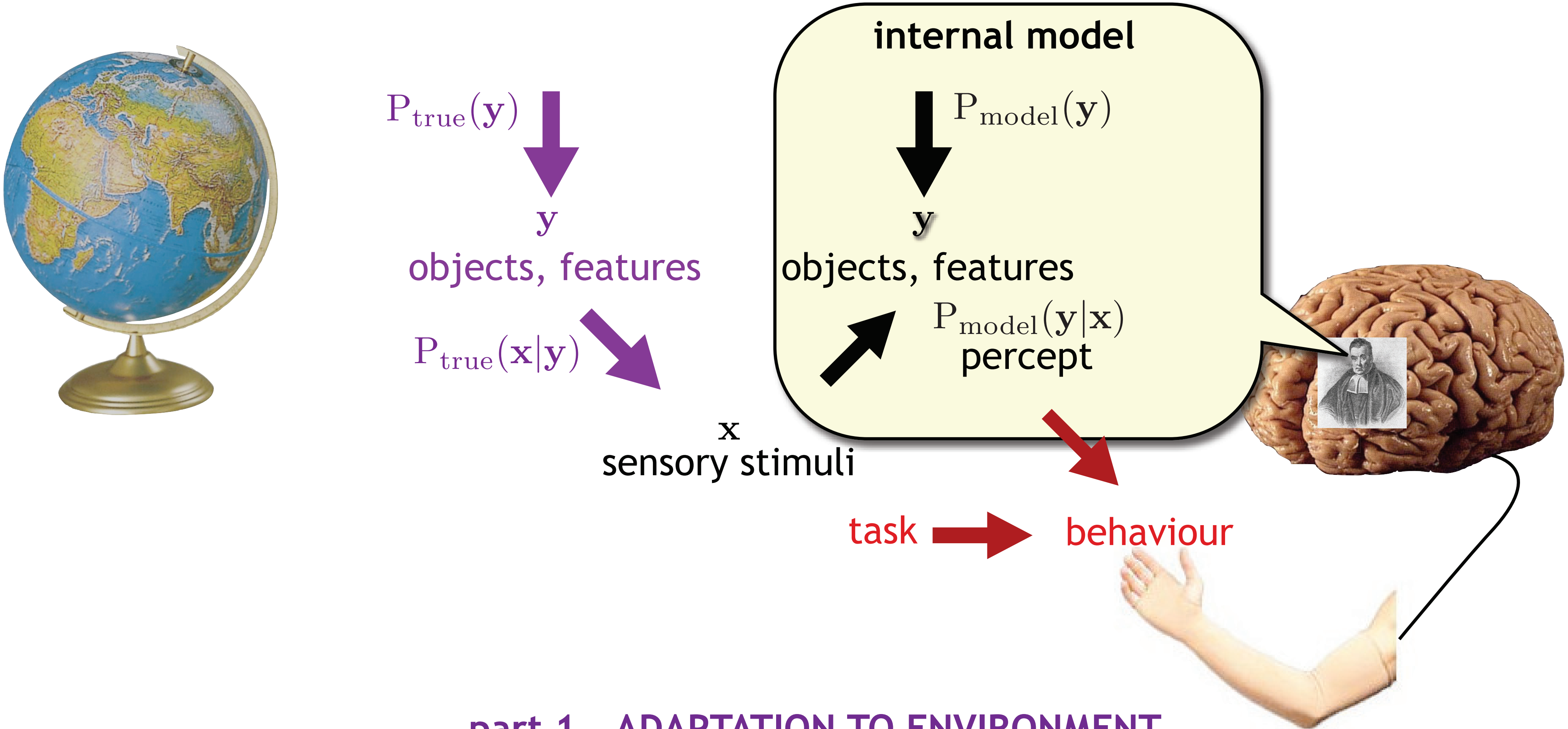
part 1. ADAPTATION TO ENVIRONMENT

PROBABILISTIC INTERNAL MODELS



part 1. ADAPTATION TO ENVIRONMENT

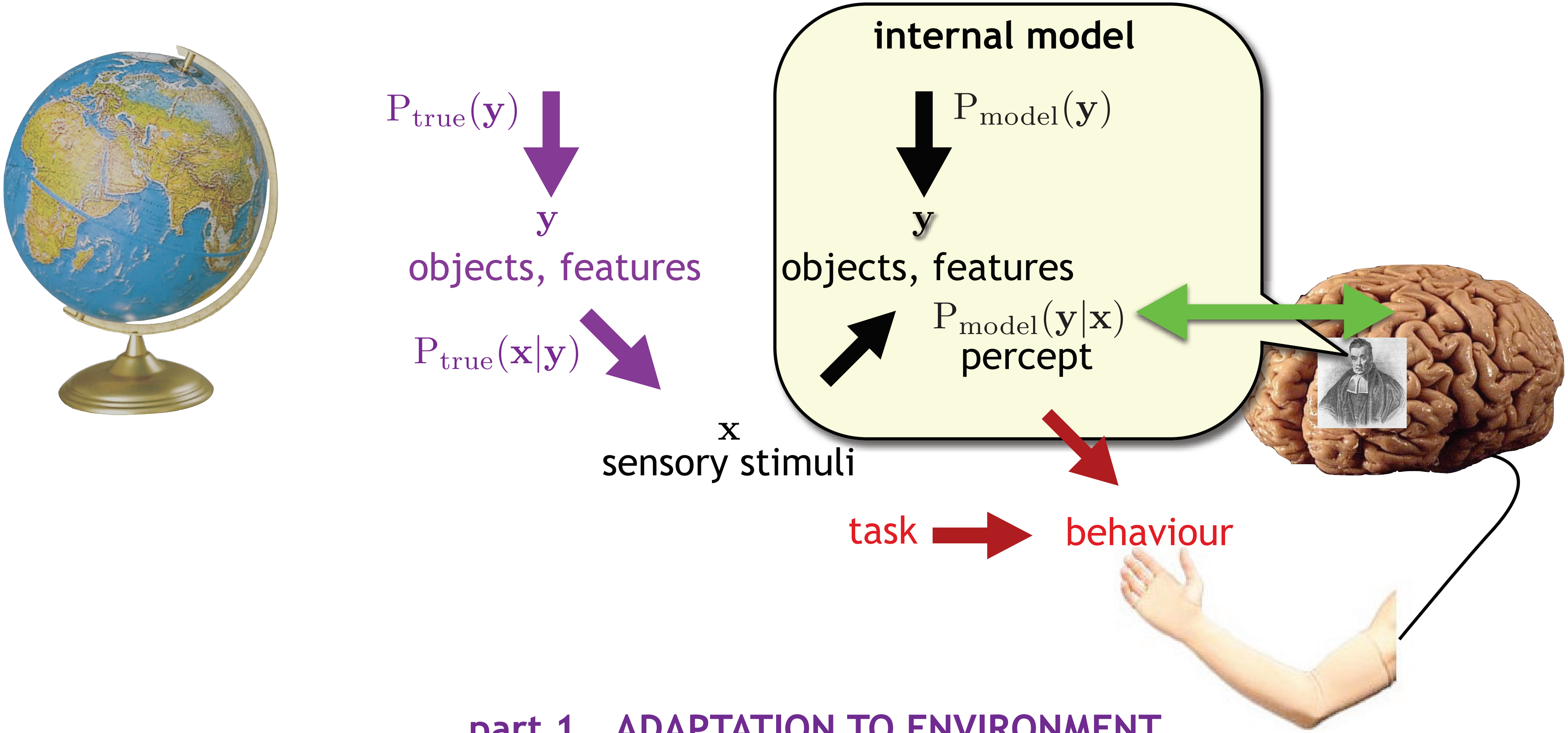
PROBABILISTIC INTERNAL MODELS



part 1. ADAPTATION TO ENVIRONMENT

part 2. TASK-INDEPENDENCE

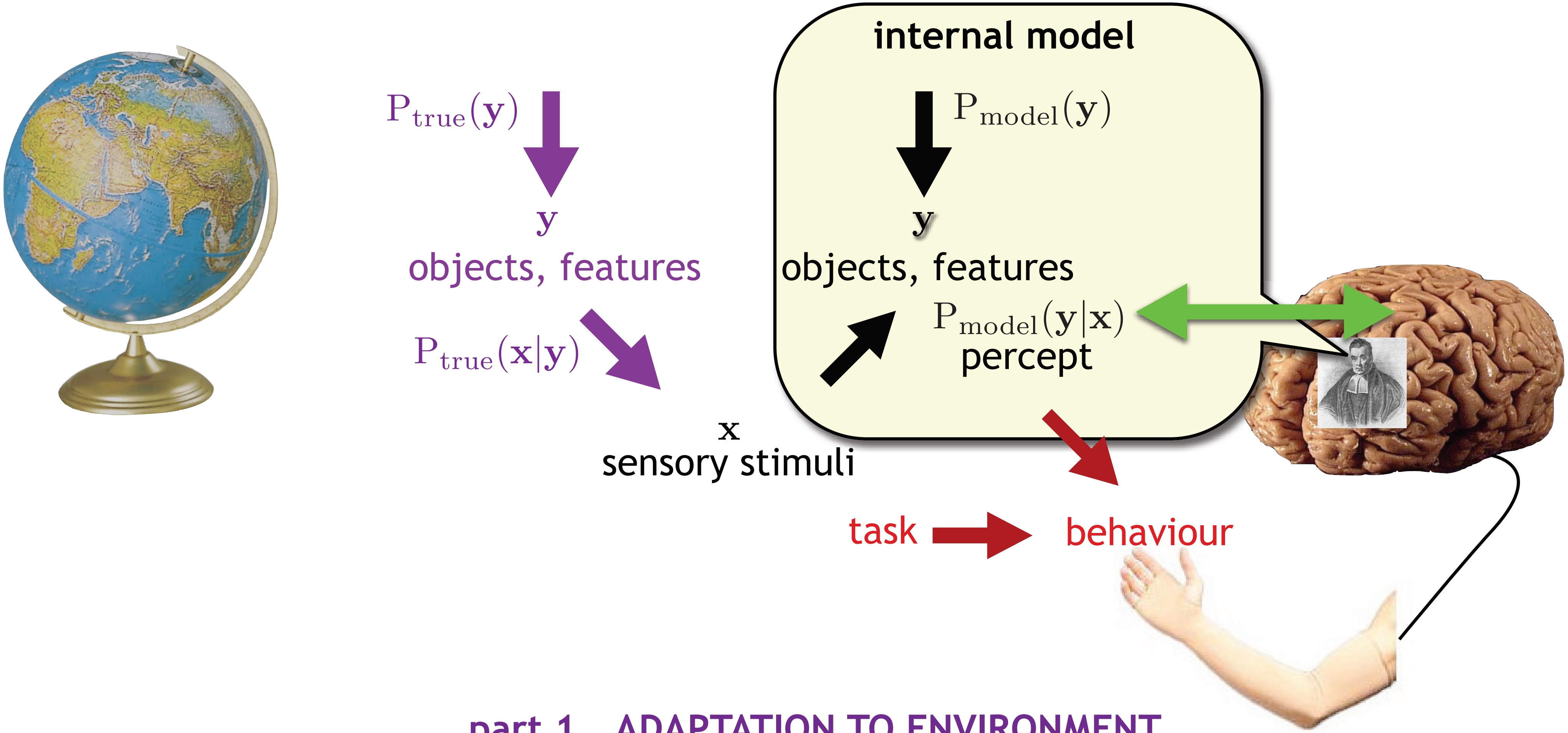
PROBABILISTIC INTERNAL MODELS



part 1. ADAPTATION TO ENVIRONMENT

part 2. TASK-INDEPENDENCE

PROBABILISTIC INTERNAL MODELS

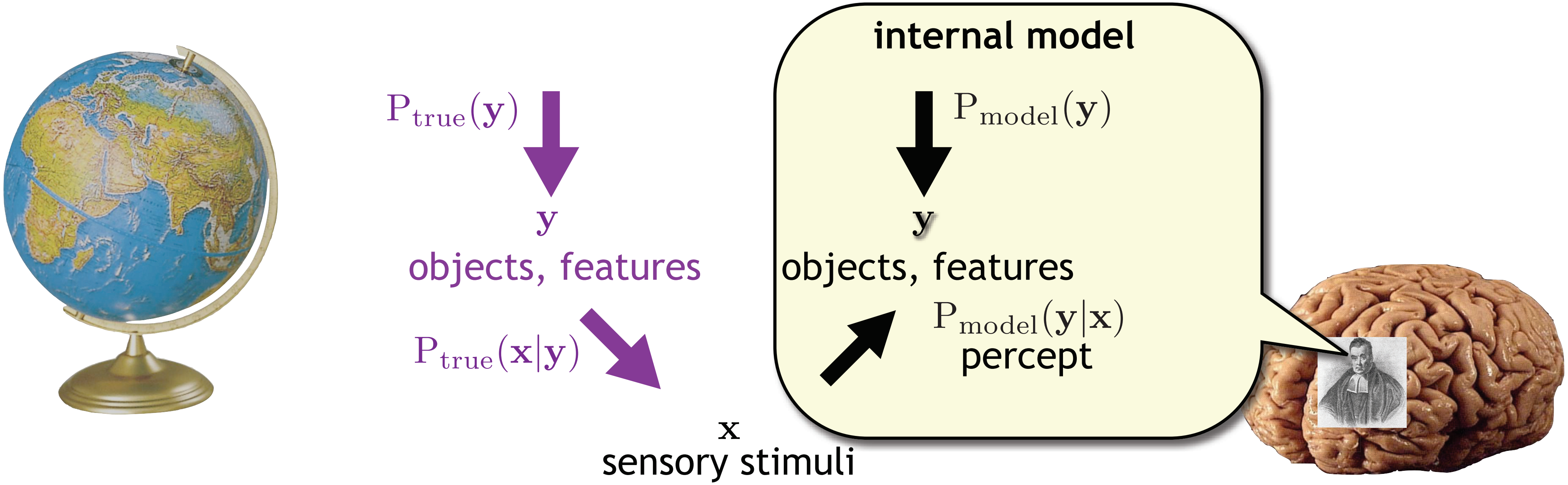


part 1. ADAPTATION TO ENVIRONMENT

part 2. TASK-INDEPENDENCE

part 3. NEURAL SUBSTRATES

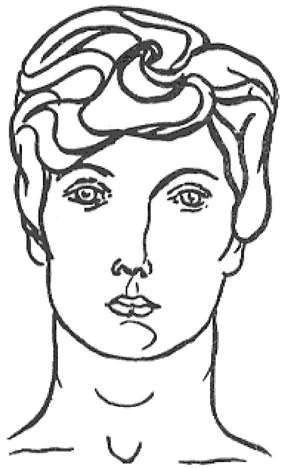
PROBABILISTIC INTERNAL MODELS



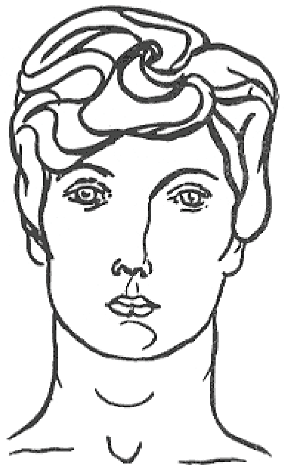
part 1. ADAPTATION TO ENVIRONMENT

CHUNK LEARNING:
HIERARCHICAL PAIR-WISE ASSOCIATIVE?

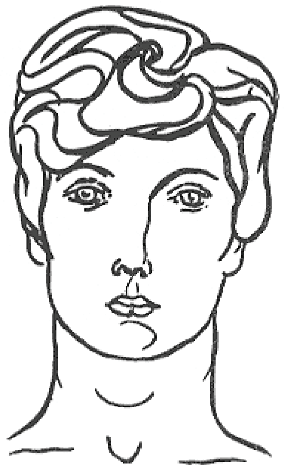
CHUNK LEARNING: HIERARCHICAL PAIR-WISE ASSOCIATIVE?



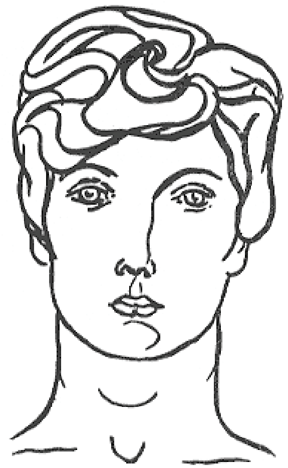
CHUNK LEARNING:
HIERARCHICAL PAIR-WISE ASSOCIATIVE?

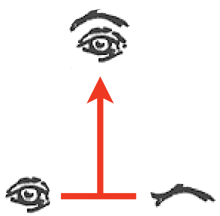


CHUNK LEARNING:
HIERARCHICAL PAIR-WISE ASSOCIATIVE?

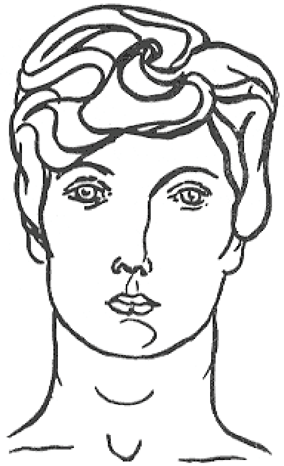


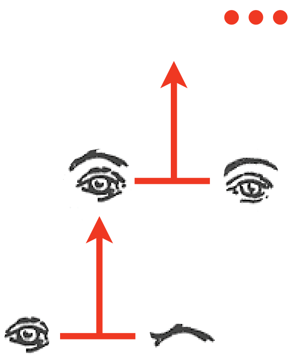
CHUNK LEARNING: HIERARCHICAL PAIR-WISE ASSOCIATIVE?



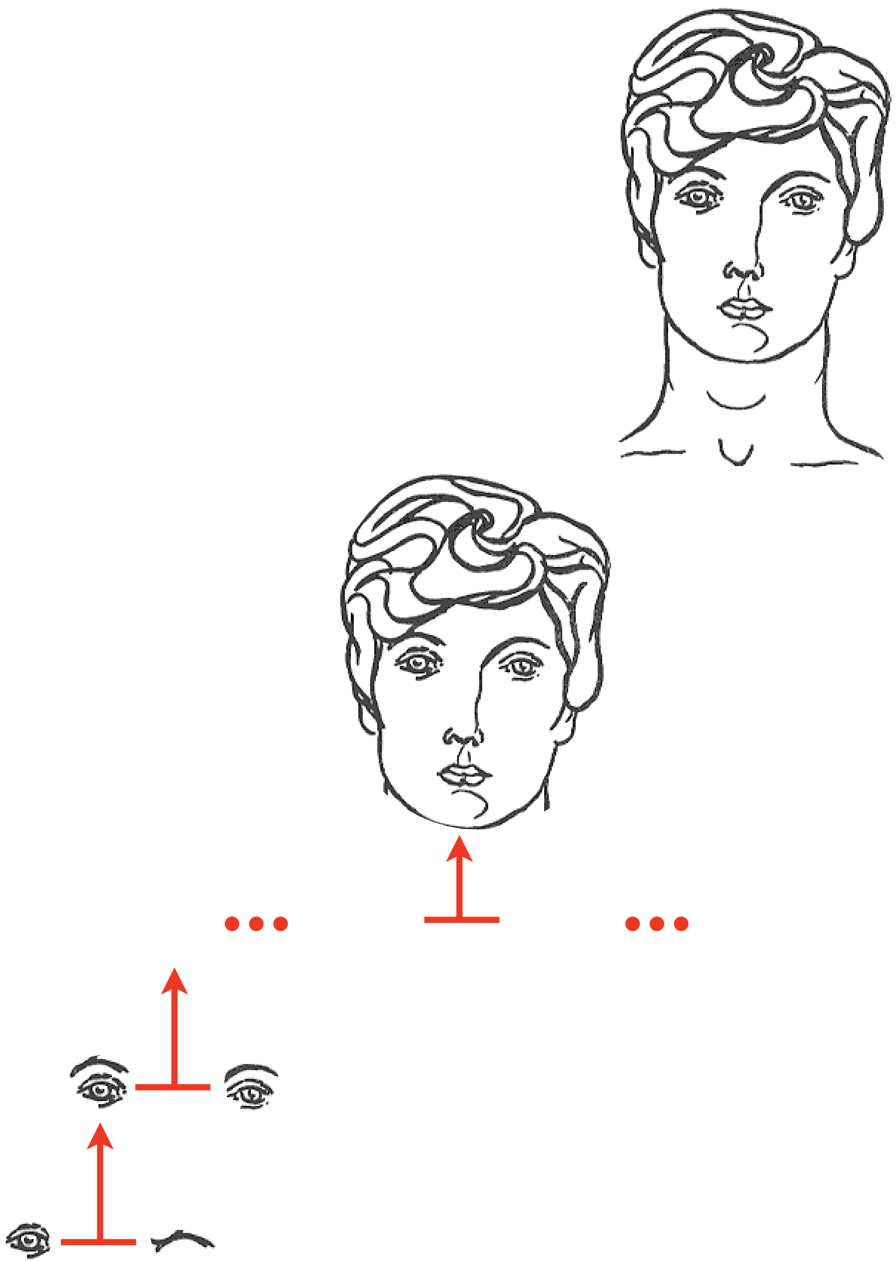


CHUNK LEARNING:
HIERARCHICAL PAIR-WISE ASSOCIATIVE?

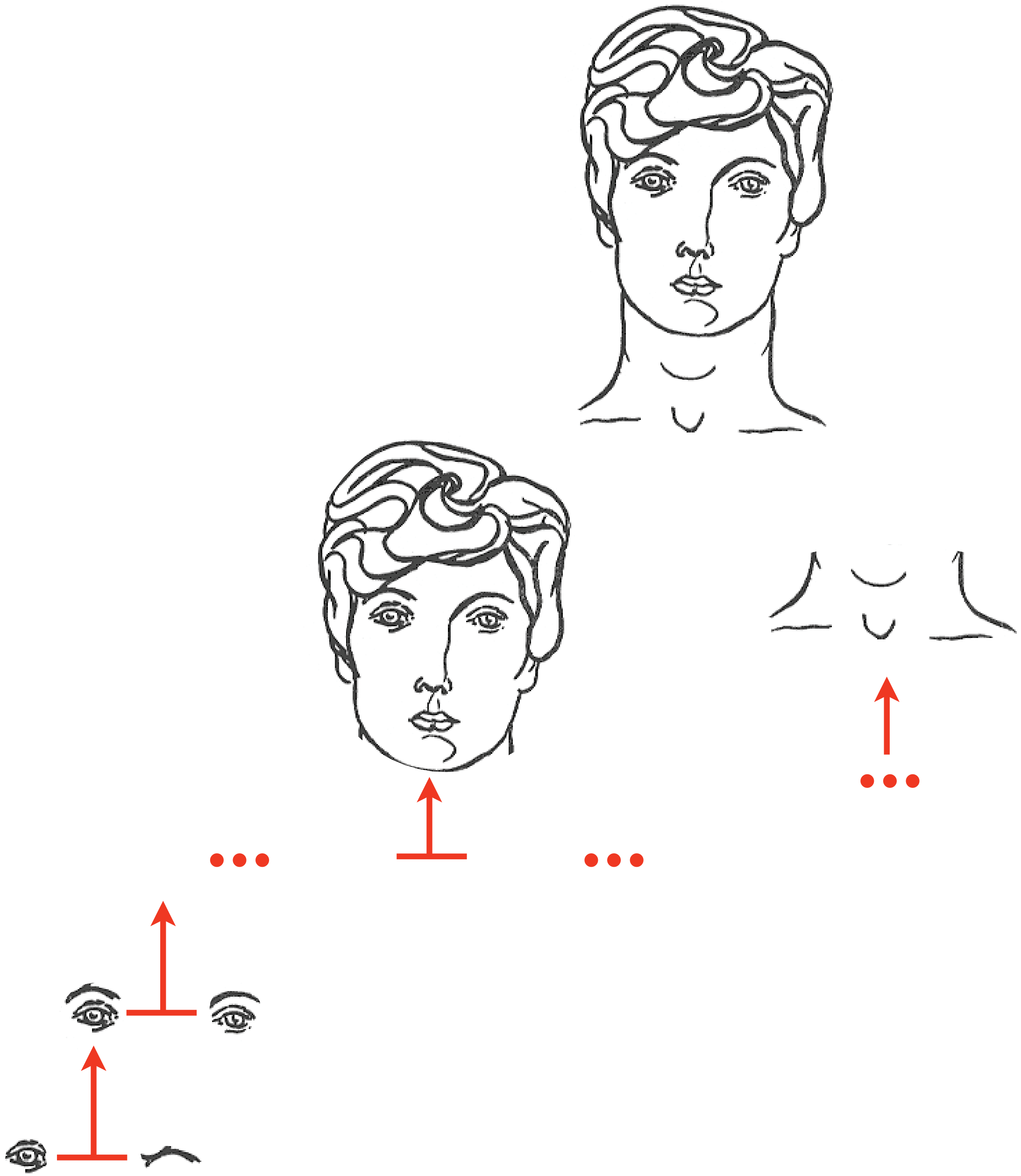




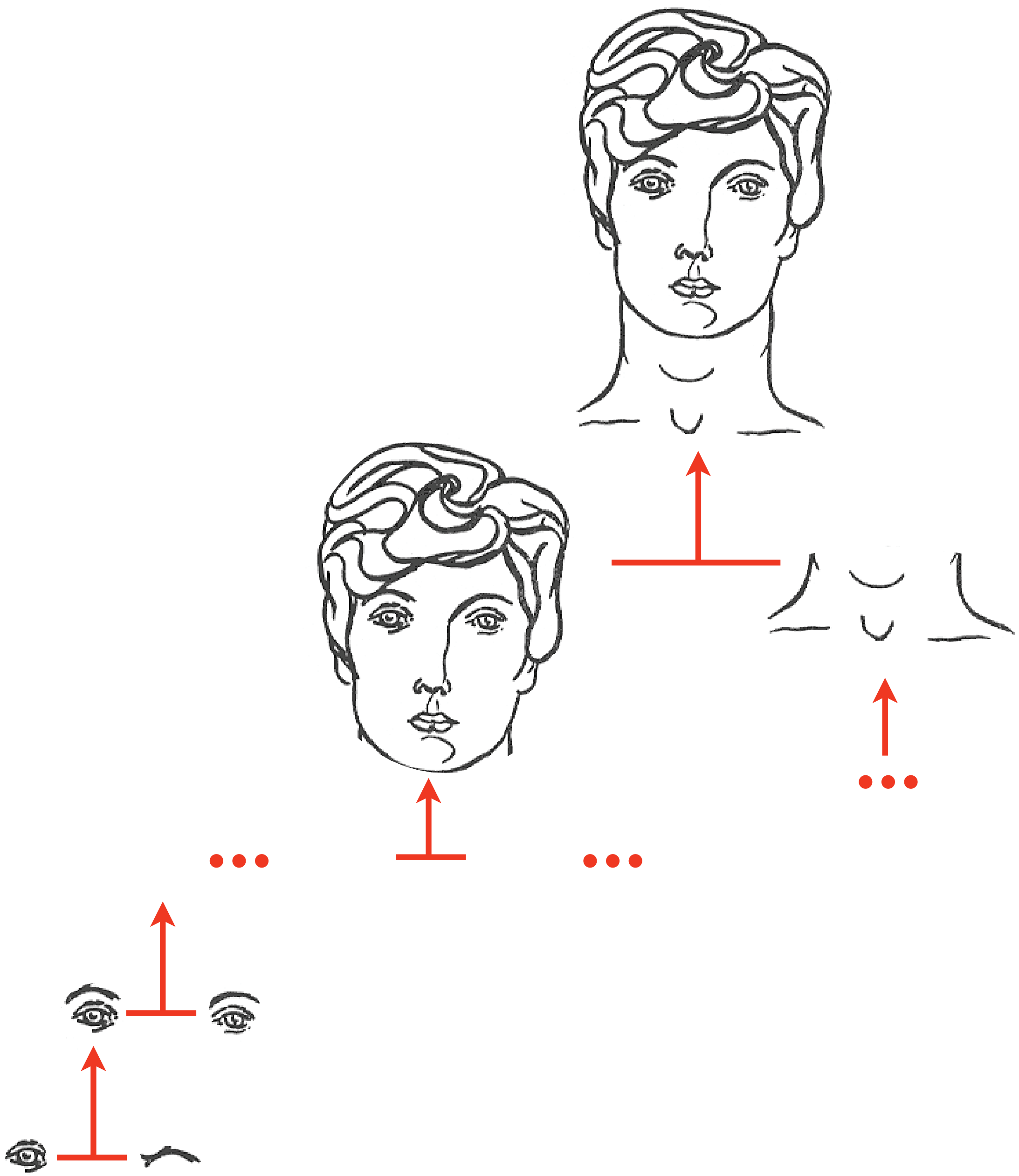
CHUNK LEARNING:
HIERARCHICAL PAIR-WISE ASSOCIATIVE?



CHUNK LEARNING:
HIERARCHICAL PAIR-WISE ASSOCIATIVE?



CHUNK LEARNING:
HIERARCHICAL PAIR-WISE ASSOCIATIVE?



THE PROBLEM WITH PAIR-WISE ASSOCIATIONS

storing pair-wise associations → learning about 2^{nd} order statistics

but the environment is much richer statistically!

THE PROBLEM WITH PAIR-WISE ASSOCIATIONS

storing pair-wise associations → learning about 2nd order statistics

but the environment is much richer statistically!

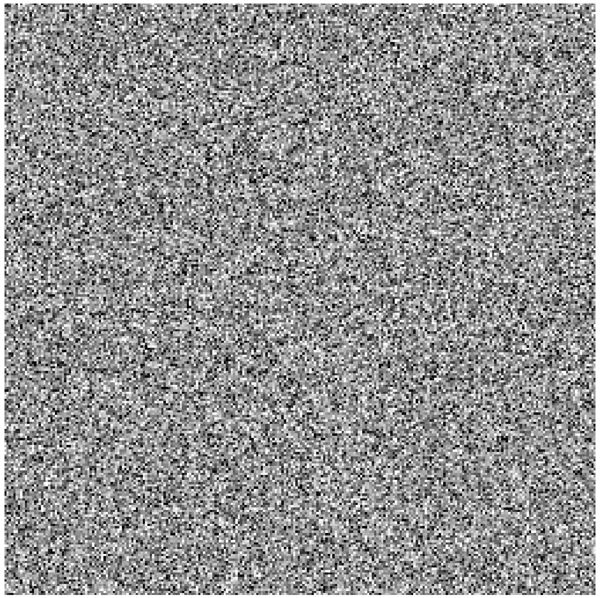
natural image

THE PROBLEM WITH PAIR-WISE ASSOCIATIONS

storing pair-wise associations → learning about 2^{nd} order statistics

but the environment is much richer statistically!

natural image



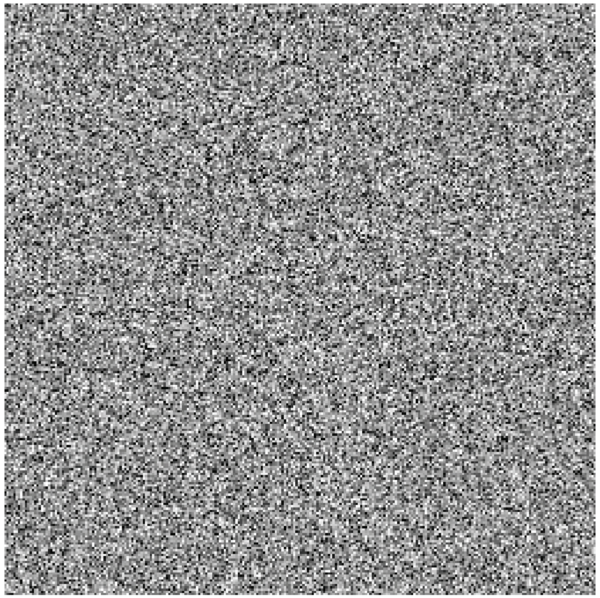
predicted image
from 1^{st} order statistics

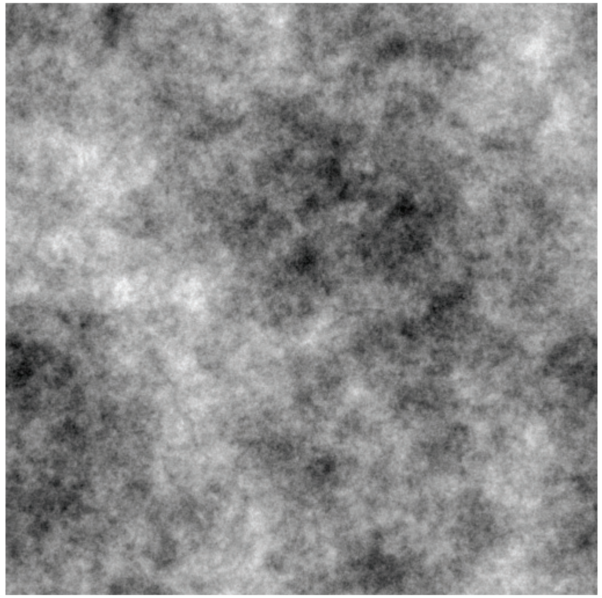
THE PROBLEM WITH PAIR-WISE ASSOCIATIONS

storing pair-wise associations → learning about 2nd order statistics

but the environment is much richer statistically!


natural image


predicted image
from 1st order statistics


Bruno Olshausen

predicted image
from 2nd order statistics

THE PROBLEM WITH LEARNING HIGHER ORDER STATISTICS

- **there are too many of them**

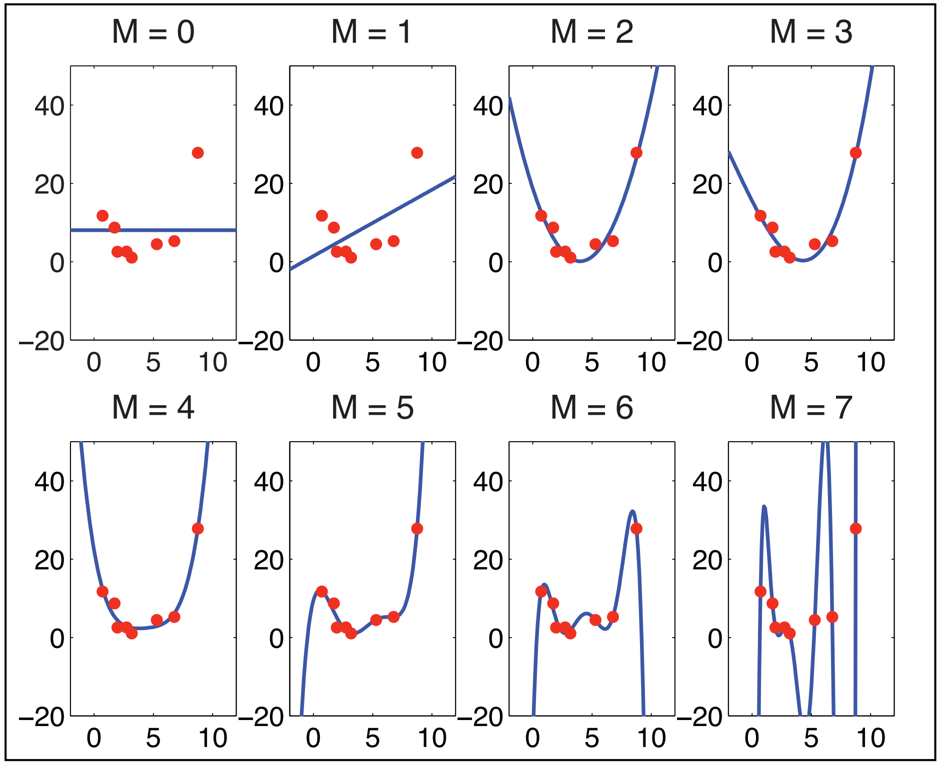
N items	order of statistics	number of statistics
	1	N
	2	N^2
	⋮	⋮
	D	N^D

THE PROBLEM WITH LEARNING HIGHER ORDER STATISTICS

- **there are too many of them**

N items	order of statistics	number of statistics
	1	N
	2	N^2
	⋮	⋮
	D	N^D

- **danger of overfitting**



Zoubin Ghahramani

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

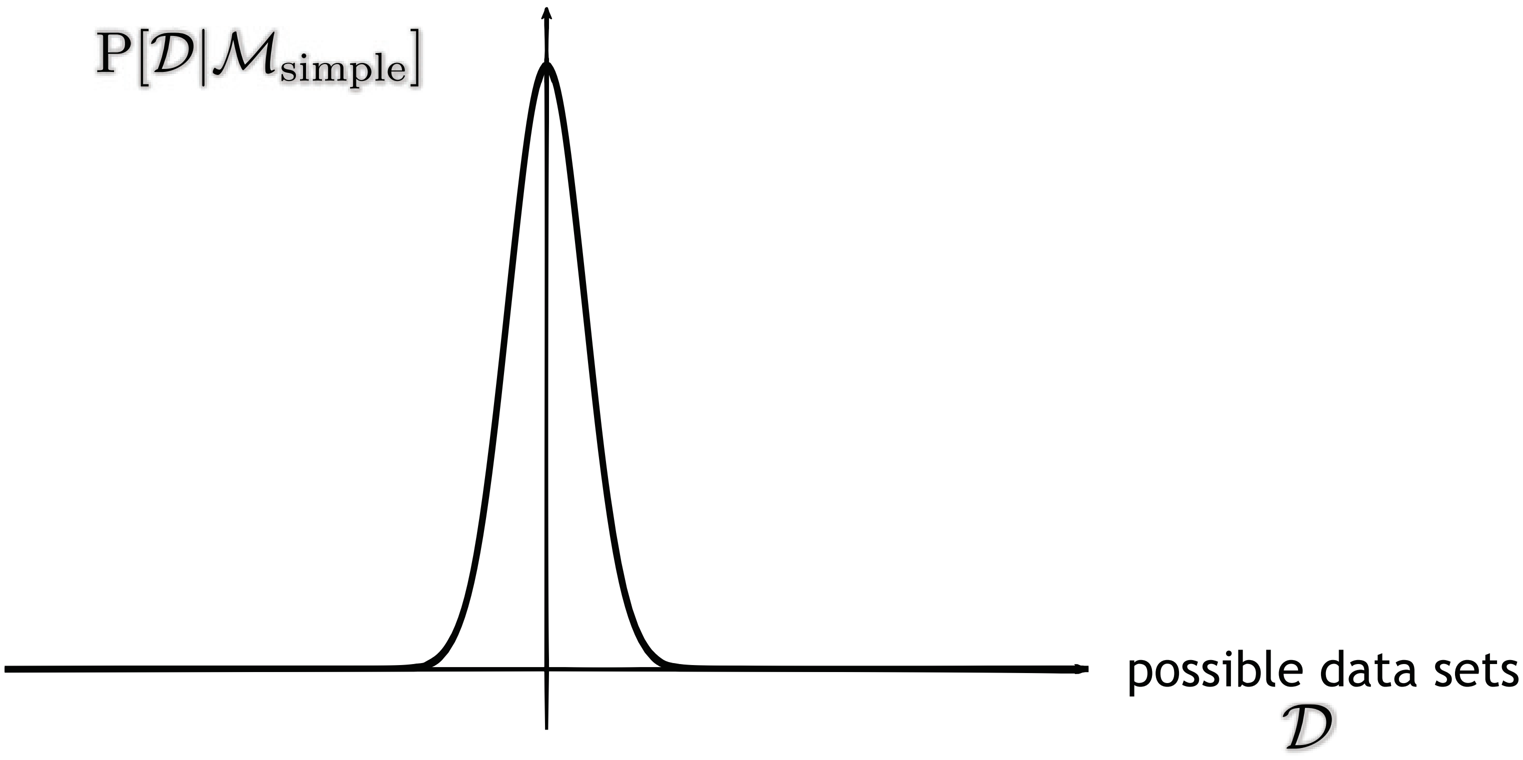
Rev. Bayes

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

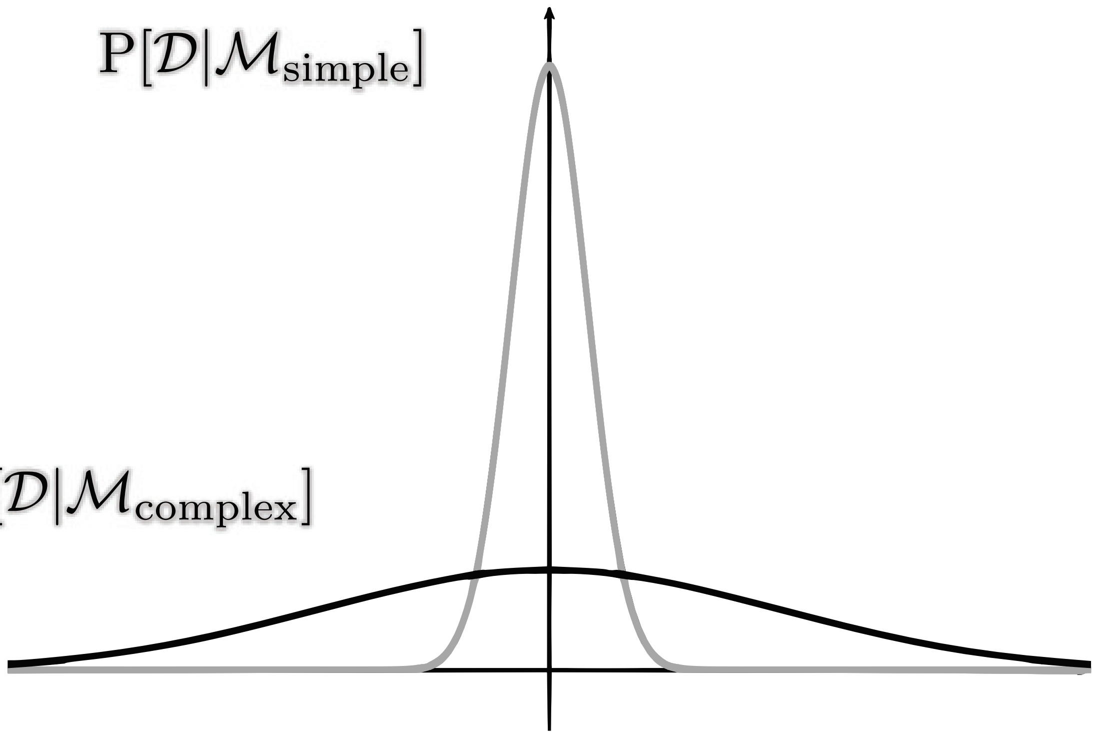
Rev. Bayes

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

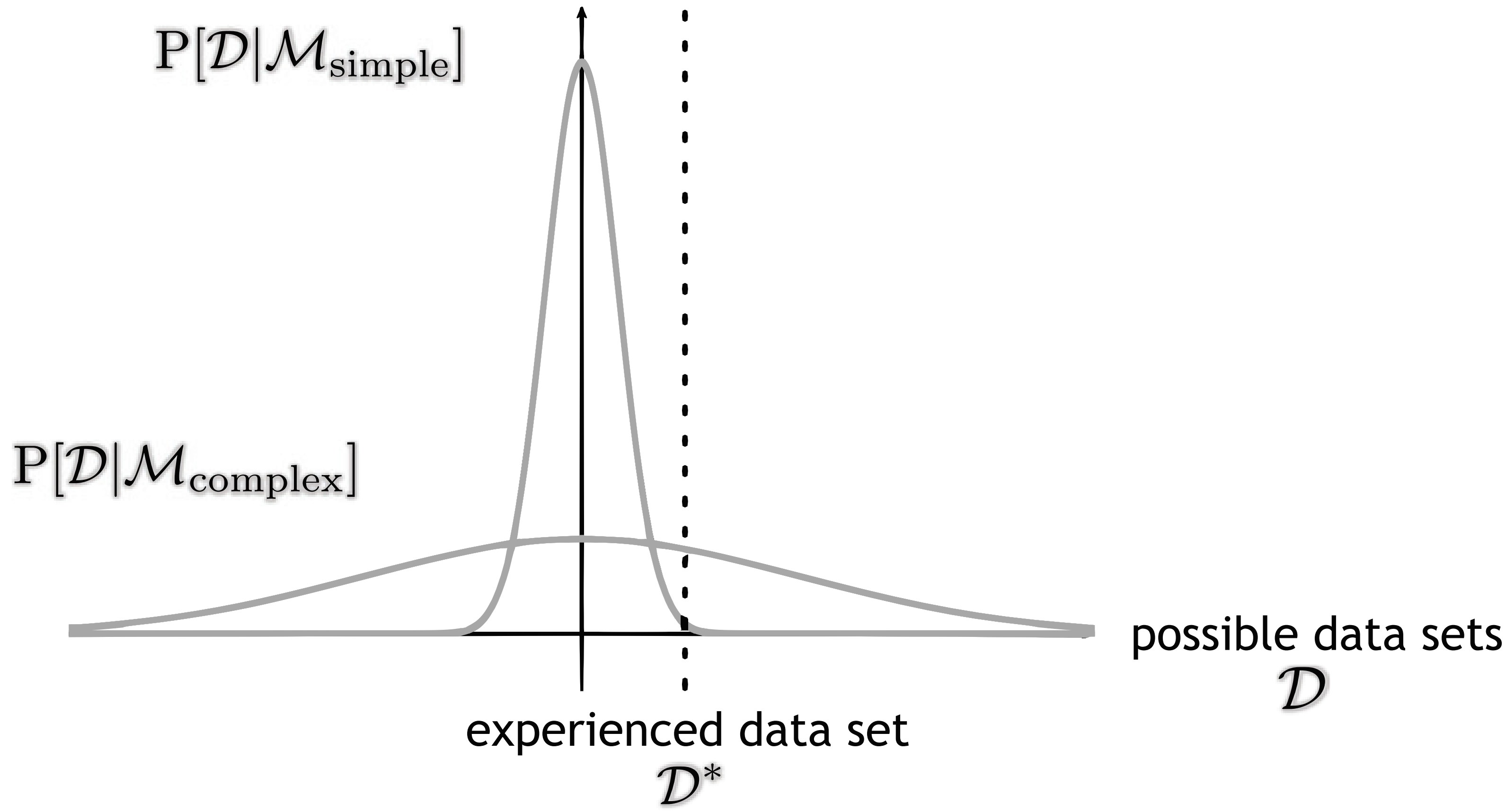
Rev. Bayes

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

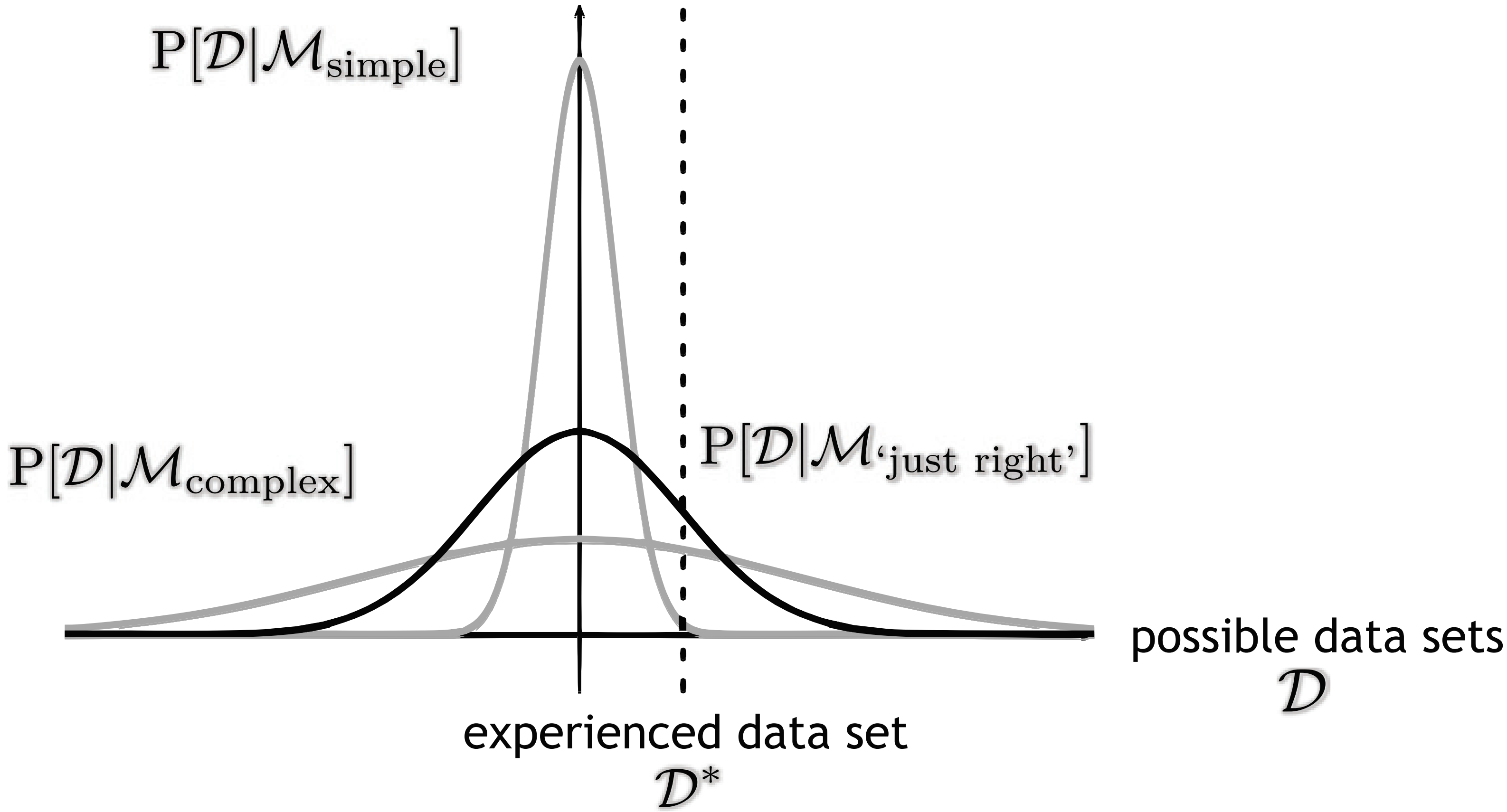
Rev. Bayes

a model defines a probability distribution over data sets

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

data coming from the environment

model stored in memory



BAYESIAN MODEL SELECTION

Rev. Bayes

a model defines a probability distribution over data sets

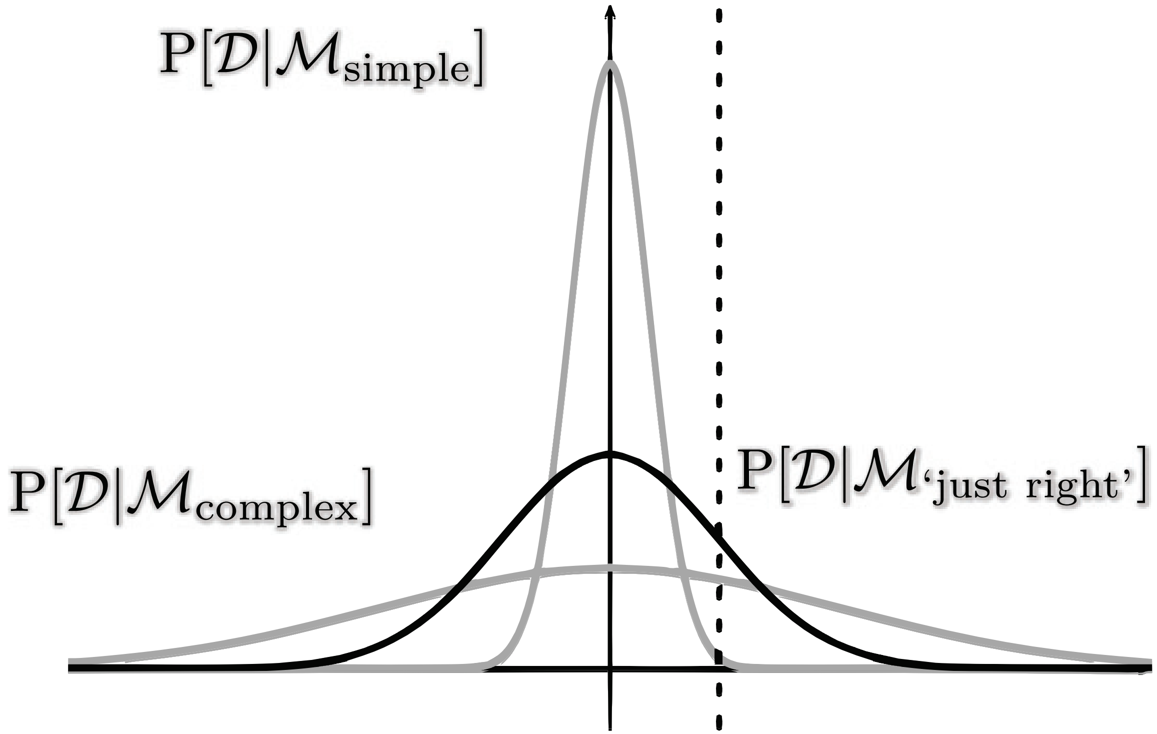
$$\mathrm{P}[\mathcal{D}|\mathcal{M}]$$

data coming from the environment

model stored in memory

Occam's razor

‘ entities should not be multiplied beyond necessity ’



VISUAL PATTERN LEARNING


Gergő Orbán


József Fiser


Richard Aslin

VISUAL PATTERN LEARNING


Gergő Orbán


József Fiser


Richard Aslin

exposure

‘pay attention’

VISUAL PATTERN LEARNING

Gergő Orbán

József Fiser

Richard Aslin

exposure

'pay attention'

VISUAL PATTERN LEARNING


Gergő Orbán


József Fiser


Richard Aslin

exposure

'pay attention'

test

'which one looks more familiar?'

VISUAL PATTERN LEARNING


Gergő Orbán


József Fiser


Richard Aslin

exposure

‘pay attention’

test

‘which one looks more familiar?’

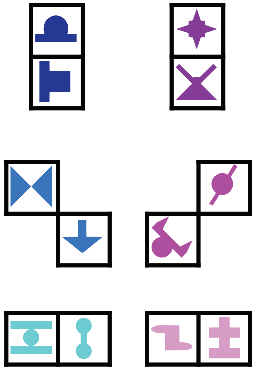
VISUAL PATTERN LEARNING

Gergő Orbán | József Fiser | Richard Aslin

exposure

‘pay attention’

inventory



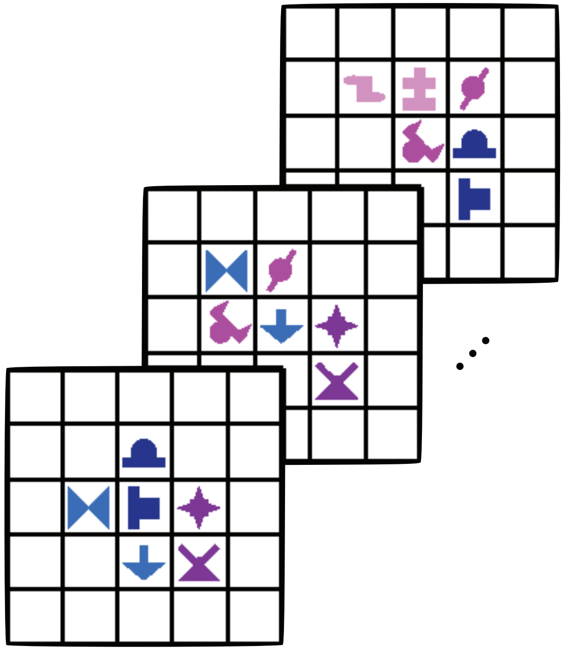
test

‘which one looks more familiar?’

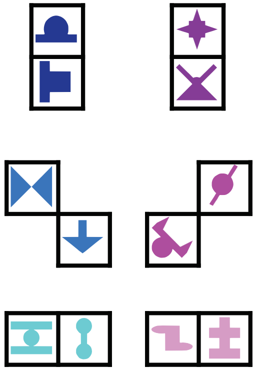
VISUAL PATTERN LEARNING

exposure

‘pay attention’



inventory



test

‘which one looks more familiar?’

VISUAL PATTERN LEARNING

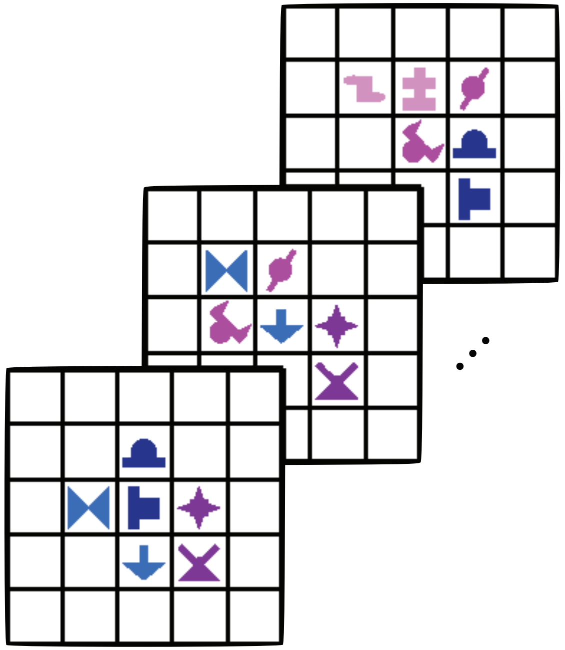

Gergő Orbán


József Fiser

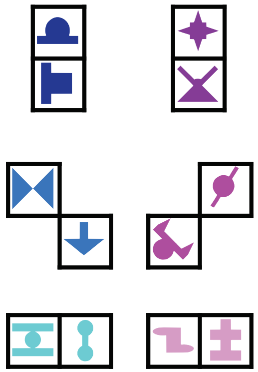

Richard Aslin

exposure

‘pay attention’

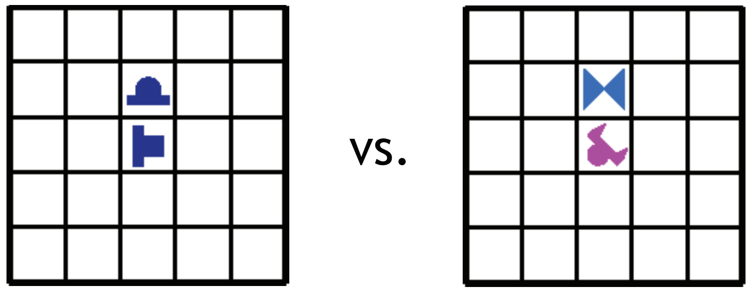


inventory



test

‘which one looks more familiar?’



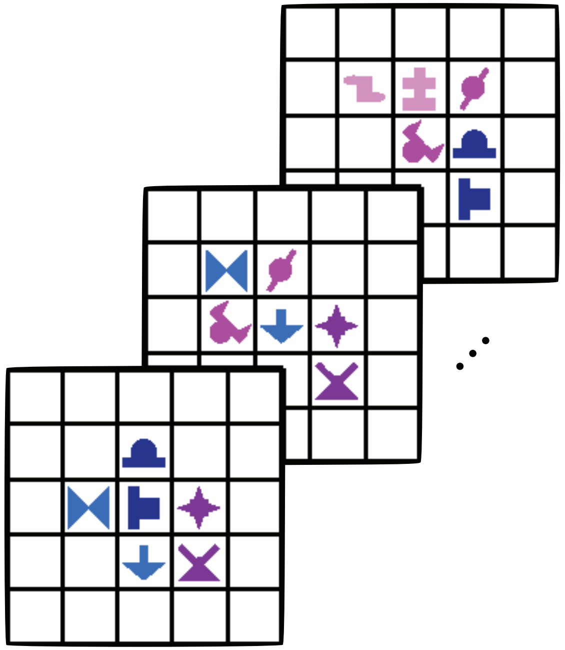
Orbán & al, PNAS 2008

VISUAL PATTERN LEARNING

Gergő Orbán | József Fiser | Richard Aslin

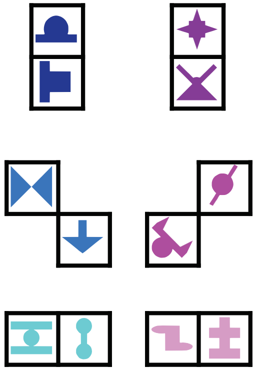
exposure

'pay attention'



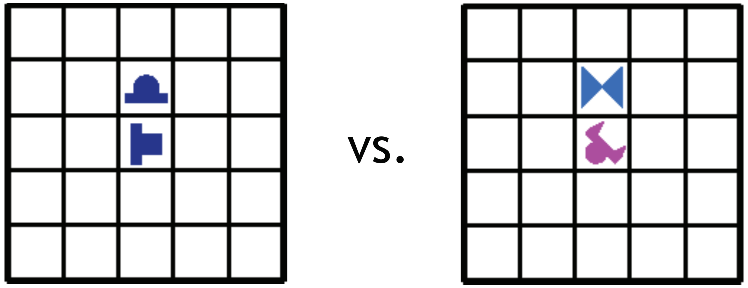
$\mathcal{D} = \mathrm{scene}_1, \ldots \mathrm{scene}_n$

inventory



test

'which one looks more familiar?'



Orbán & al, PNAS 2008

VISUAL PATTERN LEARNING

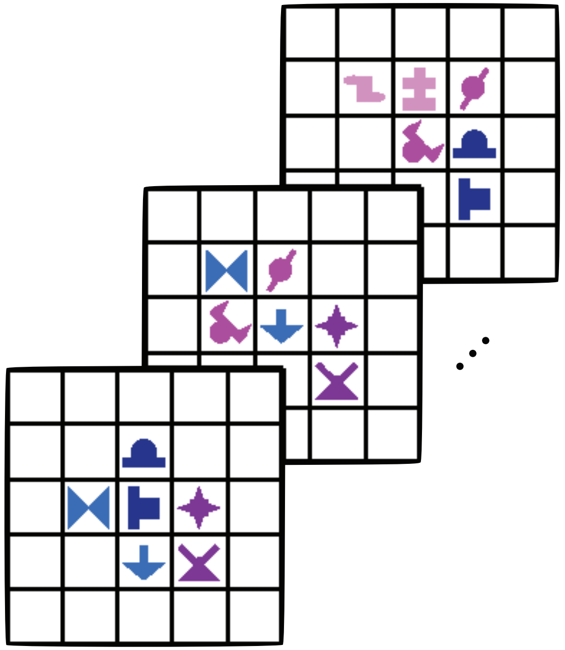

Gergő Orbán

József Fiser

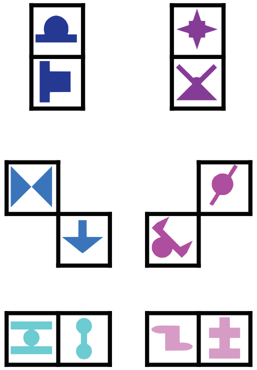
Richard Aslin

exposure

'pay attention'



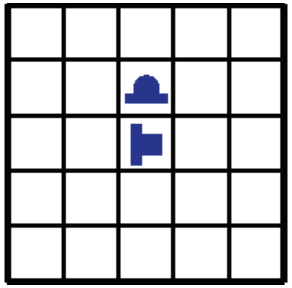
inventory



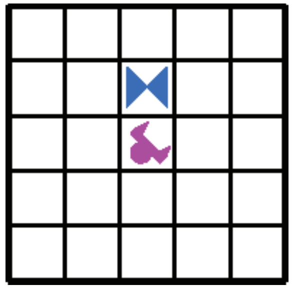
$\mathcal{M}$

test

'which one looks more familiar?'



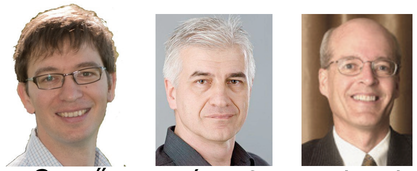
vs.



Orbán & al, PNAS 2008

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

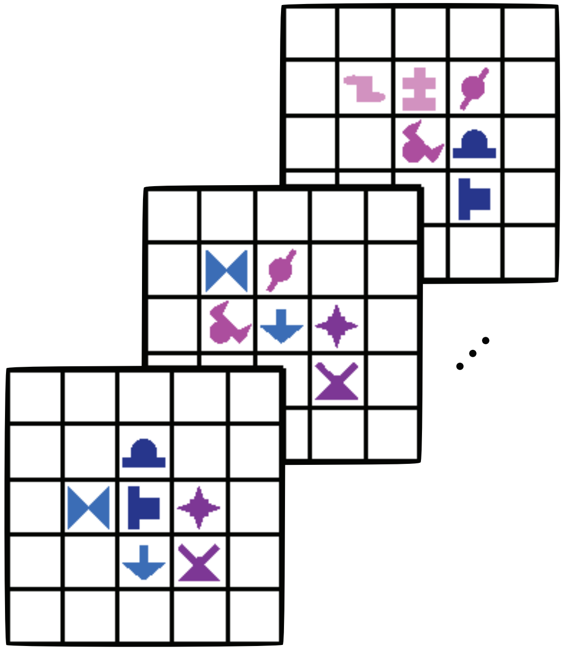
VISUAL PATTERN LEARNING



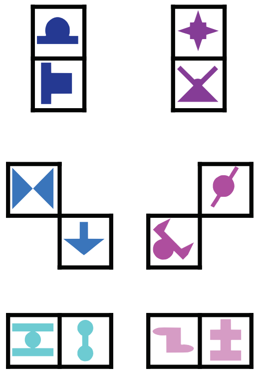
Gergő Orbán
József Fiser
Richard Aslin

exposure

‘pay attention’



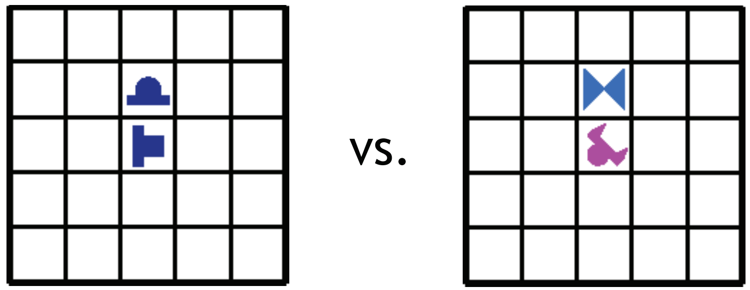
inventory



$\mathcal{M}$

test

‘which one looks more familiar?’



Orbán & al, PNAS 2008

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

VISUAL PATTERN LEARNING

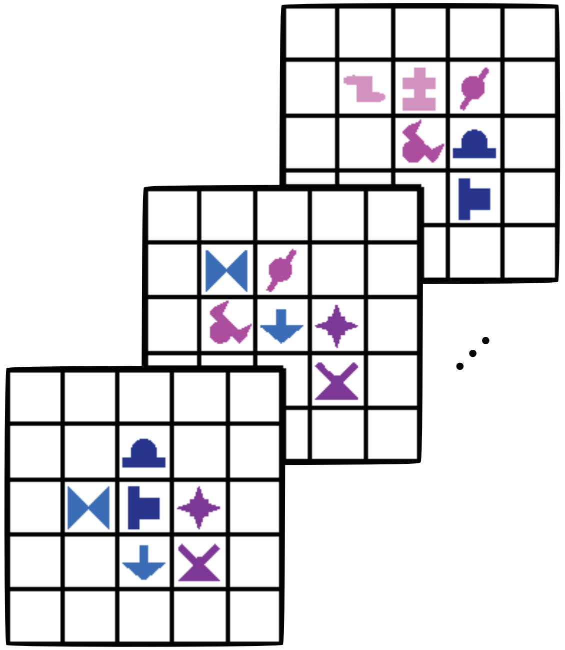

Gergő Orbán


József Fiser

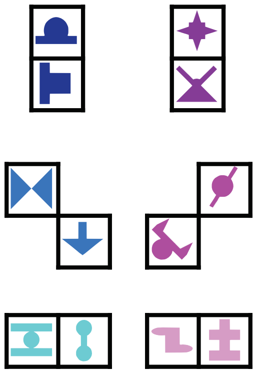

Richard Aslin

exposure

'pay attention'



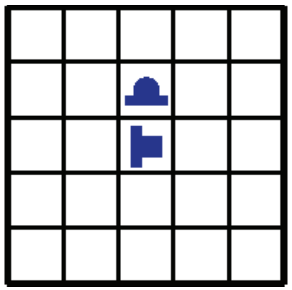
inventory



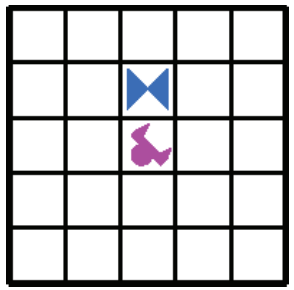
$\mathcal{M}$

test

'which one looks more familiar?'



vs.



Orbán & al, PNAS 2008

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

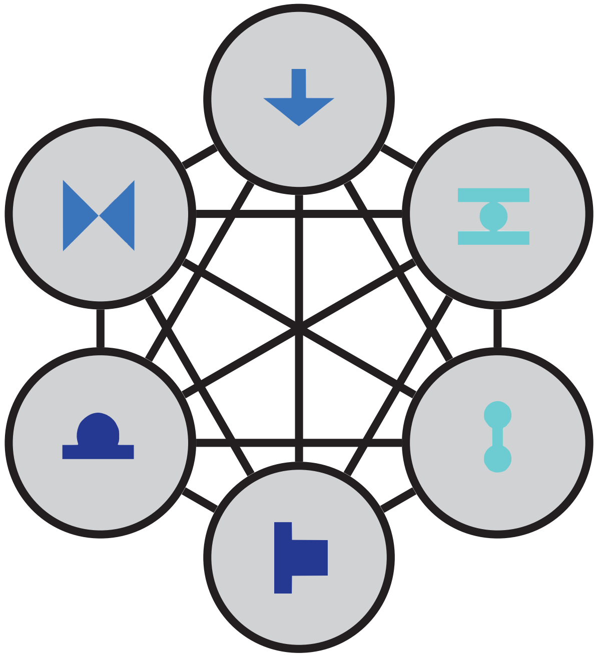
how do humans learn a statistical model of their environment?

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

ALTERNATIVE THEORIES

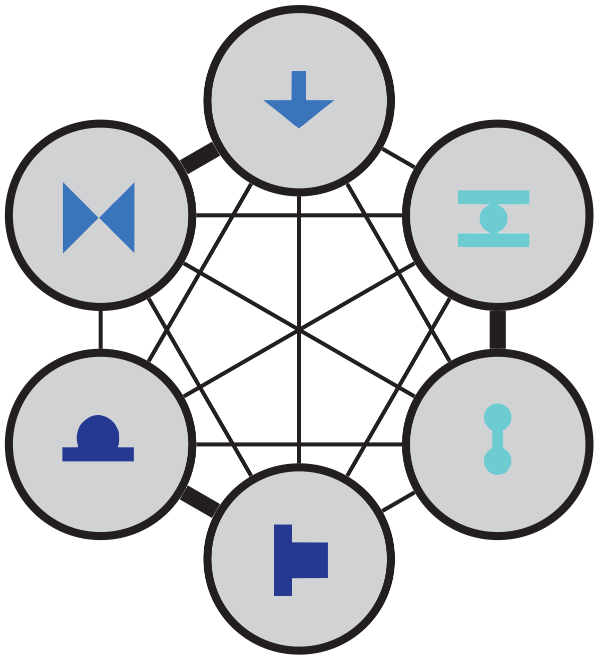
ALTERNATIVE THEORIES

associative learning



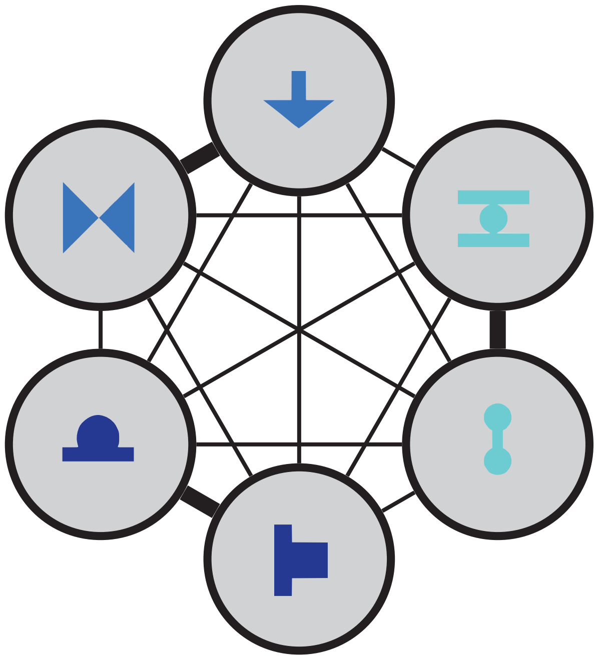
ALTERNATIVE THEORIES

associative learning

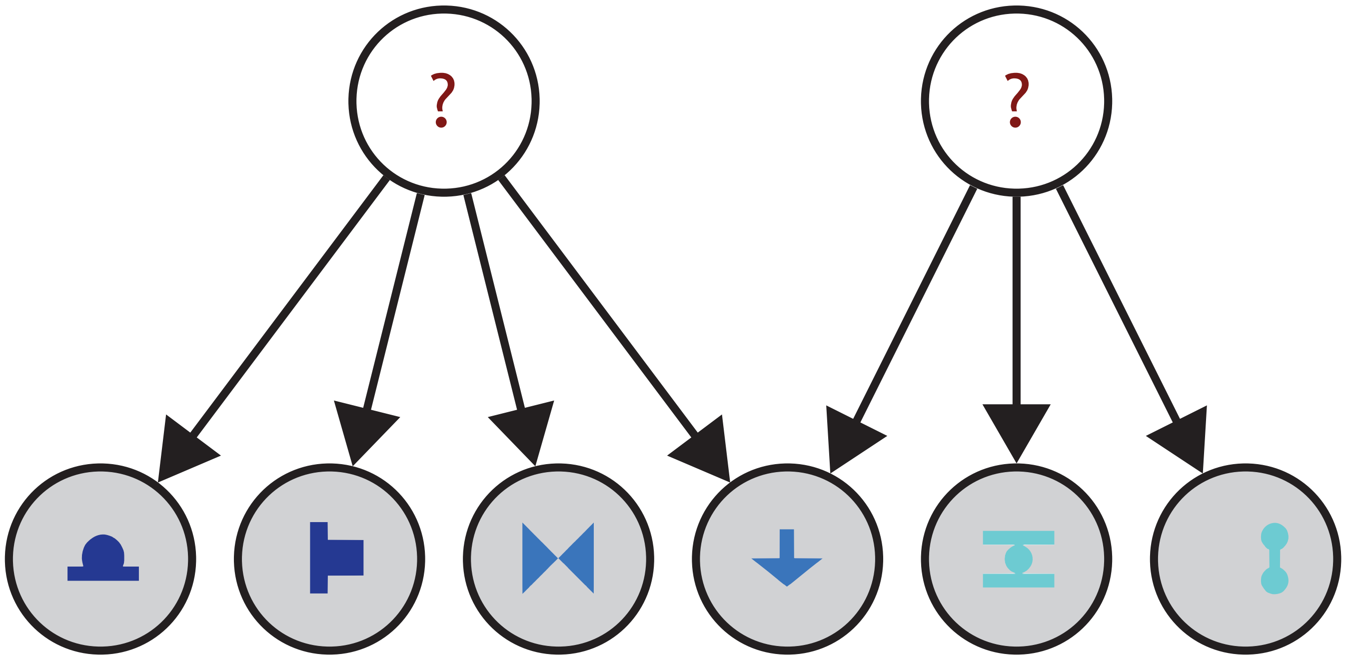


ALTERNATIVE THEORIES

associative learning

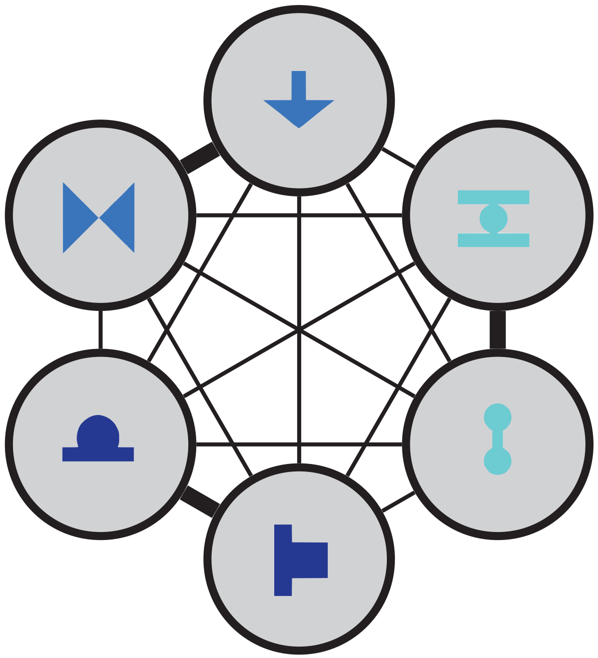


Bayesian learning

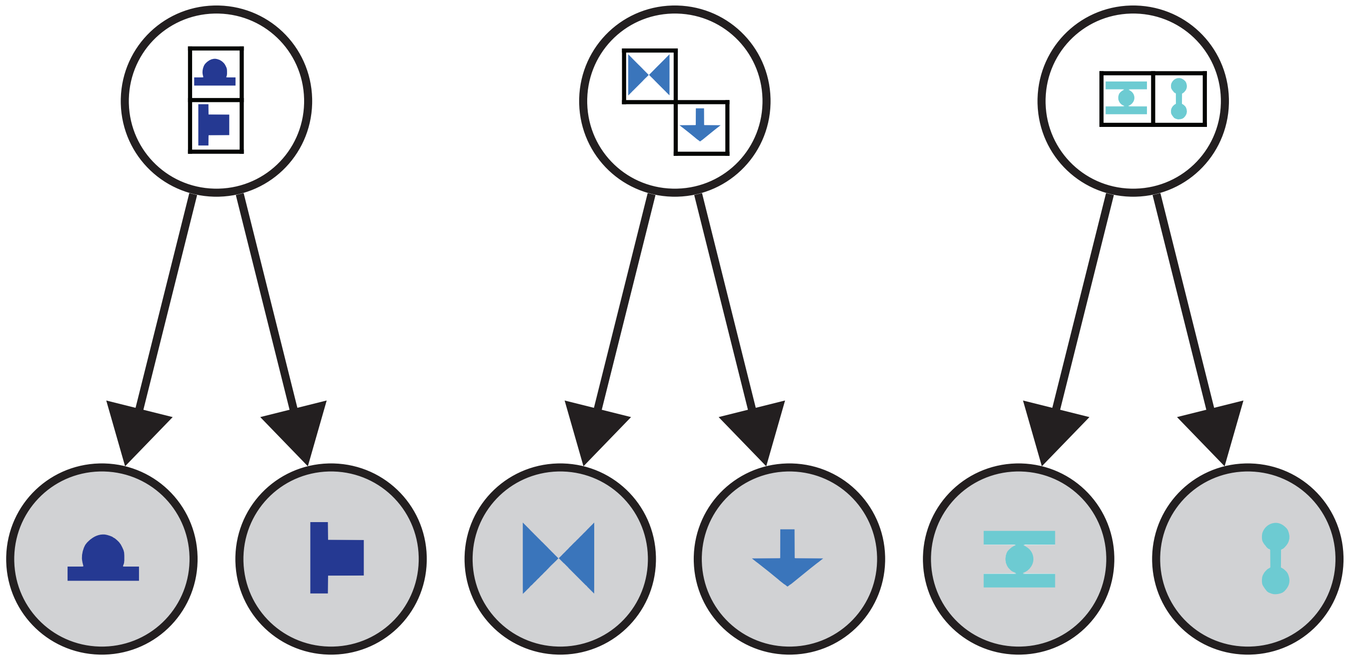


ALTERNATIVE THEORIES

associative learning



Bayesian learning

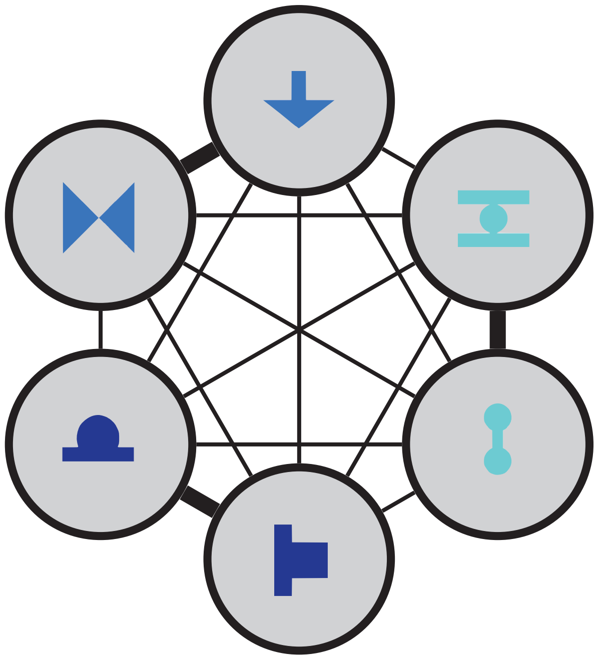


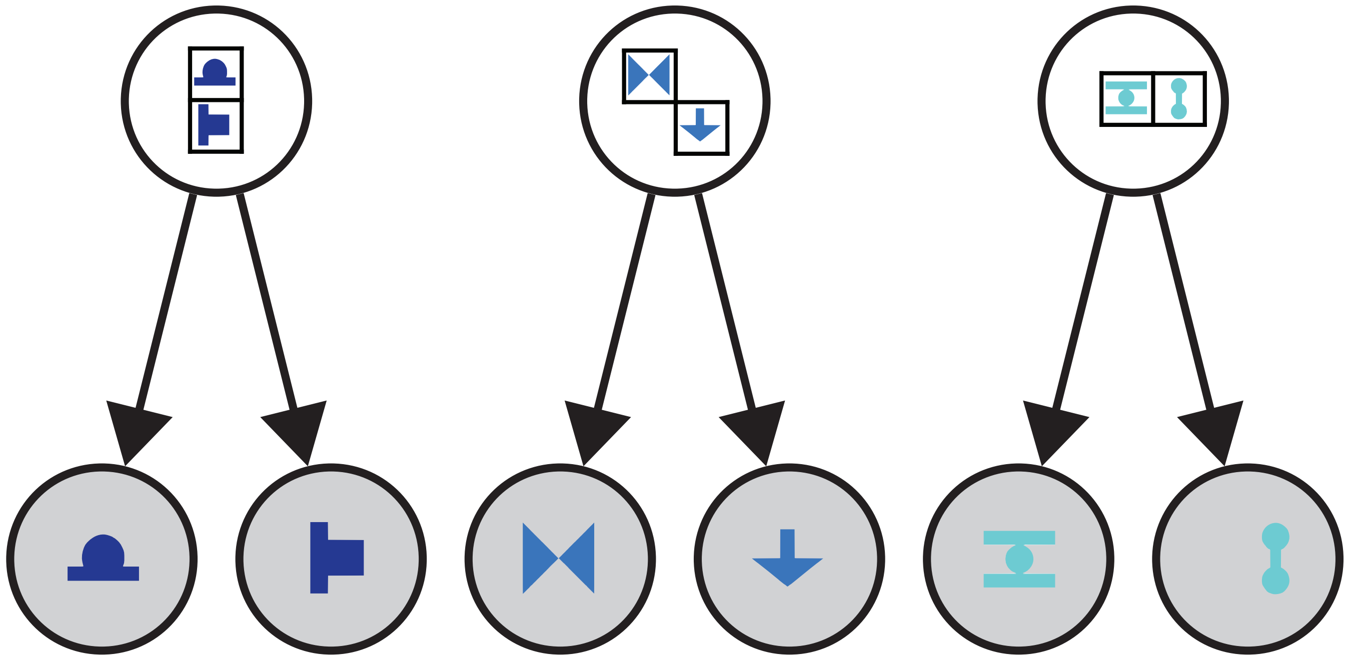
Orbán & al, PNAS 2008

ALTERNATIVE THEORIES

associative learning

Bayesian learning





Orbán & al, PNAS 2008

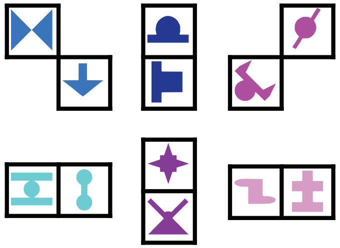
Boltzmann machine
+ Gaussian Markov random field

sigmoid belief network
+ product of (conditional) Gaussian experts

MULTIPLE EXPERIMENTS

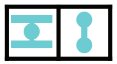
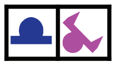
baseline

inventory



basic

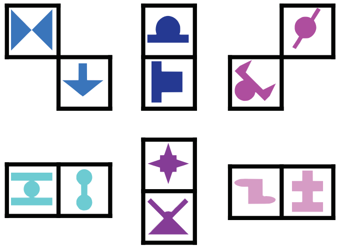
test type

 vs. 

MULTIPLE EXPERIMENTS

inventory

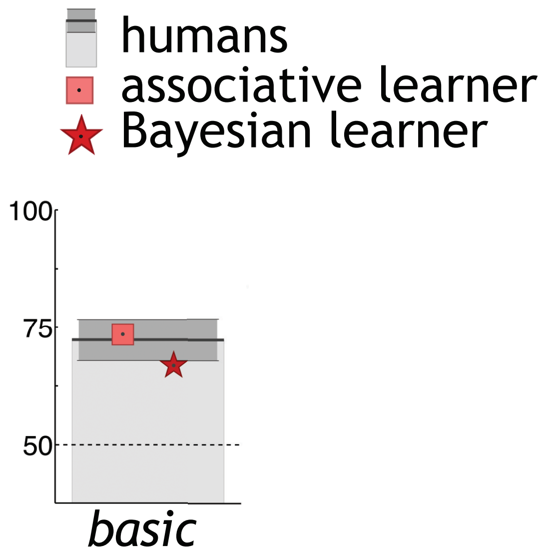
baseline



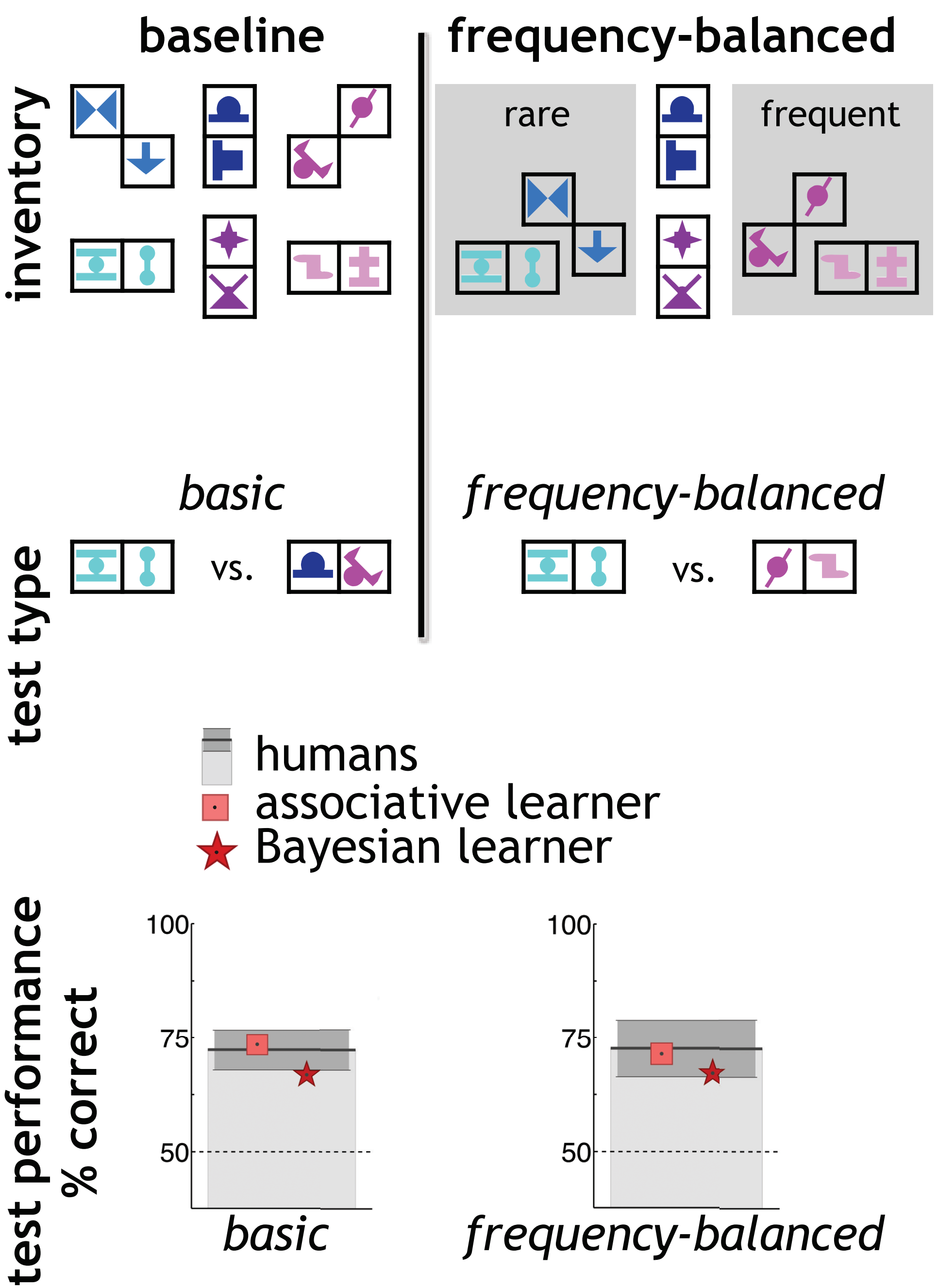
test type

basic

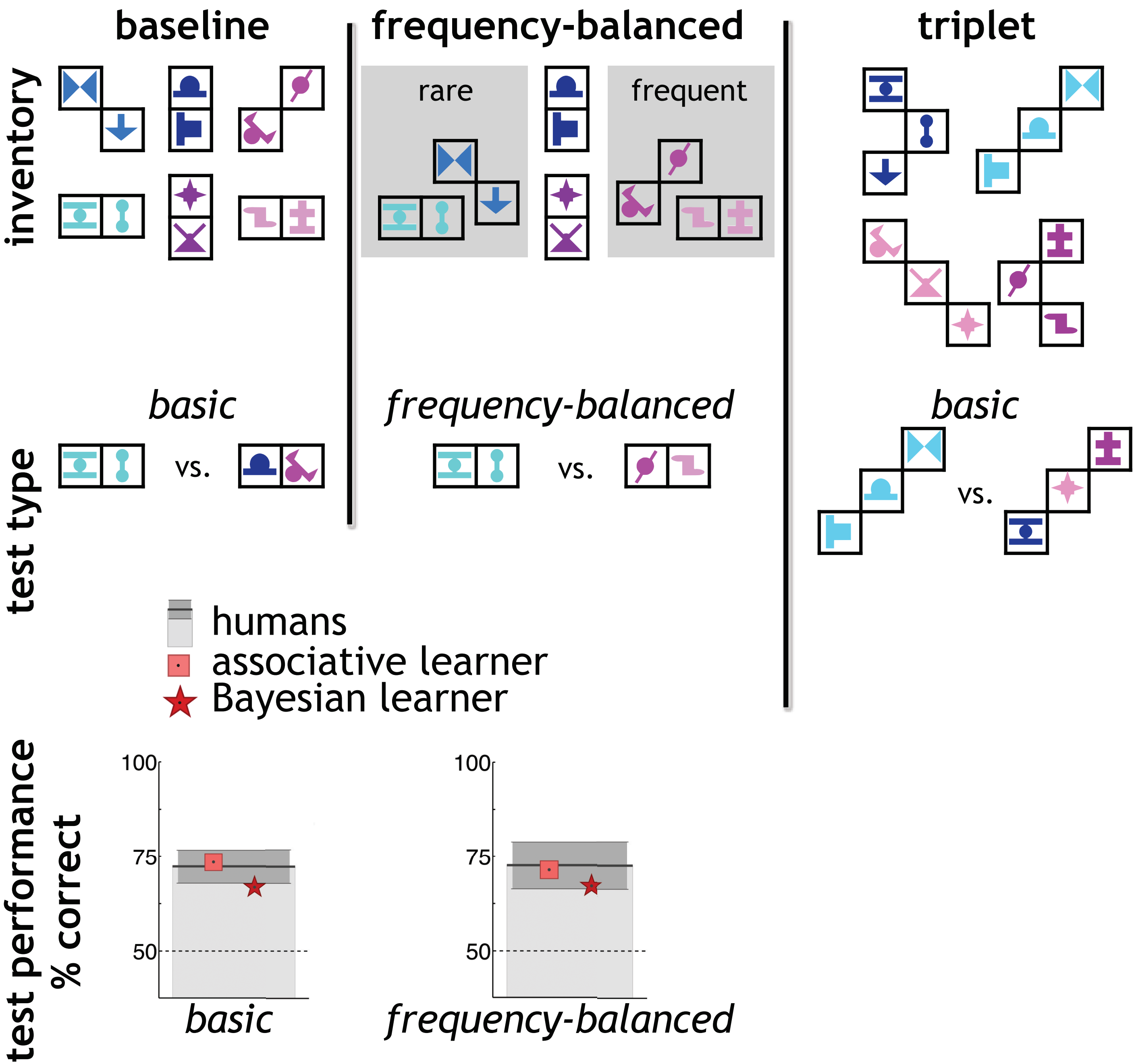
test performance
% correct



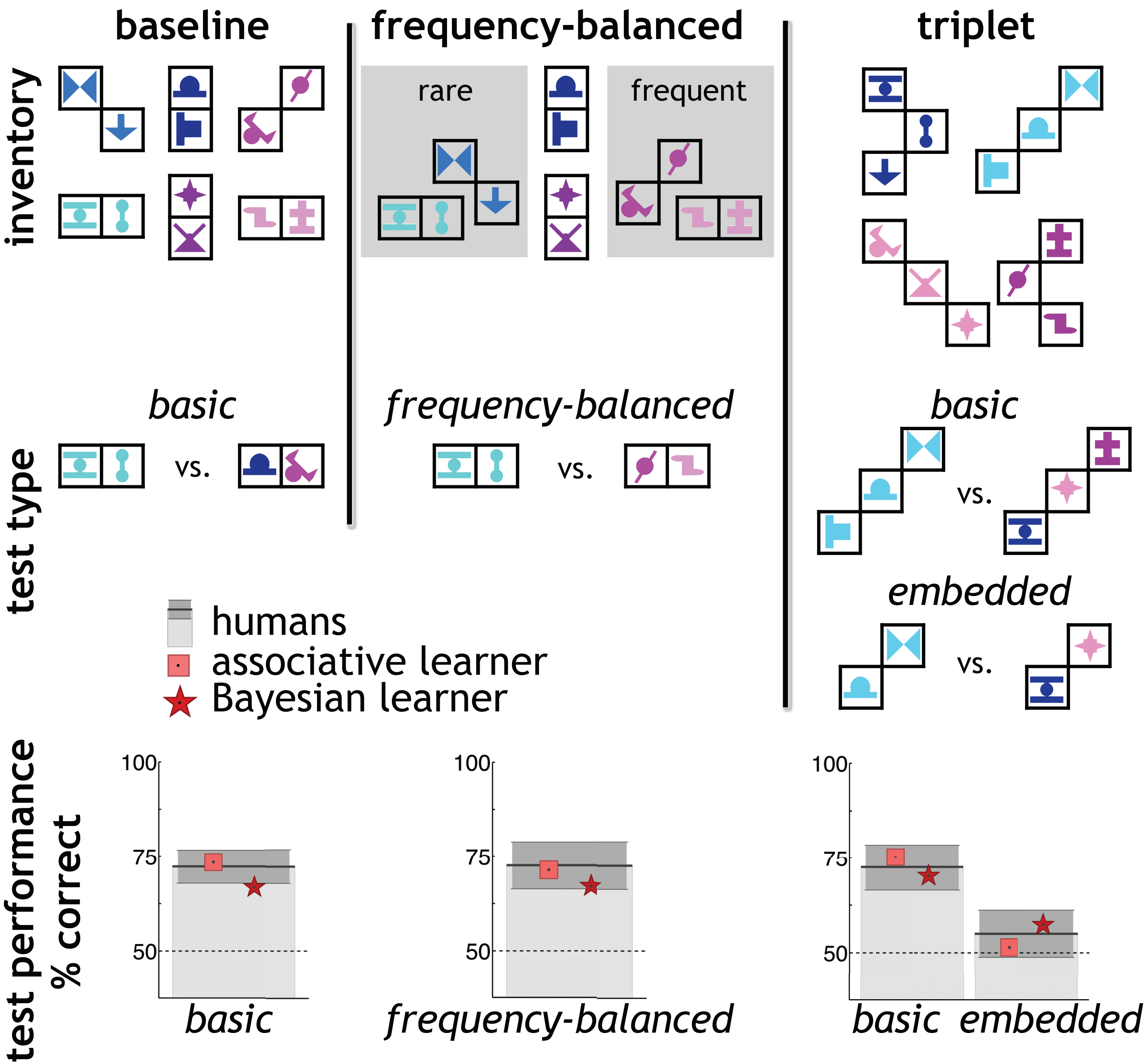
MULTIPLE EXPERIMENTS



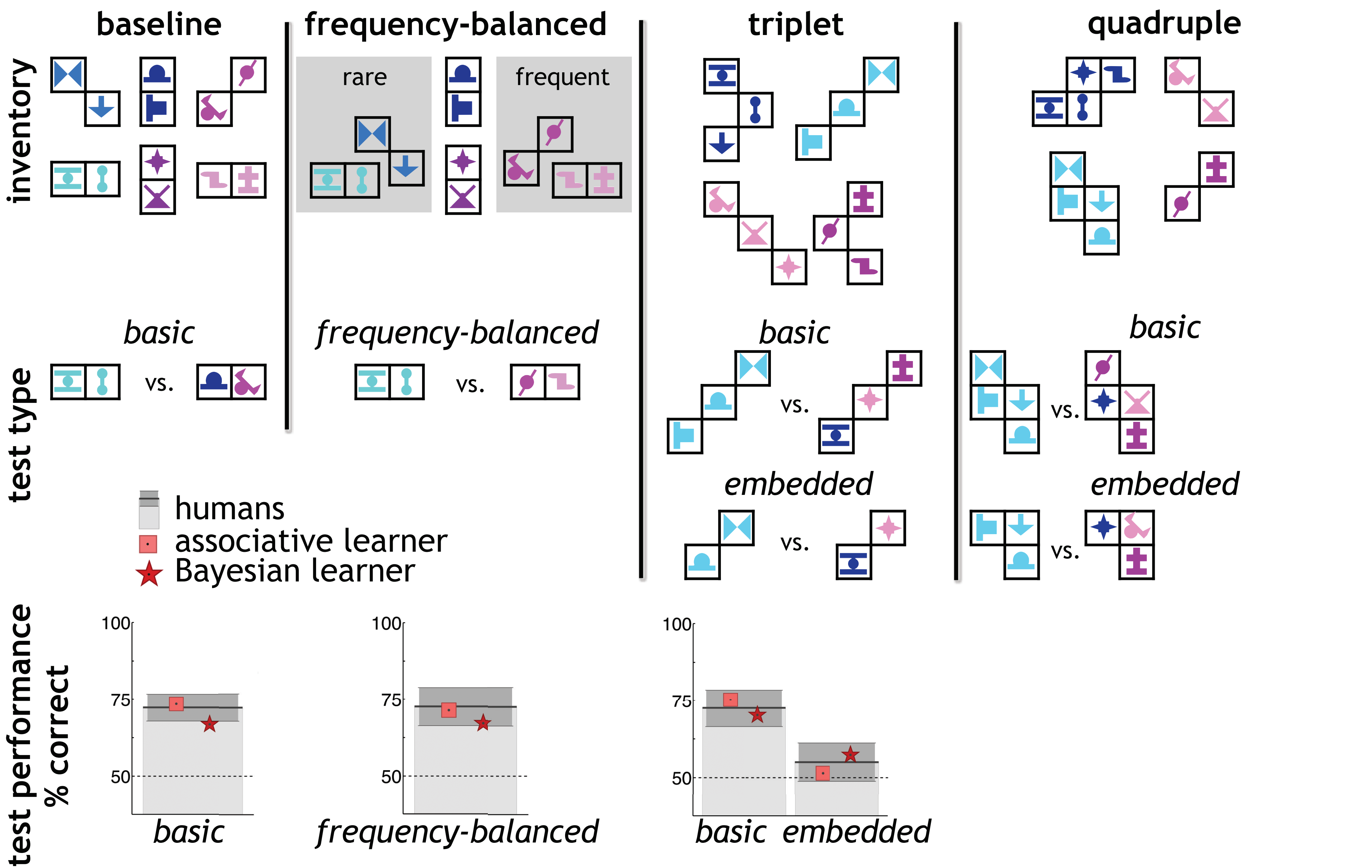
MULTIPLE EXPERIMENTS



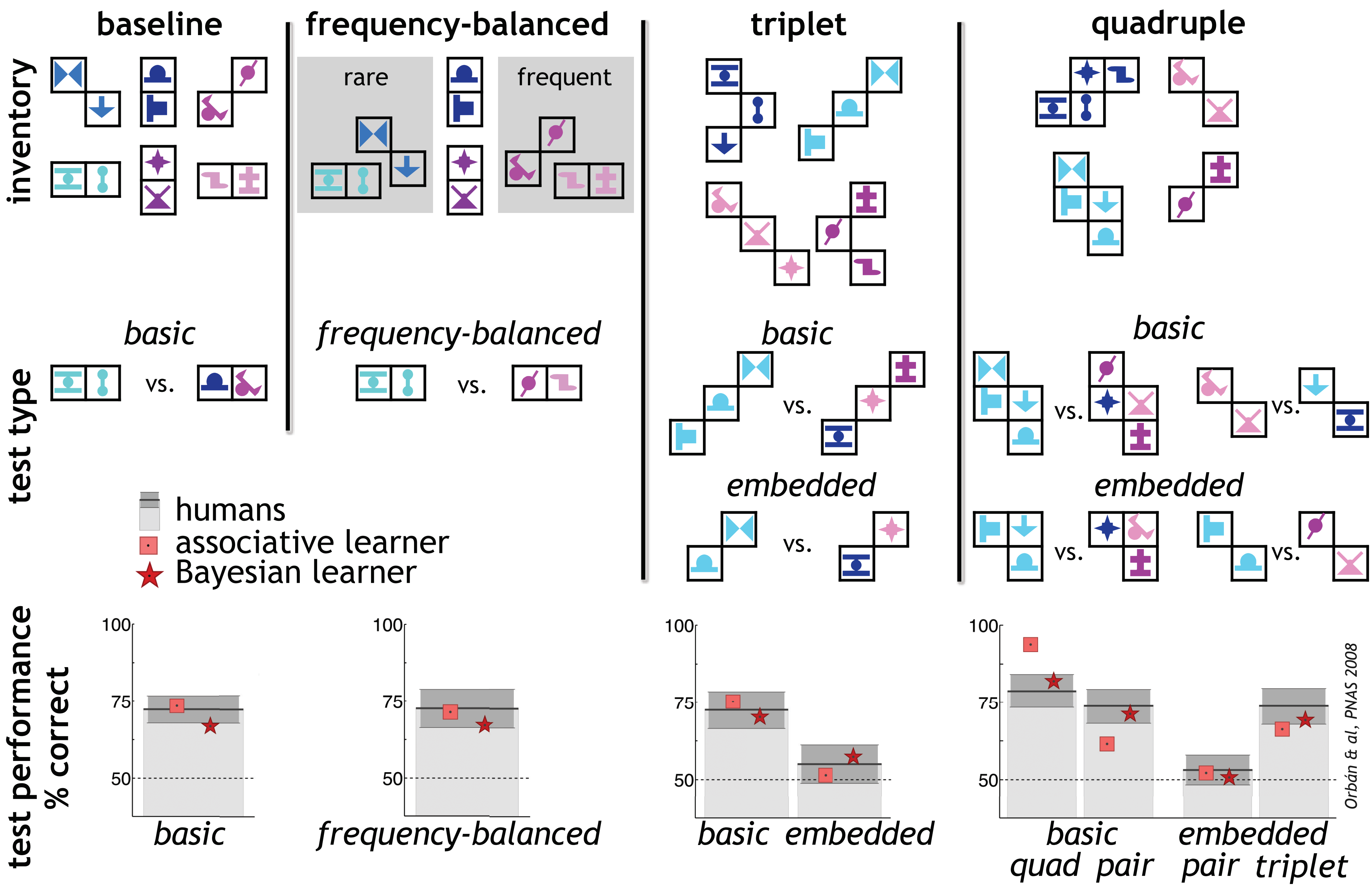
MULTIPLE EXPERIMENTS



MULTIPLE EXPERIMENTS

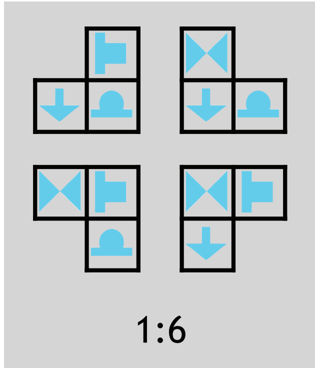


MULTIPLE EXPERIMENTS



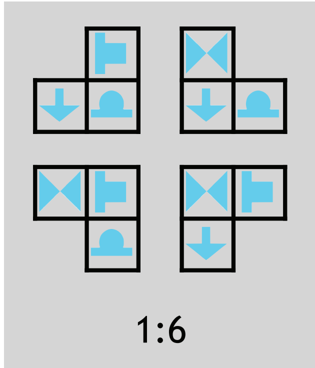
ASSOCIATIVE VS. BAYESIAN LEARNING

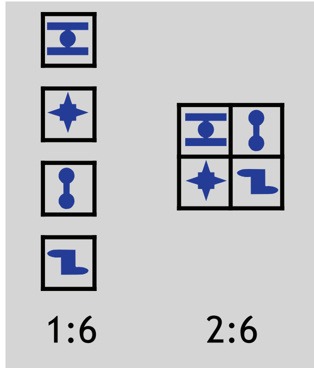
inventory:



ASSOCIATIVE VS. BAYESIAN LEARNING

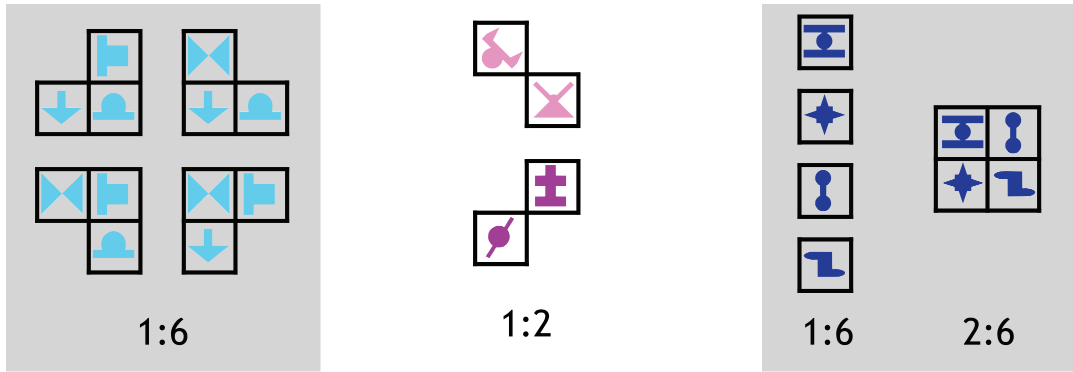
inventory:





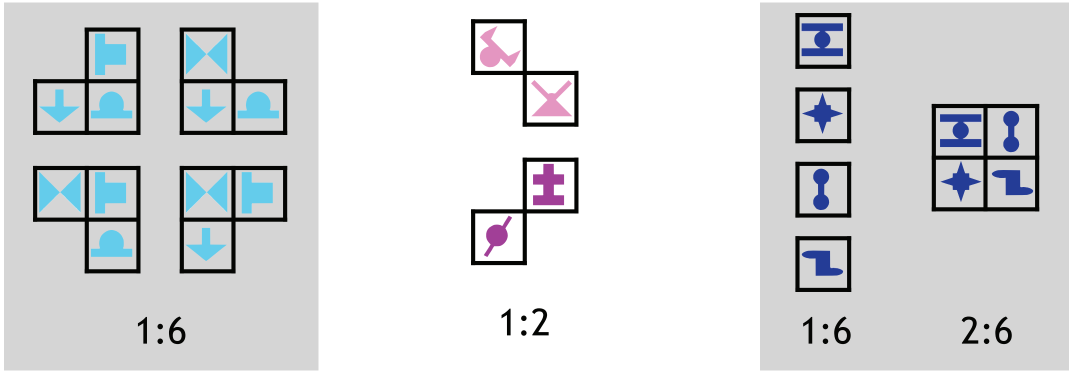
ASSOCIATIVE VS. BAYESIAN LEARNING

inventory:



ASSOCIATIVE VS. BAYESIAN LEARNING

inventory:



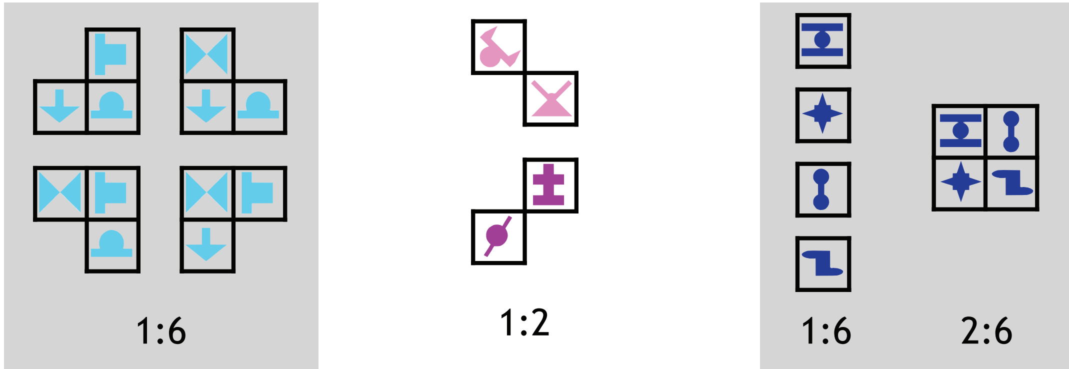
1st order statistic: shape frequencies

3 × 1/6

1/6 + 2/6

ASSOCIATIVE VS. BAYESIAN LEARNING

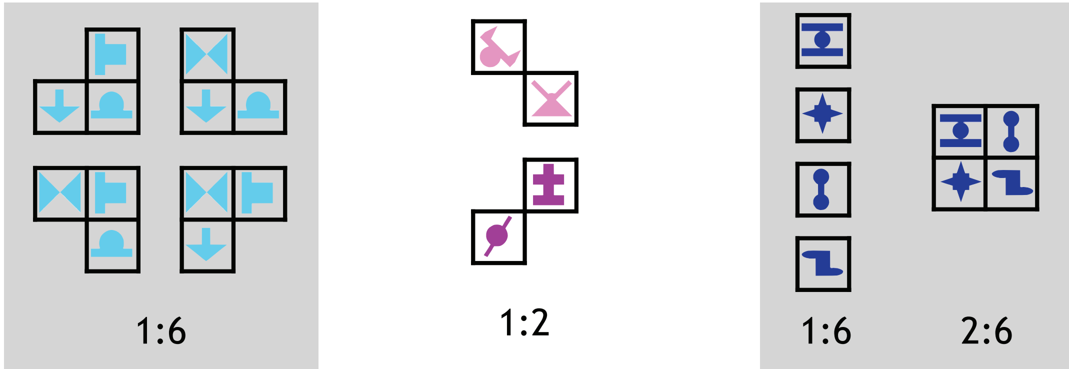
inventory:



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

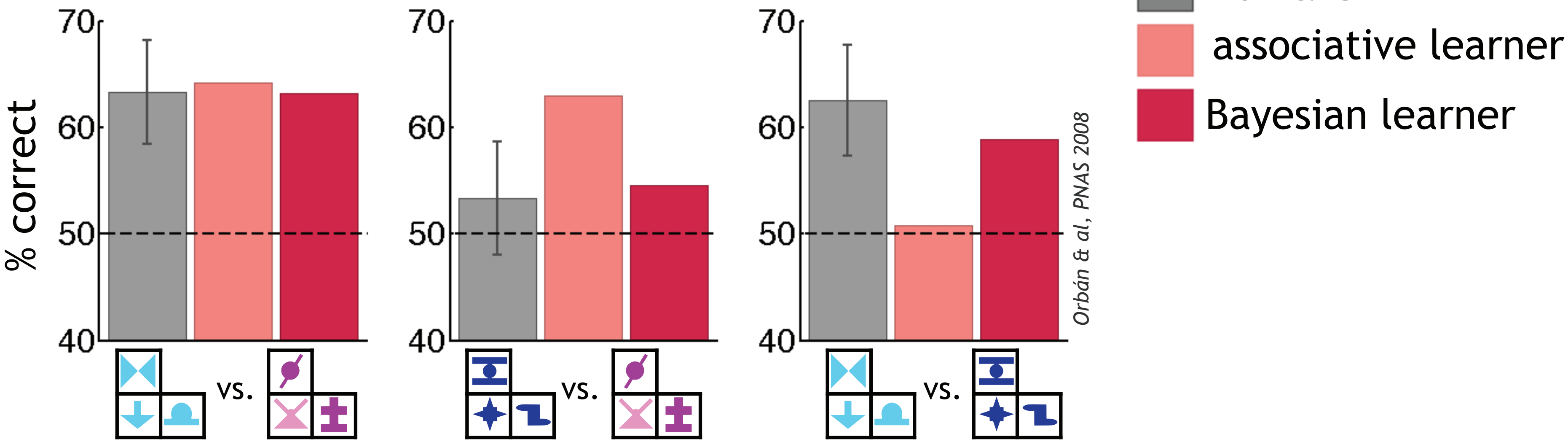
ASSOCIATIVE VS. BAYESIAN LEARNING

inventory:



1^{st} order statistic: shape frequencies		$3 \times 1/6$	$1/6 + 2/6$
2^{nd} 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:



STATISTICAL CHUNKS = OBJECTS?


Gábor Lengyel


József Fiser


Alex Pantelides

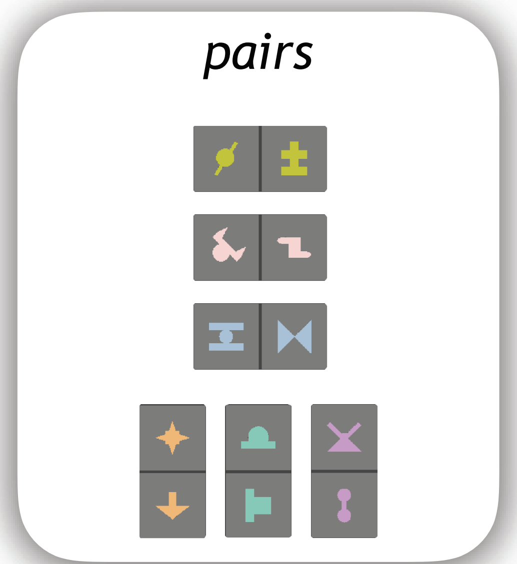

Goda Zalalyte


James Ingram

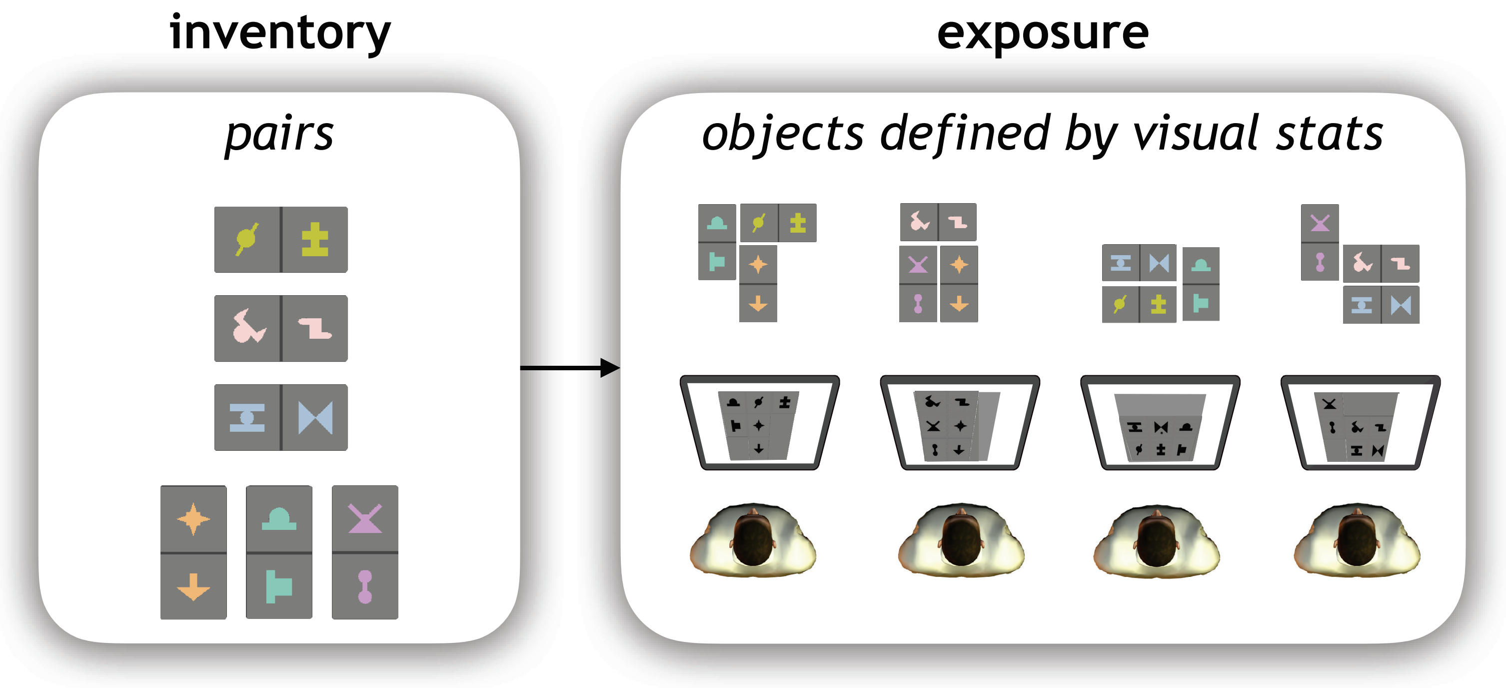

Daniel Wolpert

STATISTICAL CHUNKS = OBJECTS?

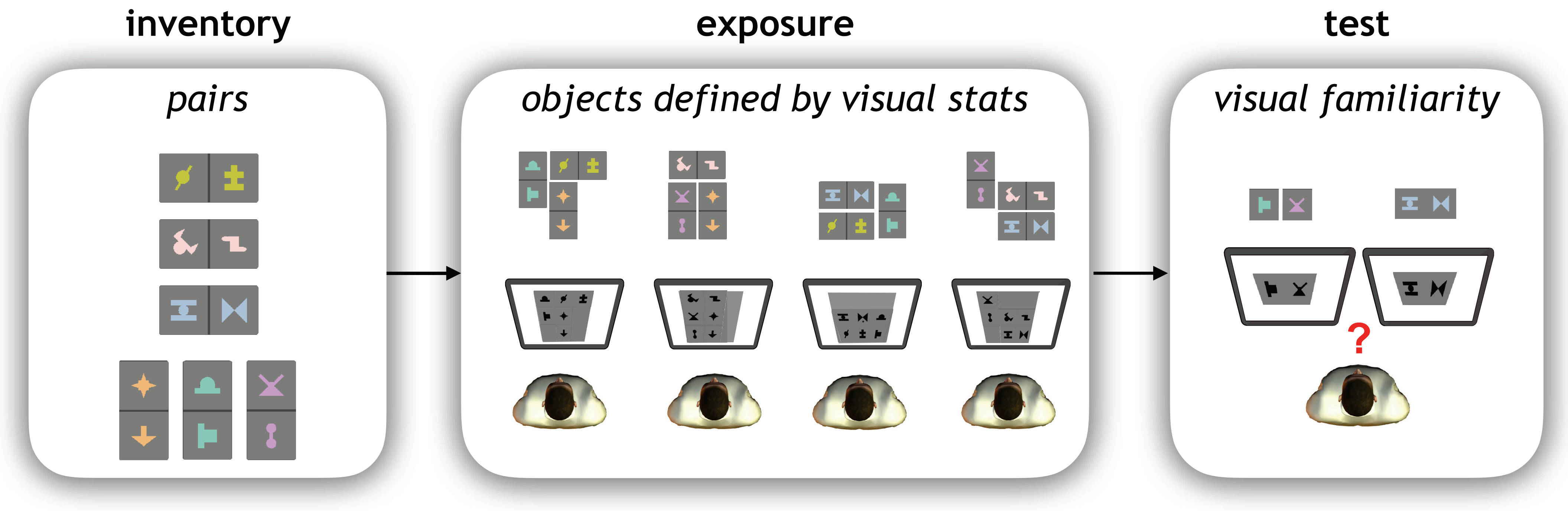
inventory



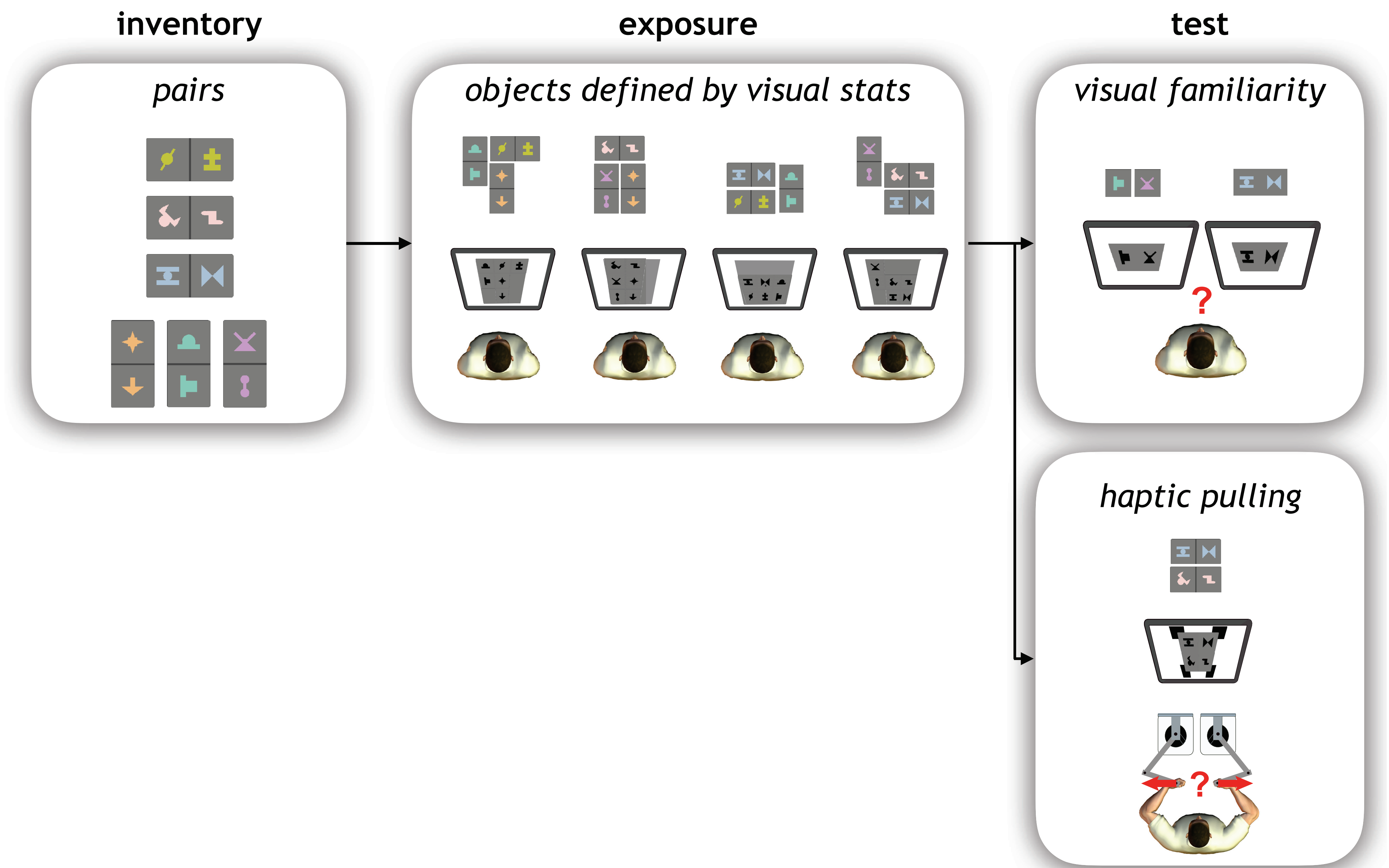
STATISTICAL CHUNKS = OBJECTS?



STATISTICAL CHUNKS = OBJECTS?

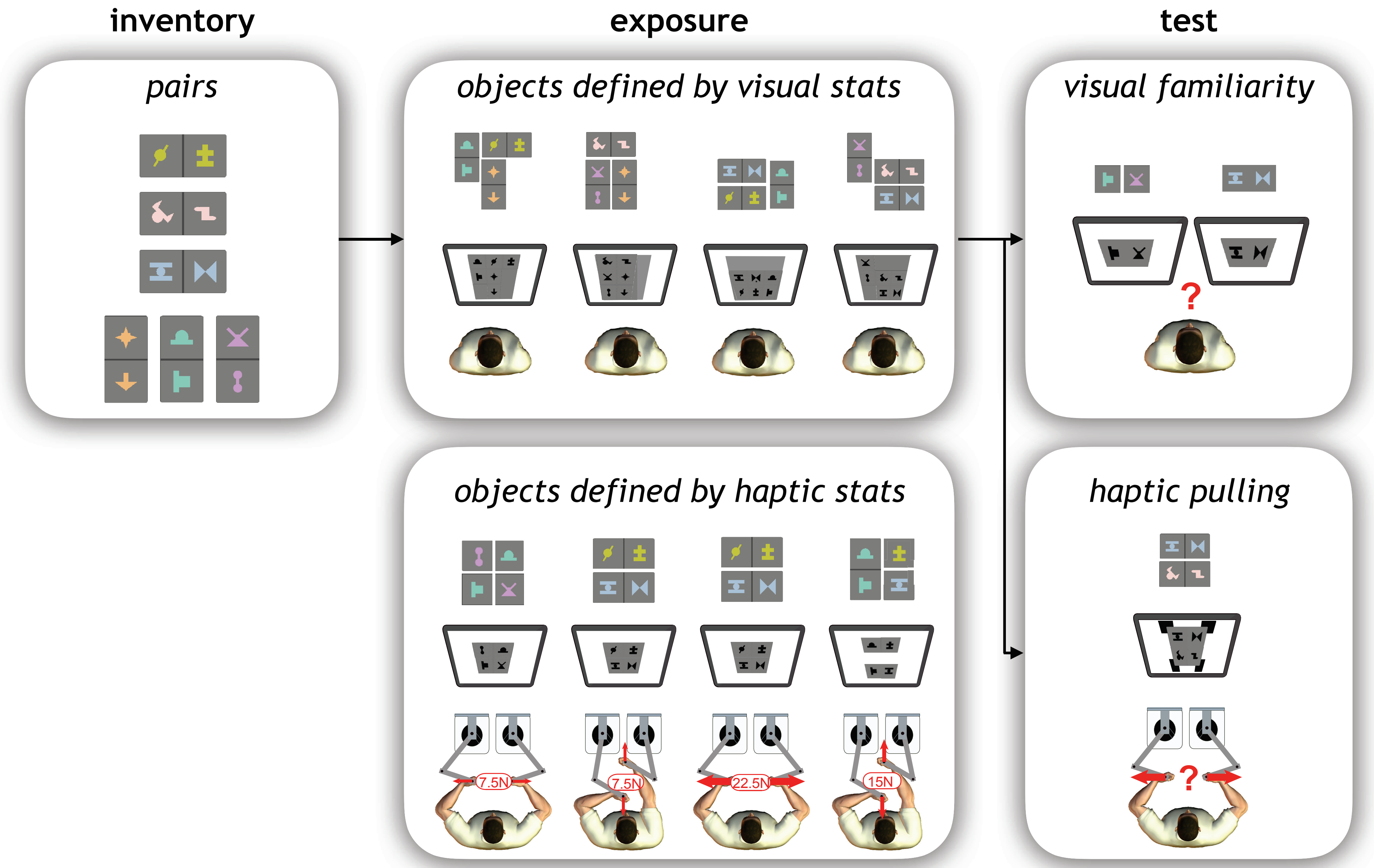


STATISTICAL CHUNKS = OBJECTS?



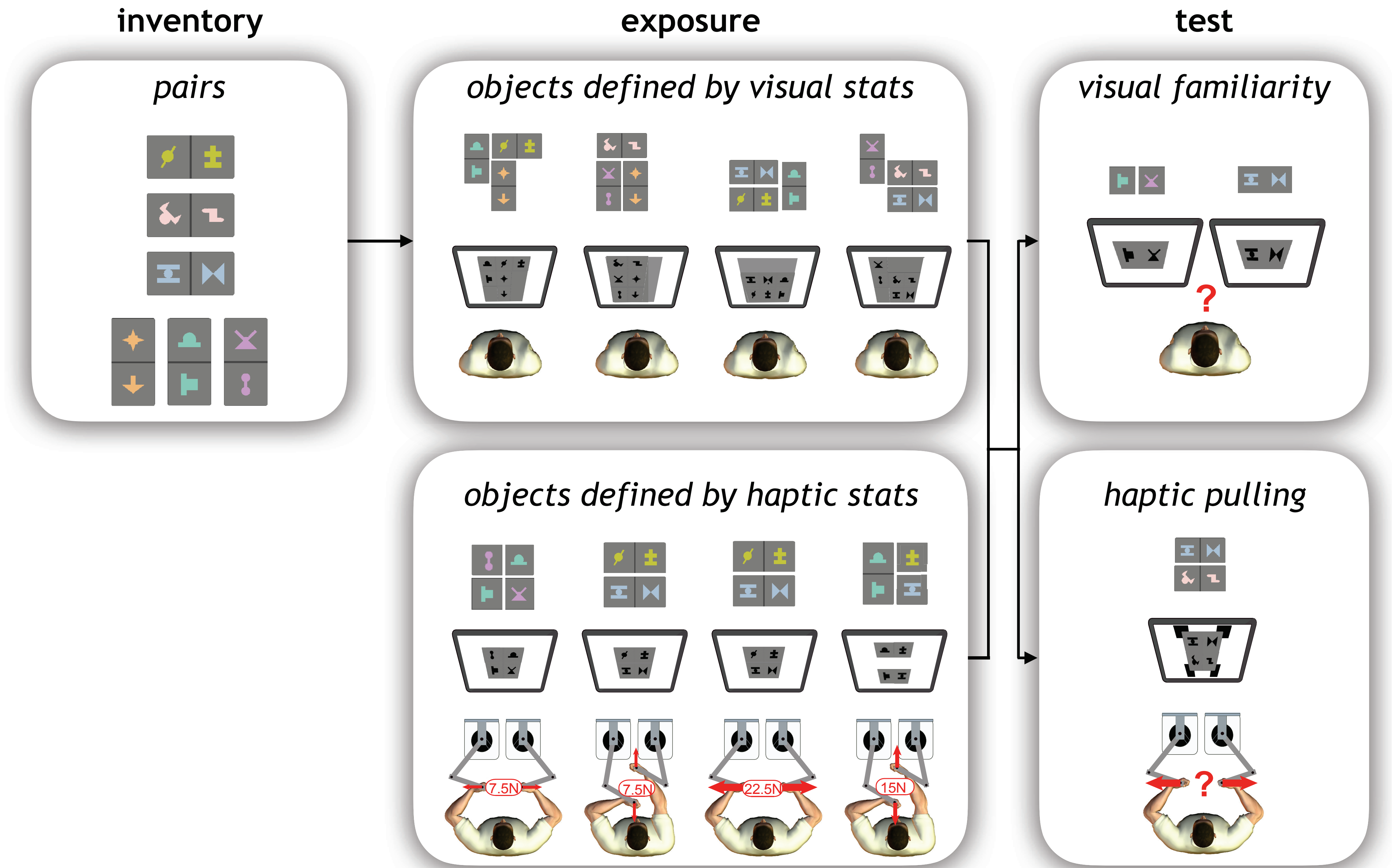
Lengyel & al, under review

STATISTICAL CHUNKS = OBJECTS?



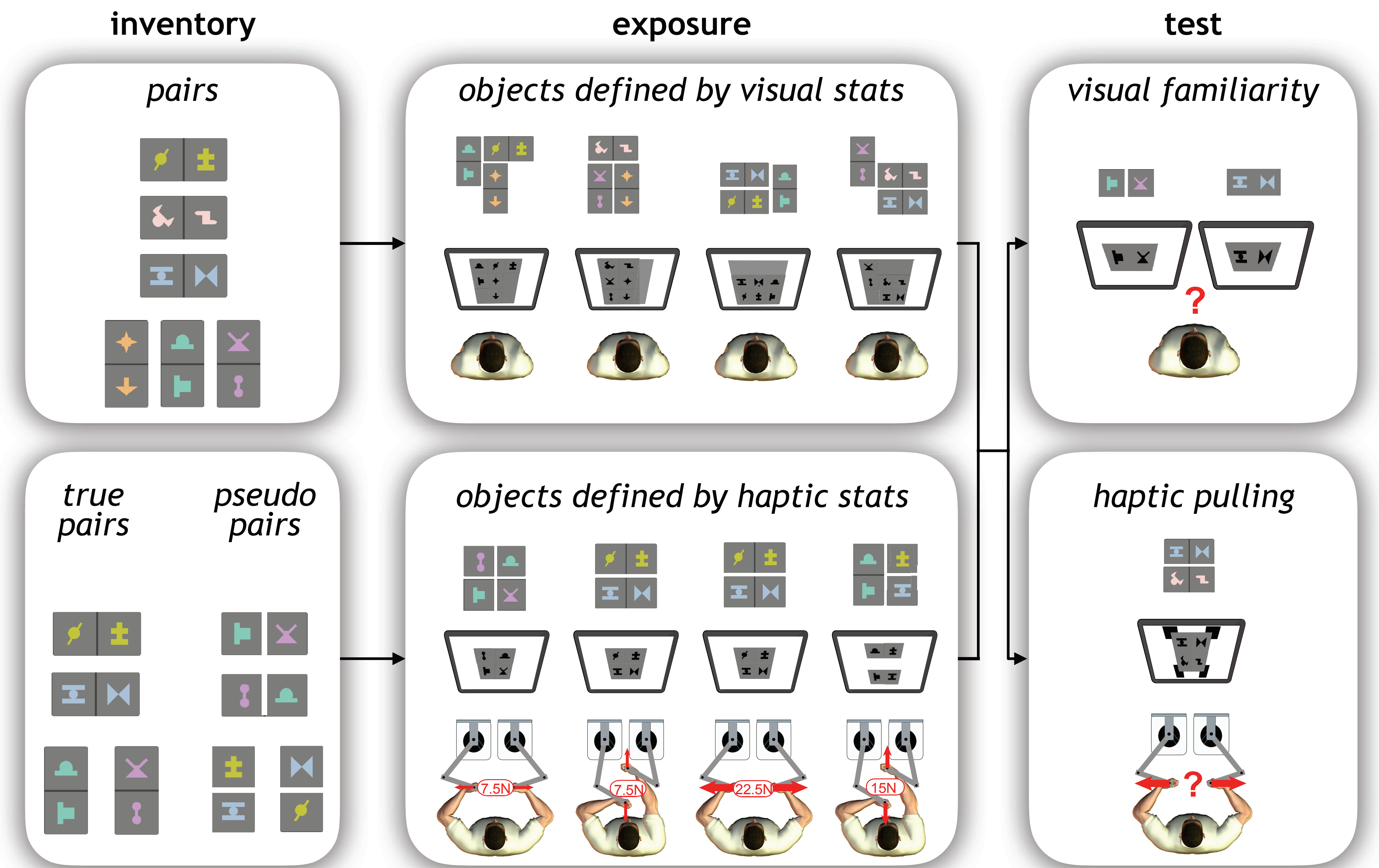
Lengyel & al, under review

STATISTICAL CHUNKS = OBJECTS?



Lengyel & al, under review

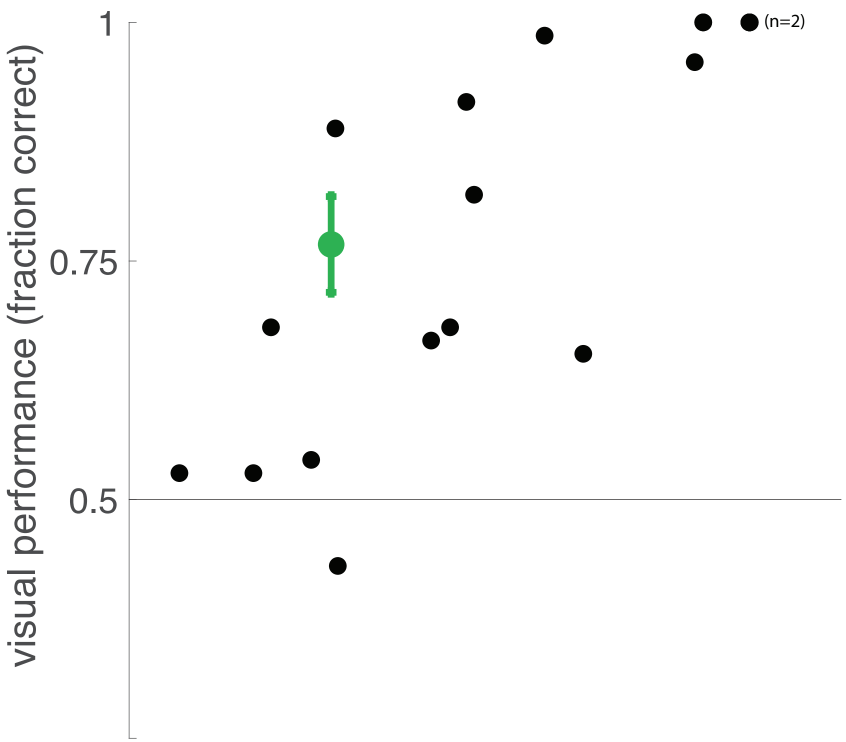
STATISTICAL CHUNKS = OBJECTS?



Lengyel & al, under review

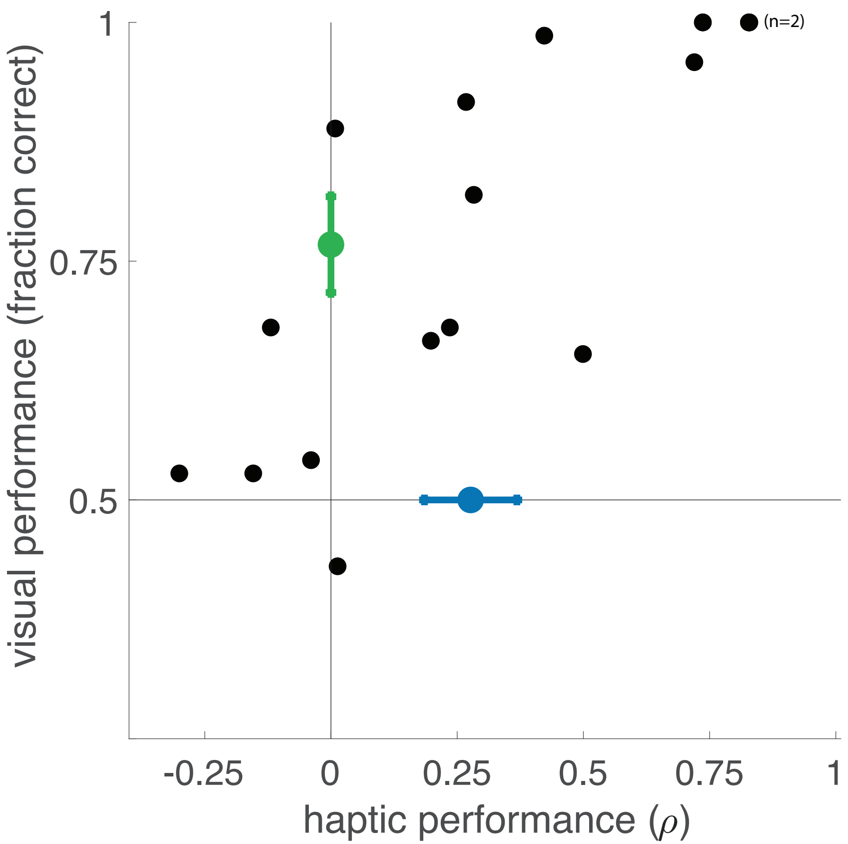
ACROSS-MODALITY GENERALISATION

*objects defined by **visual** stats*



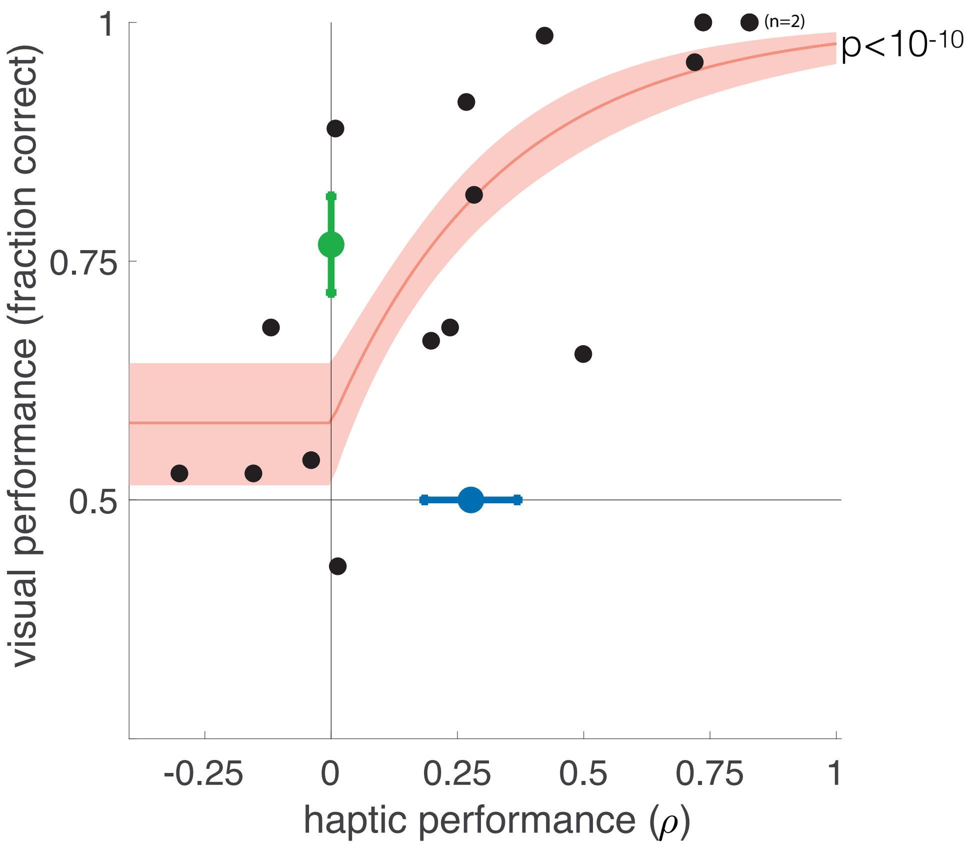
ACROSS-MODALITY GENERALISATION

*objects defined by **visual** stats*

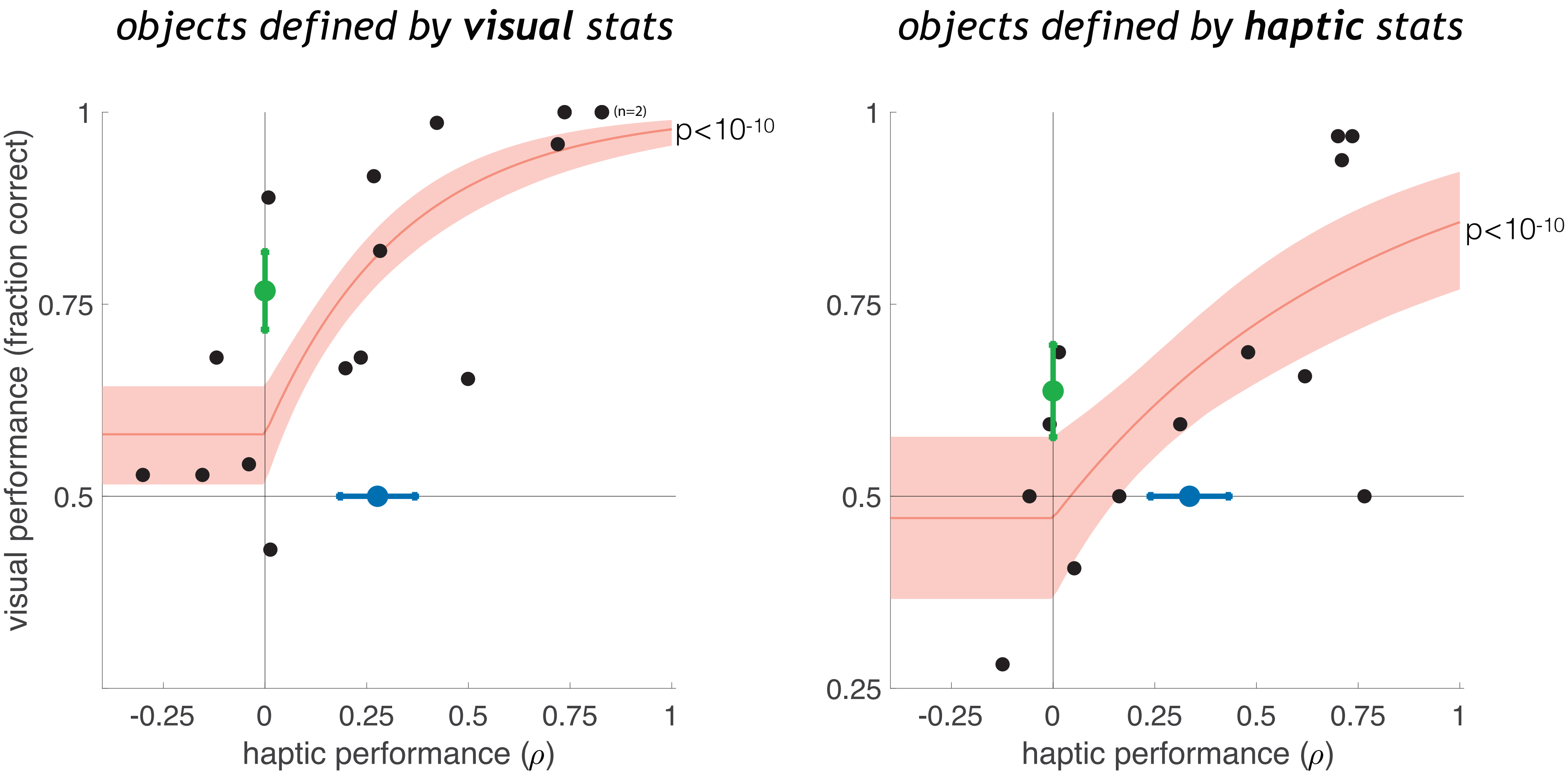


ACROSS-MODALITY GENERALISATION

*objects defined by **visual** stats*



ACROSS-MODALITY GENERALISATION



Lengyel & al, under review

PROBABILISTIC INTERNAL MODELS

$\mathrm{P}_{\mathrm{true}}(\mathbf{y})$

$\mathbf{y}$

objects, features

$\mathrm{P}_{\mathrm{true}}(\mathbf{x}|\mathbf{y})$

$\mathbf{x}$

sensory stimuli

internal model

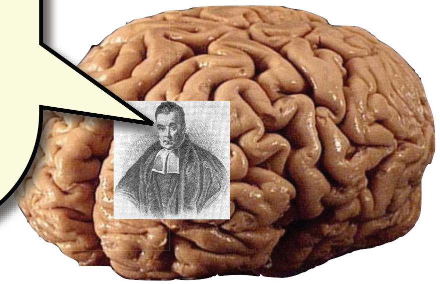
$\mathrm{P}_{\mathrm{model}}(\mathbf{y})$

$\mathbf{y}$

objects, features

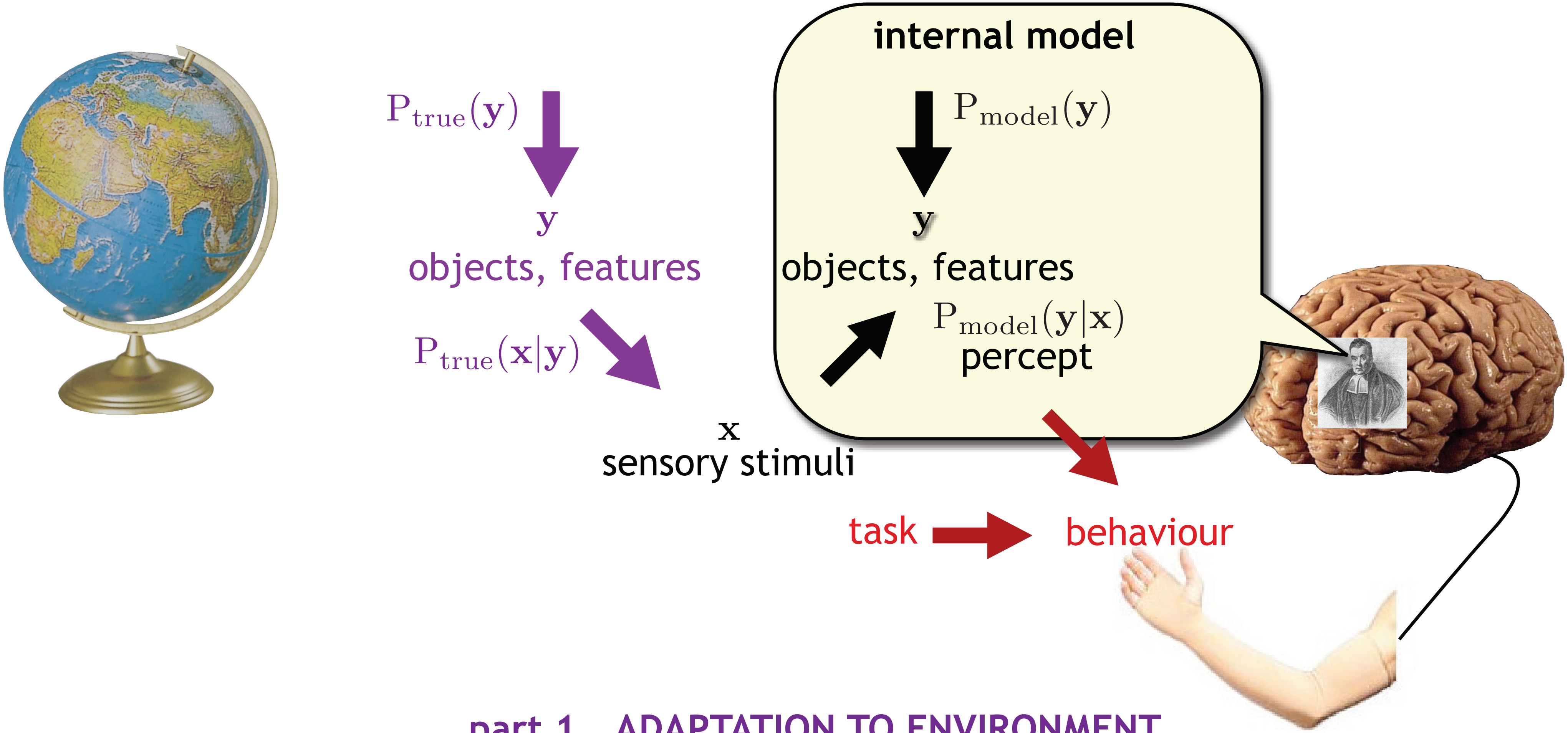
$\mathrm{P}_{\mathrm{model}}(\mathbf{y}|\mathbf{x})$

percept



part 1. ADAPTATION TO ENVIRONMENT

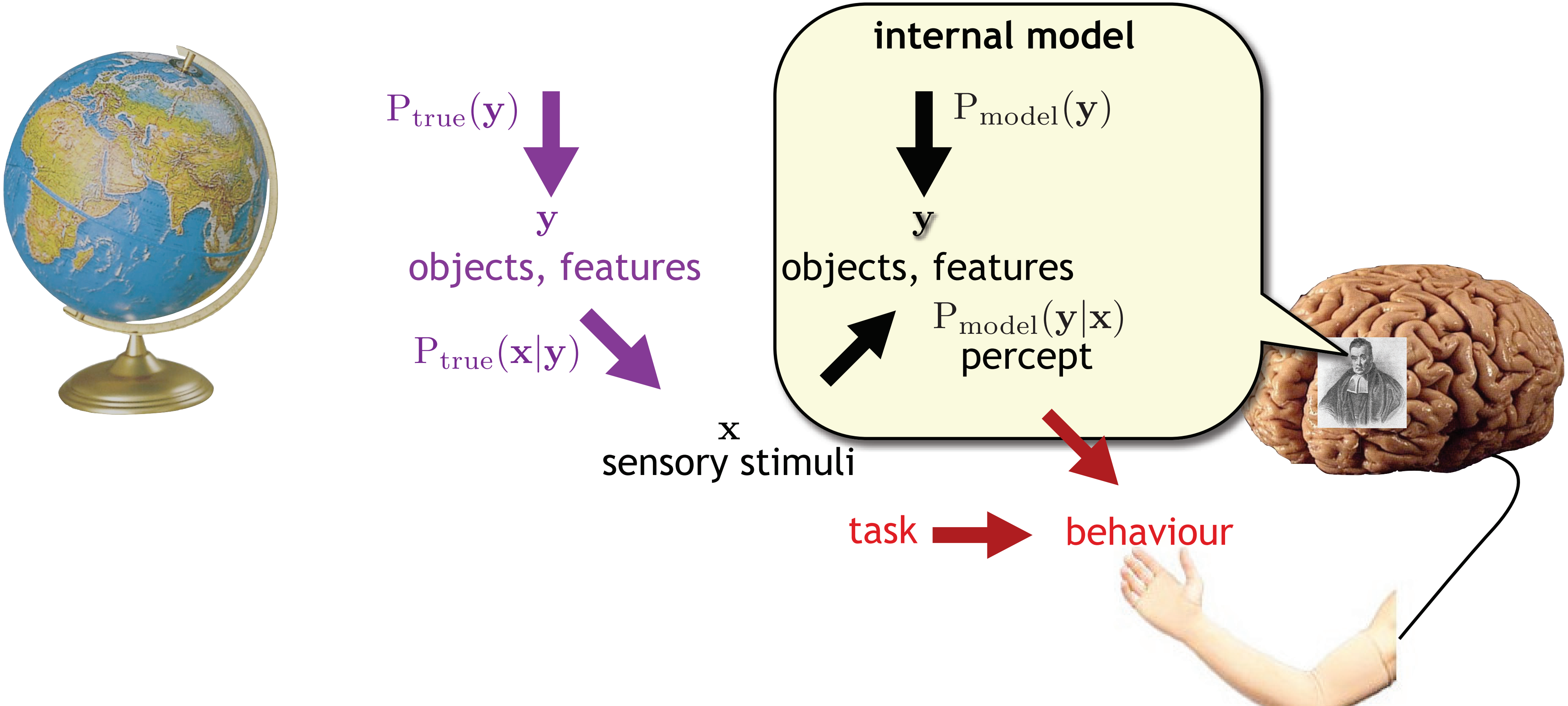
PROBABILISTIC INTERNAL MODELS



part 1. ADAPTATION TO ENVIRONMENT

part 2. TASK-INDEPENDENCE

PROBABILISTIC INTERNAL MODELS

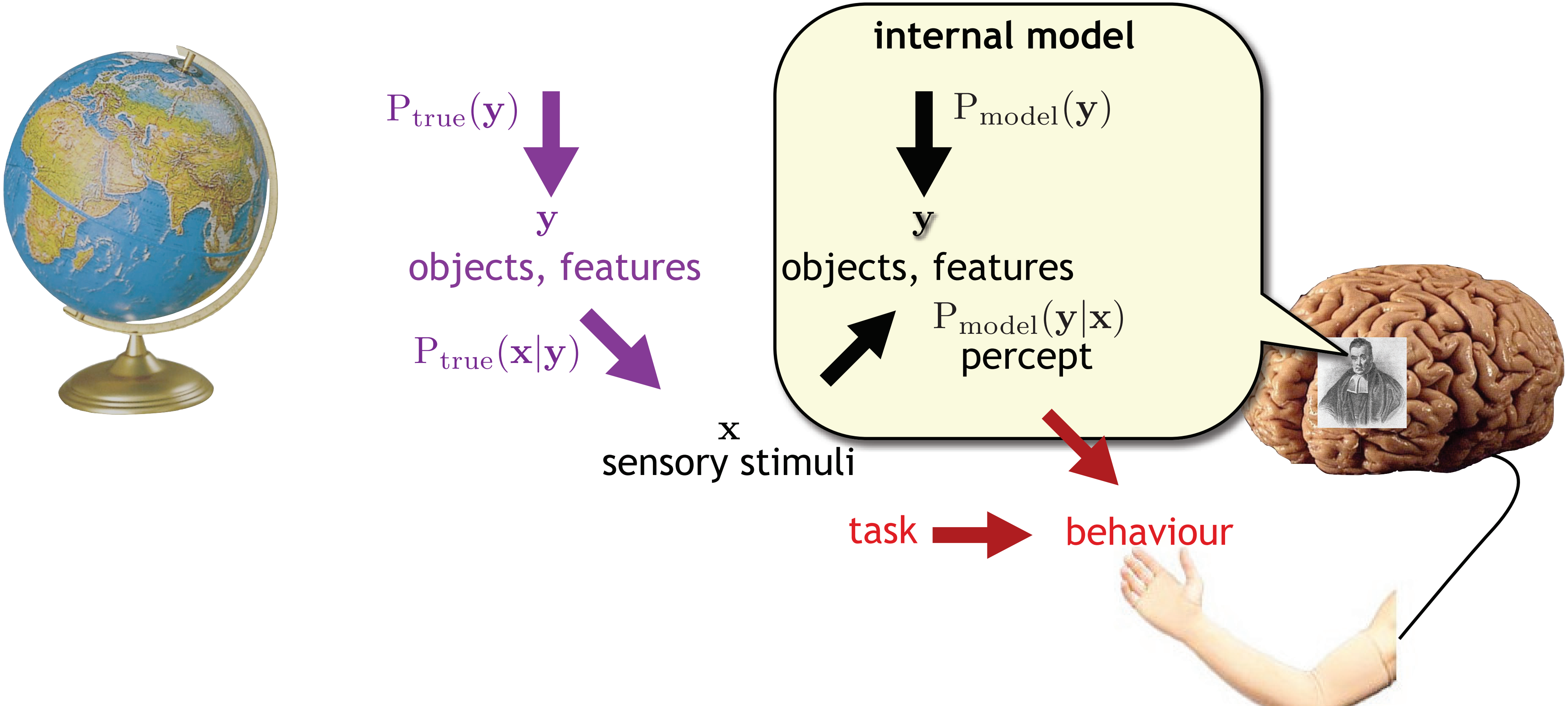


part 1. ADAPTATION TO ENVIRONMENT

part 2. TASK-INDEPENDENCE

issues with priors in single-task behavior:

PROBABILISTIC INTERNAL MODELS



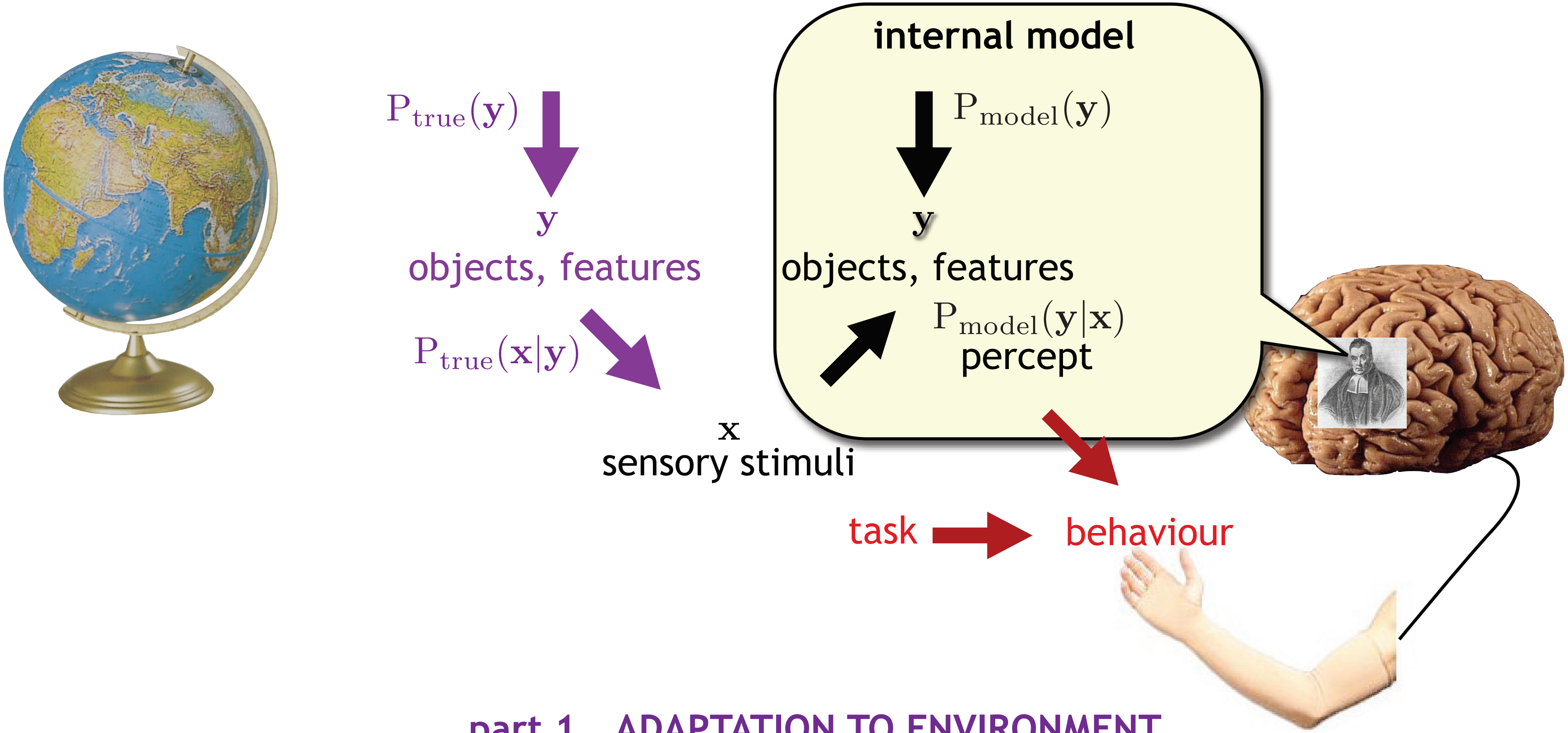
part 1. ADAPTATION TO ENVIRONMENT

part 2. TASK-INDEPENDENCE

issues with priors in single-task behavior:

- bunch of free parameters?

PROBABILISTIC INTERNAL MODELS



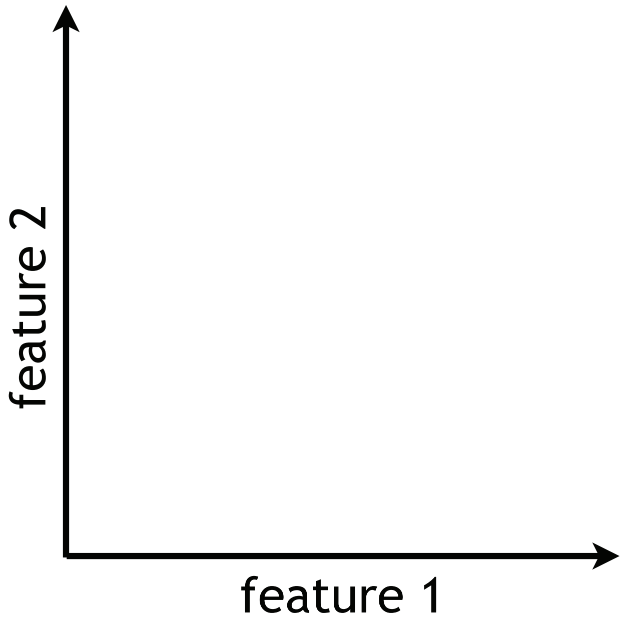
part 1. ADAPTATION TO ENVIRONMENT

part 2. TASK-INDEPENDENCE

issues with priors in single-task behavior:
- bunch of free parameters?
- generalization across tasks?

DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE

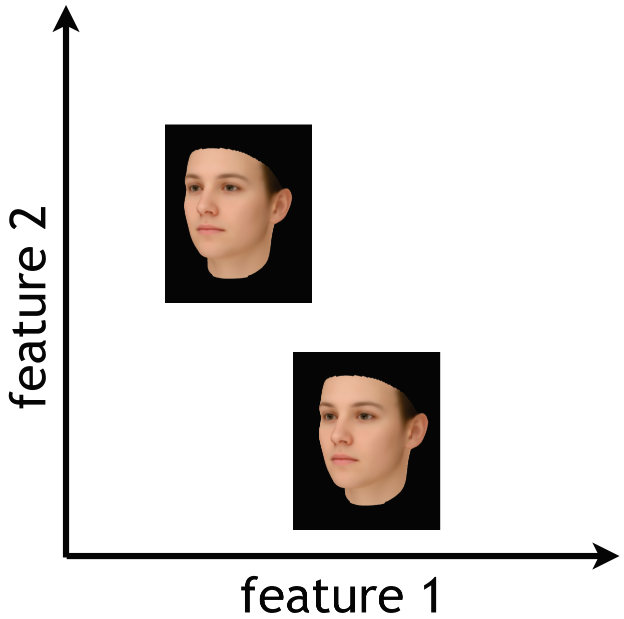


DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE

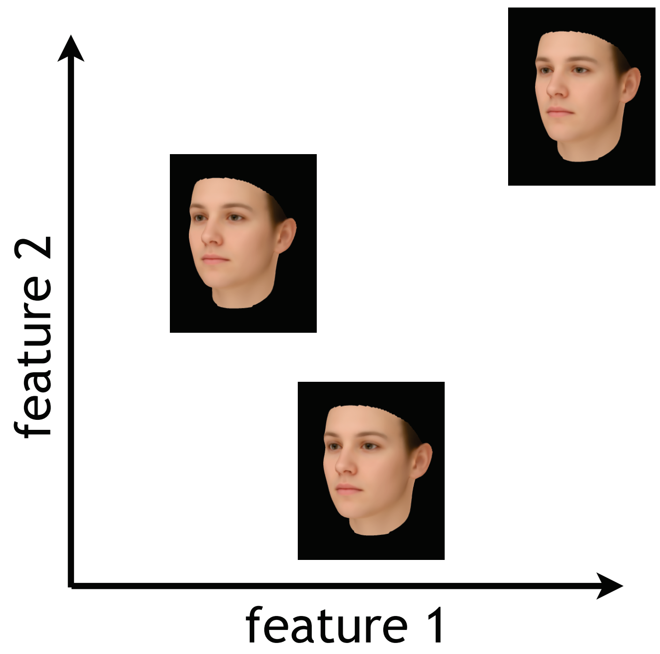
DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE



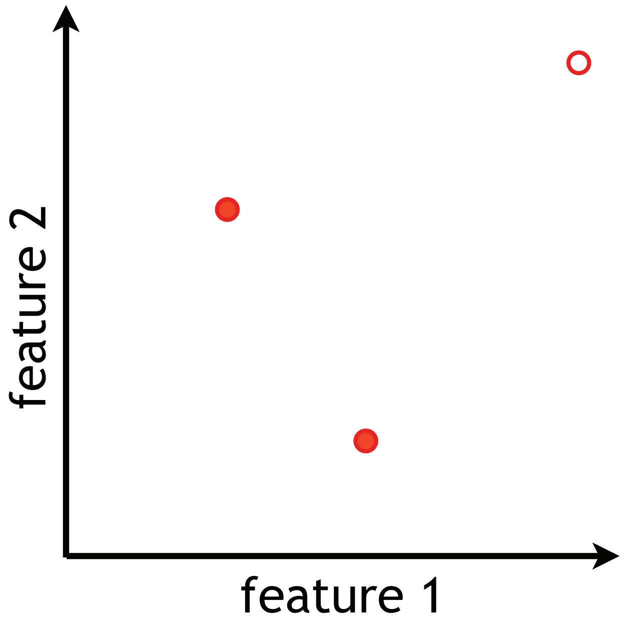
DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE



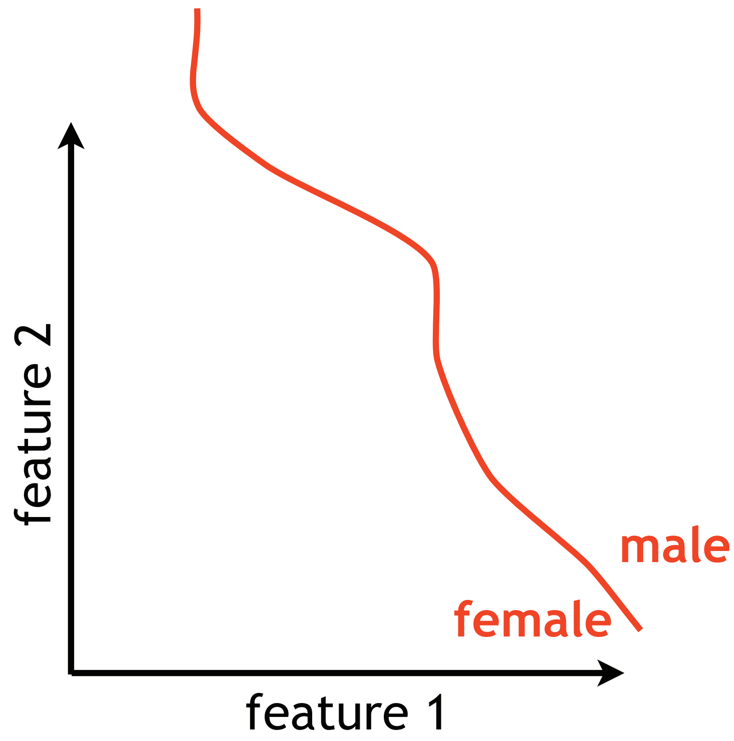
DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE



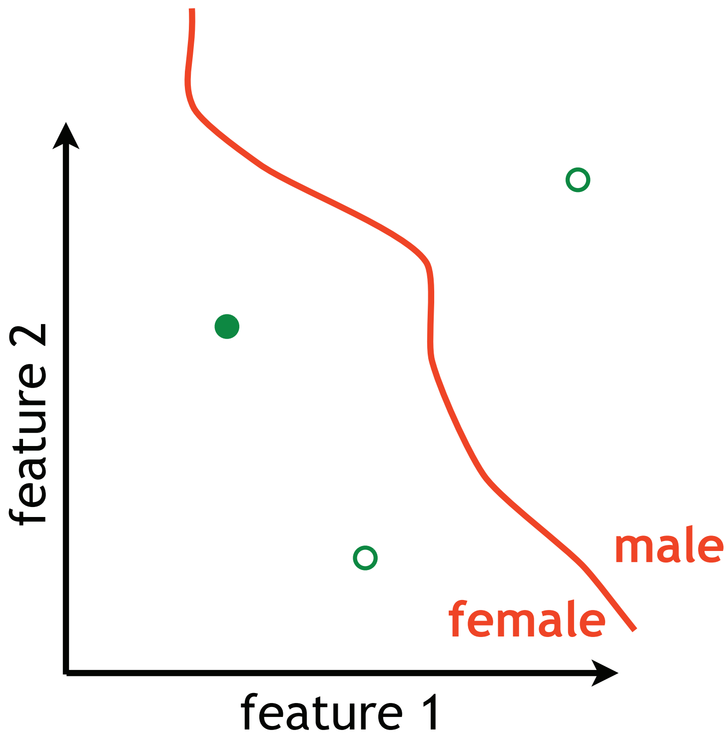
DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE



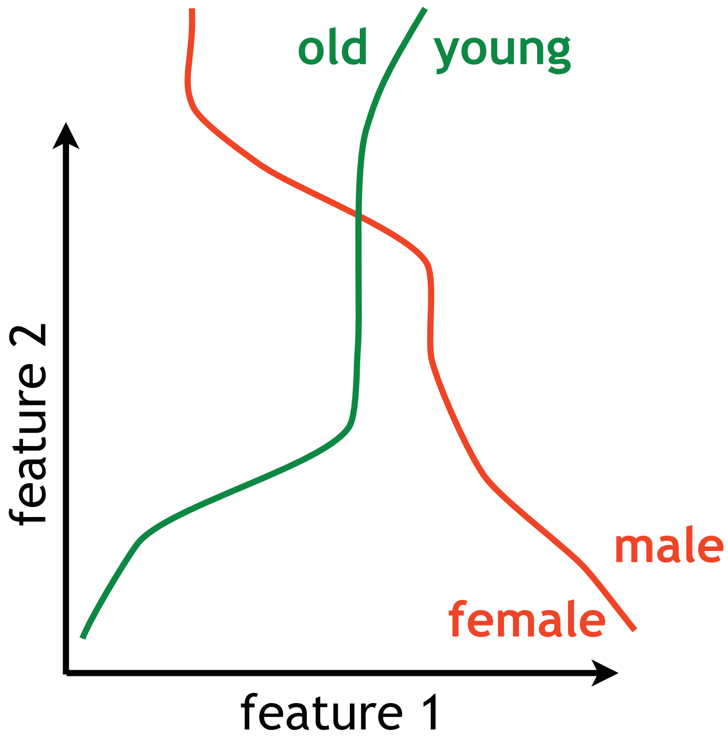
DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE



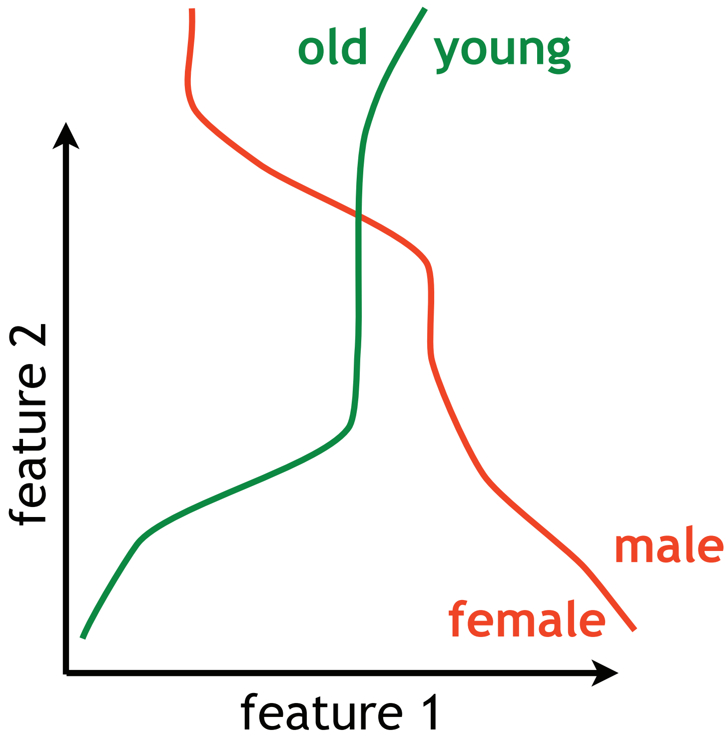
DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE

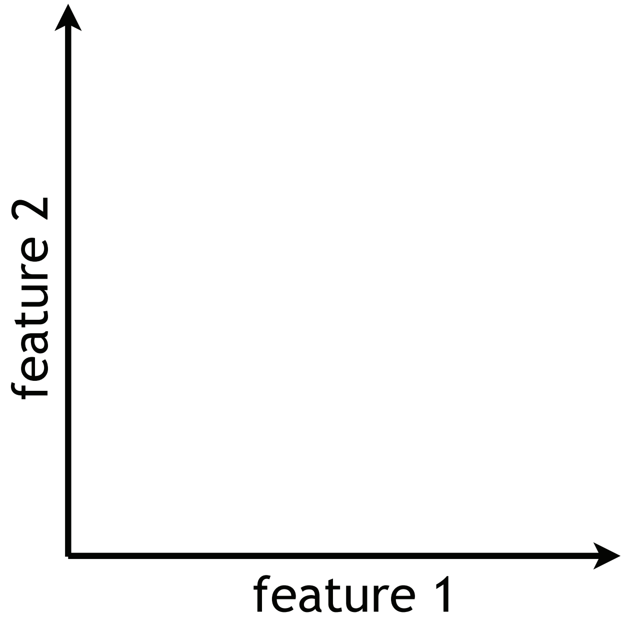


DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE

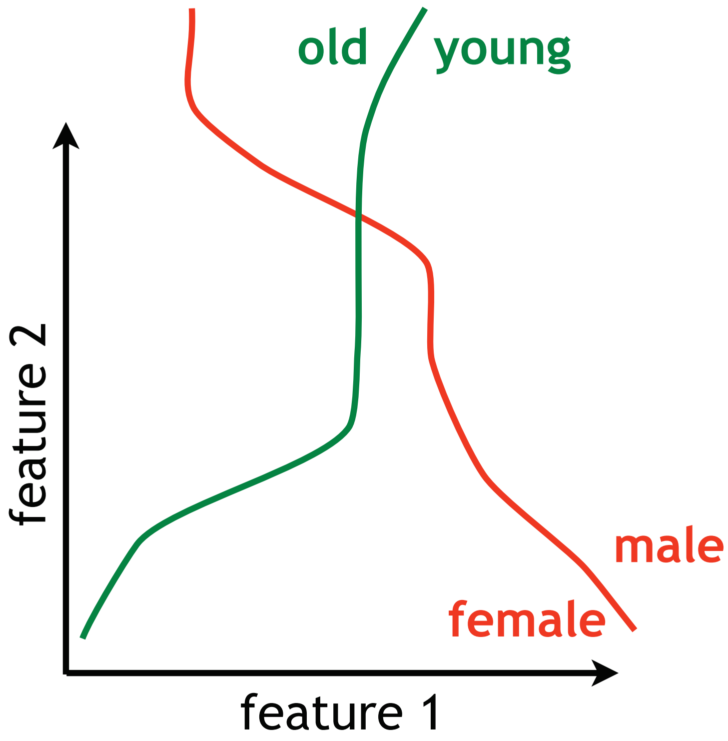


GENERATIVE

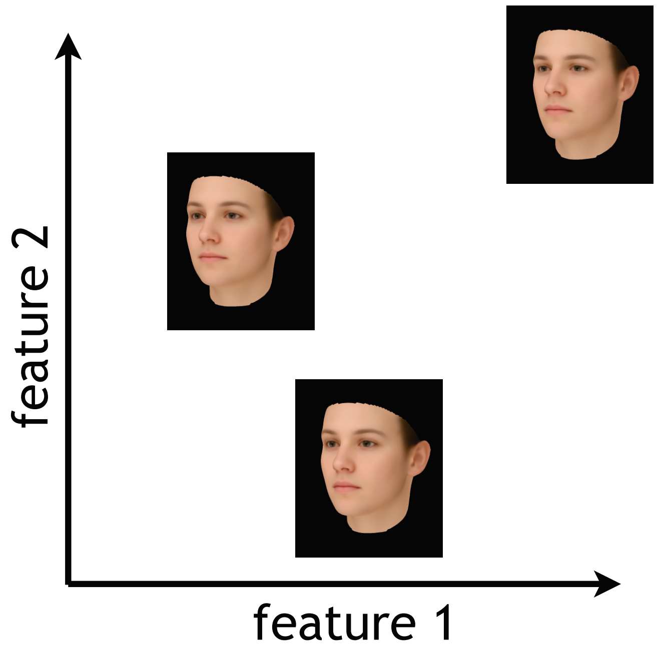


DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE

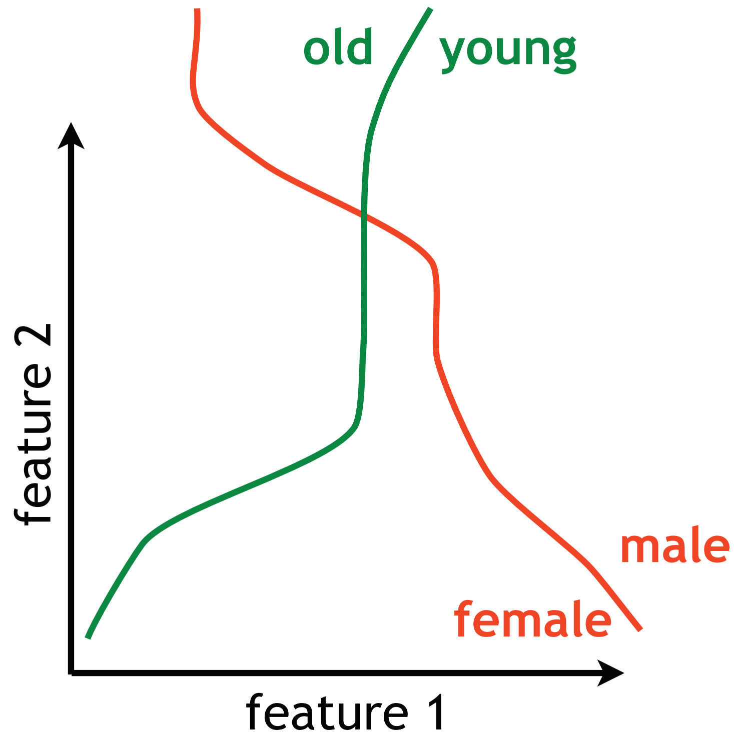


GENERATIVE

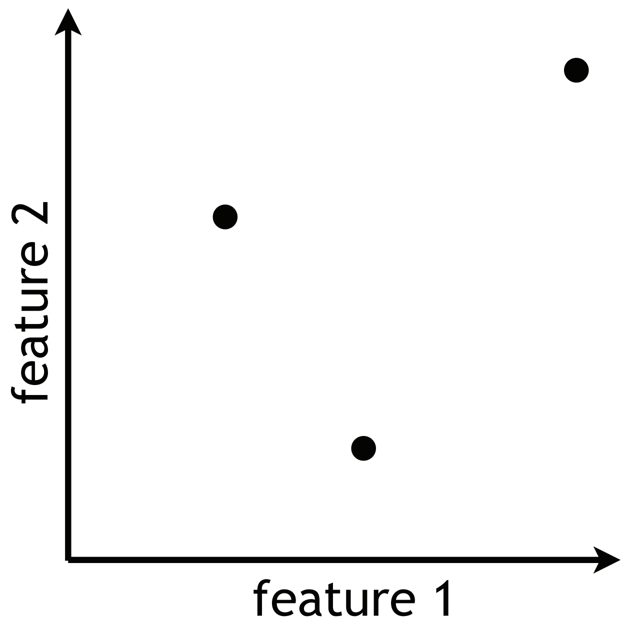


DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE

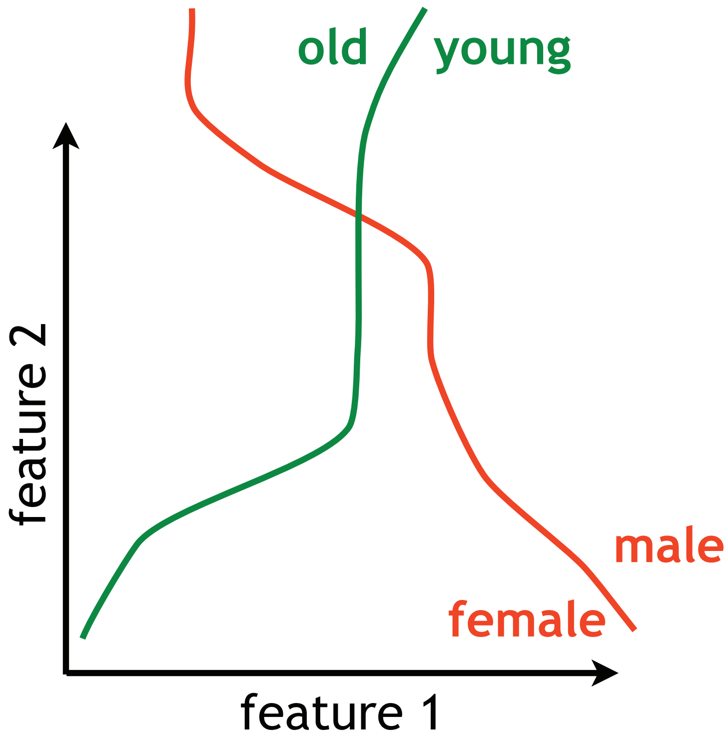


GENERATIVE

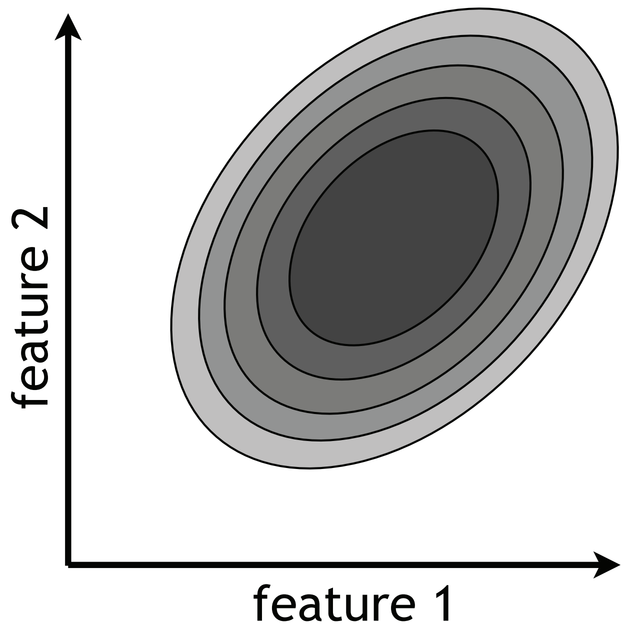


DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE



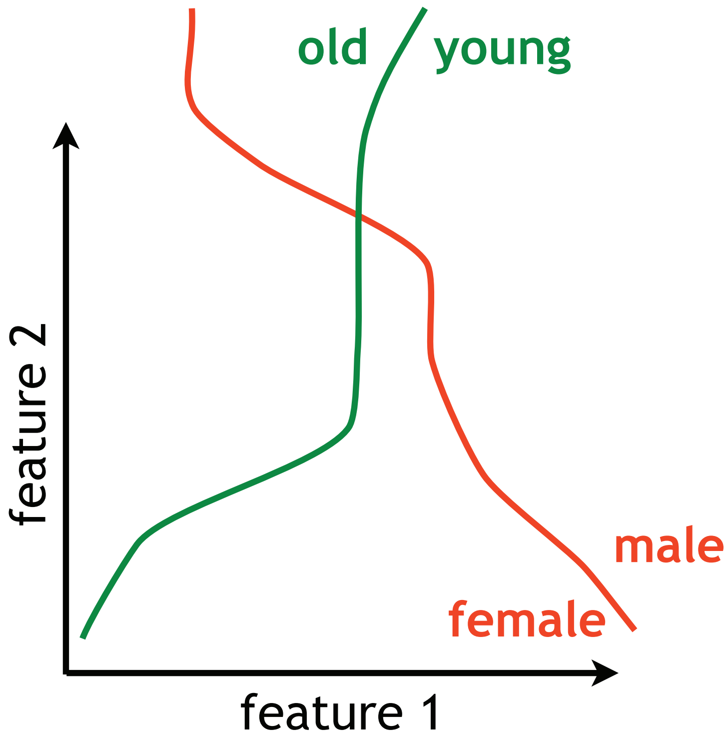
GENERATIVE



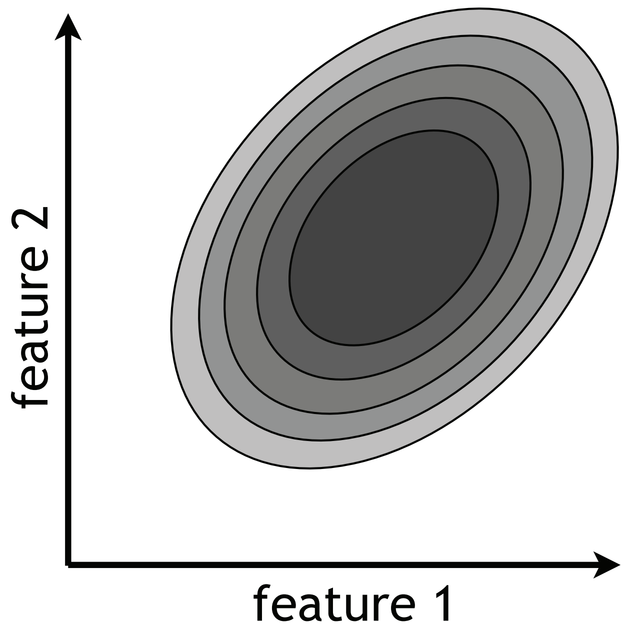
DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE

task-dependent representation

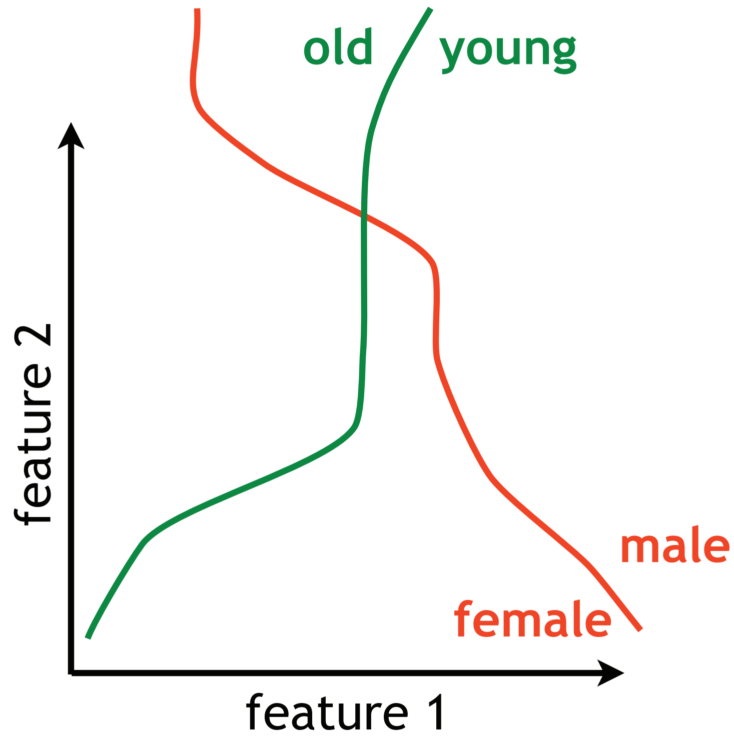


GENERATIVE

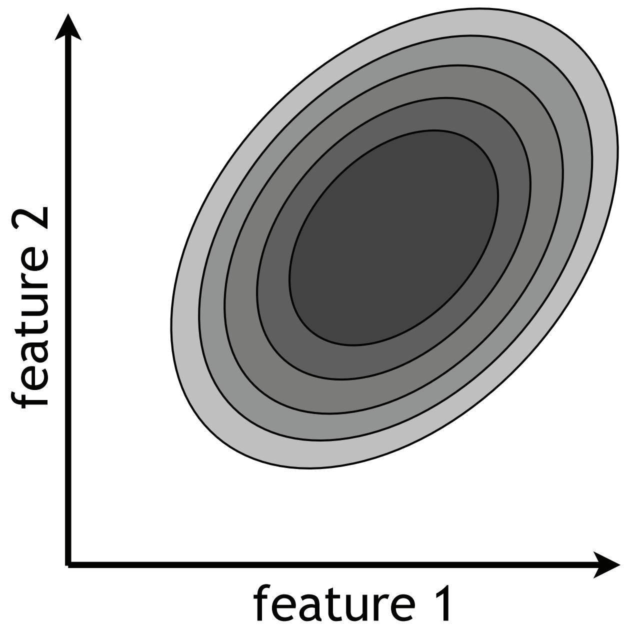


DISCRIMINATIVE VS. GENERATIVE LEARNING

DISCRIMINATIVE
task-dependent representation



GENERATIVE
task-independent representation



TWO TASKS

Neil
Houlsby

Ferenc
Huszár

Mohammad
Ghassemi

Gergő
Orbán

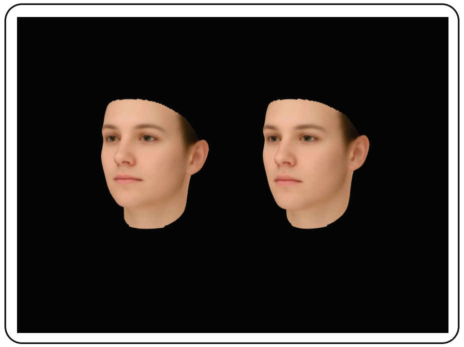
Daniel
Wolpert

TWO TASKS

Neil Houlsby   Ferenc Huszár   Mohammad Ghassemi   Gergő Orbán   Daniel Wolpert

familiarity task

"which face looks more familiar?"



TWO TASKS


Neil Houlsby

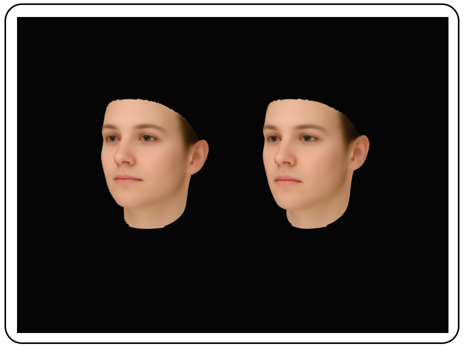

Ferenc Huszár


Mohammad Ghassemi

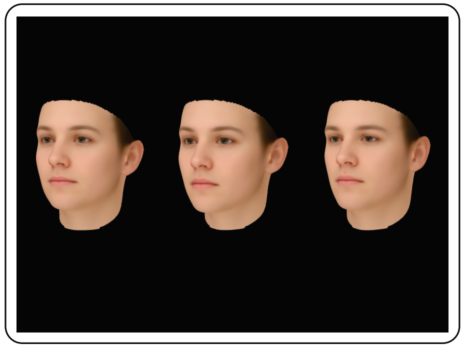

Gergő Orbán


Daniel Wolpert

familiarity task
"which face looks more familiar?"



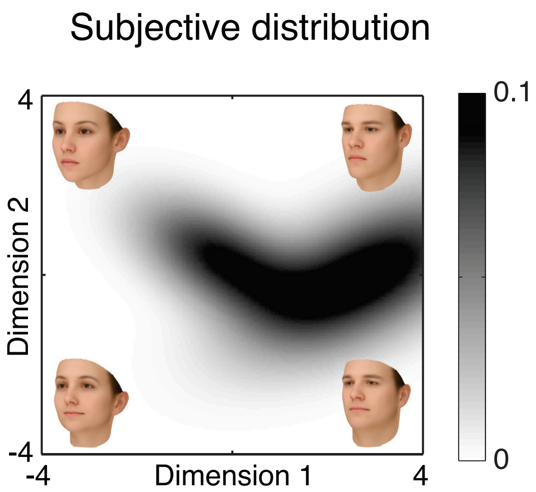
odd-one-out task
"which face is the odd-one-out?"



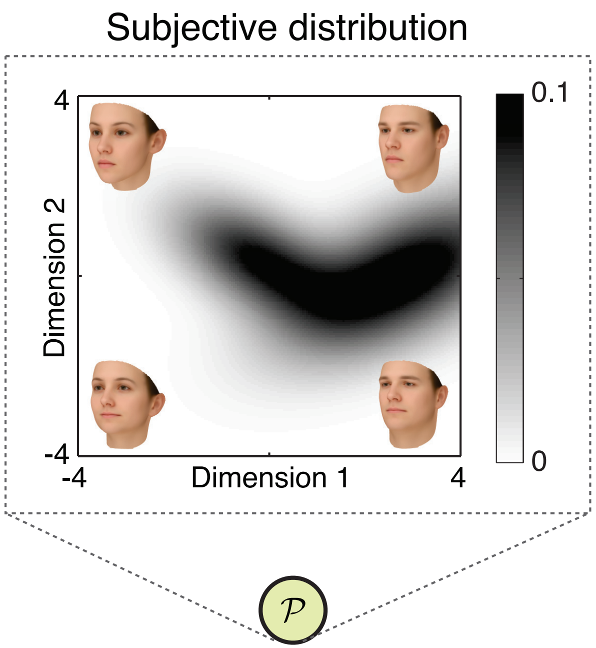
Houlsby et al, Curr Biol 2013

COGNITIVE TOMOGRAPHY

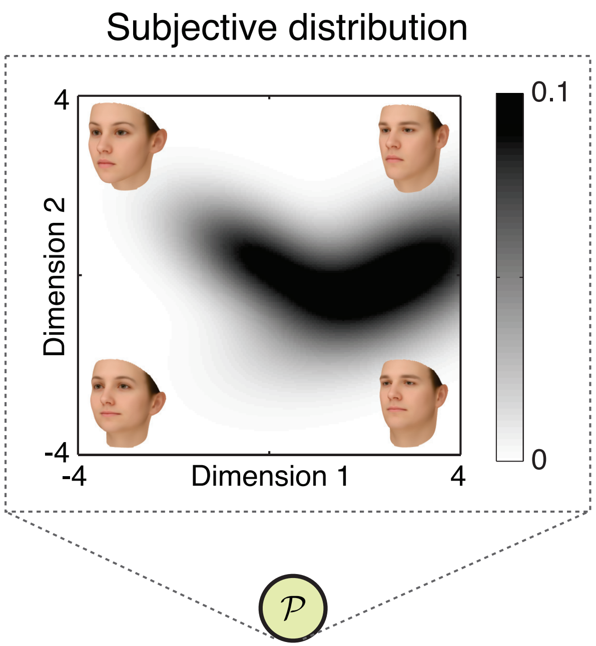
COGNITIVE TOMOGRAPHY

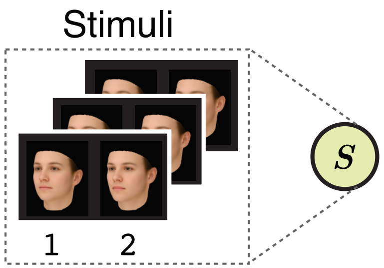


COGNITIVE TOMOGRAPHY

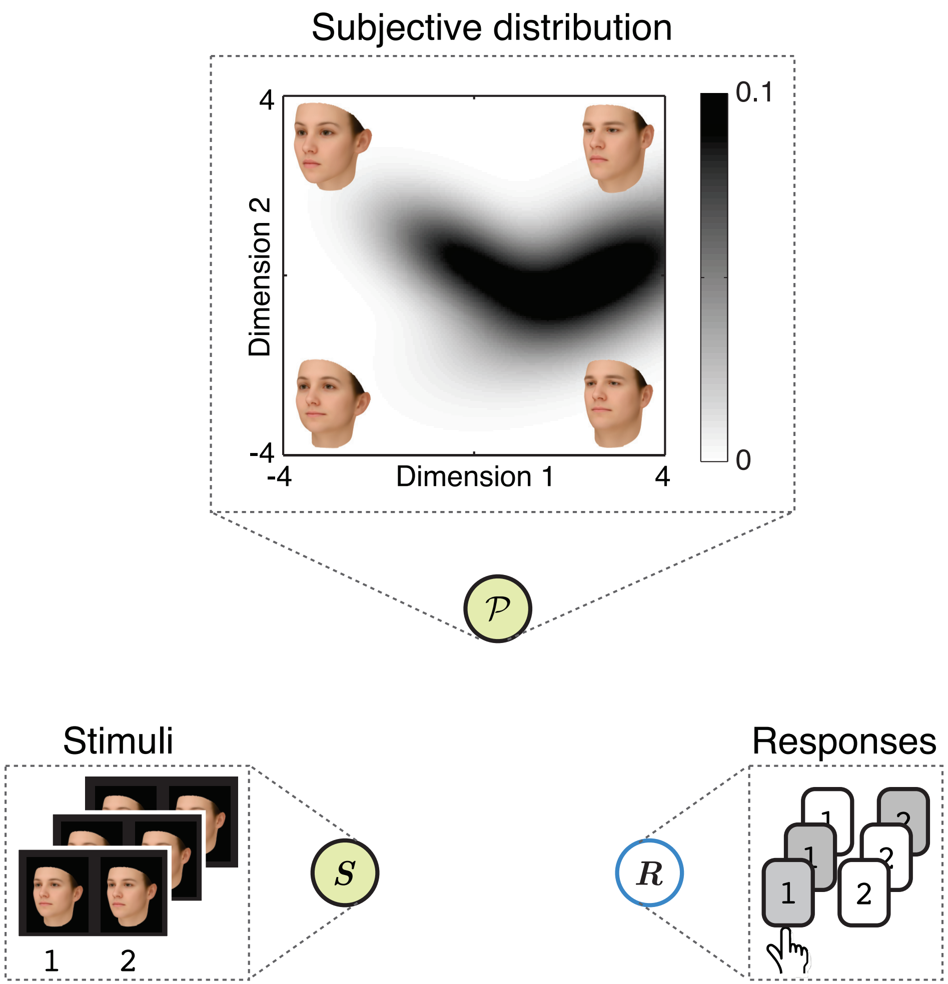


COGNITIVE TOMOGRAPHY



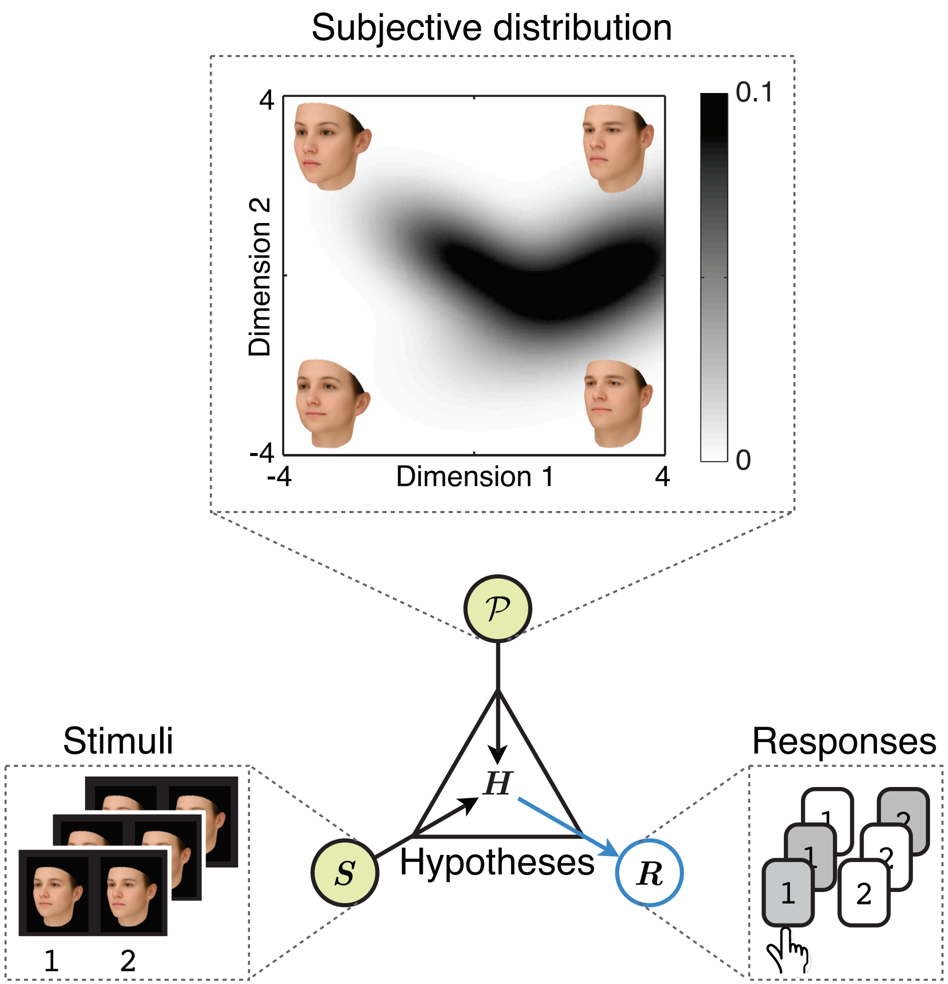


COGNITIVE TOMOGRAPHY



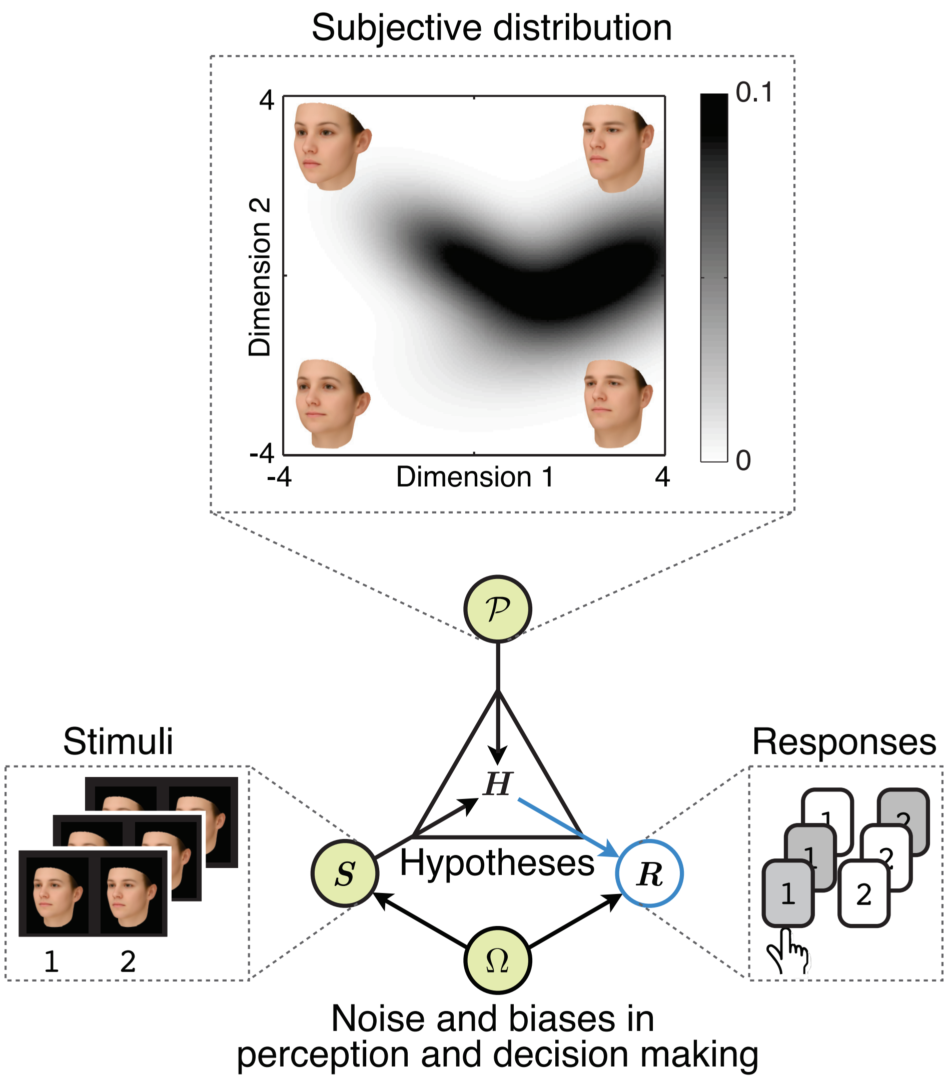
COGNITIVE TOMOGRAPHY

ideal observer model



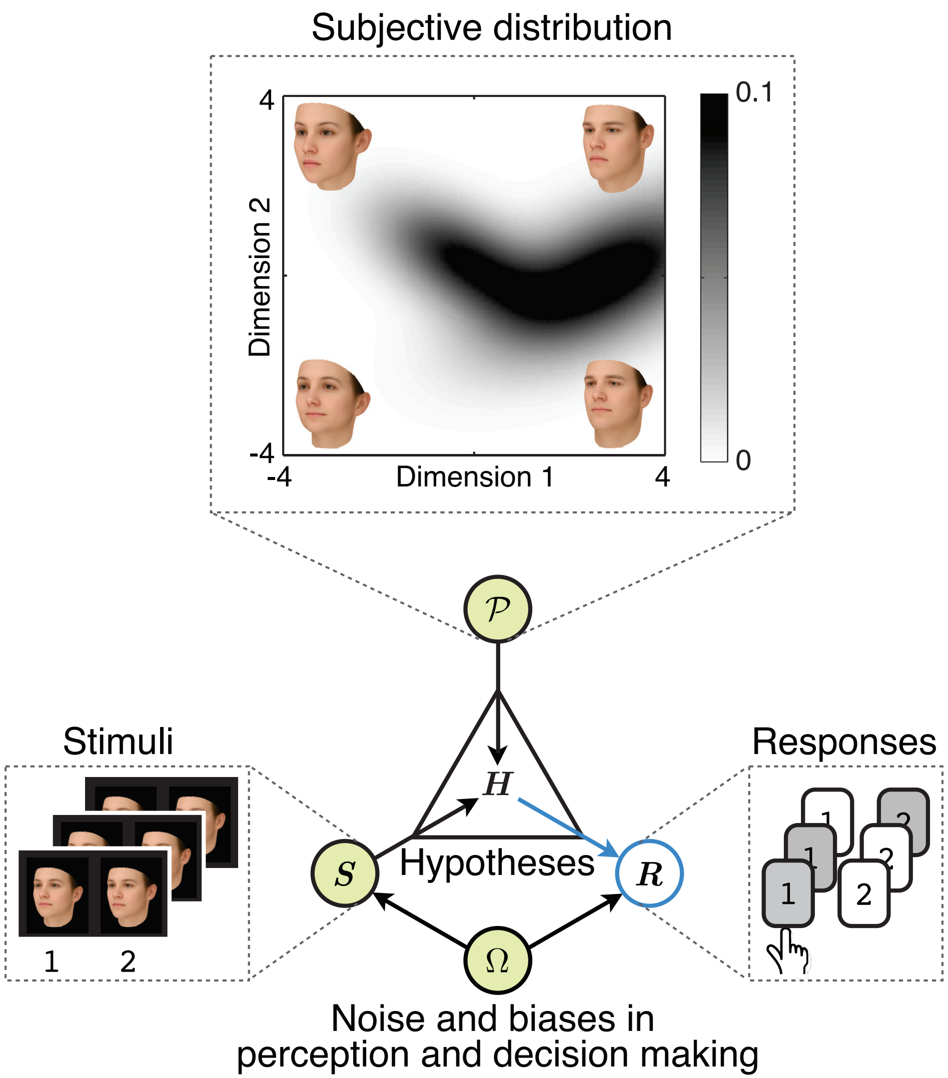
COGNITIVE TOMOGRAPHY

quasi- ideal observer model

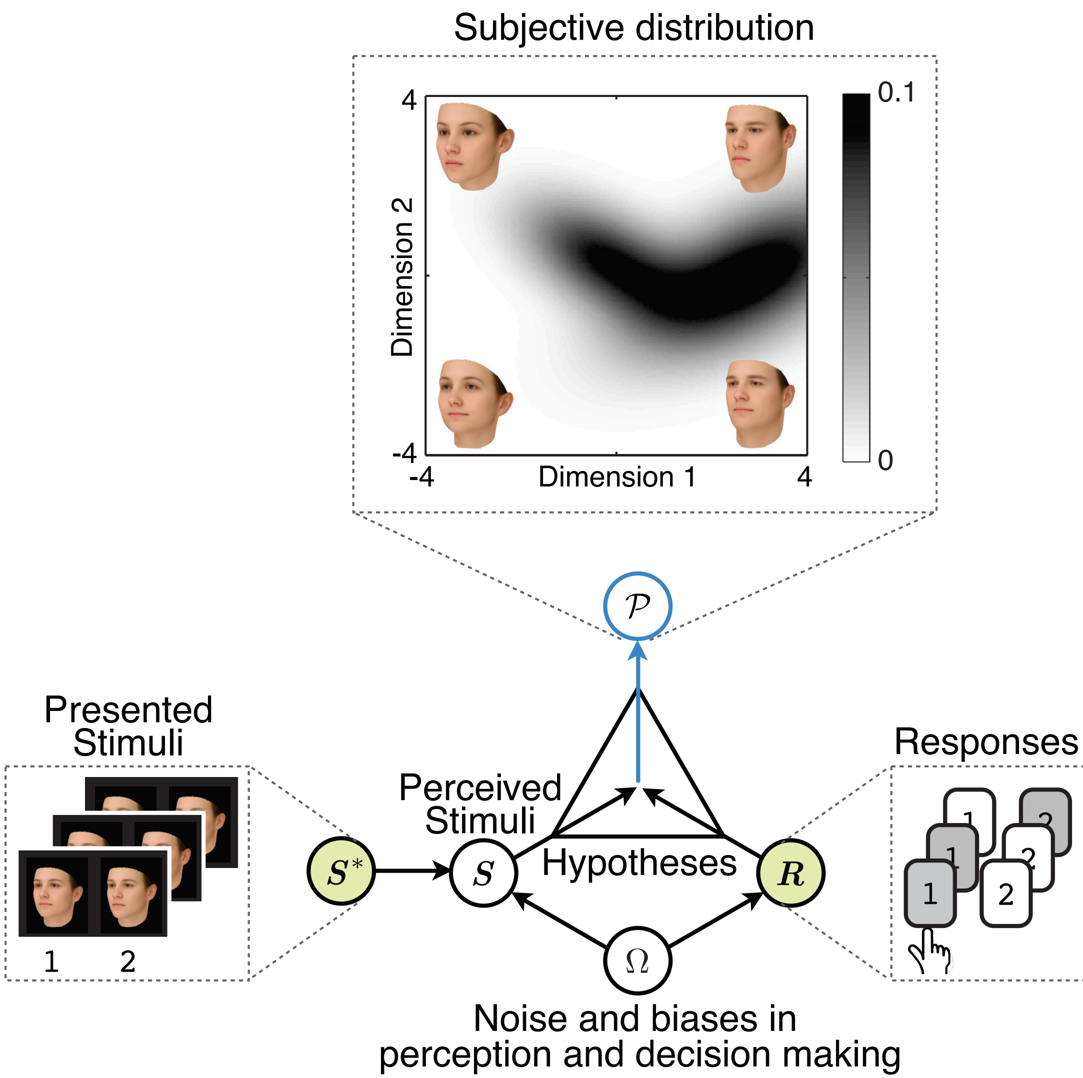


COGNITIVE TOMOGRAPHY

quasi- ideal observer model



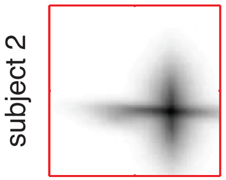
cognitive tomography
inverting the ideal observer model



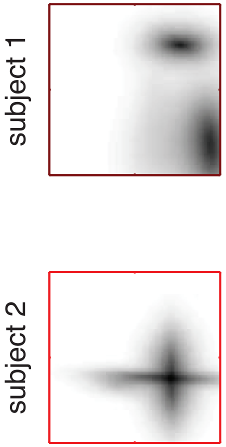
Houlsby et al, Curr Biol 2013

SUBJECTIVE DISTRIBUTIONS

SUBJECTIVE DISTRIBUTIONS

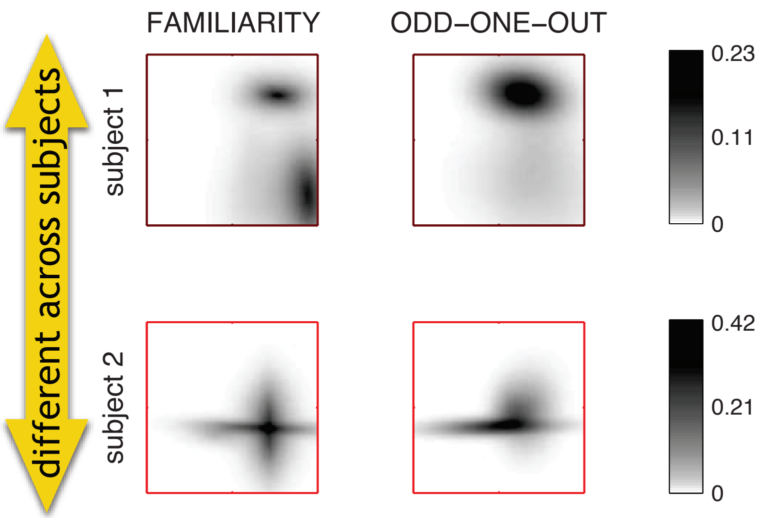


SUBJECTIVE DISTRIBUTIONS

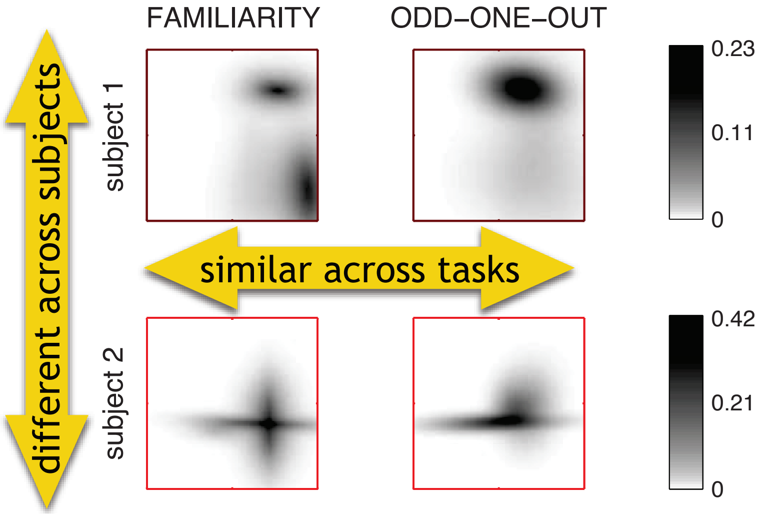


SUBJECTIVE DISTRIBUTIONS

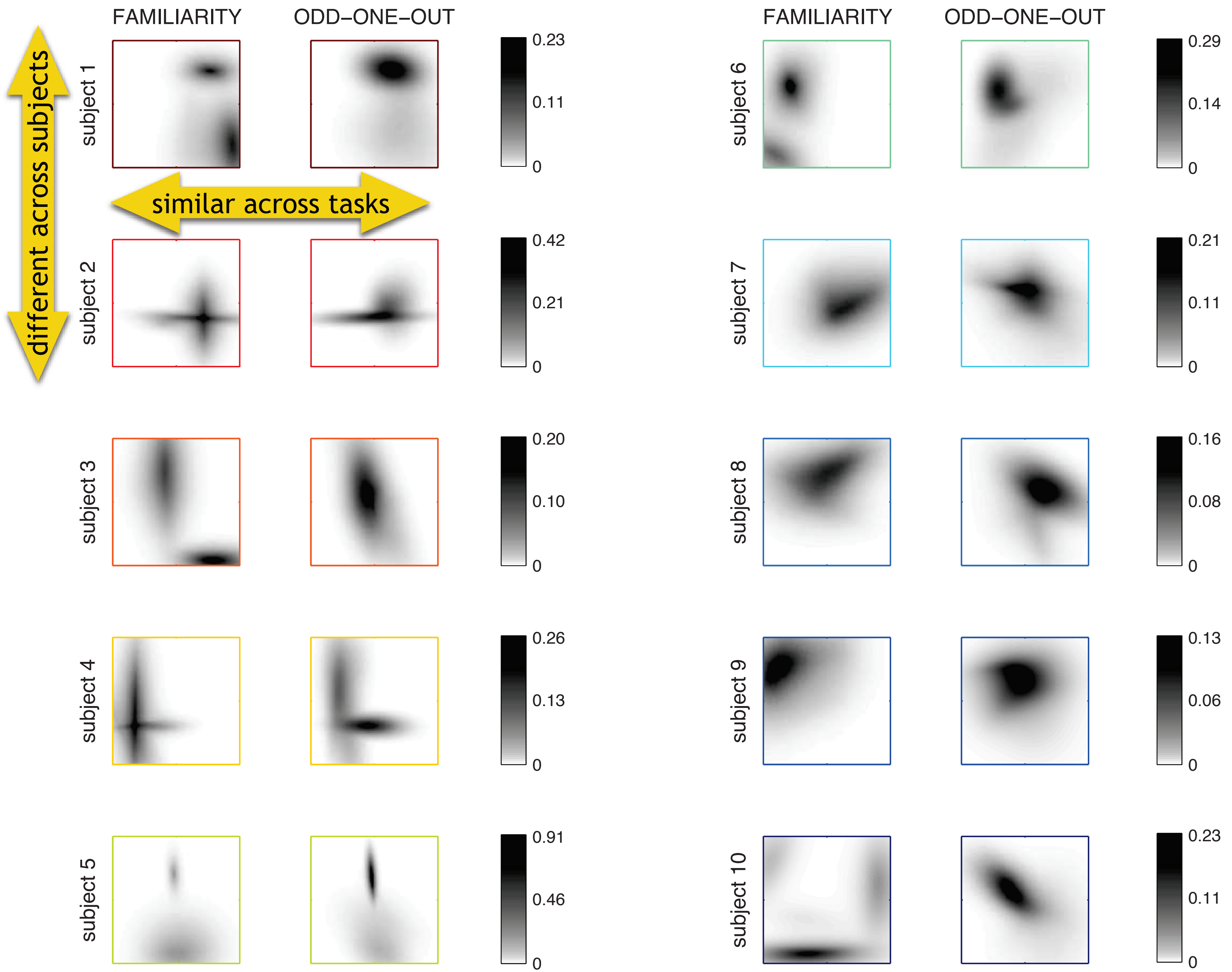
SUBJECTIVE DISTRIBUTIONS



SUBJECTIVE DISTRIBUTIONS



SUBJECTIVE DISTRIBUTIONS

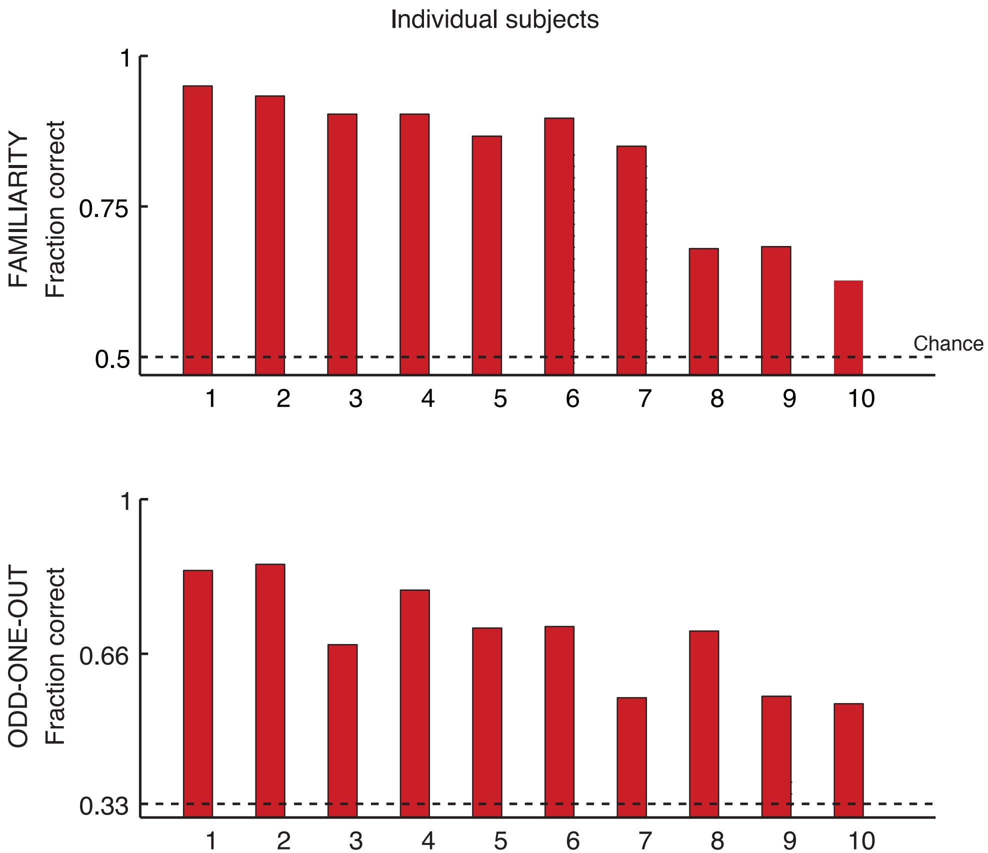


Houlsby et al, Curr Biol 2013

PREDICTING BEHAVIOUR

PREDICTING BEHAVIOUR

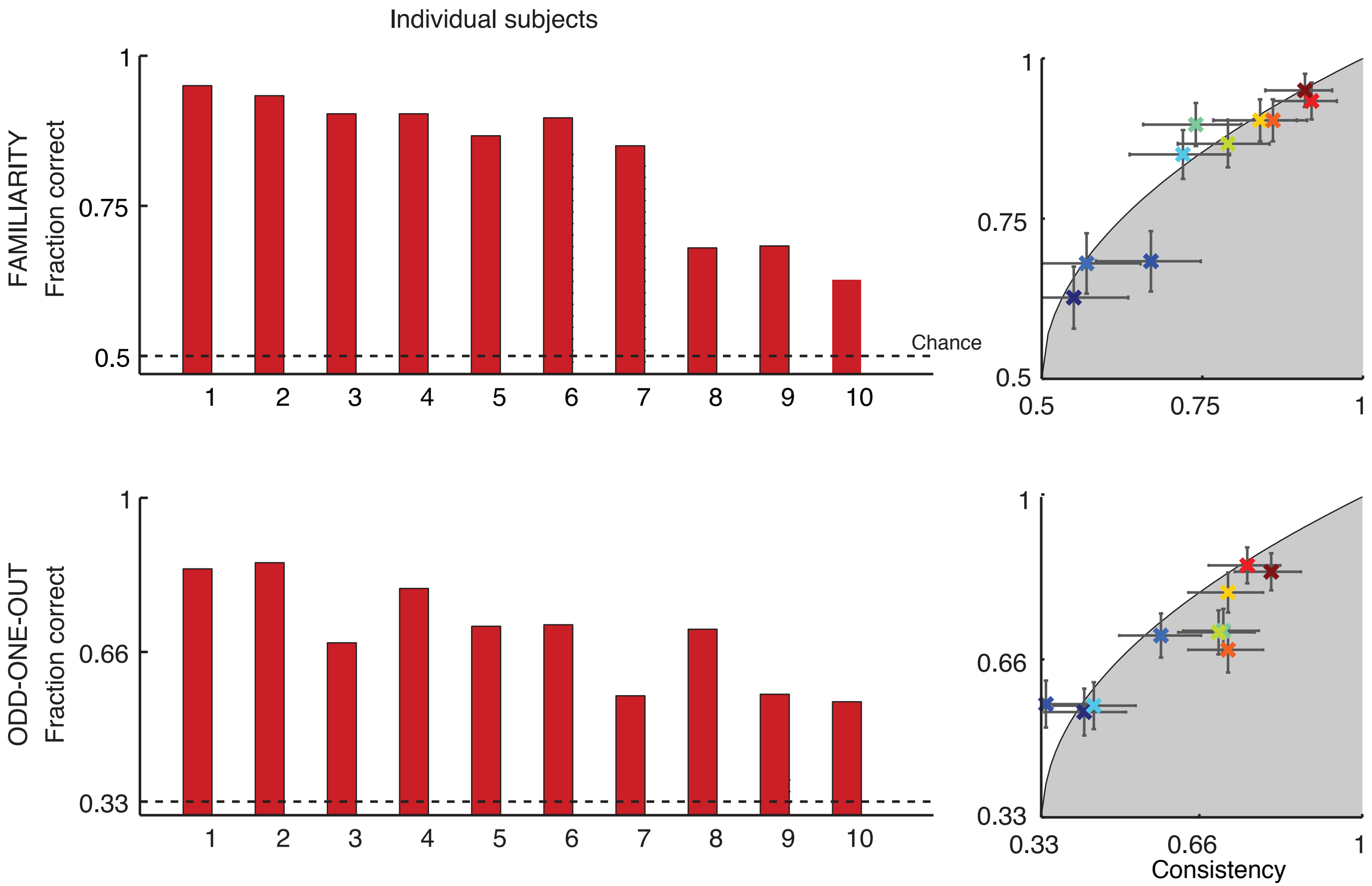
well above chance



PREDICTING BEHAVIOUR

well above chance

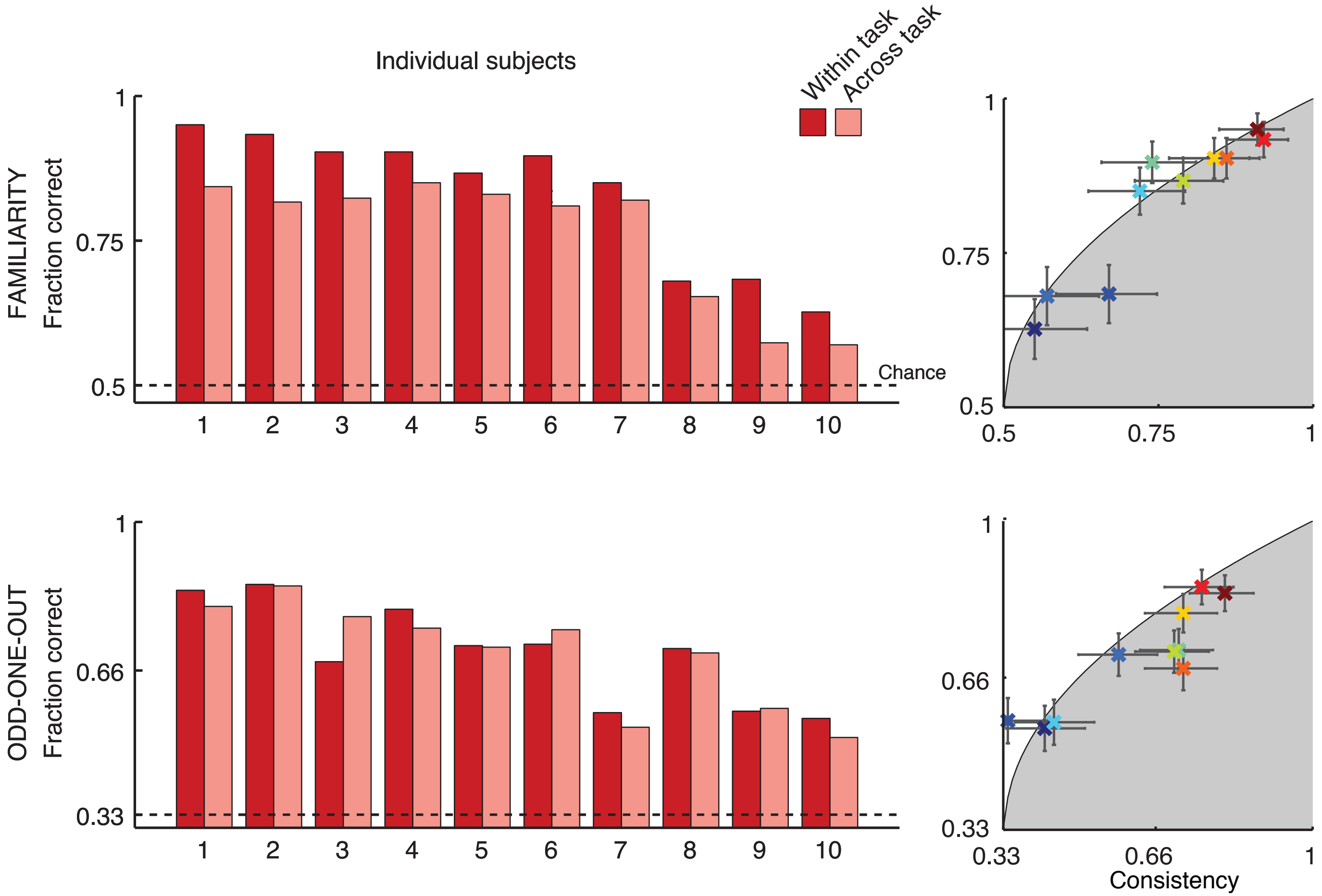
close to theoretical upper bound



PREDICTING BEHAVIOUR

well above chance both within and across tasks

close to theoretical upper bound

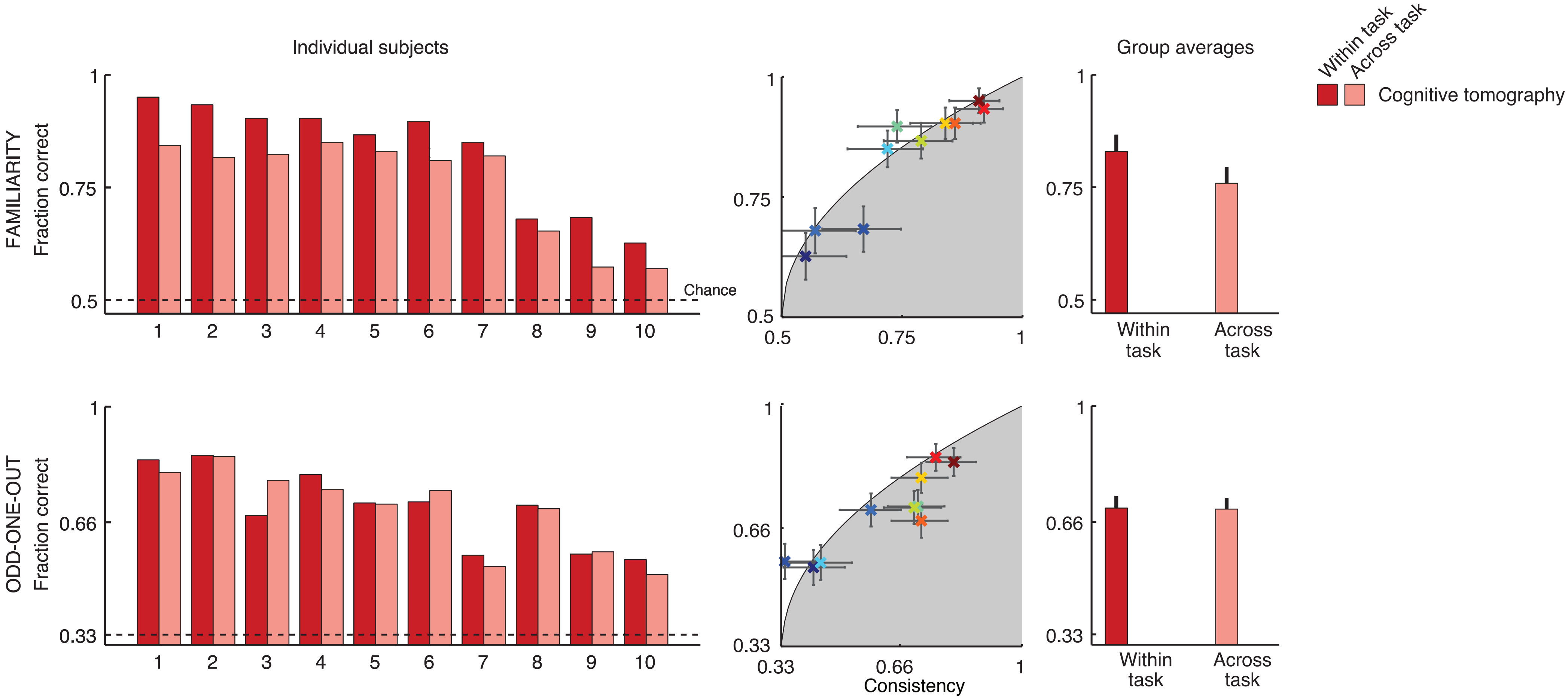


PREDICTING BEHAVIOUR

well above chance both within and across tasks

close to theoretical upper bound

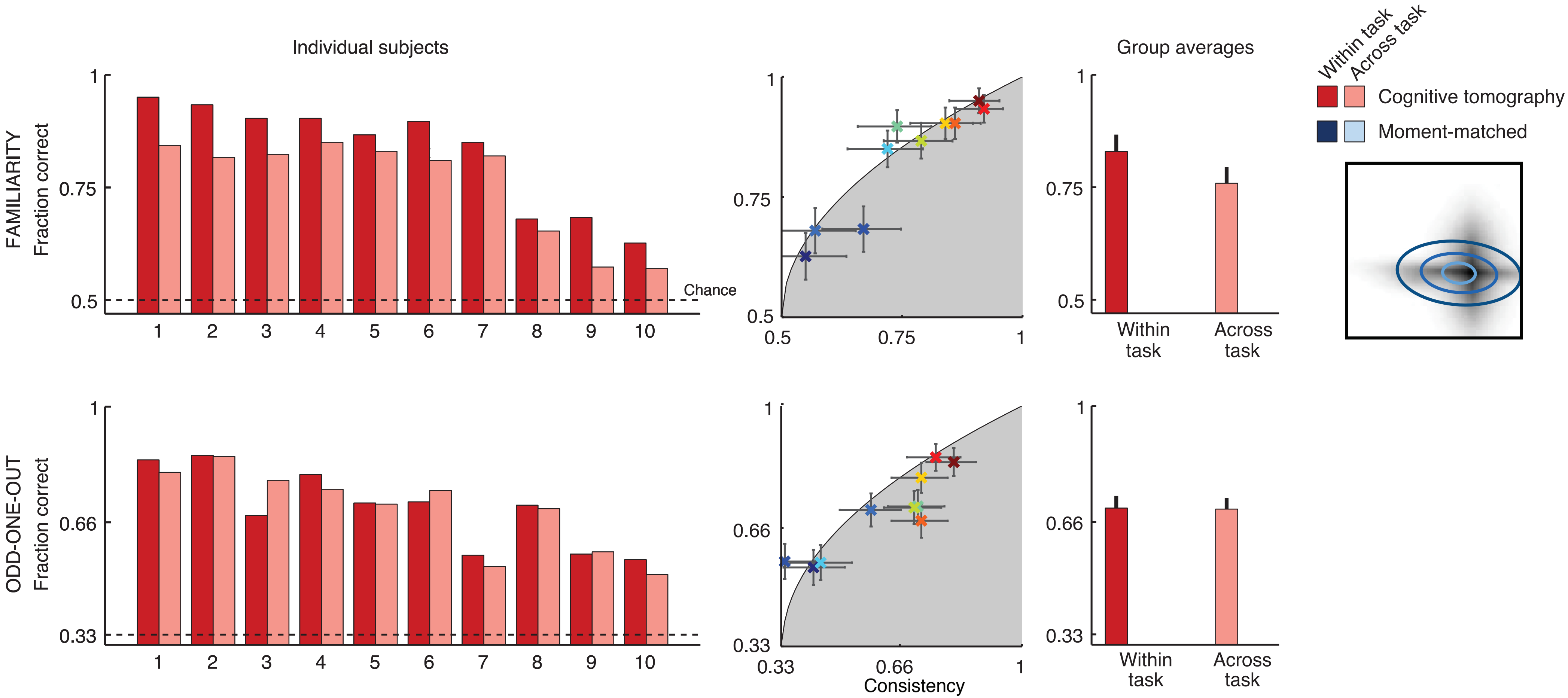
better than controls



Houlsby et al, Curr Biol 2013

PREDICTING BEHAVIOUR

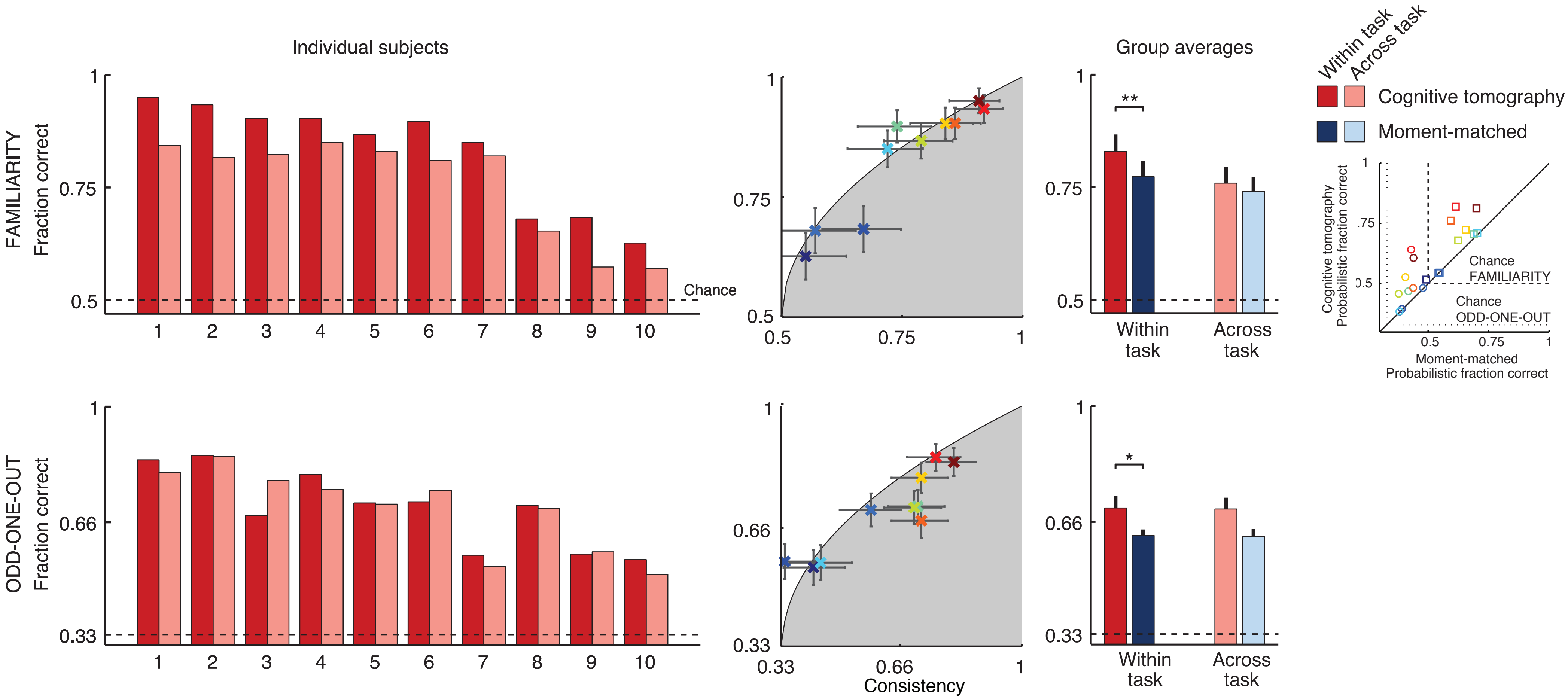
well above chance both within and across tasks | **close to theoretical upper bound** | **better than controls**



Houlsby et al, Curr Biol 2013

PREDICTING BEHAVIOUR

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



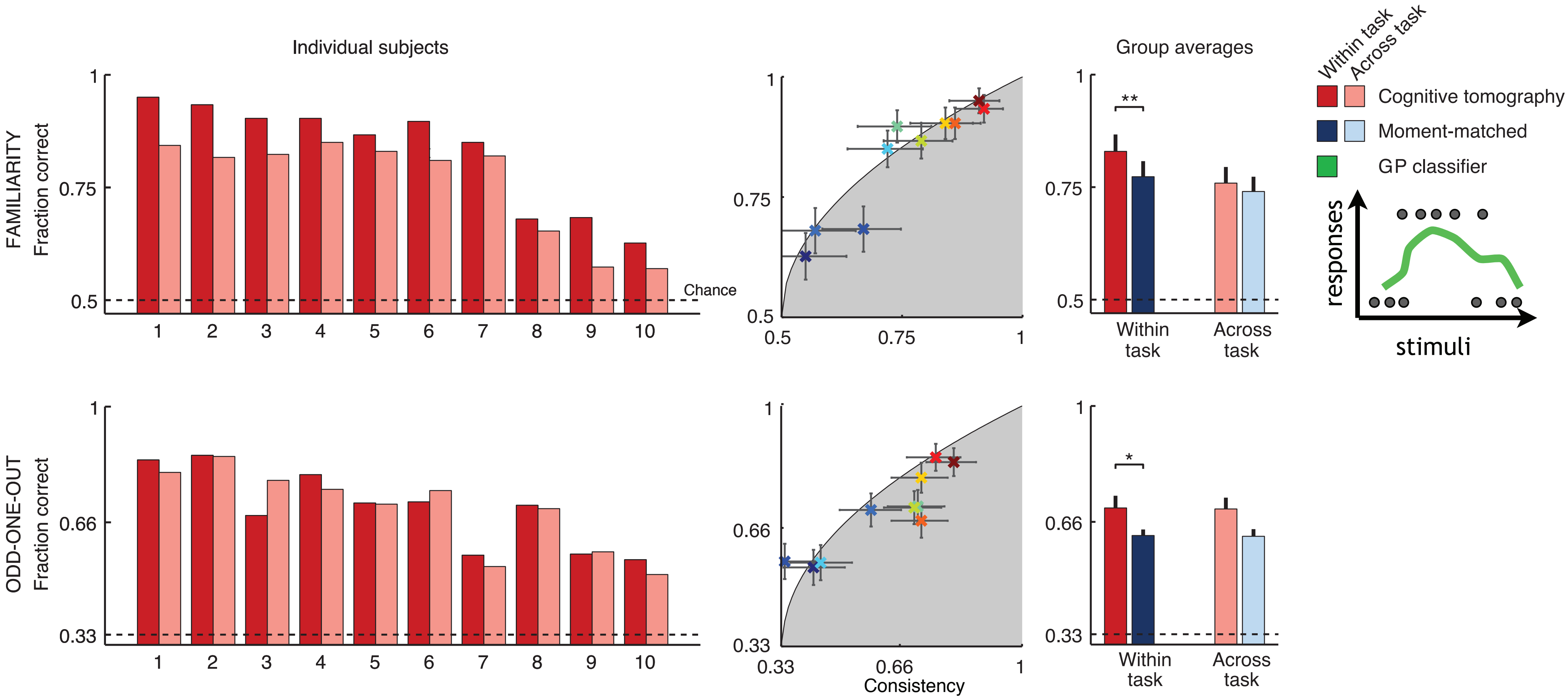
Houlsby et al, Curr Biol 2013

PREDICTING BEHAVIOUR

well above chance both within and across tasks

close to theoretical upper bound

better than controls



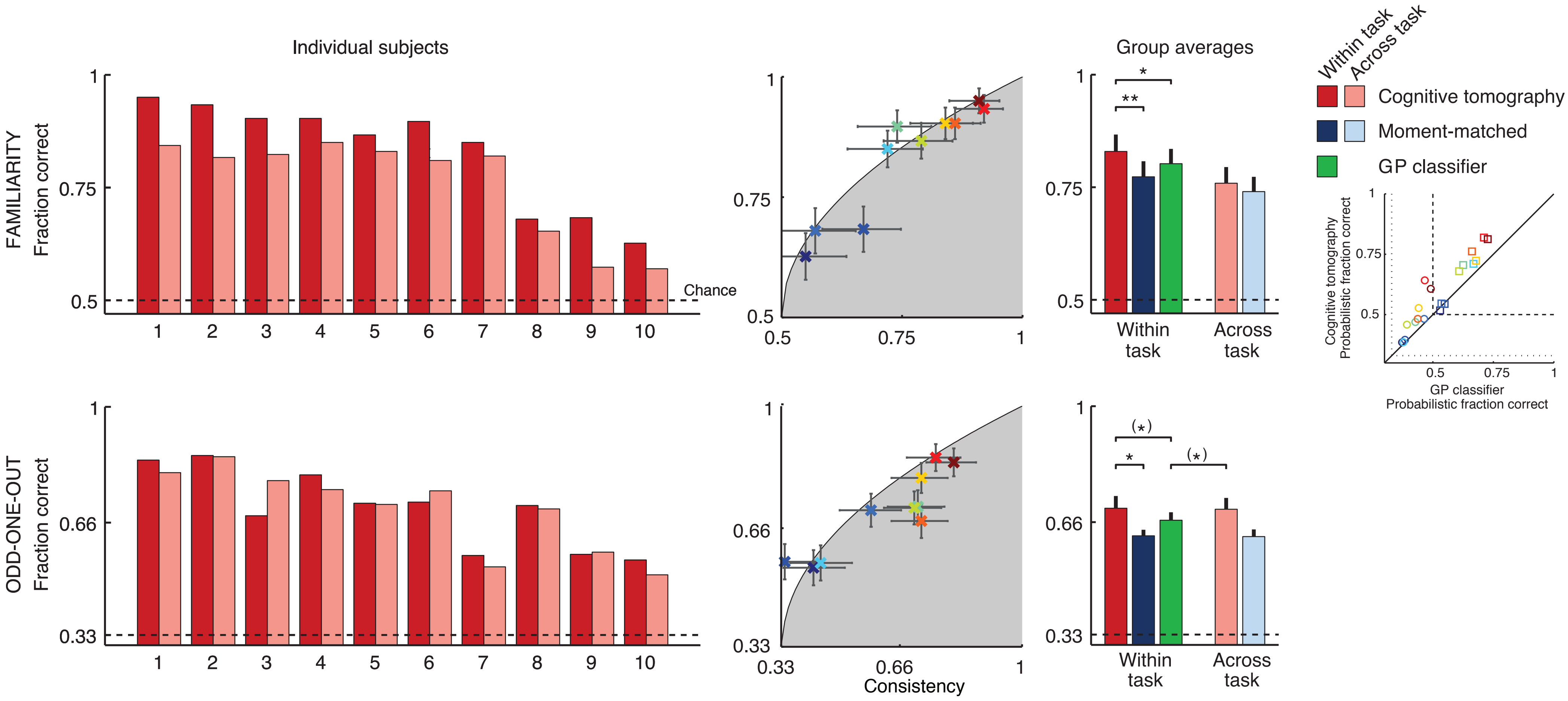
Houlsby et al, Curr Biol 2013

PREDICTING BEHAVIOUR

well above chance both within and across tasks

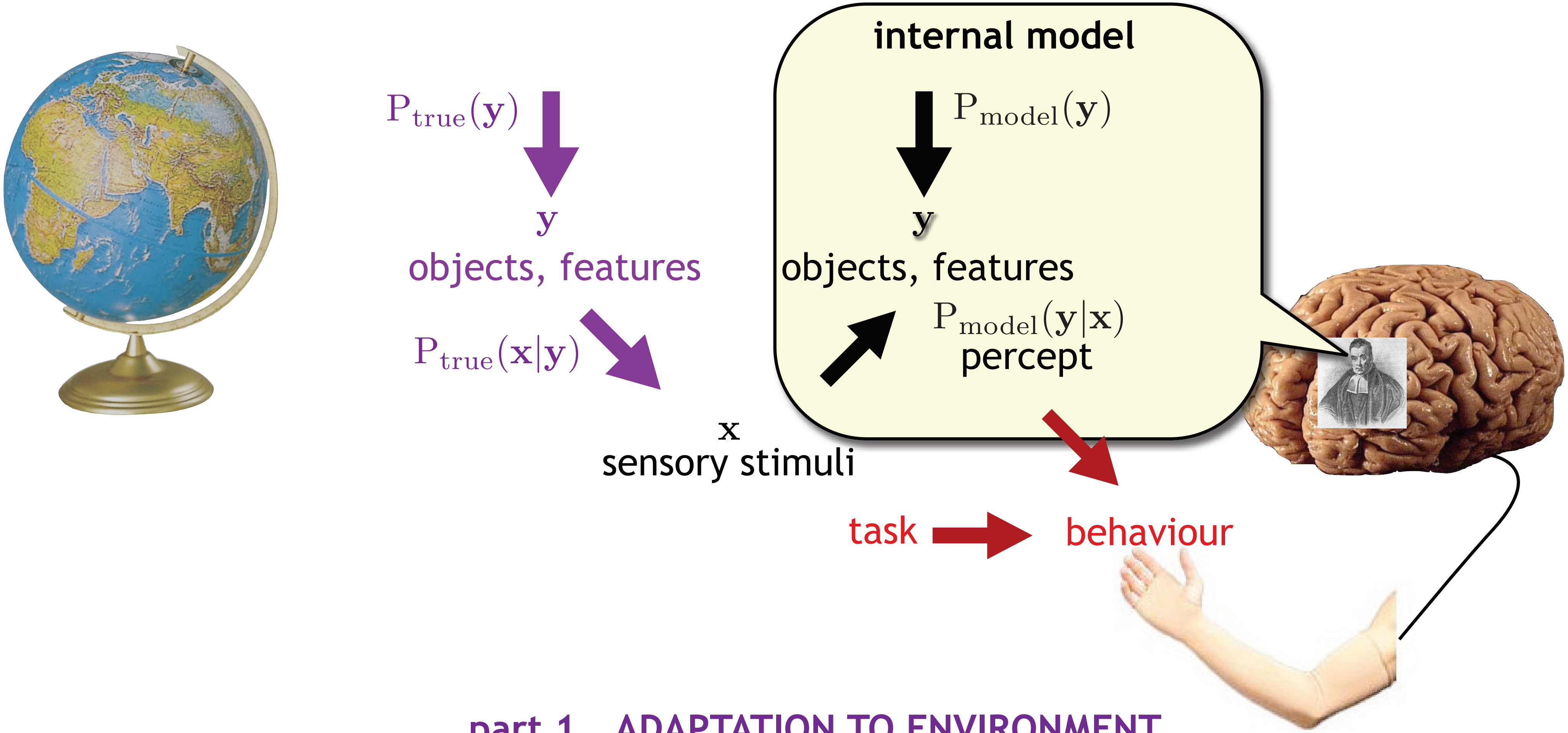
close to theoretical upper bound

better than controls



Houlsby et al, Curr Biol 2013

PROBABILISTIC INTERNAL MODELS



part 1. ADAPTATION TO ENVIRONMENT

part 2. TASK-INDEPENDENCE

PROBABILISTIC INTERNAL MODELS

$P_{\text{true}}(\mathbf{y})$

$\mathbf{y}$

objects, features

$P_{\text{true}}(\mathbf{x}|\mathbf{y})$

$\mathbf{x}$

sensory stimuli

internal model

$P_{\text{model}}(\mathbf{y})$

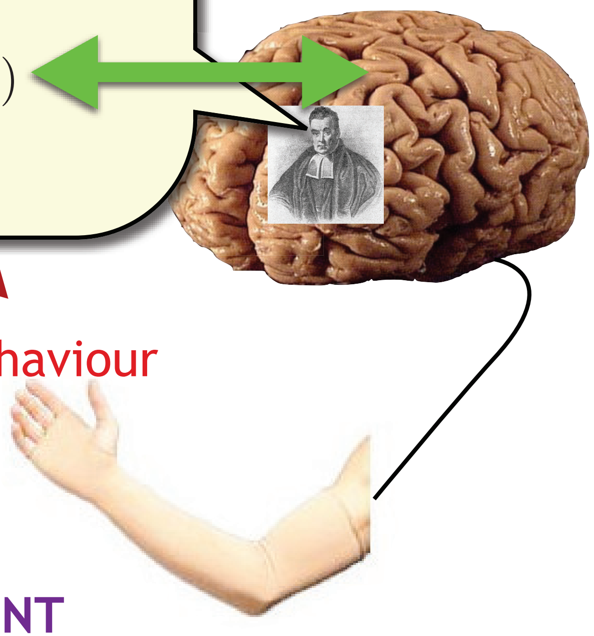
$\mathbf{y}$

objects, features

$P_{\text{model}}(\mathbf{y}|\mathbf{x})$

percept

task → behaviour



part 1. ADAPTATION TO ENVIRONMENT

part 2. TASK-INDEPENDENCE

part 3. NEURAL SUBSTRATES

BAYESIAN INFERENCE IN THE BRAIN

BAYESIAN INFERENCE IN THE BRAIN

data x

BAYESIAN INFERENCE IN THE BRAIN

latent variables y

$\uparrow$

data x

BAYESIAN INFERENCE IN THE BRAIN

parameters θ

$\uparrow$

latent variables y

$\uparrow$

data x

BAYESIAN INFERENCE IN THE BRAIN

structure $\mathcal{M}$

$\uparrow$

parameters θ

$\uparrow$

latent variables y

$\uparrow$

data x

BAYESIAN INFERENCE IN THE BRAIN

program $\mathcal{P}$

$\uparrow$

structure $\mathcal{M}$

$\uparrow$

parameters θ

$\uparrow$

latent variables y

$\uparrow$

data x

BAYESIAN INFERENCE IN THE BRAIN

cognitive science

program $\mathcal{P}$

$\uparrow$

structure $\mathcal{M}$

$\uparrow$

parameters θ

$\uparrow$

latent variables y

$\uparrow$

data x

BAYESIAN INFERENCE IN THE BRAIN

cognitive science

theory

program $\mathcal{P}$

structure $\mathcal{M}$

parameters θ

latent variables y

data x

BAYESIAN INFERENCE IN THE BRAIN

cognitive science

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

BAYESIAN INFERENCE IN THE BRAIN

cognitive science

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

BAYESIAN INFERENCE IN THE BRAIN

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

BAYESIAN INFERENCE IN THE BRAIN

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

BAYESIAN INFERENCE IN THE BRAIN

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

BAYESIAN INFERENCE IN THE BRAIN

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

BAYESIAN INFERENCE IN THE BRAIN

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

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

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

PERCEPTUAL UNCERTAINTY ↔ NEURAL VARIABILITY

PERCEPTUAL UNCERTAINTY ↔ NEURAL VARIABILITY
VIA SAMPLING

PERCEPTUAL UNCERTAINTY ↔ NEURAL VARIABILITY
VIA SAMPLING

stimulus

PERCEPTUAL UNCERTAINTY ↔ NEURAL VARIABILITY
VIA SAMPLING

stimulus

PERCEPTUAL UNCERTAINTY ↔ NEURAL VARIABILITY
VIA SAMPLING

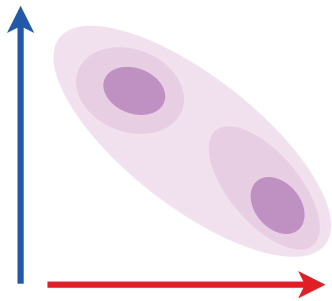
stimulus

PERCEPTUAL UNCERTAINTY ↔ NEURAL VARIABILITY

ideal observer **VIA SAMPLING**

stimulus

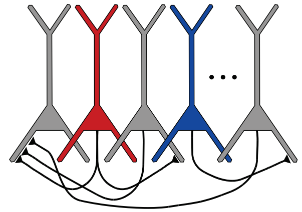
P(feature 1, feature 2 | stimulus)



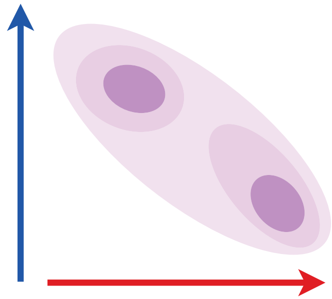
PERCEPTUAL UNCERTAINTY ↔ NEURAL VARIABILITY

ideal observer **VIA SAMPLING** *visual cortex*

stimulus



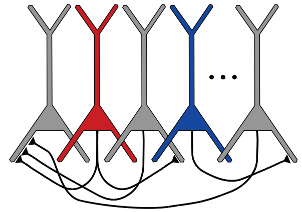
P(feature 1, feature 2 | stimulus)

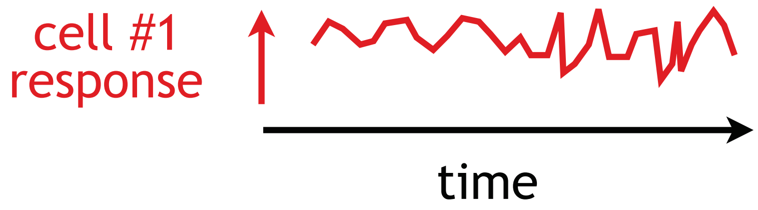


PERCEPTUAL UNCERTAINTY ↔ NEURAL VARIABILITY

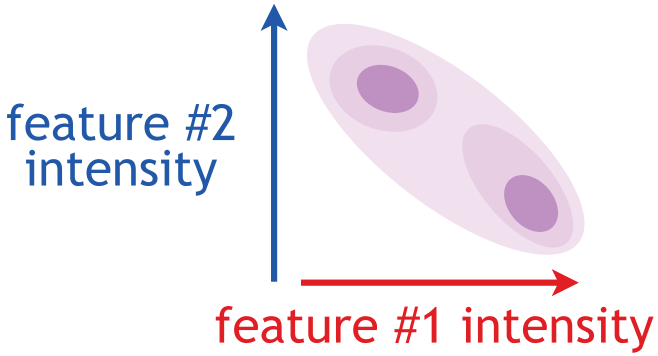
ideal observer **VIA SAMPLING** *visual cortex*

stimulus





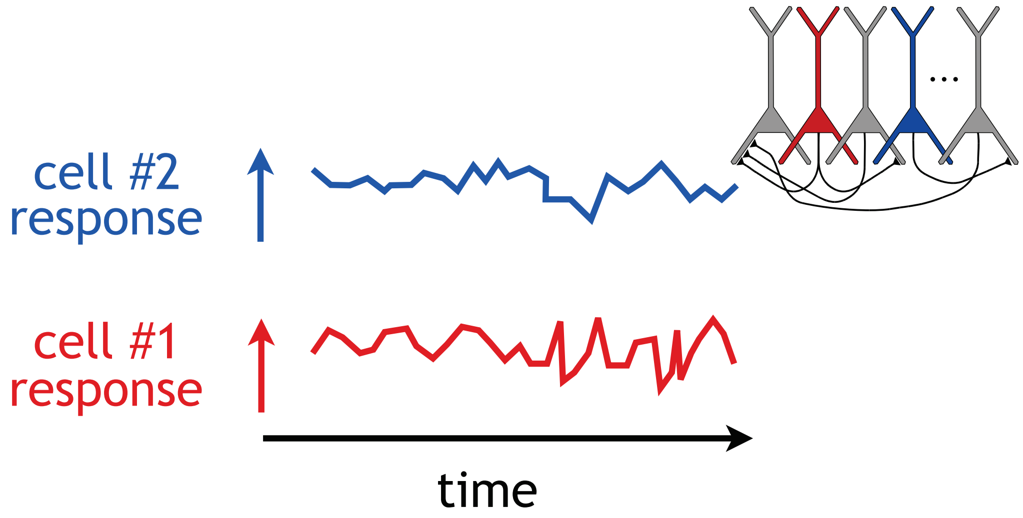
P(feature 1, feature 2 | stimulus)



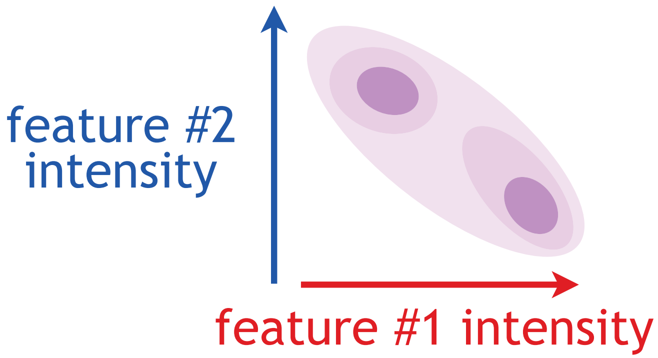
PERCEPTUAL UNCERTAINTY ↔ NEURAL VARIABILITY

ideal observer **VIA SAMPLING** *visual cortex*

stimulus



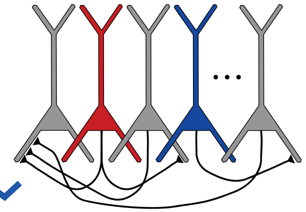
$P(\text{feature 1}, \text{feature 2} \mid \text{stimulus})$

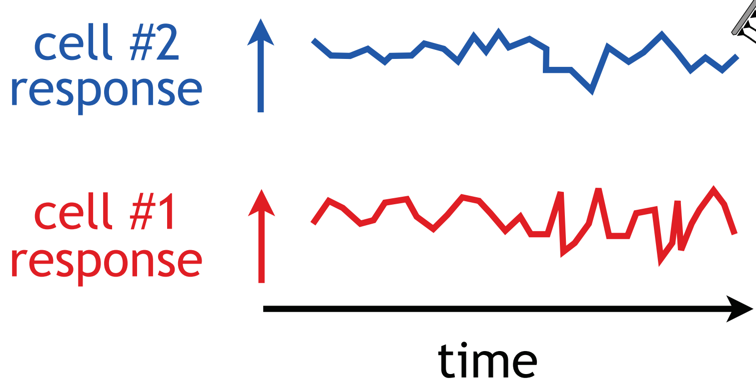


PERCEPTUAL UNCERTAINTY ↔ NEURAL VARIABILITY

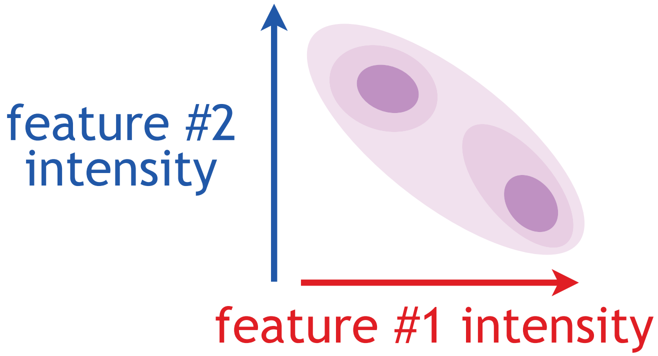
ideal observer **VIA SAMPLING** *visual cortex*

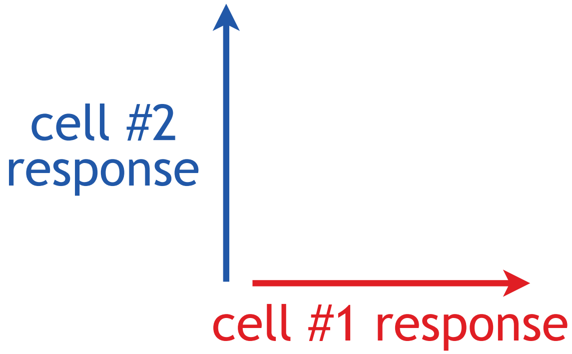
stimulus





P(feature 1, feature 2 | stimulus)

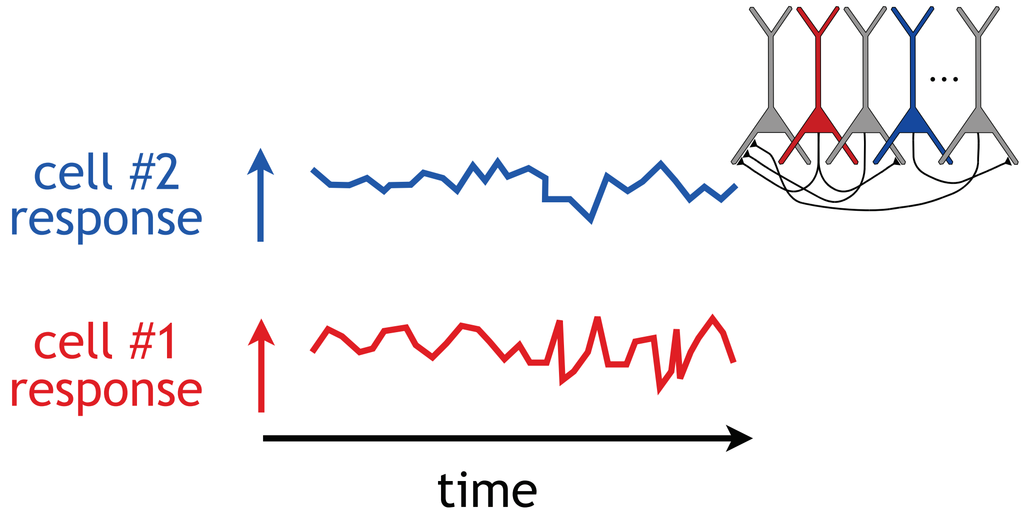




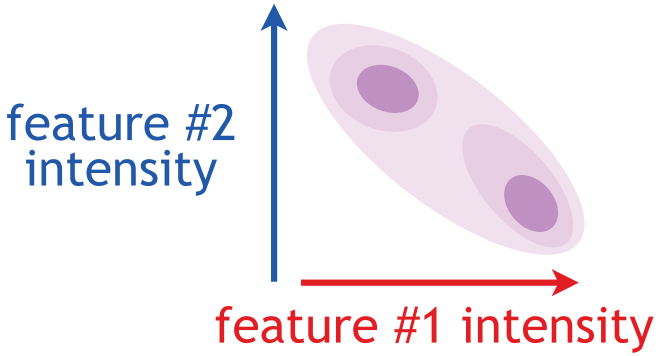
PERCEPTUAL UNCERTAINTY ↔ NEURAL VARIABILITY

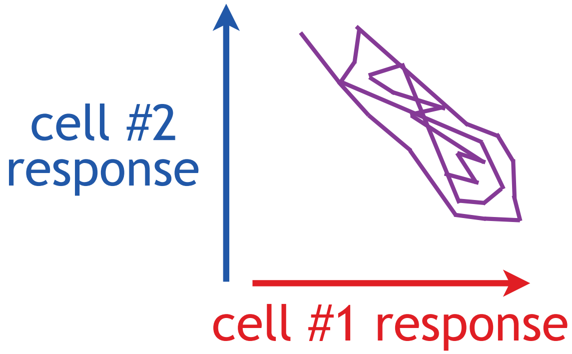
ideal observer **VIA SAMPLING** *visual cortex*

stimulus



P(feature 1, feature 2 | stimulus)

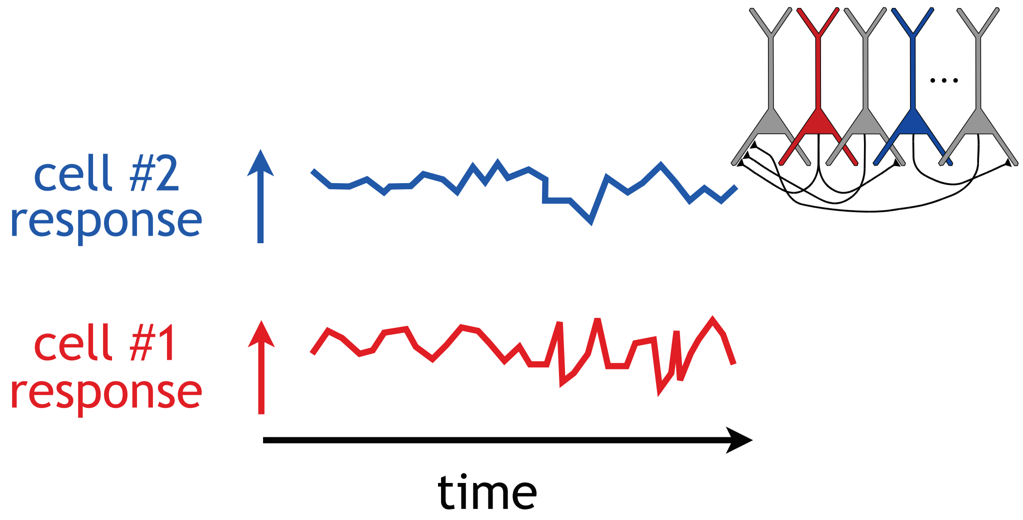




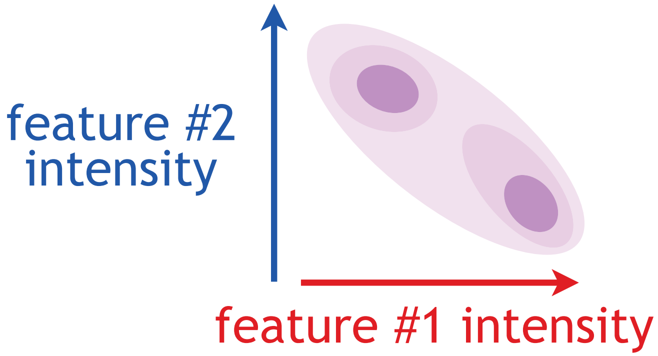
PERCEPTUAL UNCERTAINTY ↔ NEURAL VARIABILITY

ideal observer **VIA SAMPLING** *visual cortex*

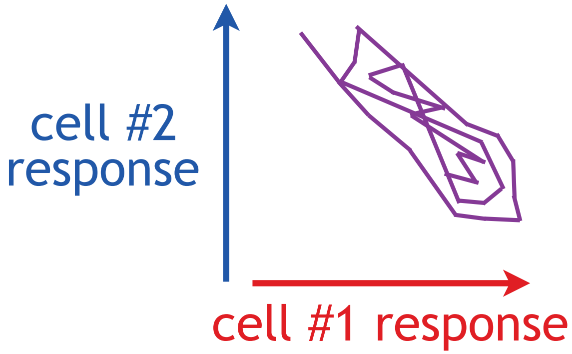
stimulus



P(feature 1, feature 2 | stimulus)



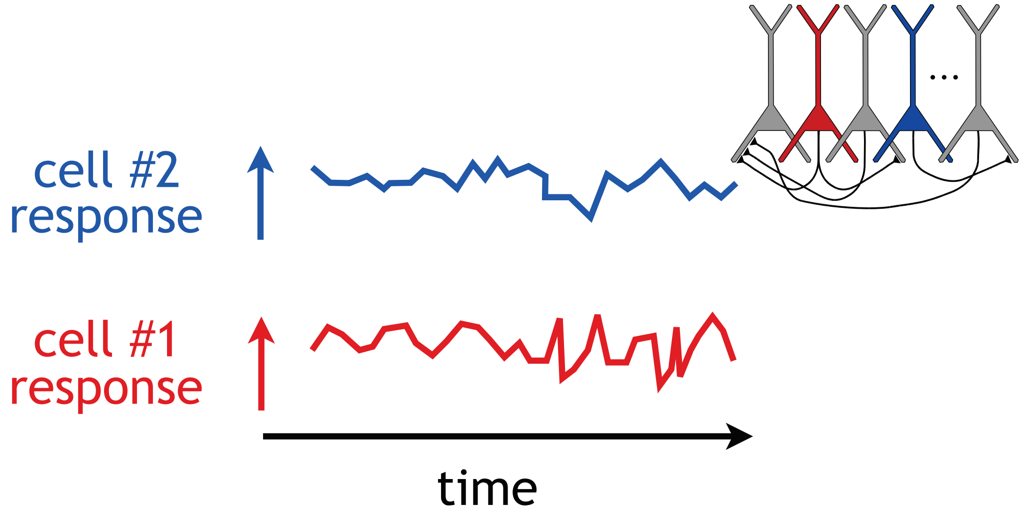
P(response 1, response 2 | stimulus)



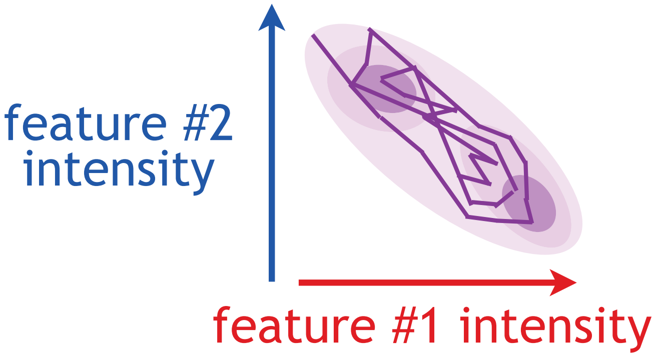
PERCEPTUAL UNCERTAINTY ↔ NEURAL VARIABILITY

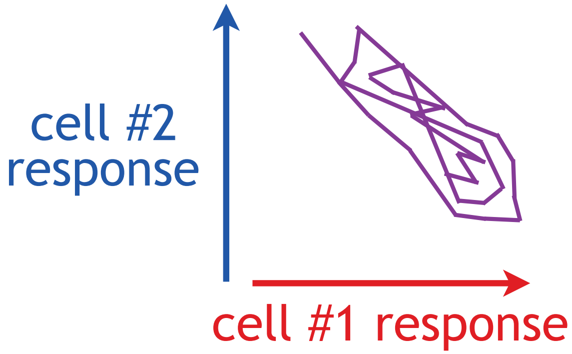
ideal observer **VIA SAMPLING** *visual cortex*

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

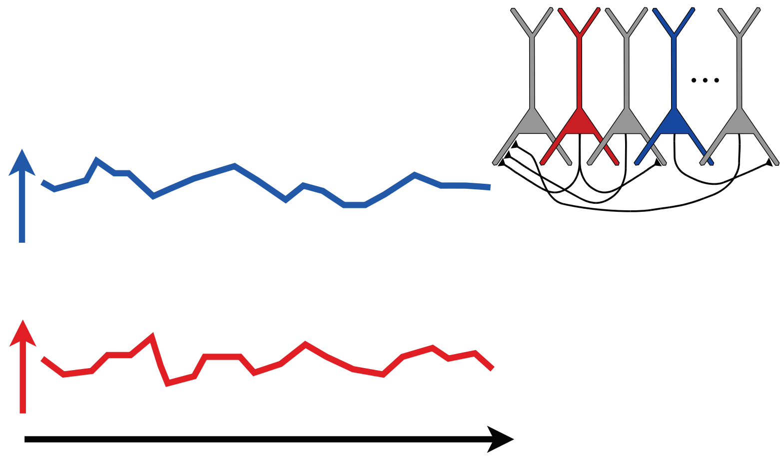
see also:
Hinton & Sejnowski, PDP 1986; Hinton et al, Science 1995; Dayan 1999; Hoyer & Hyvarinen, NIPS 2003, Lee & Mumford 2003

PERCEPTUAL UNCERTAINTY ↔ NEURAL VARIABILITY

ideal observer **VIA SAMPLING** *visual cortex*

stimulus

cell #2
response

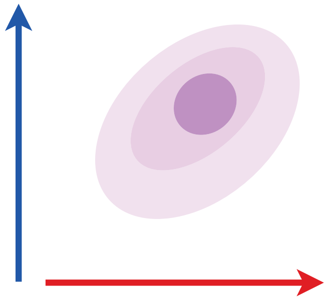


cell #1
response

time

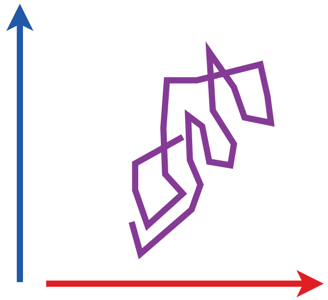
$$P(\text{feature 1}, \text{feature 2} \mid \text{stimulus}) = P(\text{response 1}, \text{response 2} \mid \text{stimulus})$$

feature #2
intensity



feature #1 intensity

cell #2
response



cell #1 response

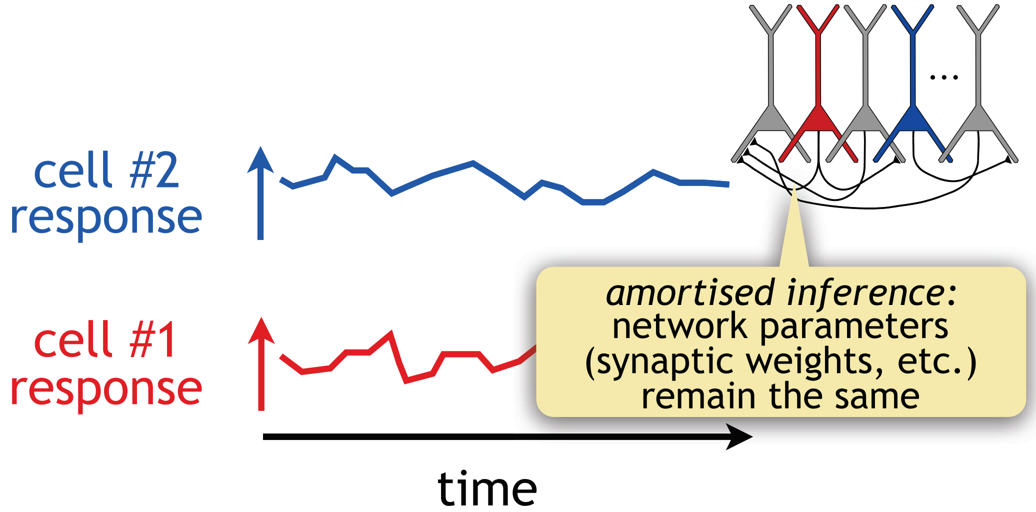
Fiser et al, TICS 2010

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

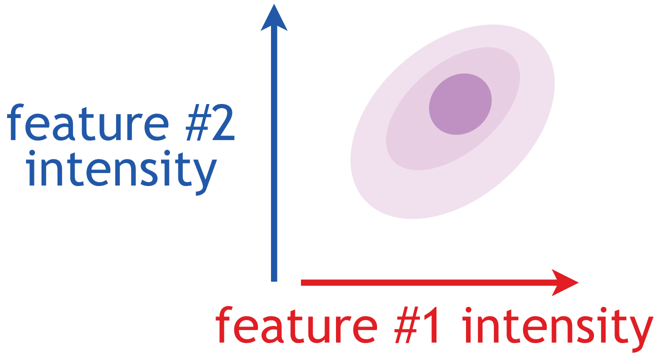
PERCEPTUAL UNCERTAINTY ↔ NEURAL VARIABILITY

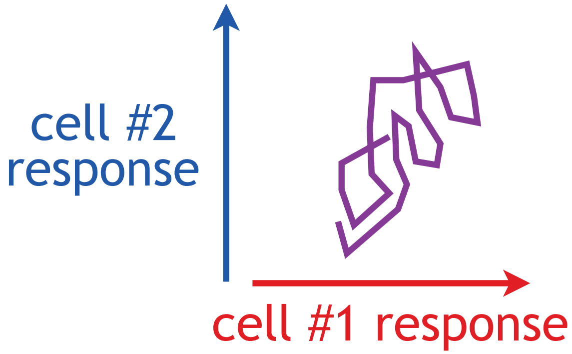
ideal observer **VIA SAMPLING** *visual cortex*

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

NEURAL HALLMARKS OF SAMPLING

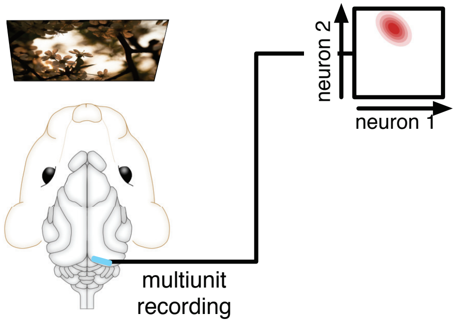
Gergő Orbán

Pietro Berkes

József Fiser

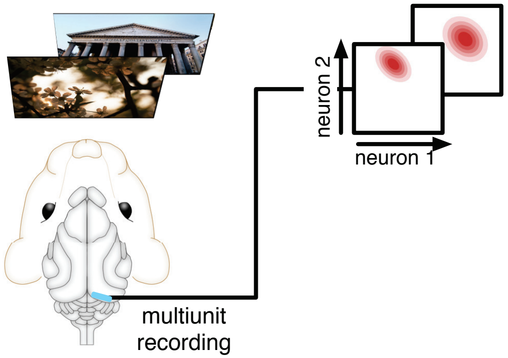
NEURAL HALLMARKS OF SAMPLING

Gergő Orbán | Pietro Berkes | József Fiser



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

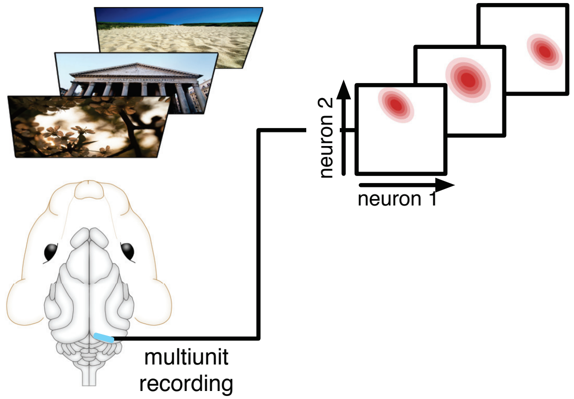
NEURAL HALLMARKS OF SAMPLING



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

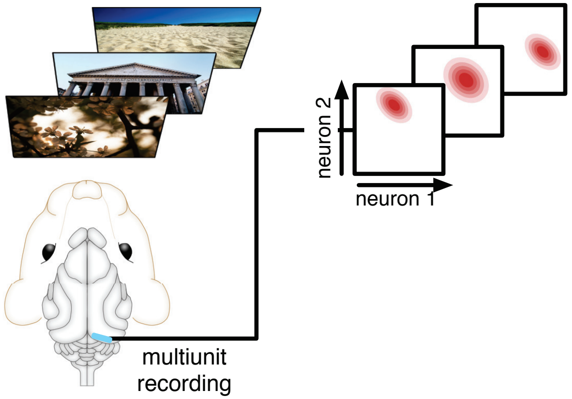
NEURAL HALLMARKS OF SAMPLING

Gergő Orbán
Pietro Berkes
József Fiser

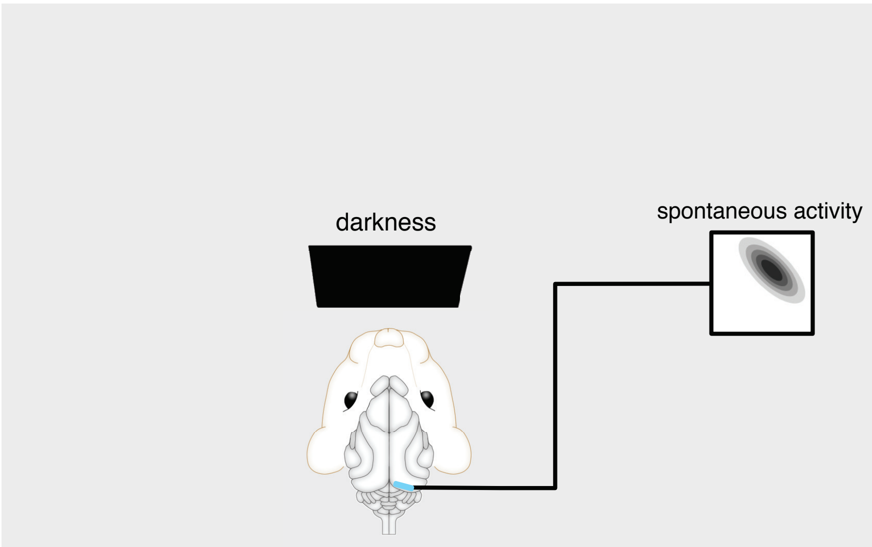


$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}_3)$$
$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}_2)$$
$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}_1)$$

NEURAL HALLMARKS OF SAMPLING



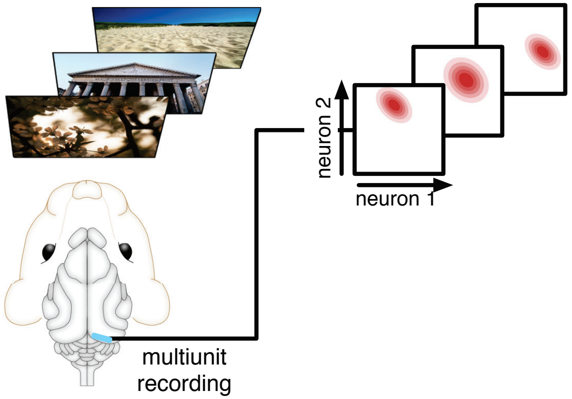
$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}_3)$$
$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}_2)$$
$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}_1)$$



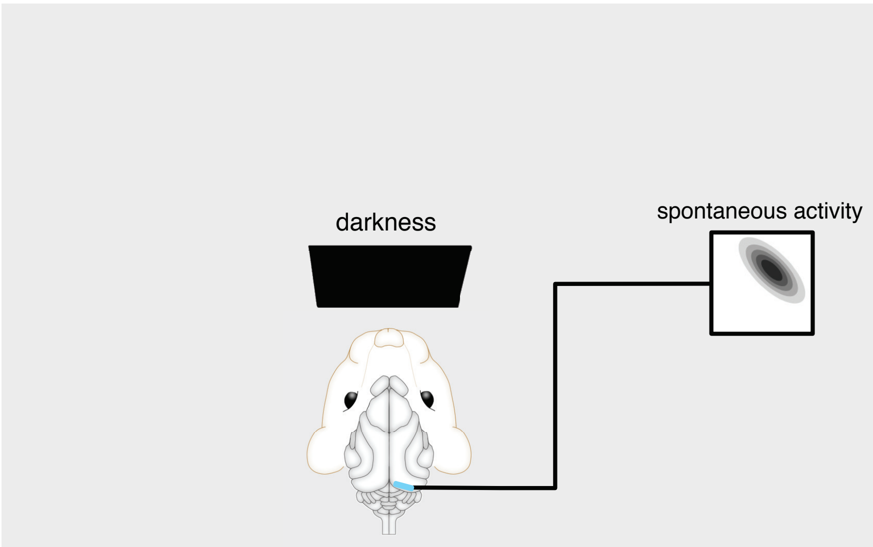
$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim} = 0)$$

Berkes et al, Science 2011

NEURAL HALLMARKS OF SAMPLING



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

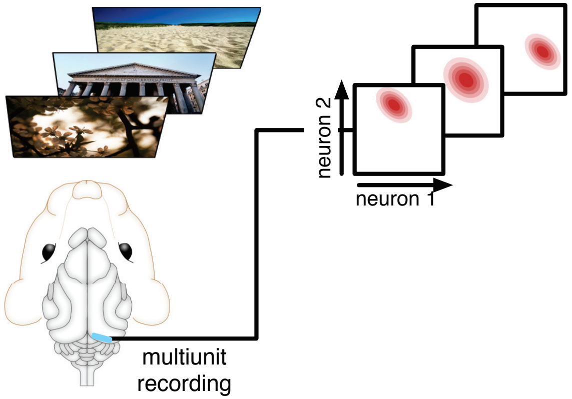


$$\mathrm{P}_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}=0) \simeq \mathrm{P}_{\mathrm{model}}(\mathsf{feat})$$

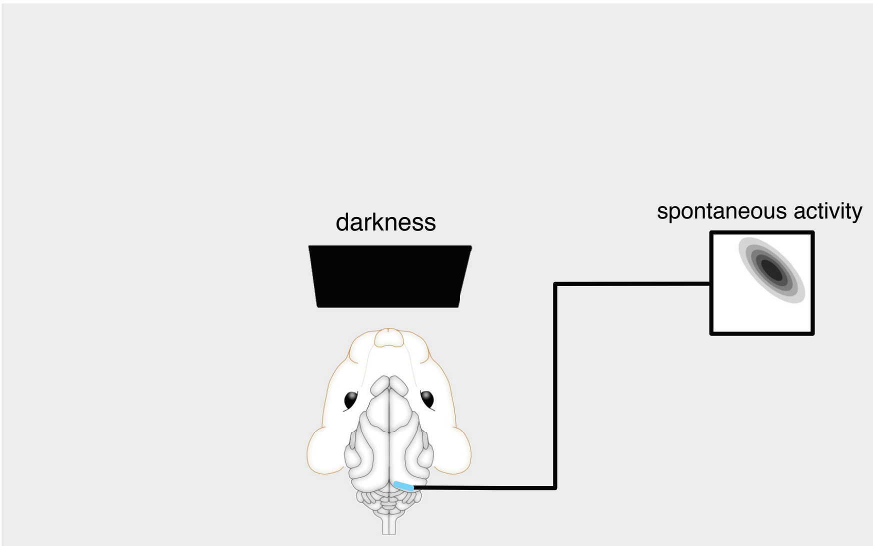
Berkes et al, Science 2011

NEURAL HALLMARKS OF SAMPLING

Gergő Orbán | Pietro Berkes | József Fiser



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

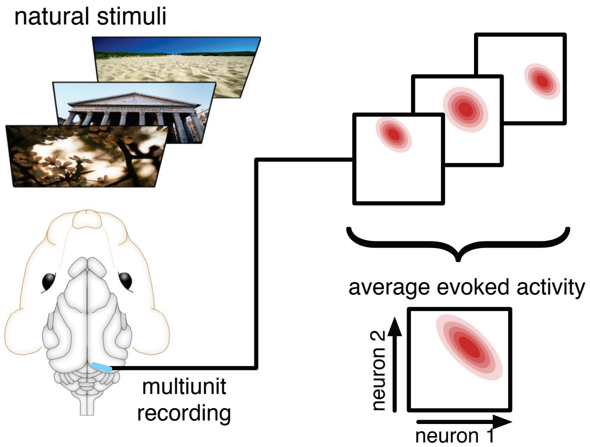


$$P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim}=0) \simeq P_{\mathrm{model}}(\mathsf{feat})$$
$$\simeq \int P_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim})\, P_{\mathrm{model}}(\mathsf{stim}) \;\; d\mathsf{stim}$$

Berkes et al, Science 2011

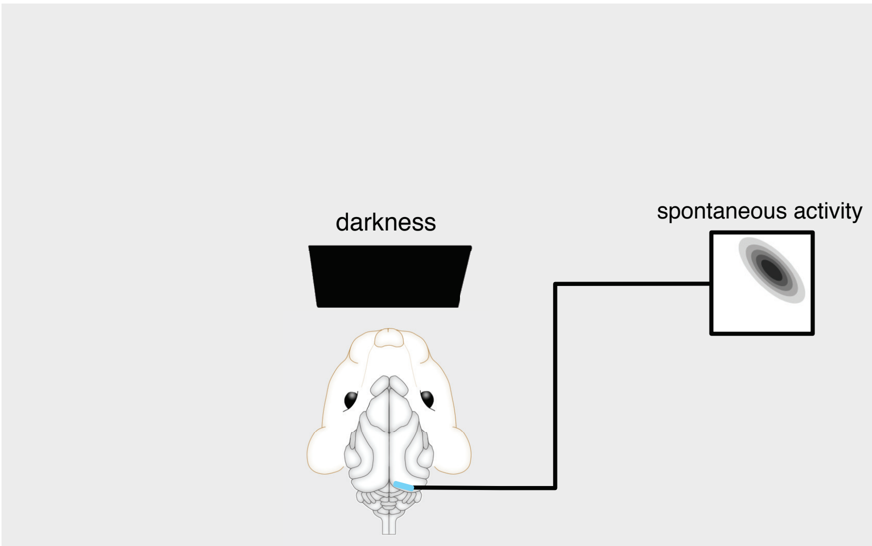
NEURAL HALLMARKS OF SAMPLING

Gergő Orbán | Pietro Berkes | József Fiser



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

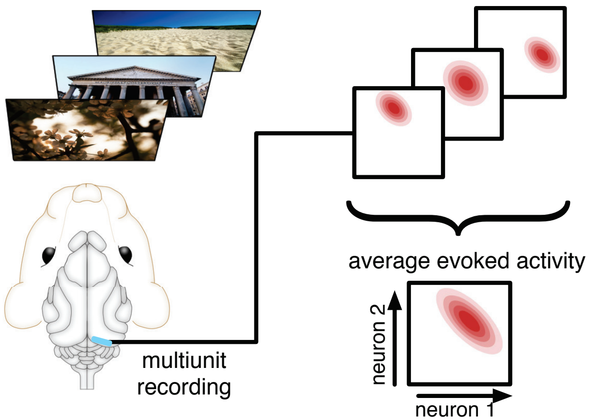
$$\int \mathrm{P}_{\mathrm{model}}(\mathsf{feat}|\mathsf{stim})\,\mathrm{P}_{\mathrm{natural}}(\mathsf{stim})\,d\mathsf{stim}$$



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

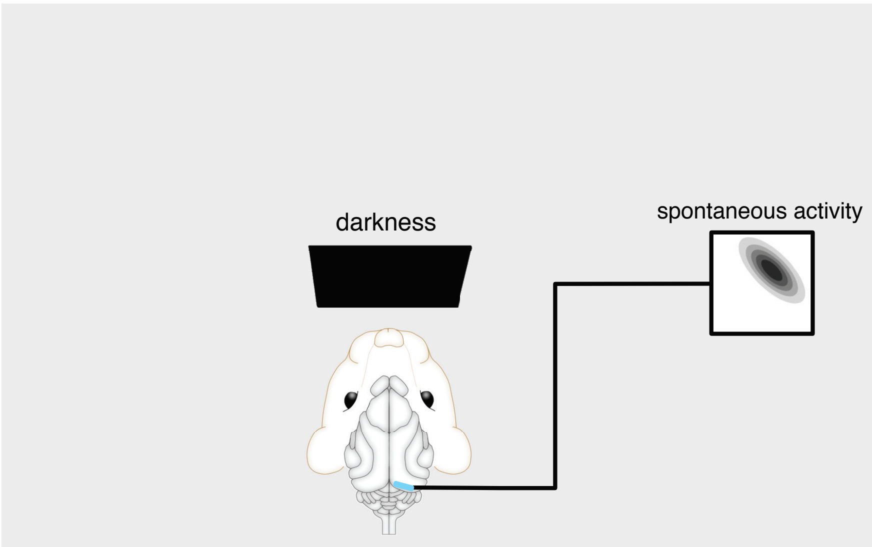
Berkes et al, Science 2011

NEURAL HALLMARKS OF SAMPLING



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

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



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

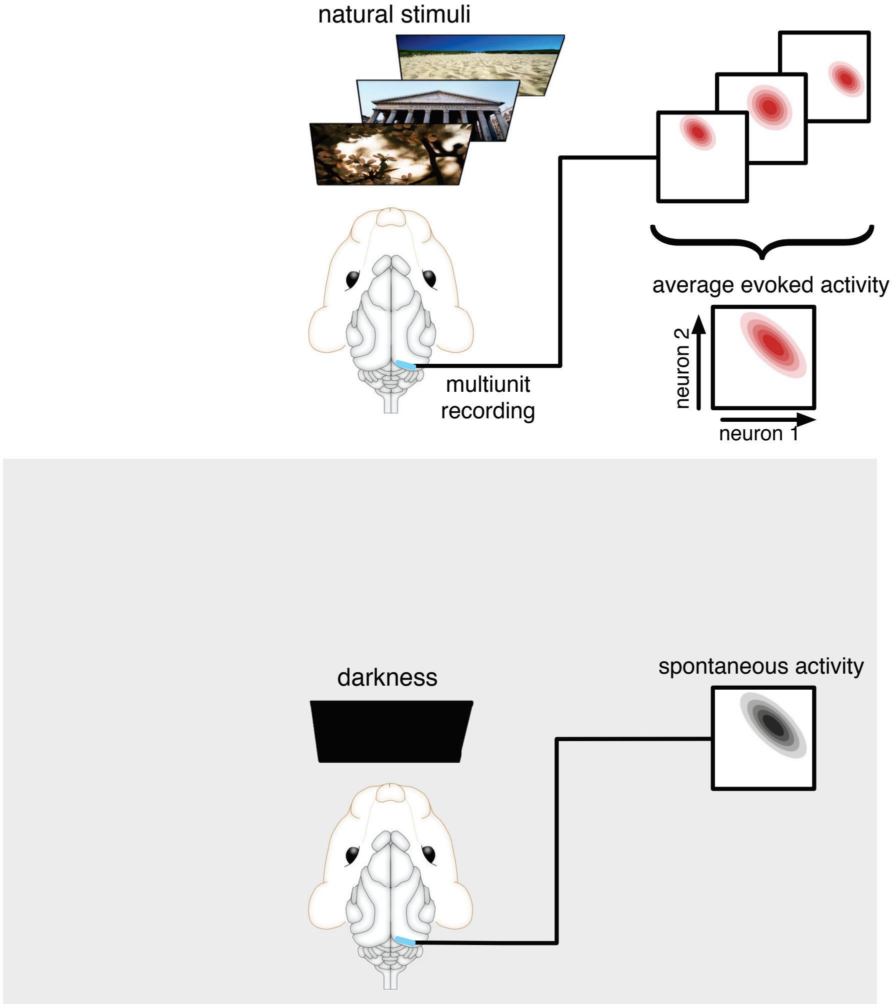
Berkes et al, Science 2011

NEURAL HALLMARKS OF SAMPLING

+ STATISTICALLY OPTIMAL ADAPTATION

$$P_{\mathrm{model}}(\mathsf{stim}) = P_{\mathrm{natural}}(\mathsf{stim})$$

Gergő Orbán
Pietro Berkes
József Fiser



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

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

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

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

Berkes et al, Science 2011

NEURAL HALLMARKS OF SAMPLING

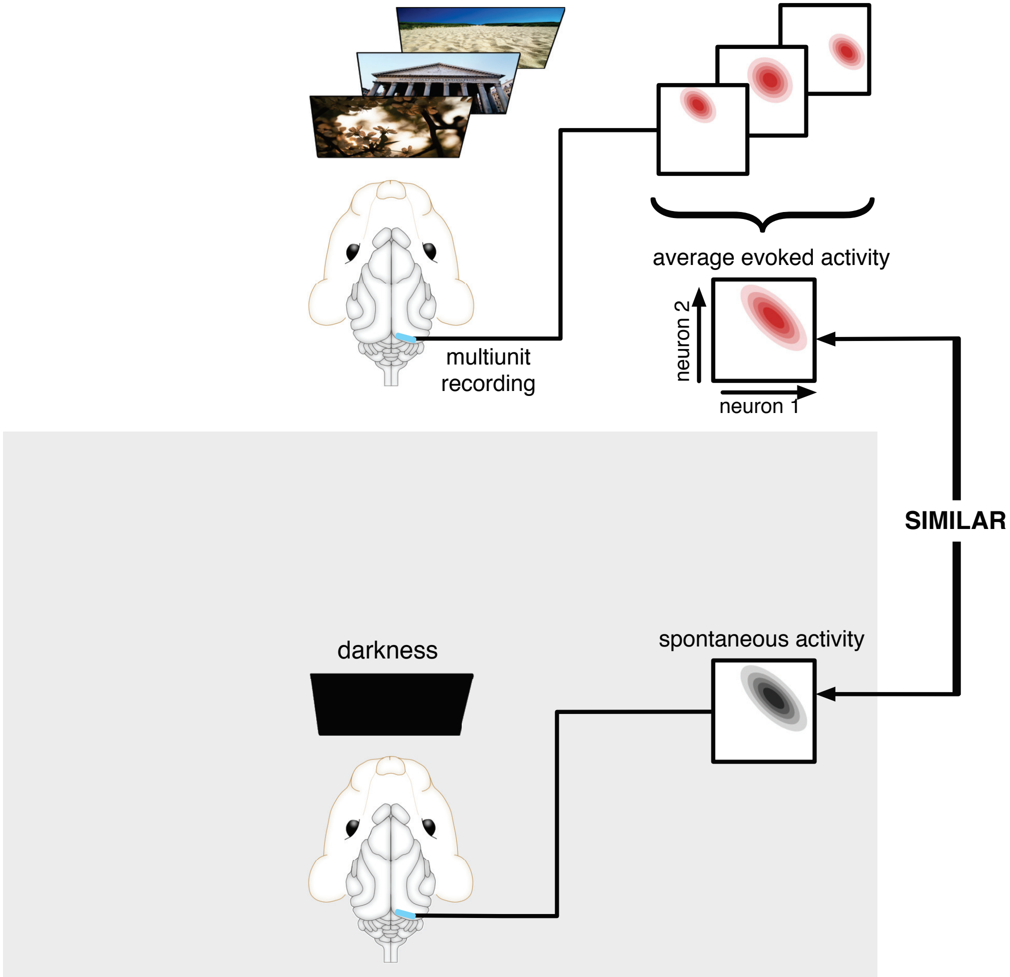
+ STATISTICALLY OPTIMAL ADAPTATION

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


Gergő Orbán


Pietro Berkes


József Fiser



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

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

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

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

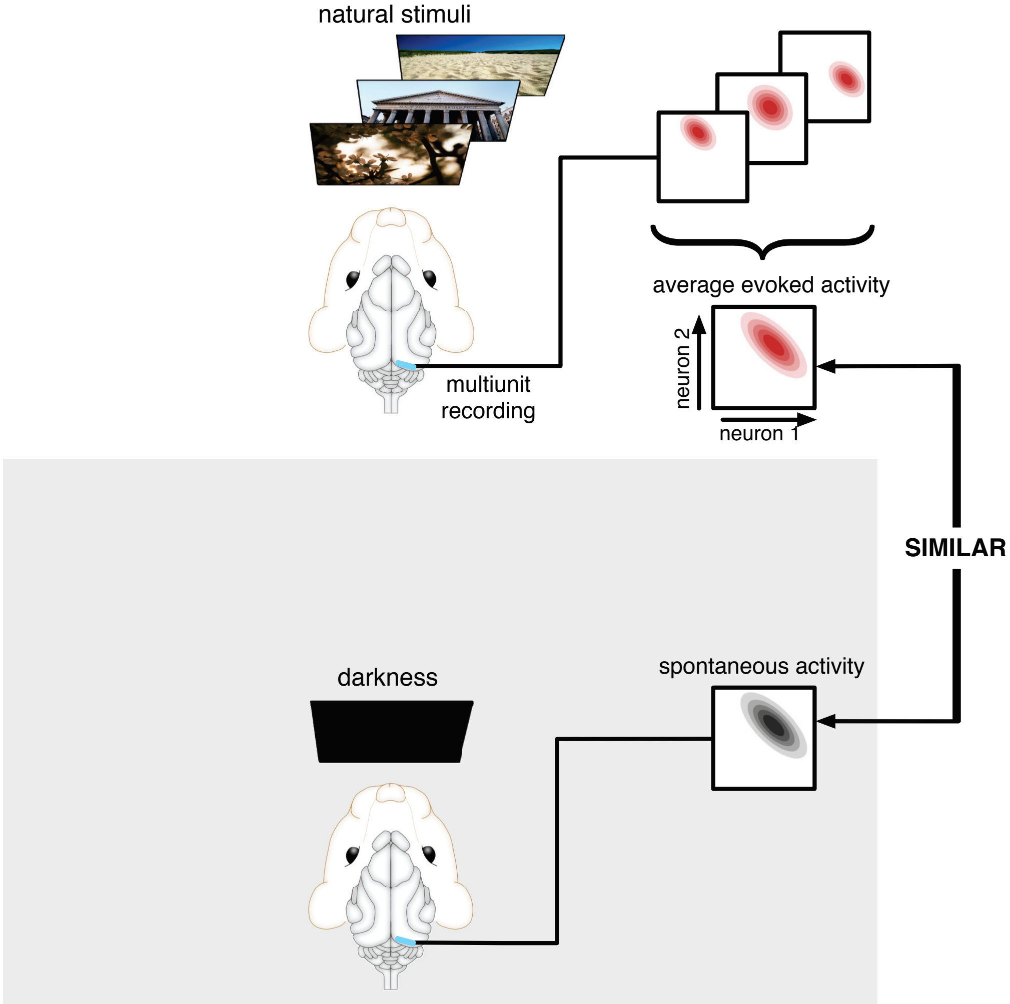
Berkes et al, Science 2011

NEURAL HALLMARKS OF SAMPLING

+ STATISTICALLY OPTIMAL ADAPTATION

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

Gergő Orbán | Pietro Berkes | József Fiser



Berkes et al, Science 2011

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

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

average inference

=

prior expectations

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

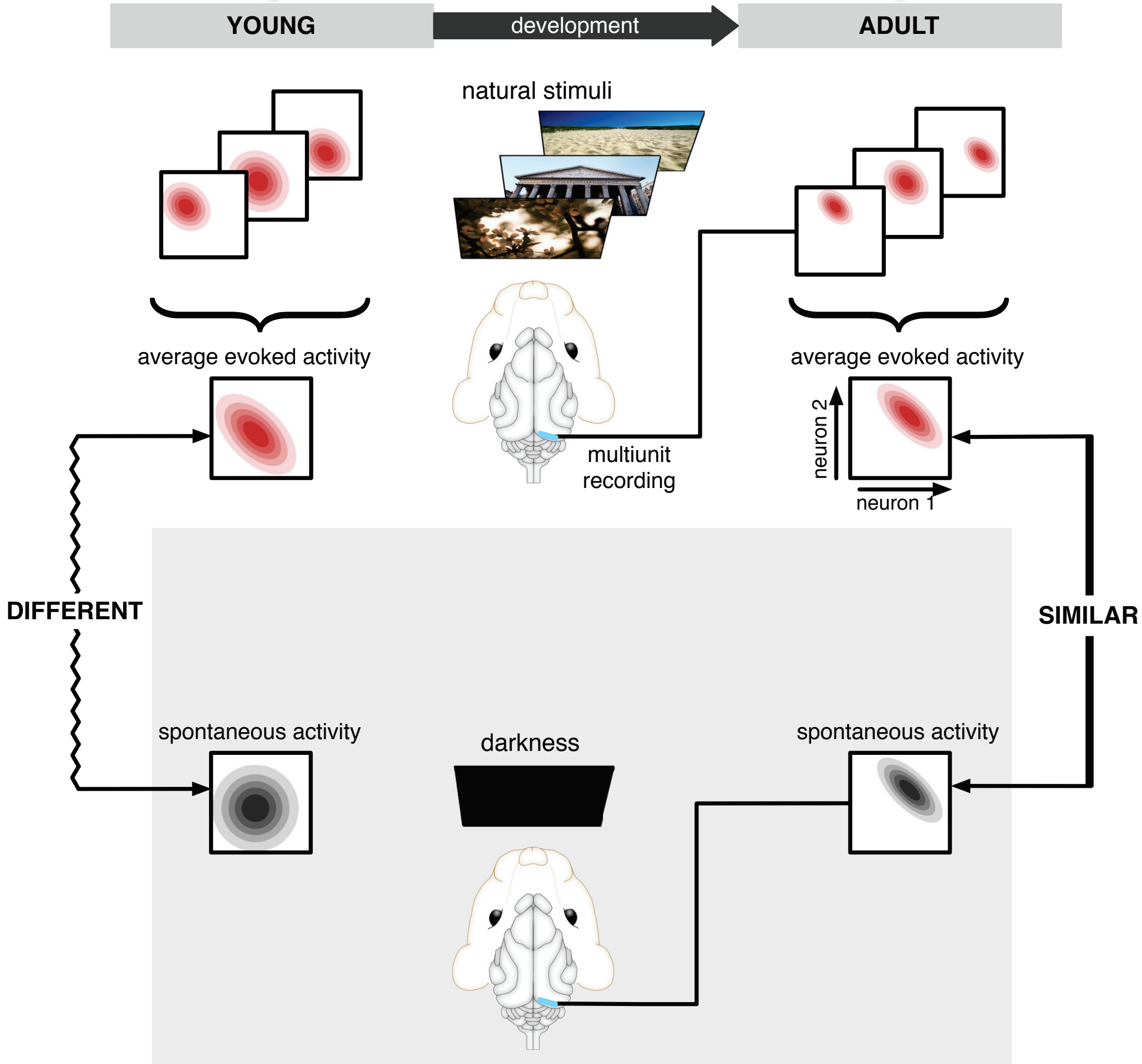
$$\simeq \boxed{\int \mathrm{P_{model}}(\text{feat}|\text{stim})\, \mathrm{P_{model}}(\text{stim})\, d\text{stim}}$$

NEURAL HALLMARKS OF SAMPLING

+ STATISTICALLY OPTIMAL ADAPTATION

$P_{\mathrm{model}}(\mathrm{stim}) \neq P_{\mathrm{natural}}(\mathrm{stim})$

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



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

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

average inference

=

prior expectations

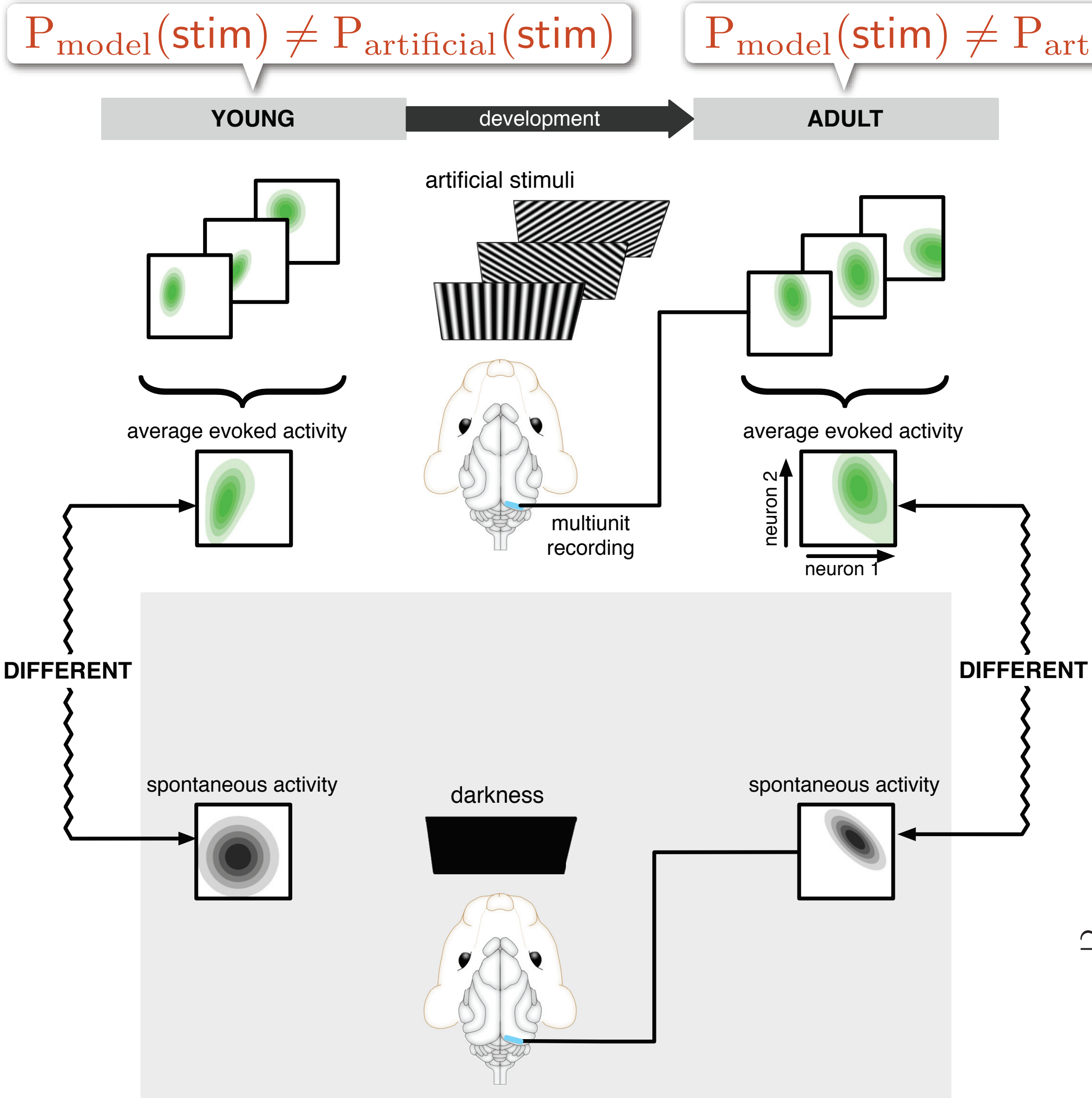
$$P_{\mathrm{model}}(\mathrm{feat}|\mathrm{stim}=0) \simeq P_{\mathrm{model}}(\mathrm{feat})$$

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

Berkes et al, Science 2011

NEURAL HALLMARKS OF SAMPLING

+ STATISTICALLY OPTIMAL ADAPTATION



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

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

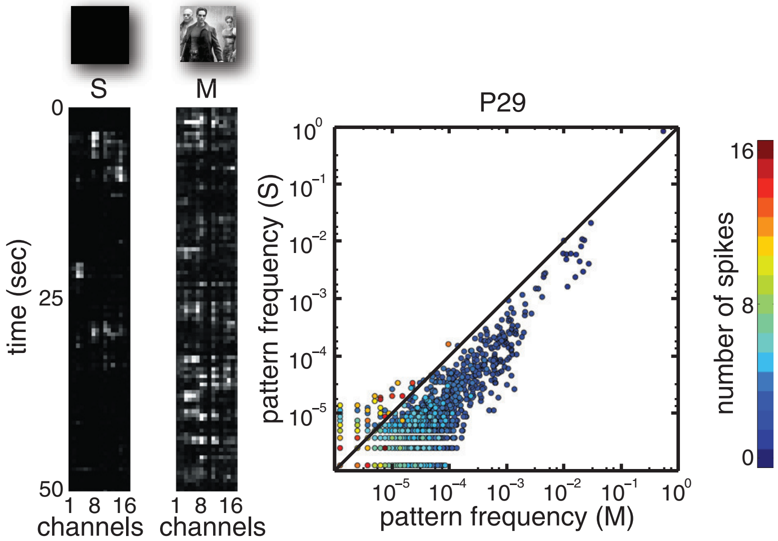
average inference
≠
prior expectations

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

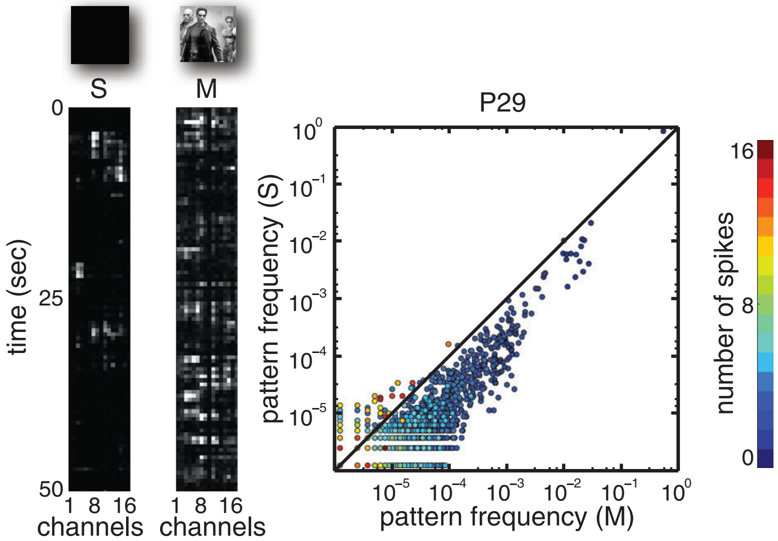
$$\simeq \int \mathrm{P_{model}(feat|stim)\, P_{model}(stim)}\, d\mathrm{stim}$$

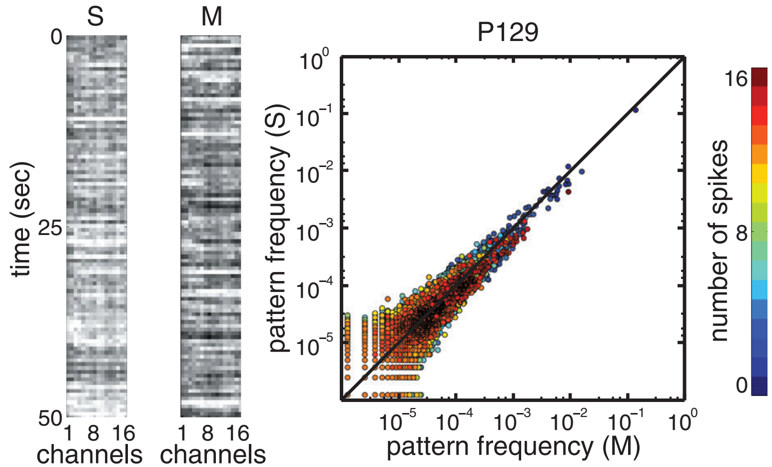
Berkes et al, Science 2011

DEVELOPMENTAL CHANGES

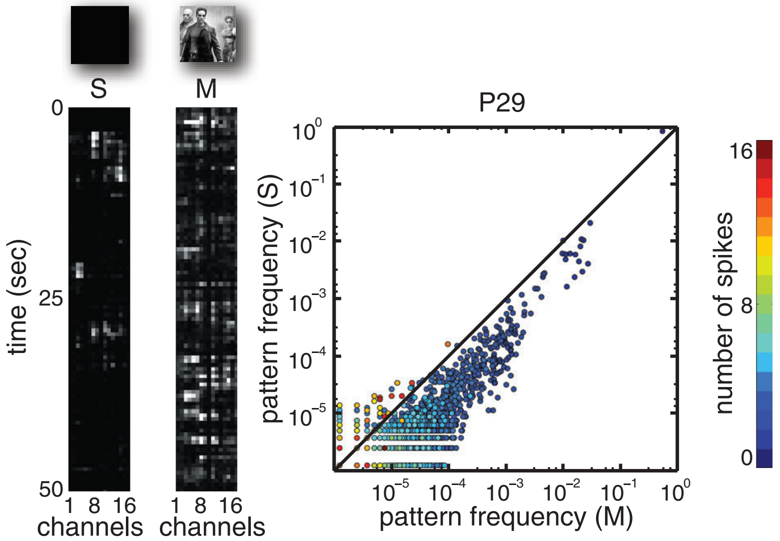


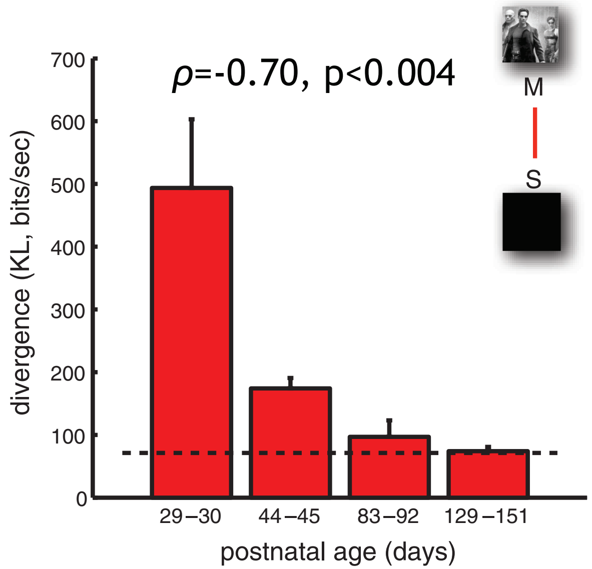
DEVELOPMENTAL CHANGES

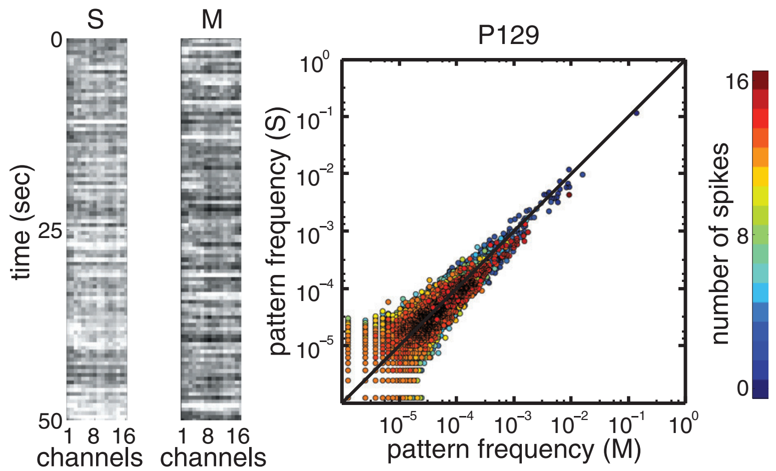




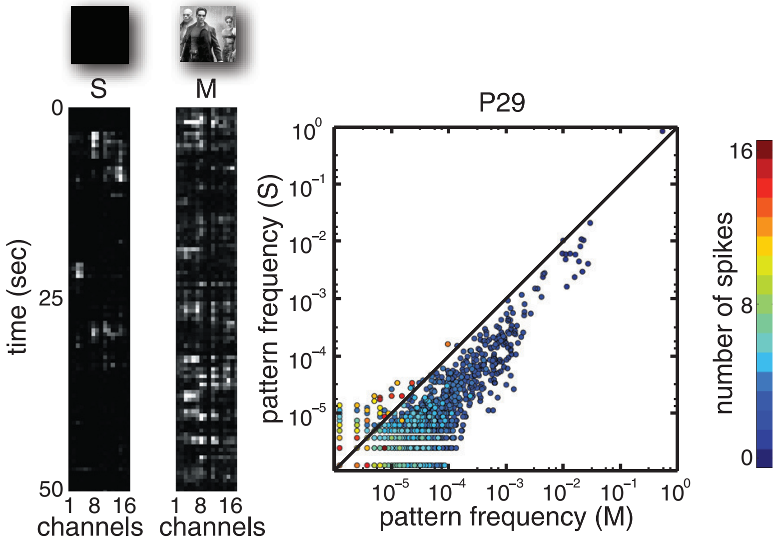
DEVELOPMENTAL CHANGES

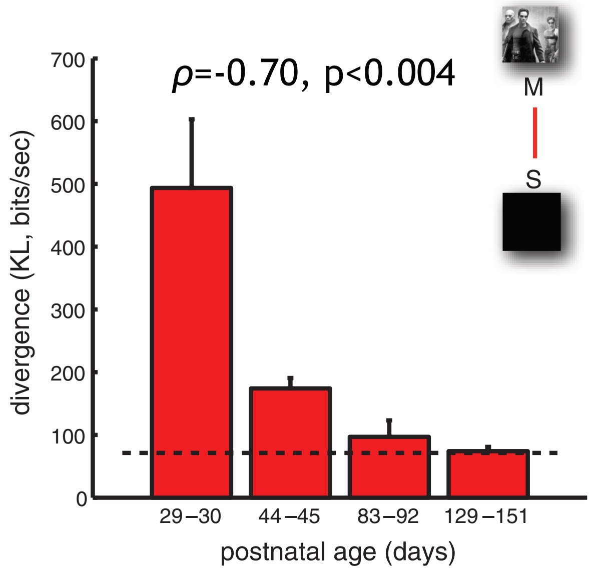


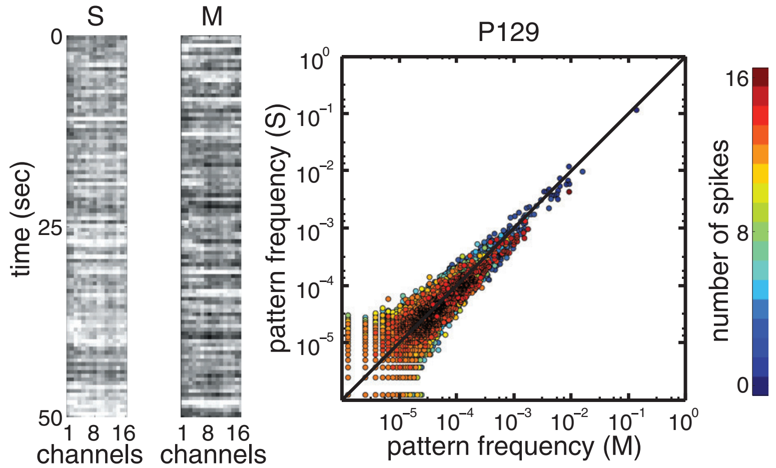


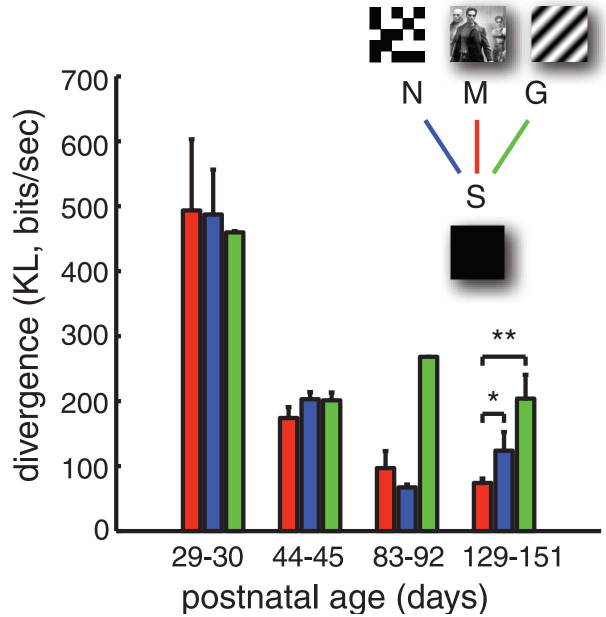


DEVELOPMENTAL CHANGES



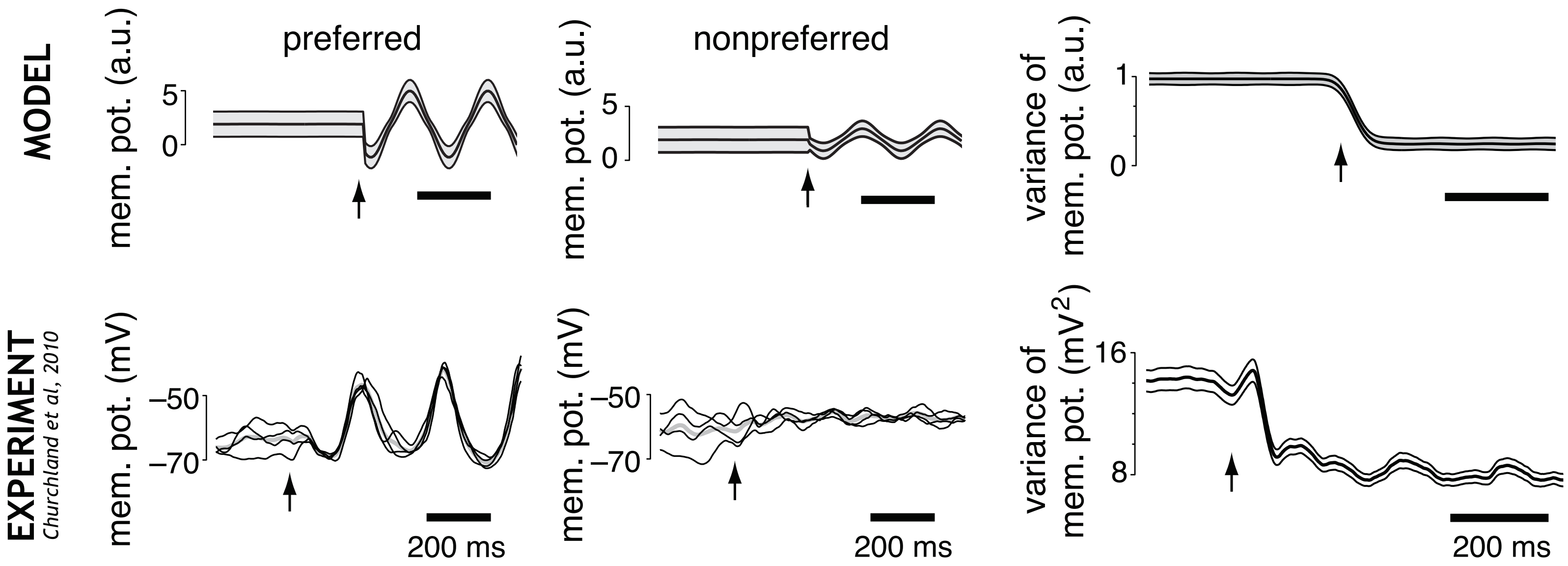




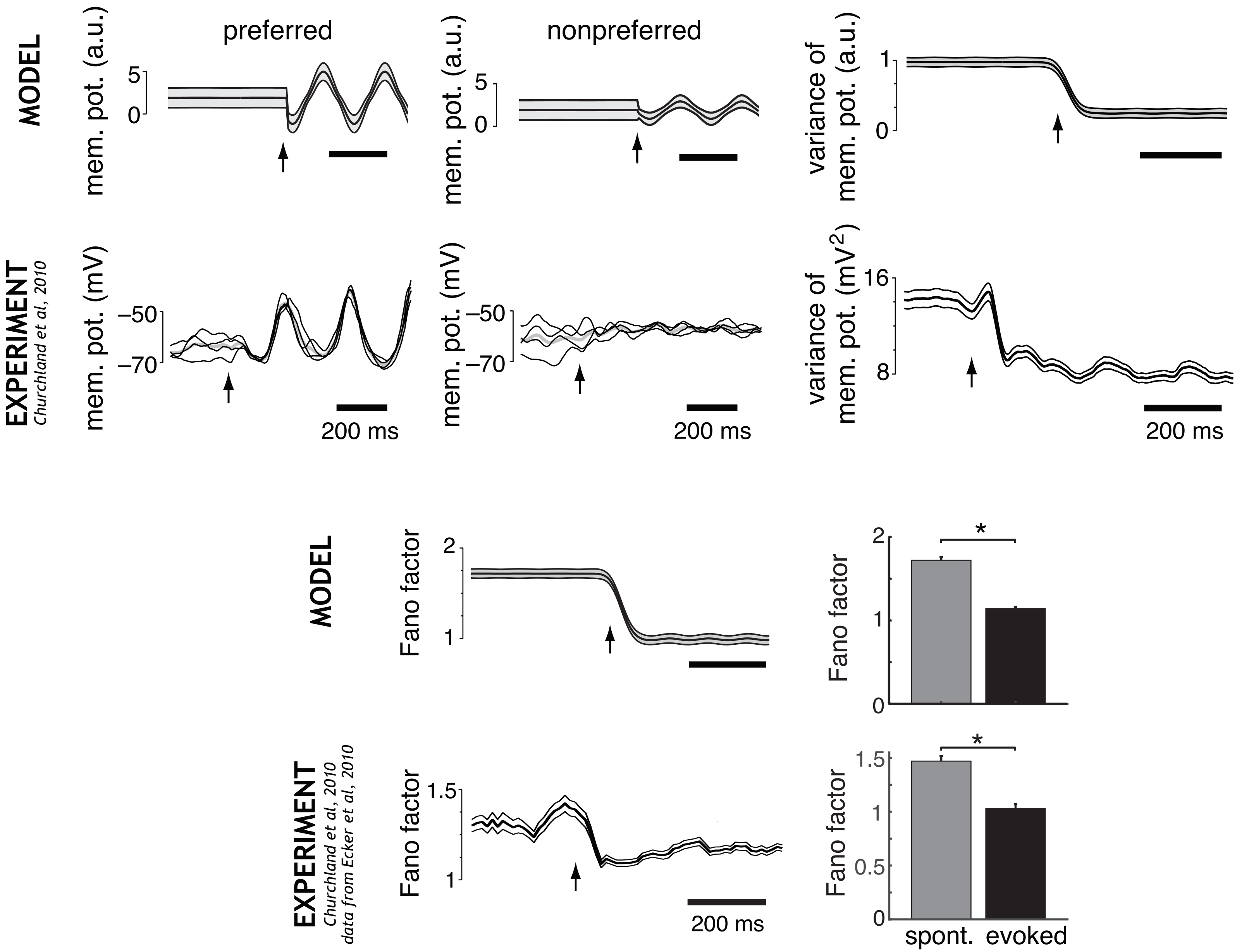


Berkes et al, Science 2011

STIMULUS ONSET ⇒ CONTRAST ↑ ⇒ NOISE VARIABILITY ↓

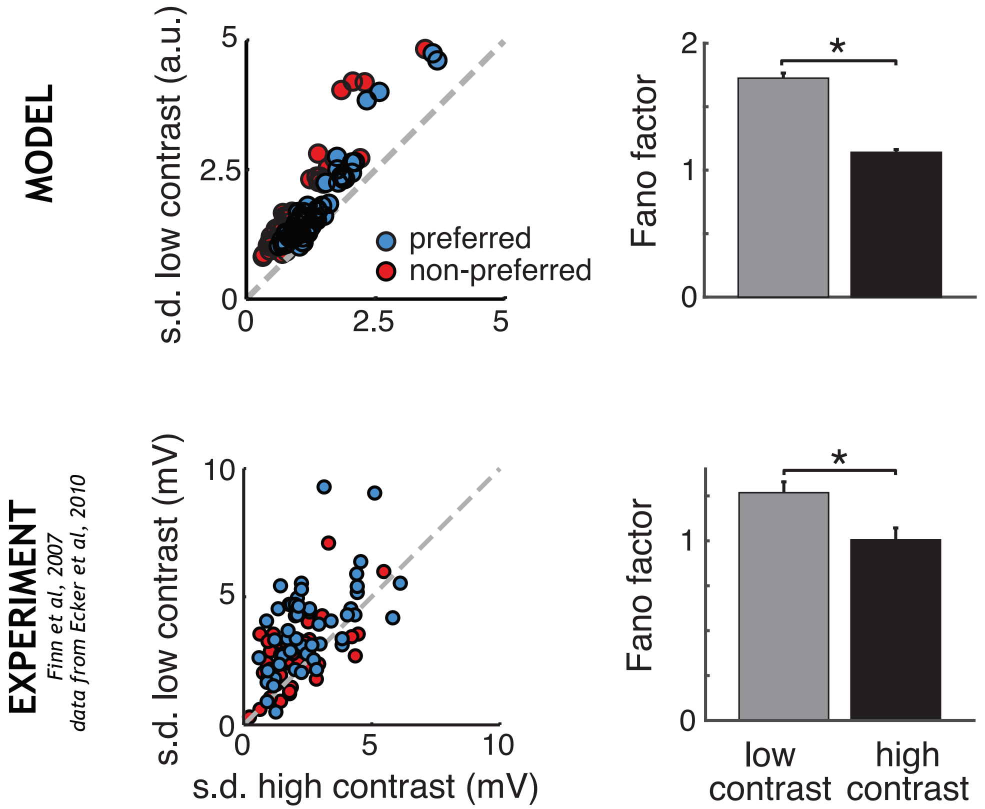


STIMULUS ONSET ⇒ CONTRAST ↑ ⇒ NOISE VARIABILITY ↓



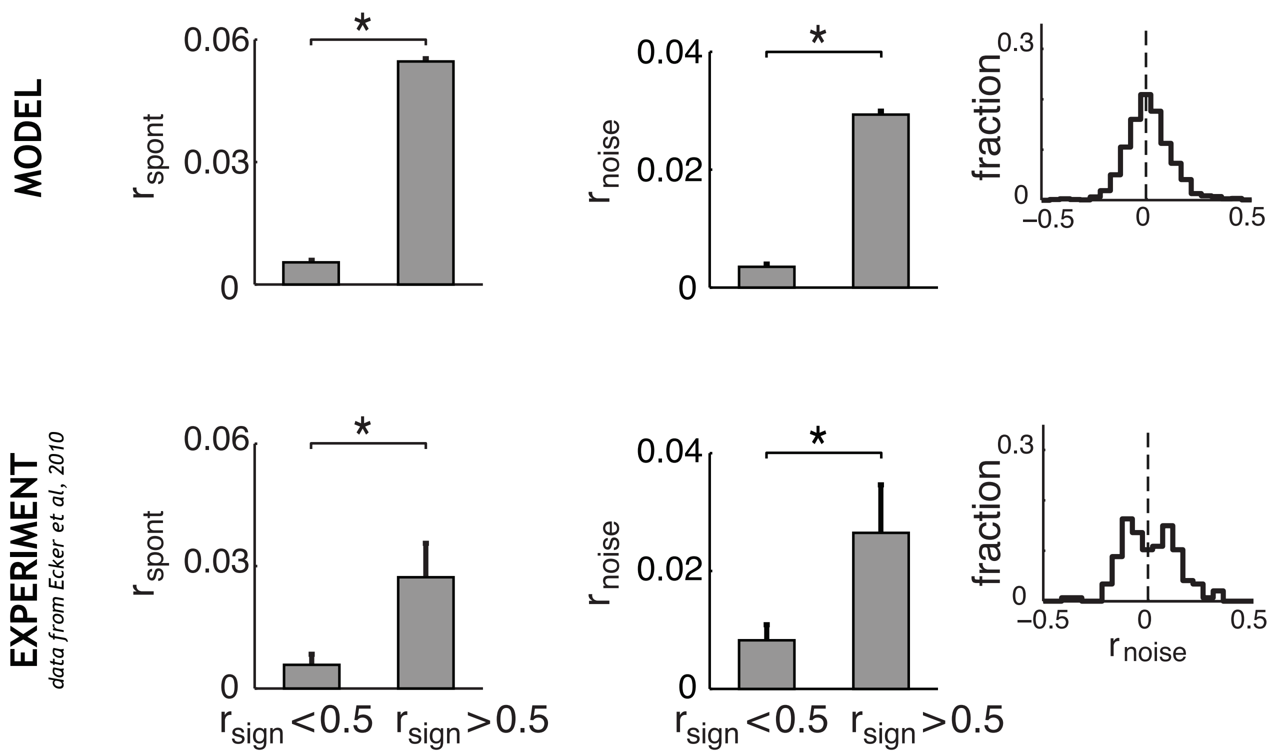
Orbán et al, Neuron 2016

CONTRAST ↑ ⇒ NOISE VARIABILITY ↓



Orbán et al, Neuron 2016

SPONTANEOUS ≈ SIGNAL ≈ NOISE CORRELATIONS



Orbán et al, Neuron 2016

INTERNAL MODELS

INTERNAL MODELS

- ubiquitous
  low-level visual features, abstract shapes, faces

INTERNAL MODELS

- ubiquitous
  low-level visual features, abstract shapes, faces
- adapted in a Bayes-optimal way
  statistical learning, visual development

INTERNAL MODELS

- ubiquitous
  low-level visual features, abstract shapes, faces
- adapted in a Bayes-optimal way
  statistical learning, visual development
- task-, and modality-independent
  familiarity, visual discrimination, haptic affordances

INTERNAL MODELS

- ❖ ubiquitous
  low-level visual features, abstract shapes, faces
- ❖ adapted in a Bayes-optimal way
  statistical learning, visual development
- ❖ task-, and modality-independent
  familiarity, visual discrimination, haptic affordances
- ❖ directly quantifiable in neural activity
  evoked and spontaneous neural activity
  (so far: stationary statistics → circuit dynamics?)

INTERNAL MODELS

- ubiquitous
  low-level visual features, abstract shapes, faces
- adapted in a Bayes-optimal way
  statistical learning, visual development
- task-, and modality-independent
  familiarity, visual discrimination, haptic affordances
- directly quantifiable in neural activity
  evoked and spontaneous neural activity
  (so far: stationary statistics → circuit dynamics?)

ACKNOWLEDGEMENTS


Neil Houlsby


Ferenc Huszár


Mohammad Ghassemi


Alex Pantelides


Goda Zalalyte


James Ingram

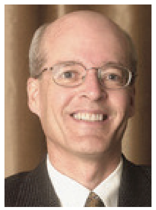

Daniel Wolpert


Gergő Orbán


Pietro Berkes


Gábor Lengyel


József Fiser


Richard Aslin

data sharing:

Alexander Ecker
Matthias Bethge
Andreas Tolias

INTERNAL MODELS

- ubiquitous
  low-level visual features, abstract shapes, faces
- adapted in a Bayes-optimal way
  statistical learning, visual development
- task-, and modality-independent
  familiarity, visual discrimination, haptic affordances
- directly quantifiable in neural activity
  evoked and spontaneous neural activity
  (so far: stationary statistics → circuit dynamics?)

ACKNOWLEDGEMENTS


Neil Houlsby


Ferenc Huszár


Mohammad Ghassemi


Alex Pantelides


Goda Zalalyte


James Ingram


Daniel Wolpert

Wigner Research Centre for Physics

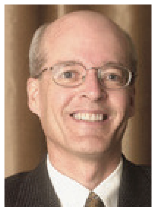

Gergő Orbán

CEU


Pietro Berkes


Gábor Lengyel


József Fiser


Richard Aslin

data sharing:

Alexander Ecker
Matthias Bethge
Andreas Tolias

wellcome trust Investigator

erc


POSTDOC POSITION AVAILABLE
estimate humans' complex, high dimensional,
dynamically changing internal models from behaviour
collaborate with world-leading
experimental cognitive scientists
CEU
erc