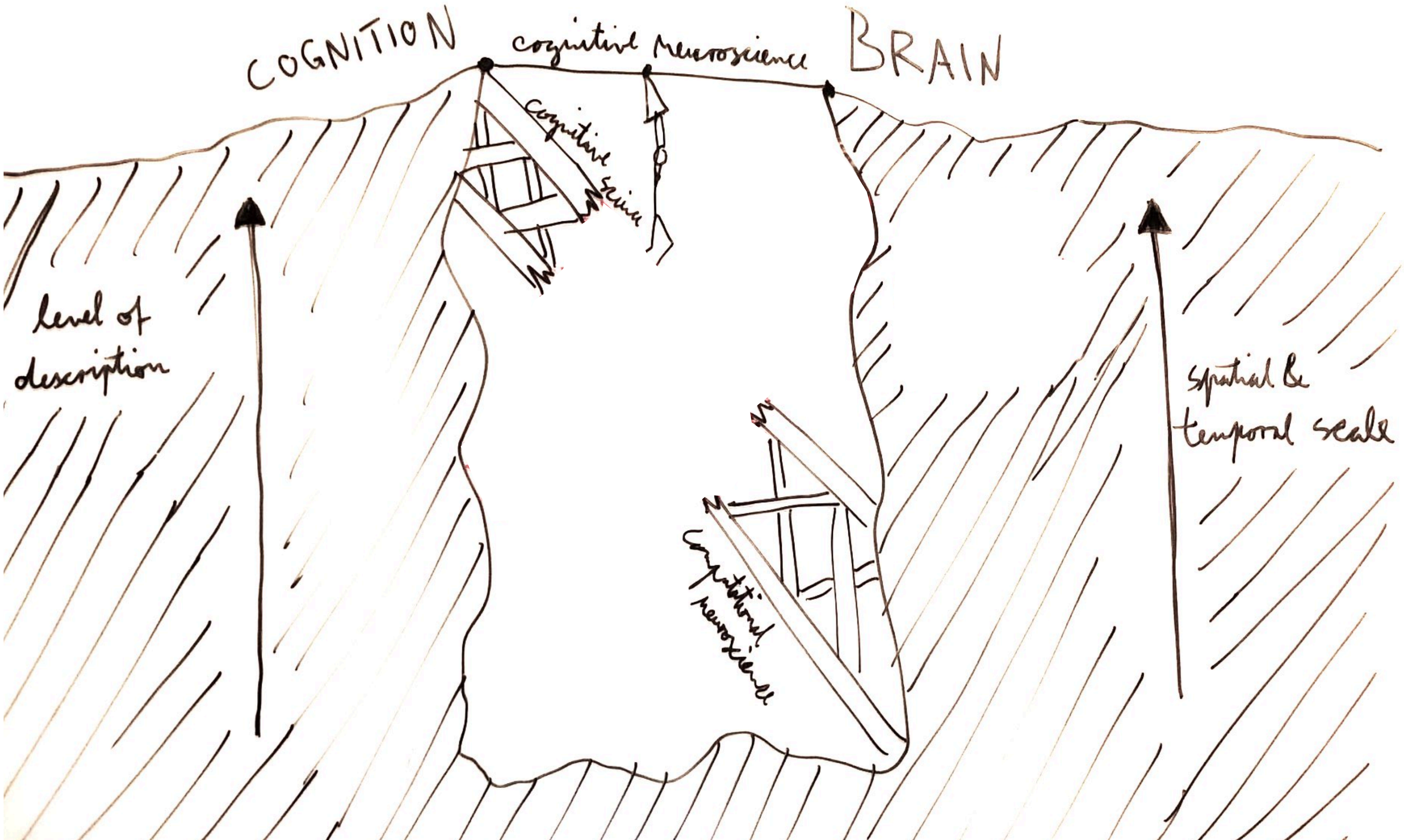
Lecture 14: Computational Cognitive Modeling

Computational Cognitive Neuroscience

course website:
https://brendenlake.github.io/CCM-site/

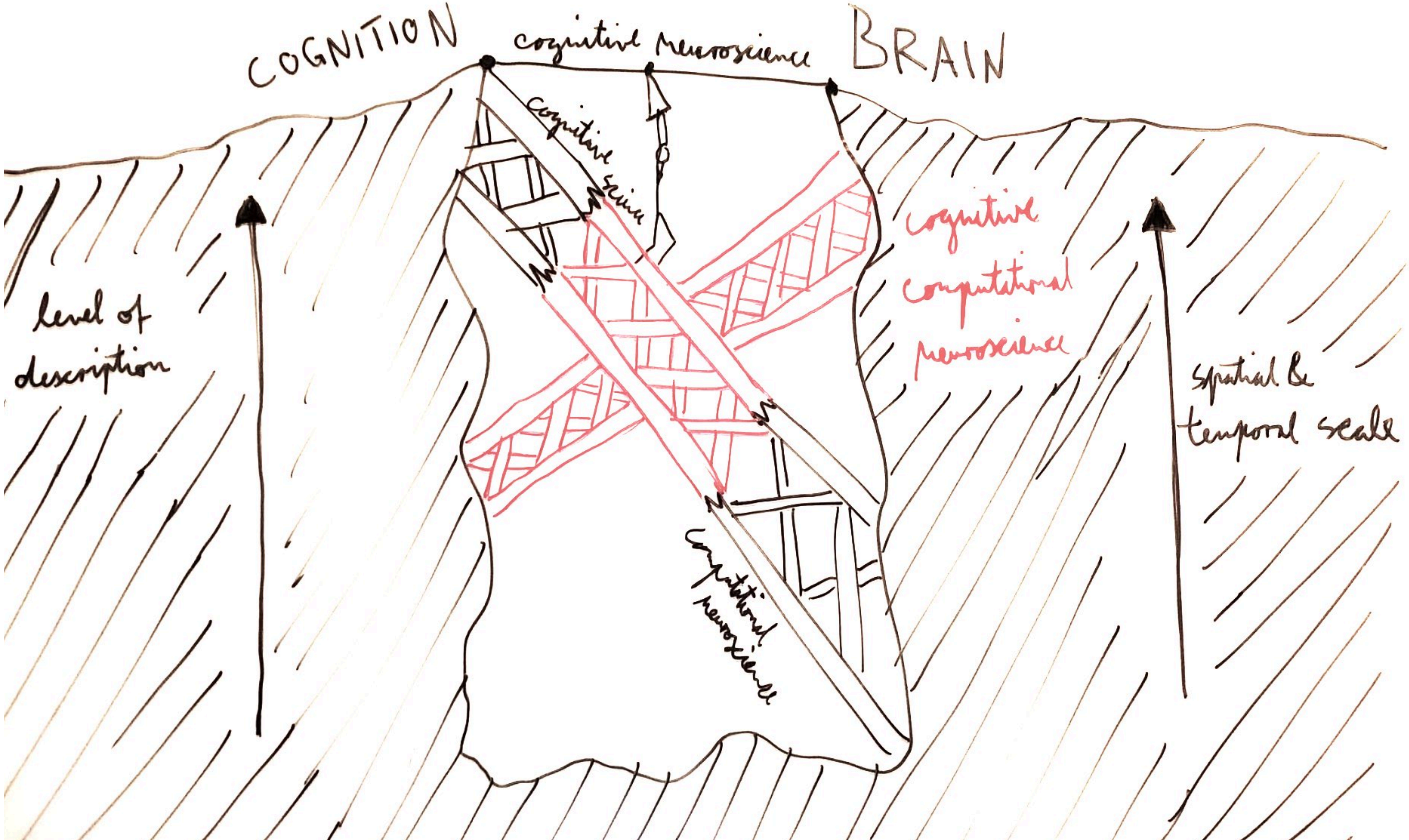
The “Chasm”

Illustration by @NKriegeskorte



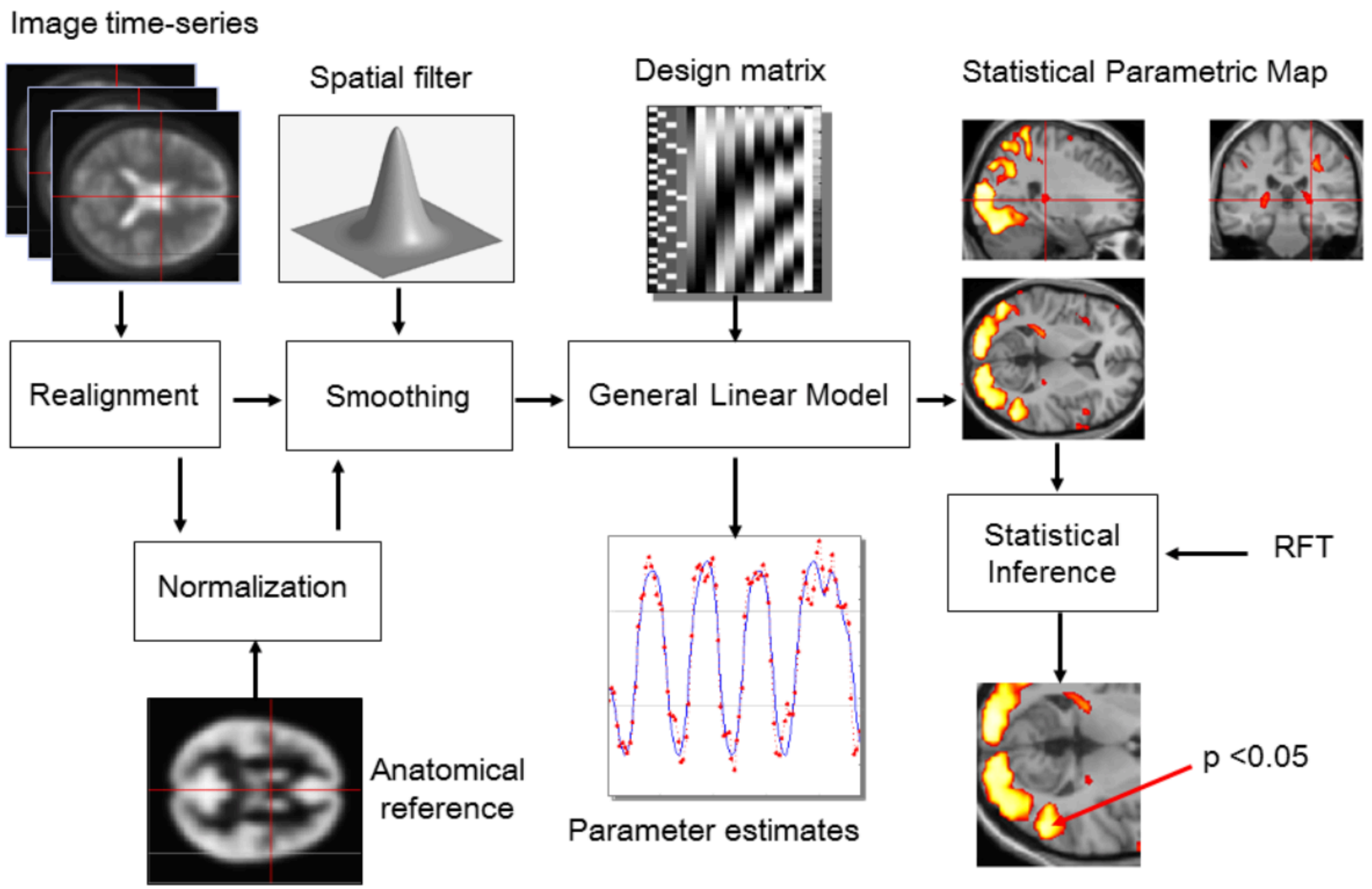
The “Chasm”

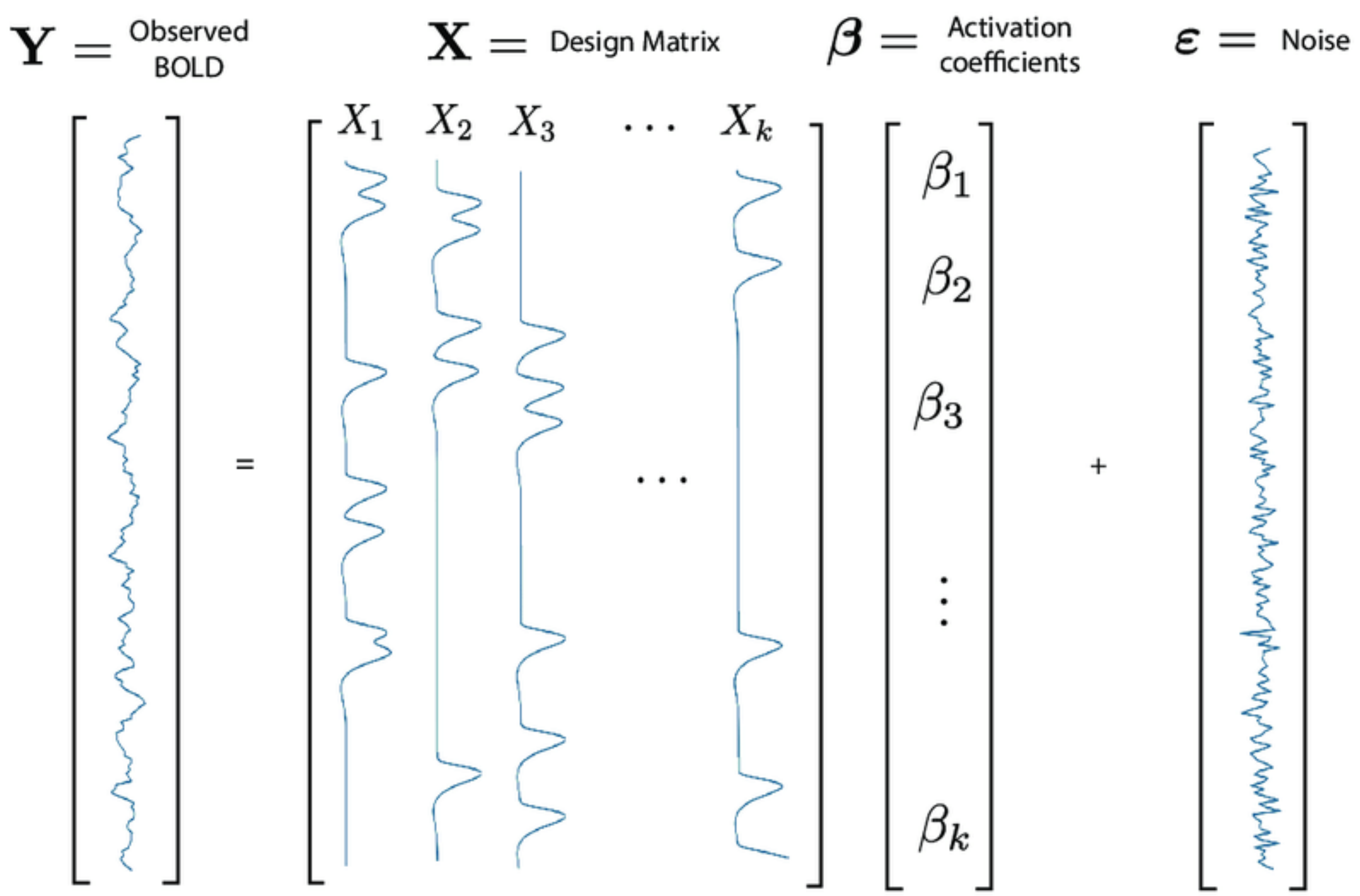
Illustration by @NKriegeskorte



Classic approach

Directly relate operational definitions of psychological constructs to the brain:





Novelty and Reward

- Novelty bonuses (Kakade & Dayan, 2002)
- Daw Novelty study

Striatal Activity Underlies Novelty-Based Choice in Humans

Bianca C. Wittmann,[1,3,*] Nathaniel D. Daw,[2,3,4] Ben Seymour,[1] and Raymond J. Dolan[1]
[1]Wellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, London WC1N 3BG, UK
[2]Gatsby Computational Neuroscience Unit, University College London, 17 Queen Square, London WC1N 3AR, UK
[3]These authors contributed equally to this work
[4]Present address: Center for Neural Science and Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA
*Correspondence: b.wittmann@ucl.ac.uk
DOI 10.1016/j.neuron.2008.04.027

Table 1. Parameter Estimates for the Behavioral Model, Shown as Mean (Over Subjects) ± 1 SE

Learning rate ν	0.23 ± 0.038
Softmax inv. temperature β	8.5 ± 1.2
Initial value, familiarized Q_f	0.37 ± 0.071
Initial value, novel Q_n	0.41 ± 0.076

Due to poor identification of β and ν, one subject is omitted from these averages.

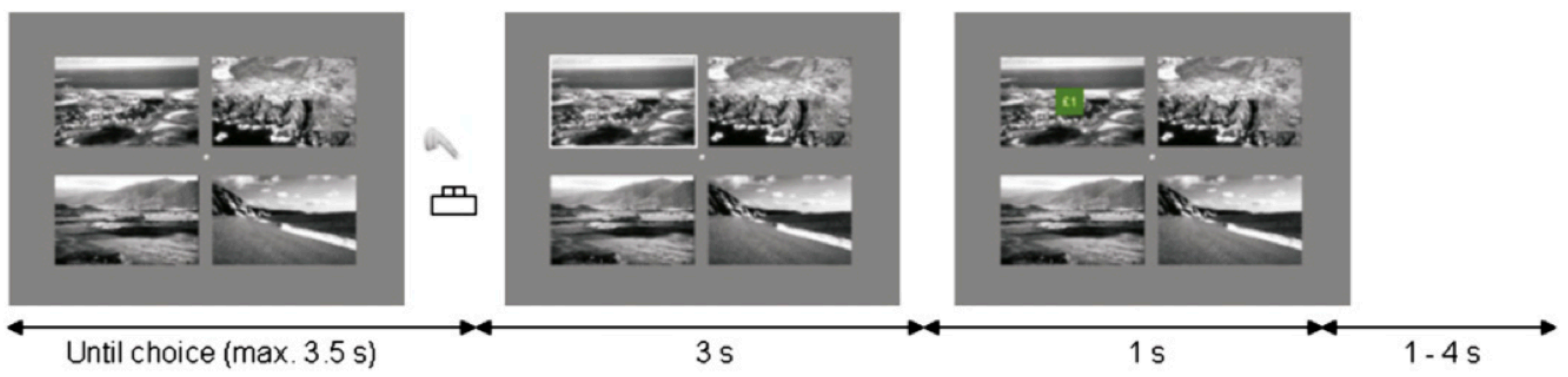


Figure 1. Experimental Design
Following a familiarization phase, participants were shown four pictures on each trial and asked to choose one. Both familiarized and novel pictures were presented at randomized locations that changed on each trial. Each picture was repeated for an average of 20 trials and then replaced. Participants were informed that each picture had been assigned a unique probability of winning £1 that would not change as long as that picture was repeated. They were given feedback at the end of each trial indicating whether they had won or received nothing.

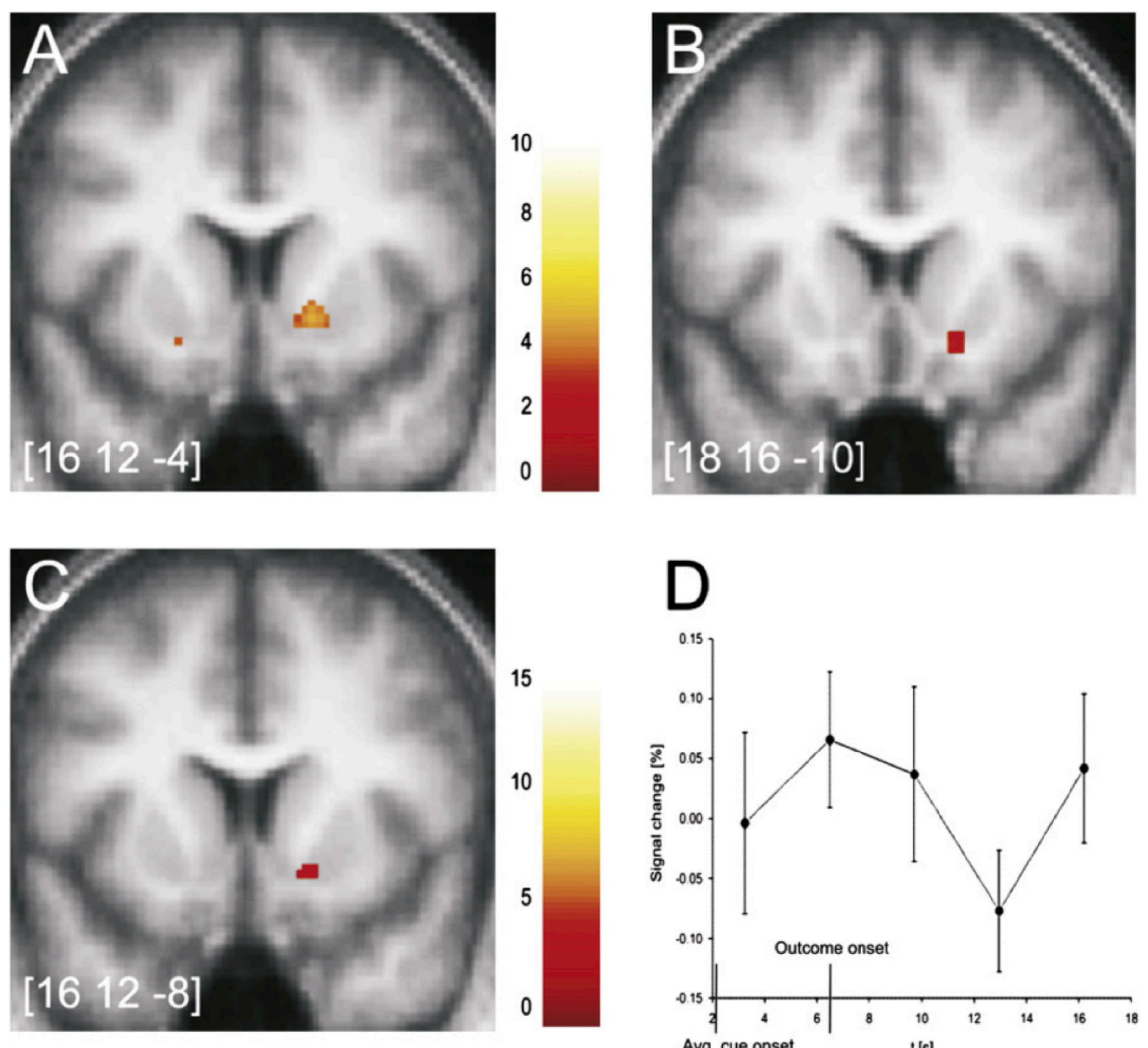
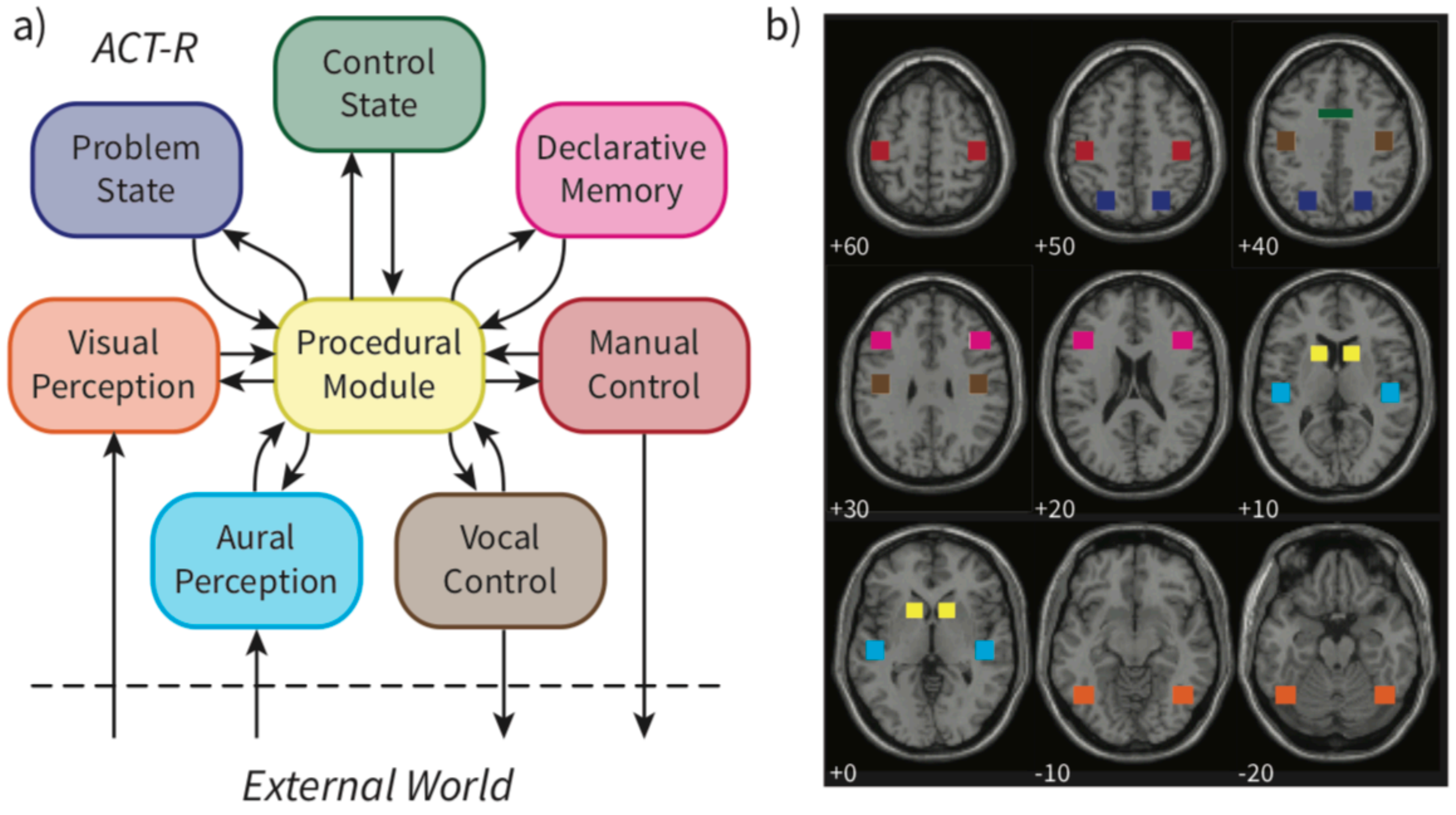
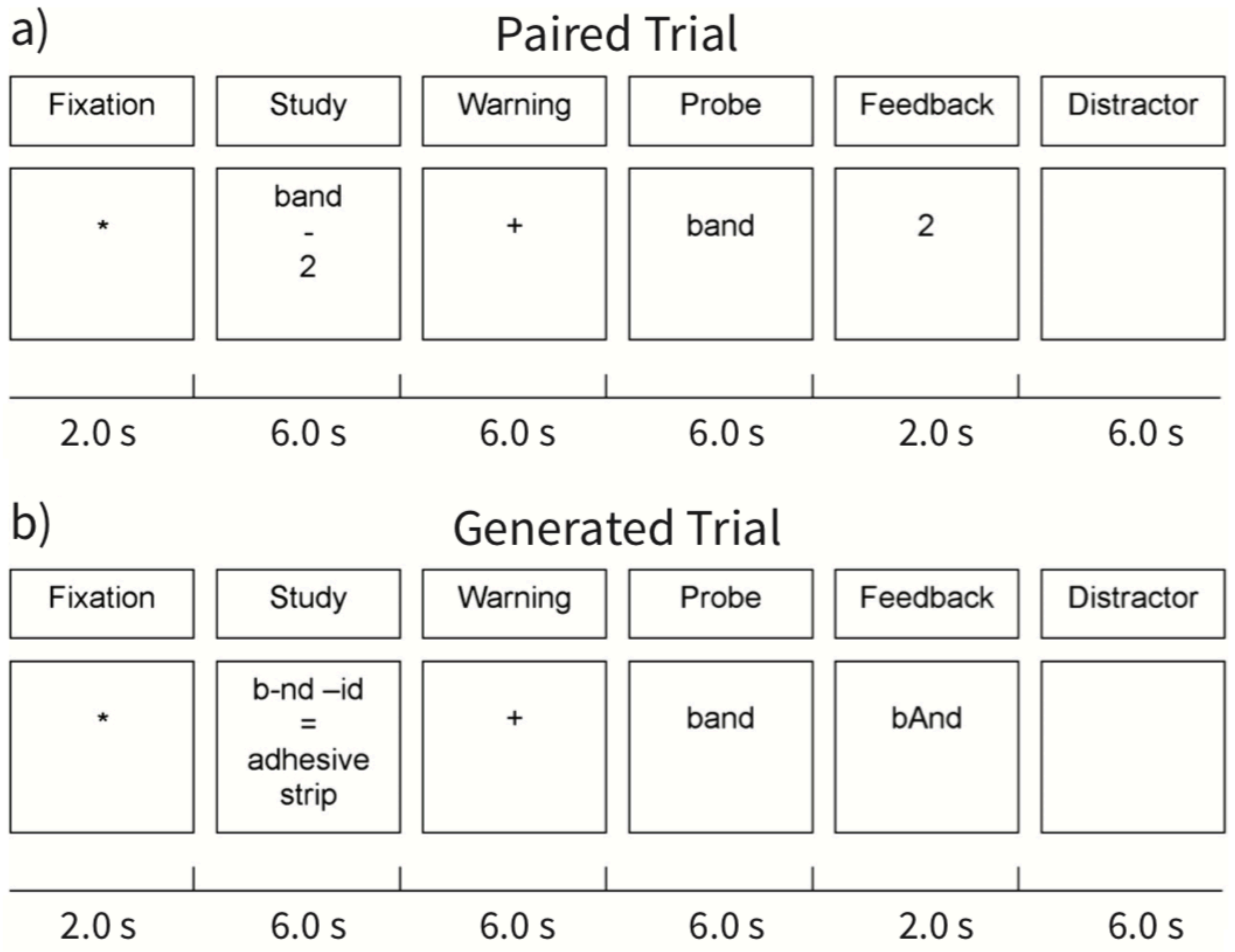
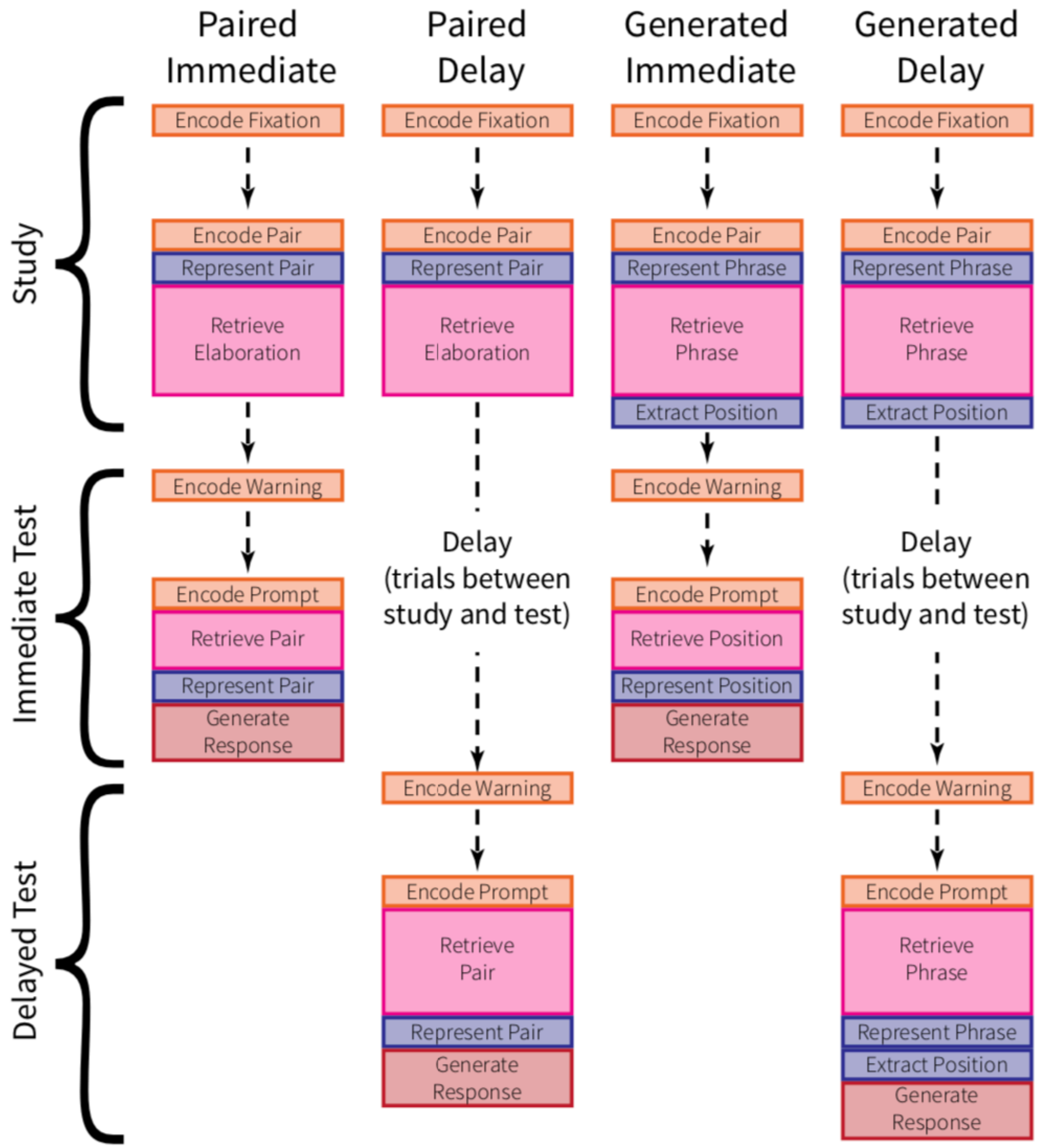


Figure 2. Ventral Striatal Response to Prediction Error and Novelty
Peak coordinates are given in MNI space on all images. Color bars indicate T values.
(A) Activation in right ventral striatum correlated significantly with reward prediction errors generated by the standard TD model ($p < 0.001$ uncorrected, $p < 0.05$ SVC, cluster > 5 voxels).
(B) Activation in right ventral striatum correlated significantly with additional prediction error due to inclusion of a novelty bonus ($p < 0.001$ uncorrected, $p < 0.05$ SVC, cluster > 5 voxels).
(C) Significant overlap between activation in right ventral striatum for the novelty bonus (see [B]) and activation obtained for standard model (see [A]) derived by inclusively masking (B) with (A) ($p < 0.005$, uncorrected, for both contrasts, cluster > 5 voxels).
(D) Striatal activation time courses calculated for the first two trials a novel stimulus is chosen minus the first two choices of familiar stimuli, shown for the peak voxel correlating with the novelty bonus (MNI coordinates: 14, 20, −10). Trials are aligned by the time of reward outcome at 6.5 s; the average stimulus onset time is also indicated. Error bars indicate SEM.

Process decoding







Process decoding

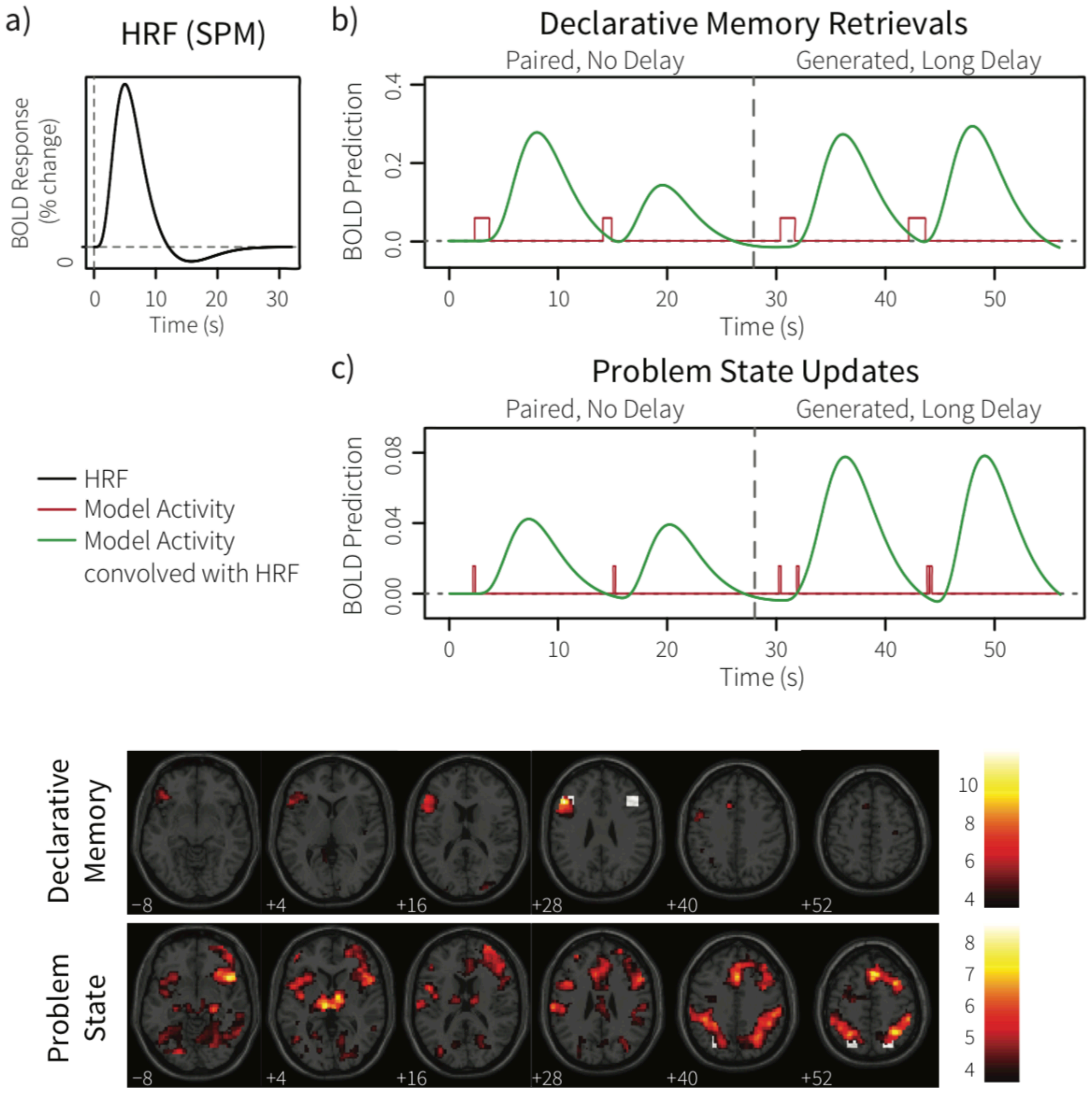


Figure 8. Model-based fMRI results. Statistical maps were thresholded at $p < .001$ (uncorrected). White squares indicate predefined ACT-R regions.

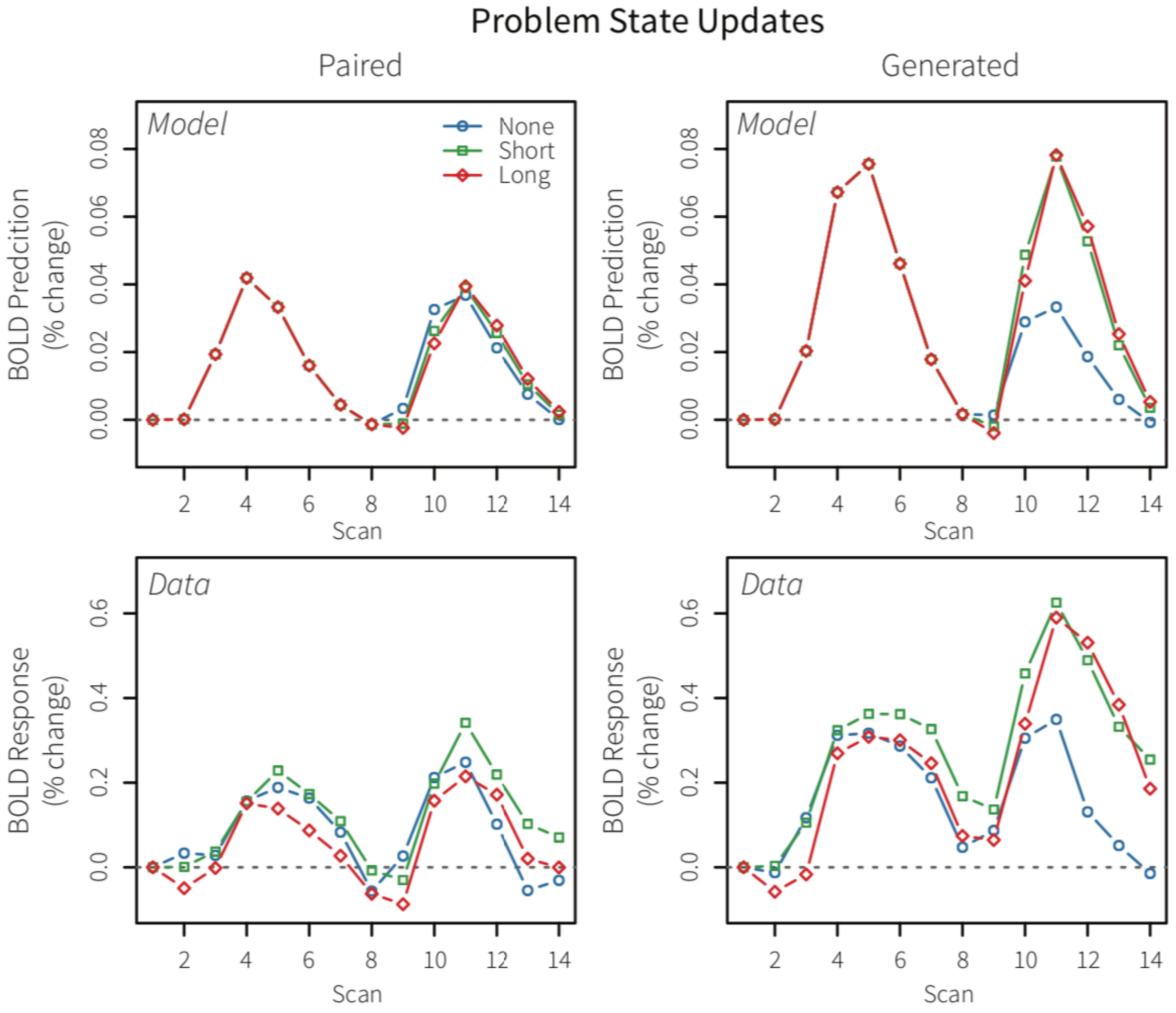
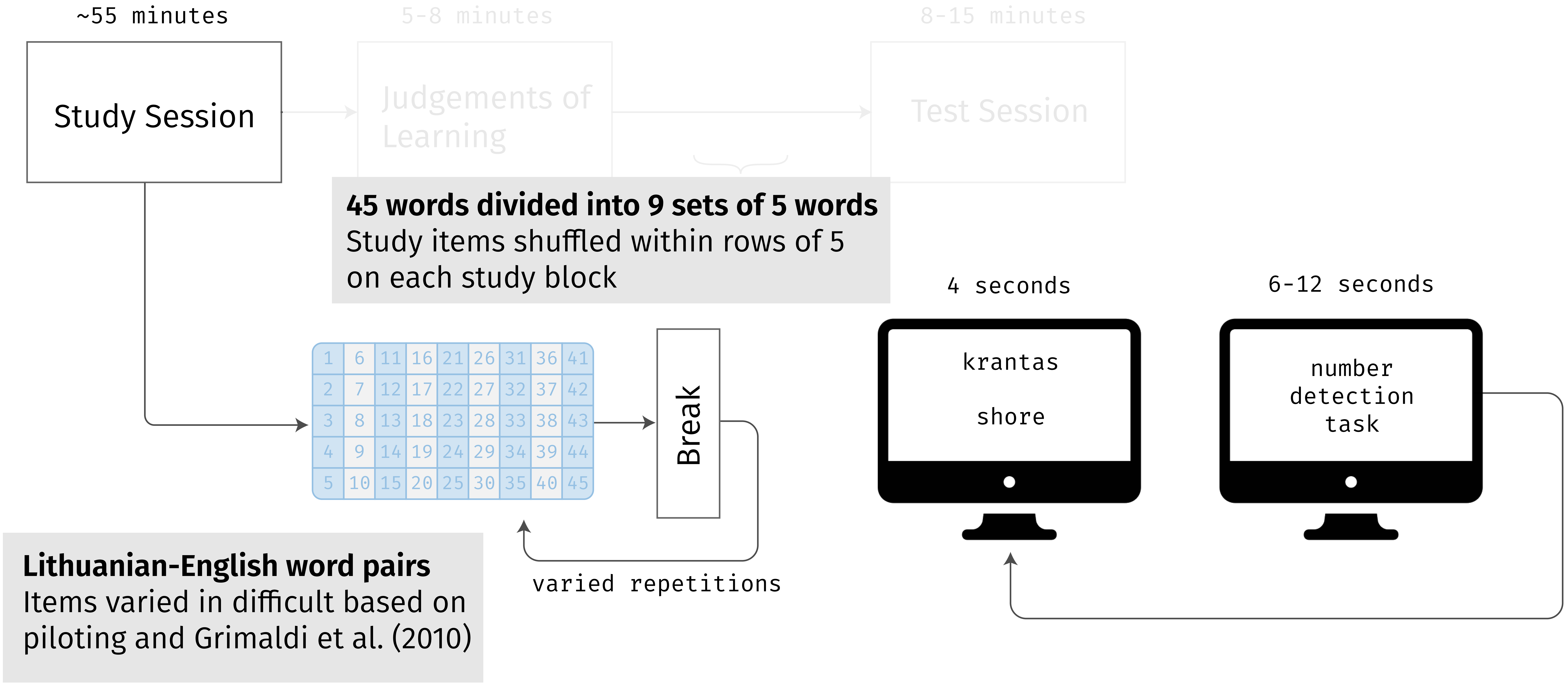
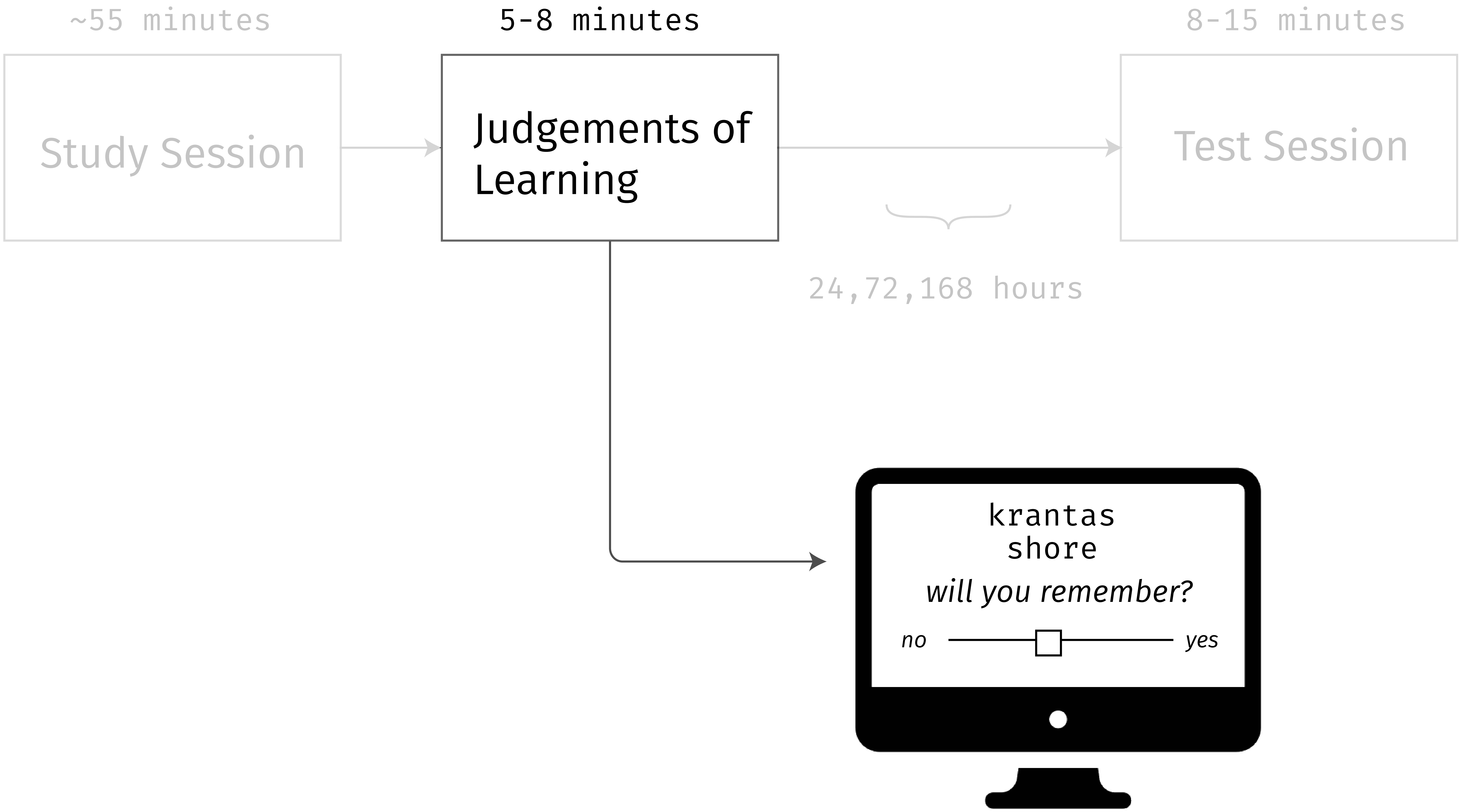


Figure 7. ROI results for problem state updates. Top panels show model predictions; bottom panels data. 1 scan = 2 seconds.

Experiment Design



Experiment Design



Experiment Design

~55 minutes

Study Session

5-8 minutes

Judgements of Learning

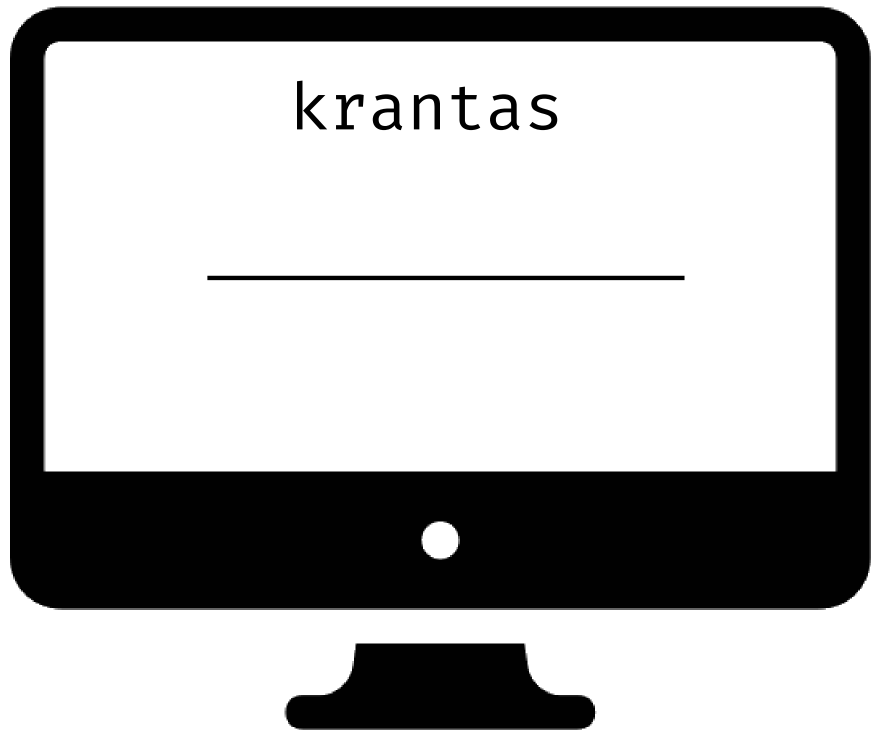
24,72,168 hours

8-15 minutes

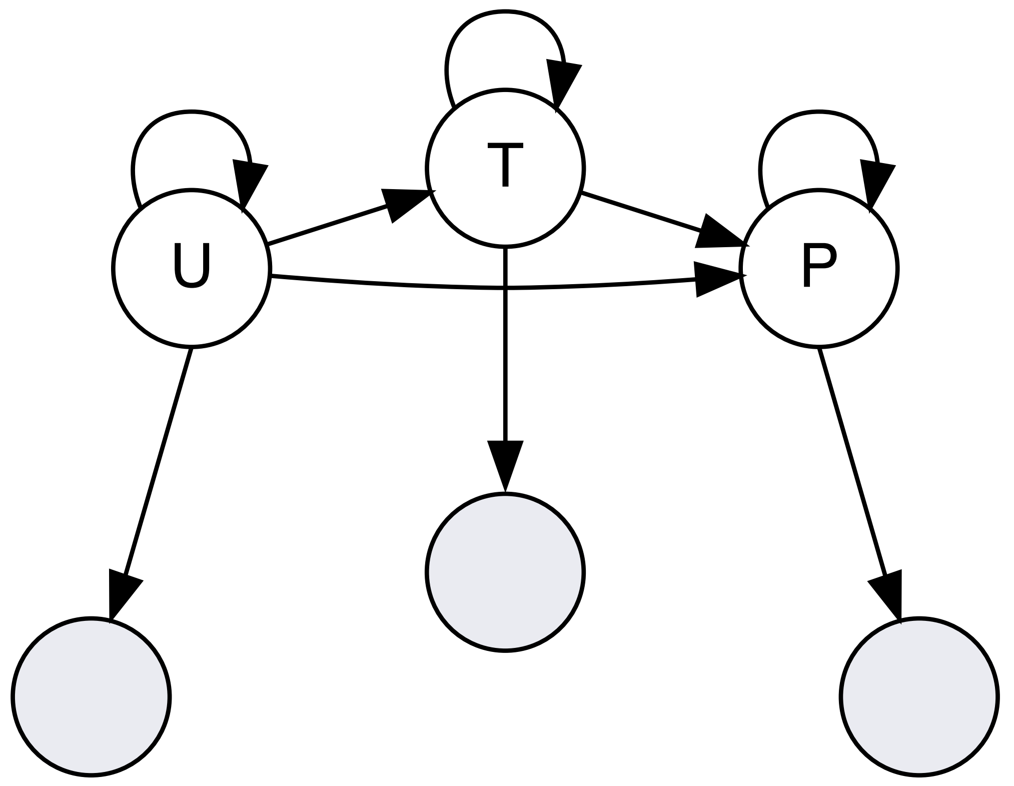
Test Session

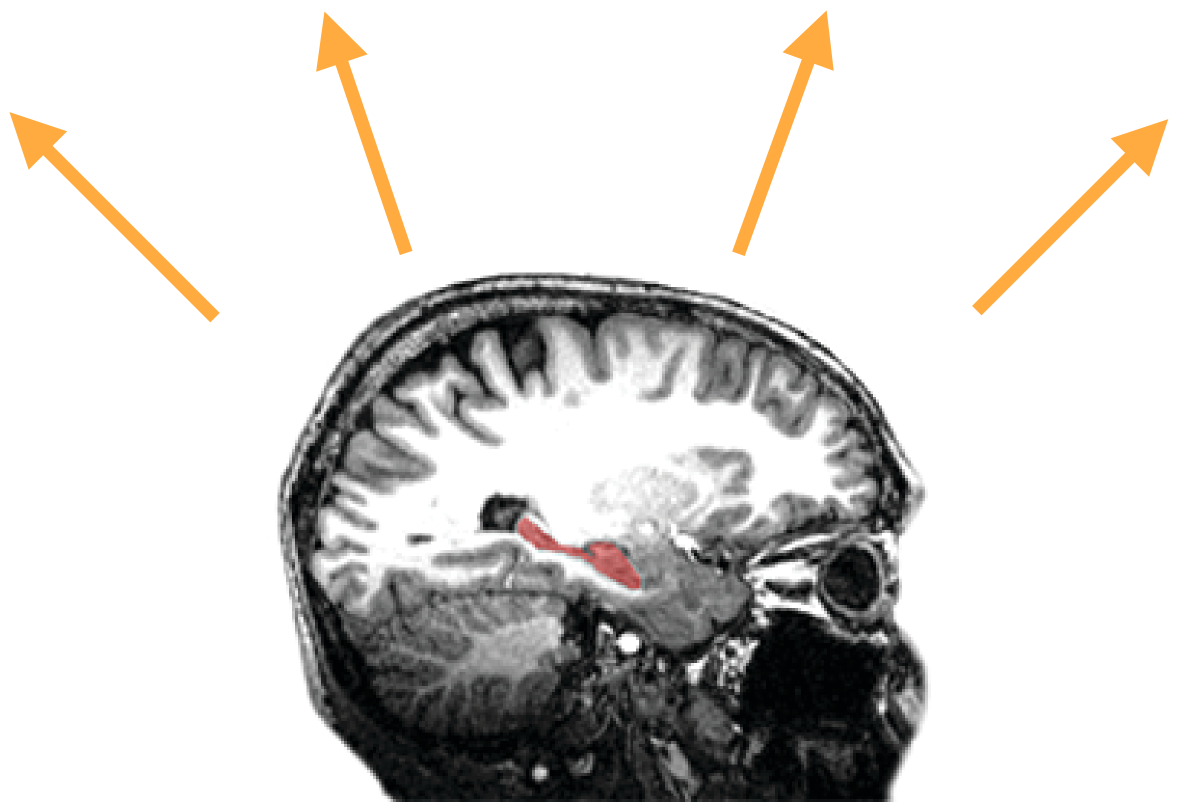
Delay between study and test varied
Different subjects came back on different days, all > 24 hours.

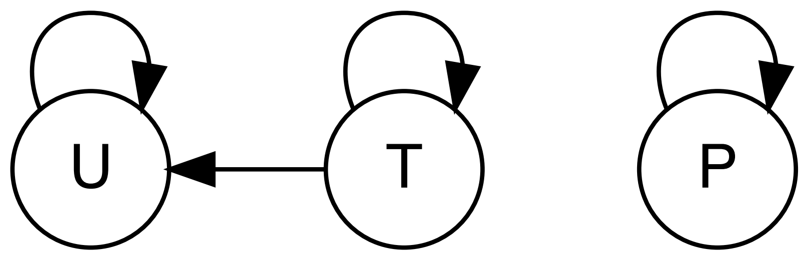
Cued Recall Test
Type out response with keyboard within 12 seconds



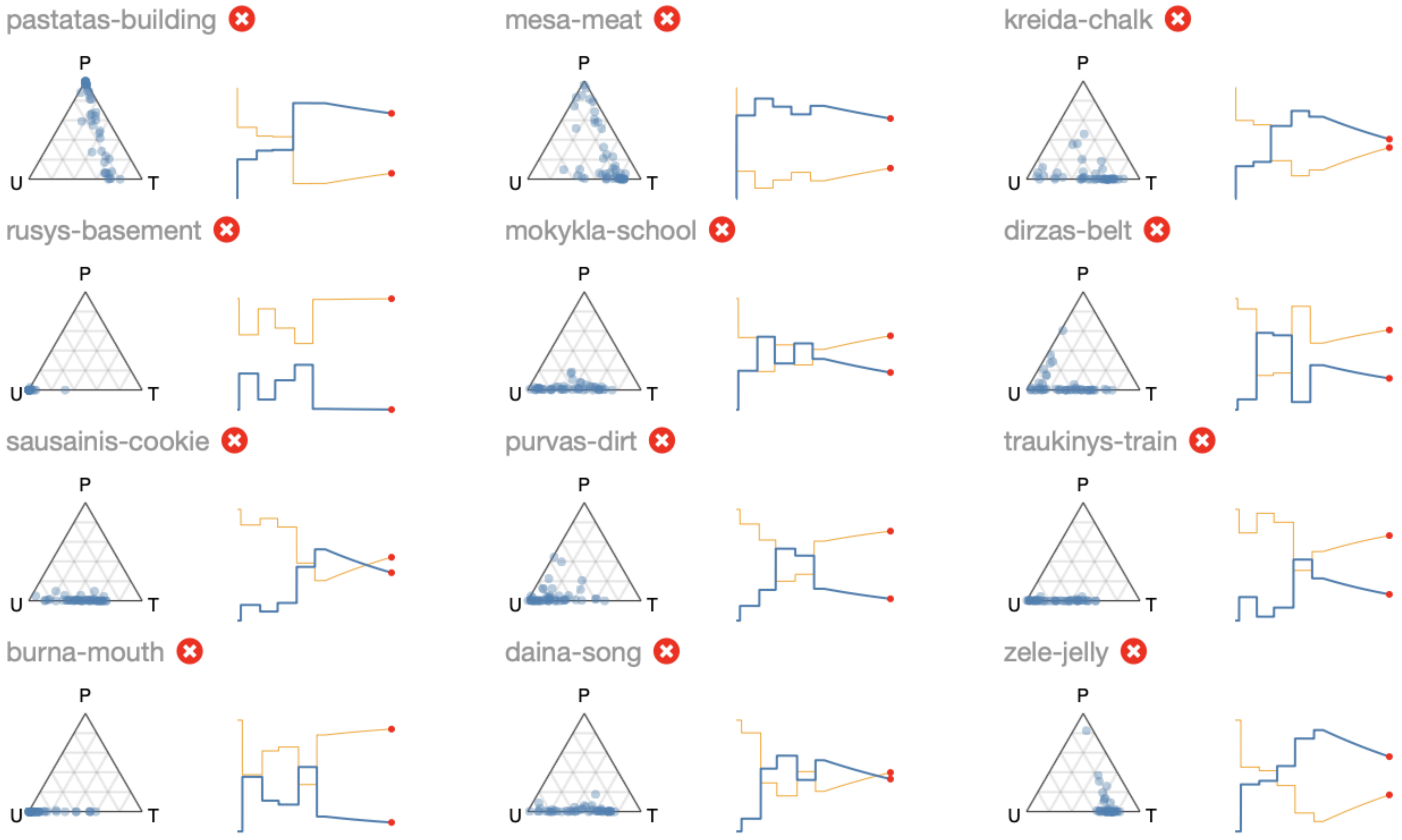
Process decoding





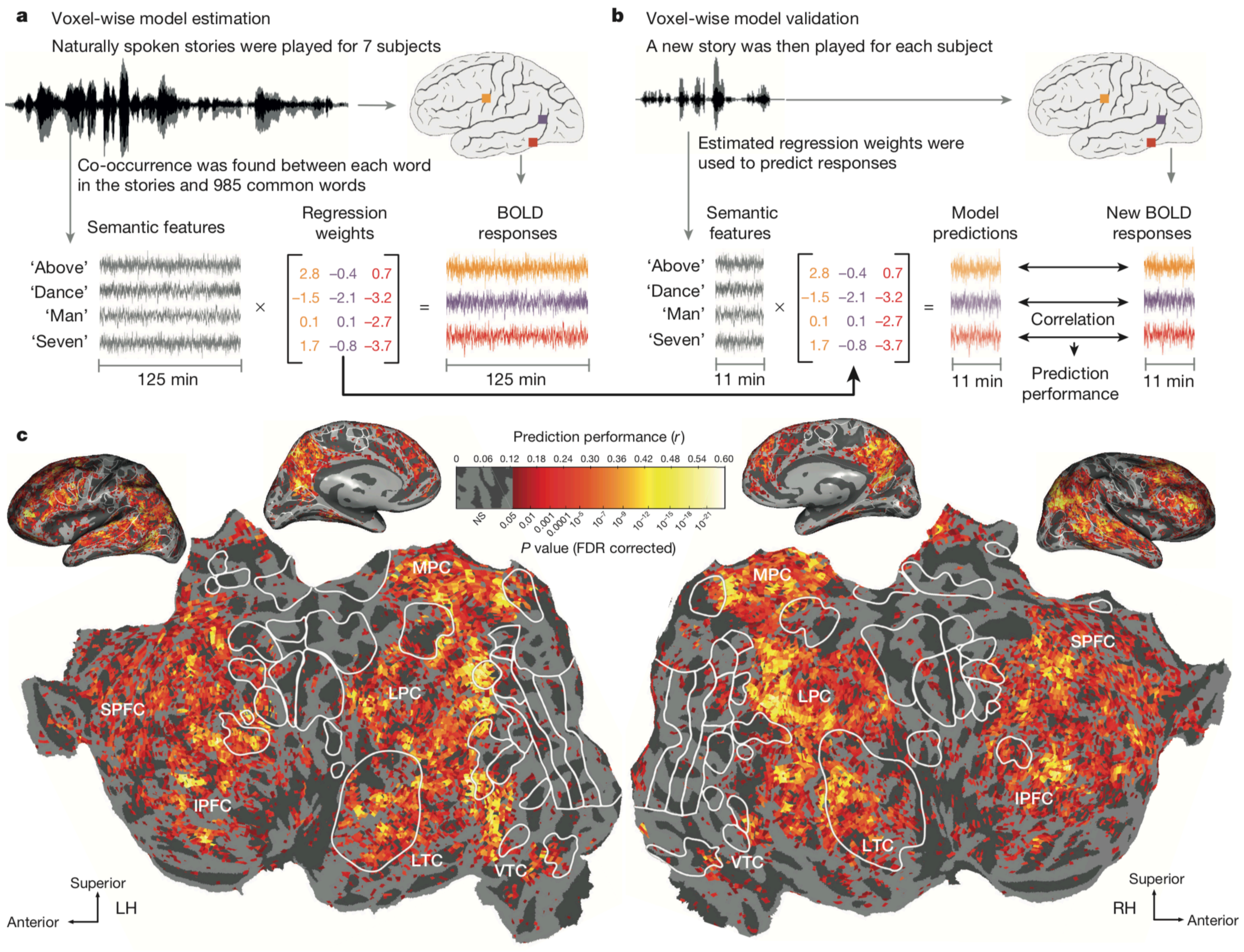


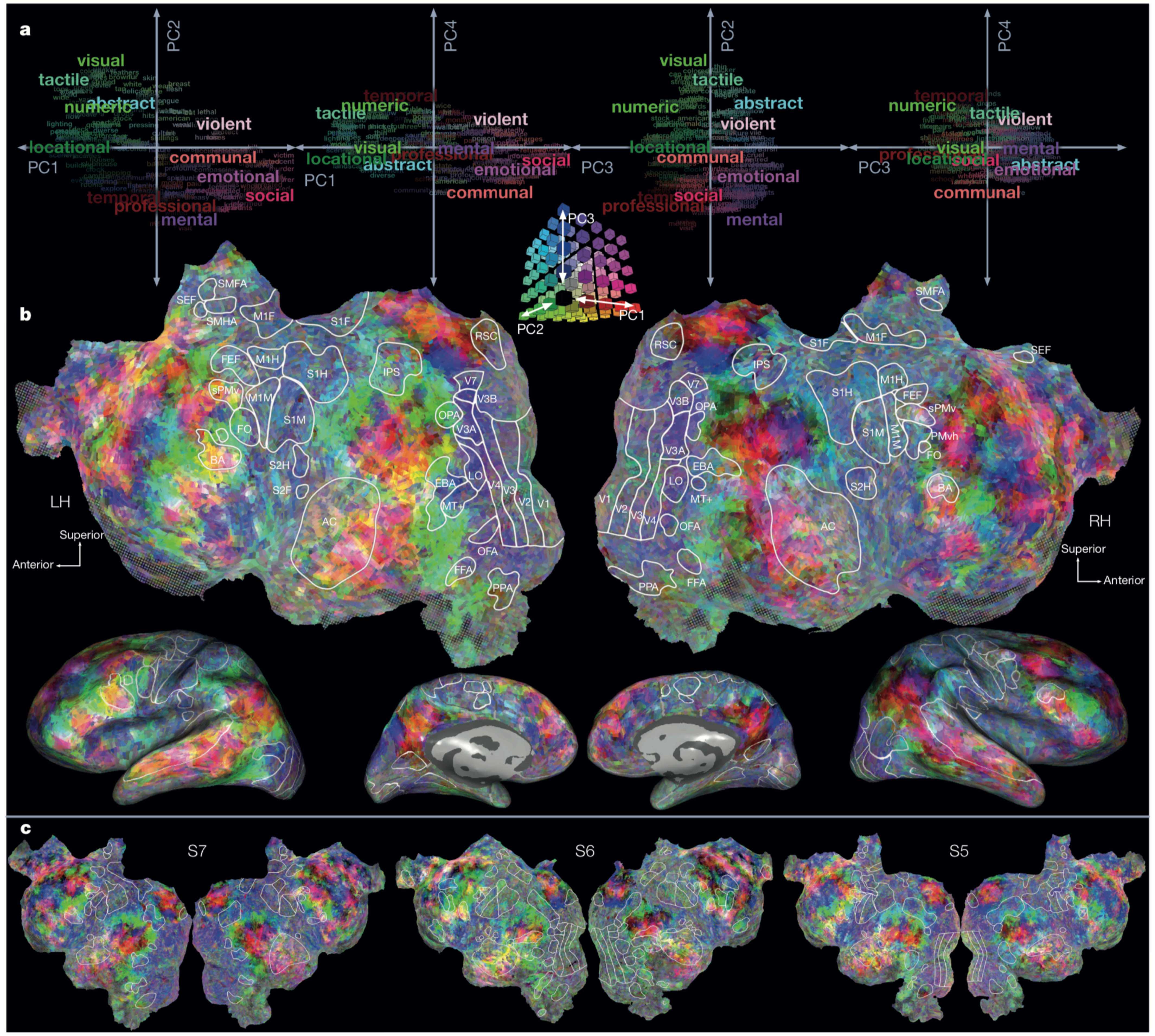
gureckislab.org/omni



Halpern, D., Tubridy, S., Wang, H.Y., Gasser, C., Popp, P.J.O., Davachi, L., & Gureckis, T.M. (2018). Knowledge Tracing Using the Brain. Educational Data Mining 2018. Buffalo, NY.

Brain mapping of semantic space





Natural speech reveals the semantic maps that tile human cerebral cortex

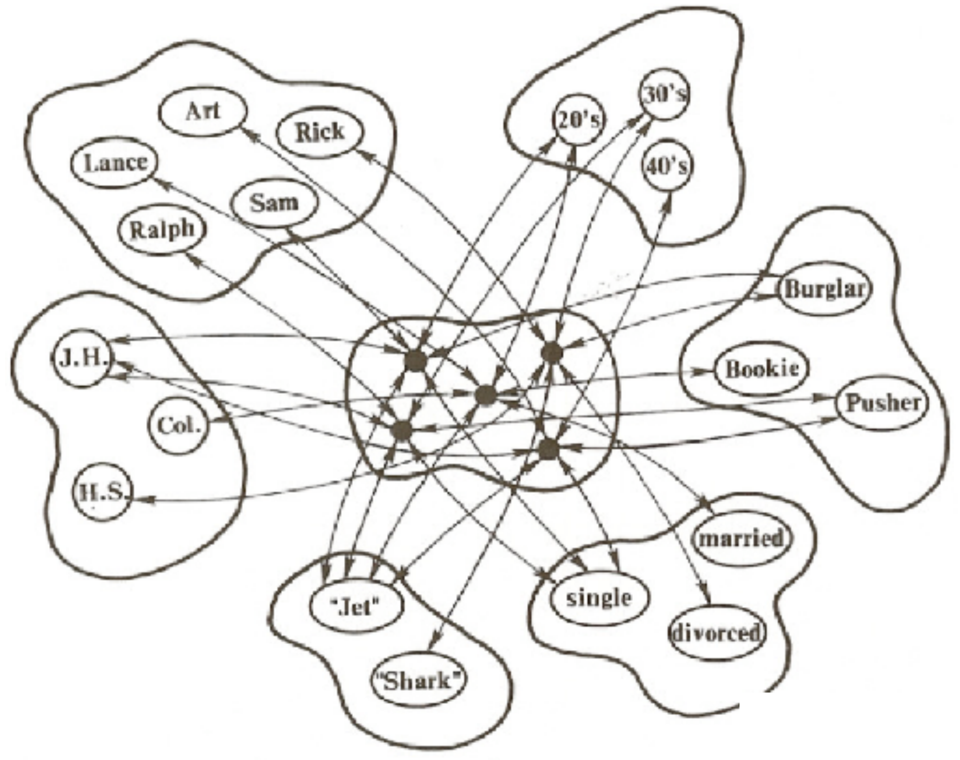
Alexander G. Huth[1], Wendy A. de Heer[2], Thomas L. Griffiths[1,2], Frédéric E. Theunissen[1,2] & Jack L. Gallant[1,2]

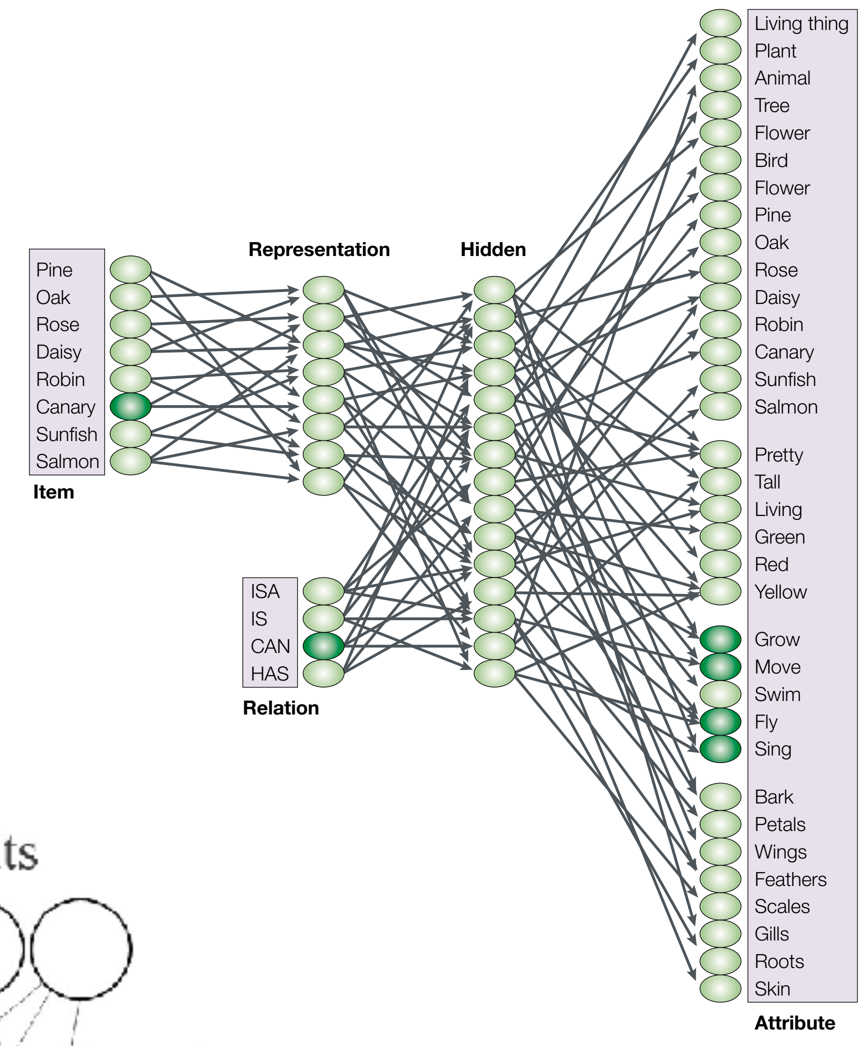
Summary

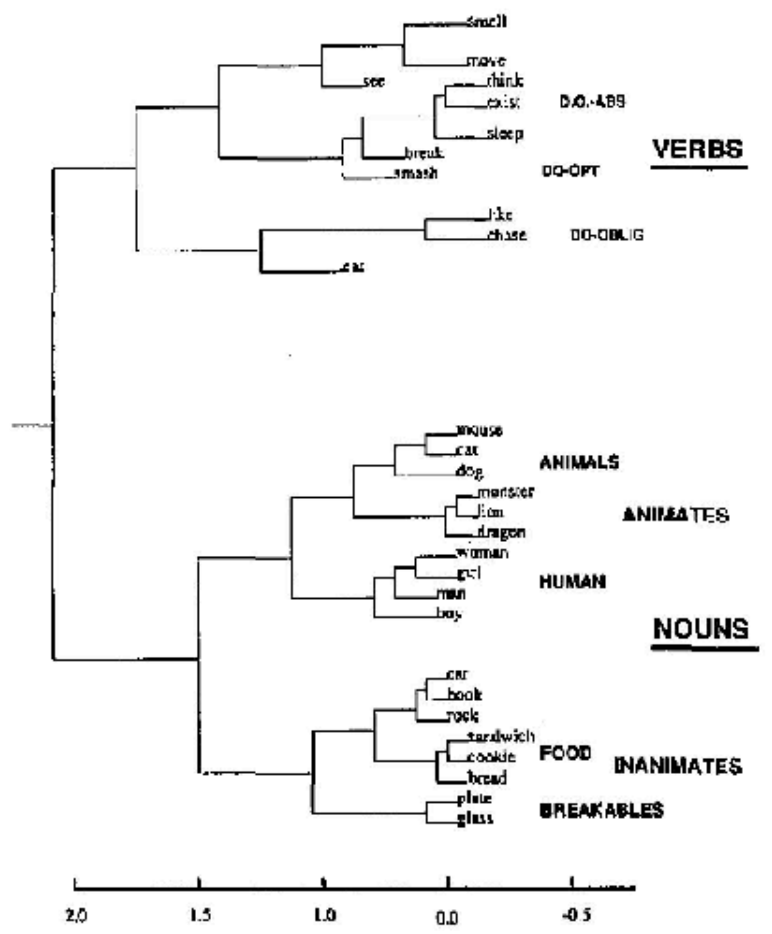
- Cognitive models might provide the bridge between Marr's levels of analyses
- Cognitive models are able to account of behavior (e.g., choices, reaction time) and thus provide strong targets for localizing and interpreting brain data
- Can possibly use brain data to adjust predictions of behavior for individual subjects
- Large scale mapping studies provide insight into the organization of semantic memory in the brain

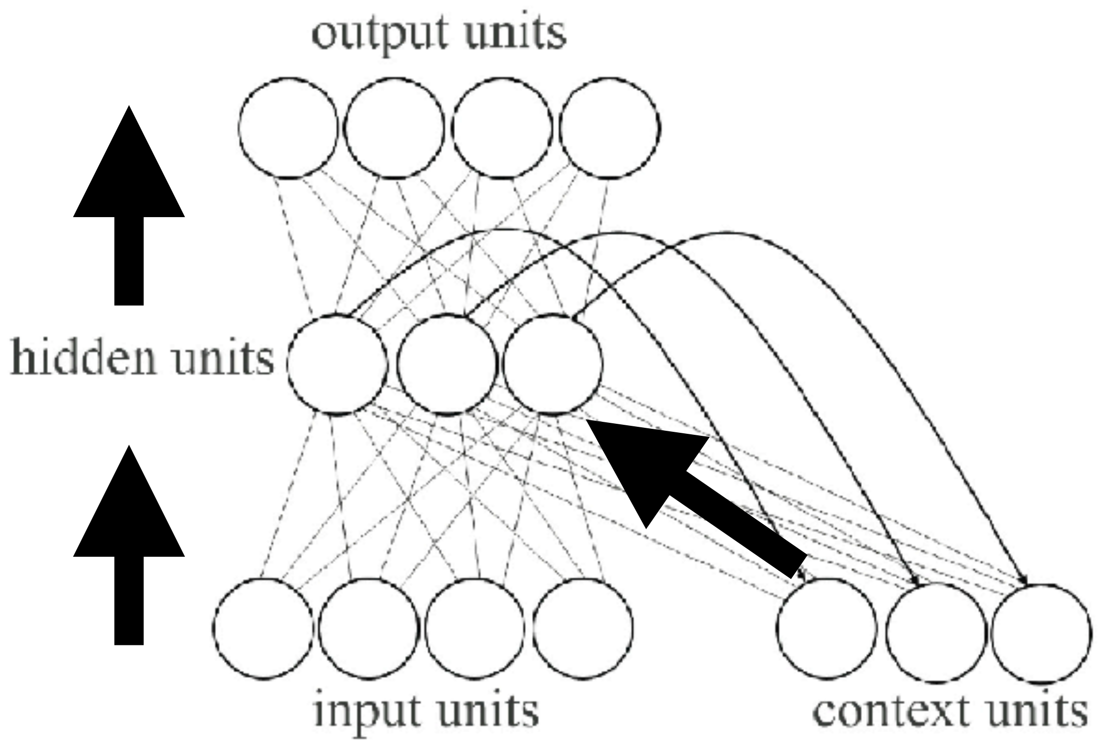
Summary

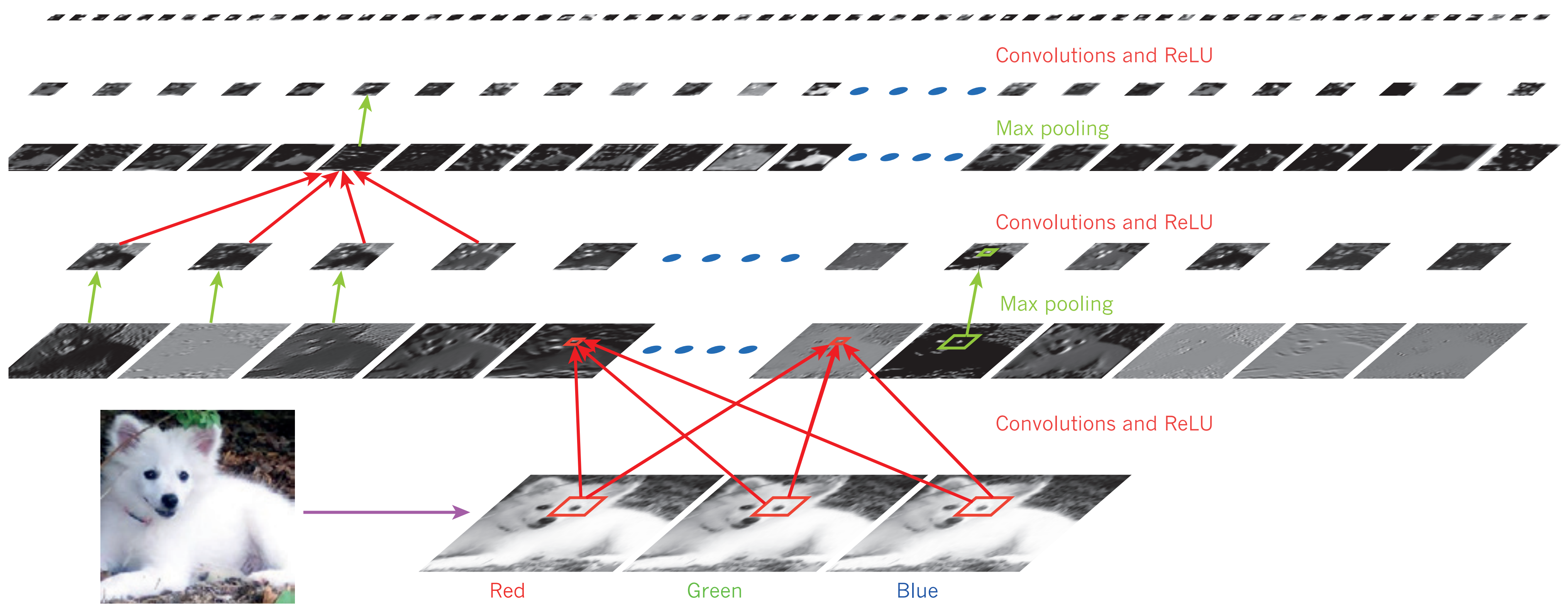
neural networks / deep learning

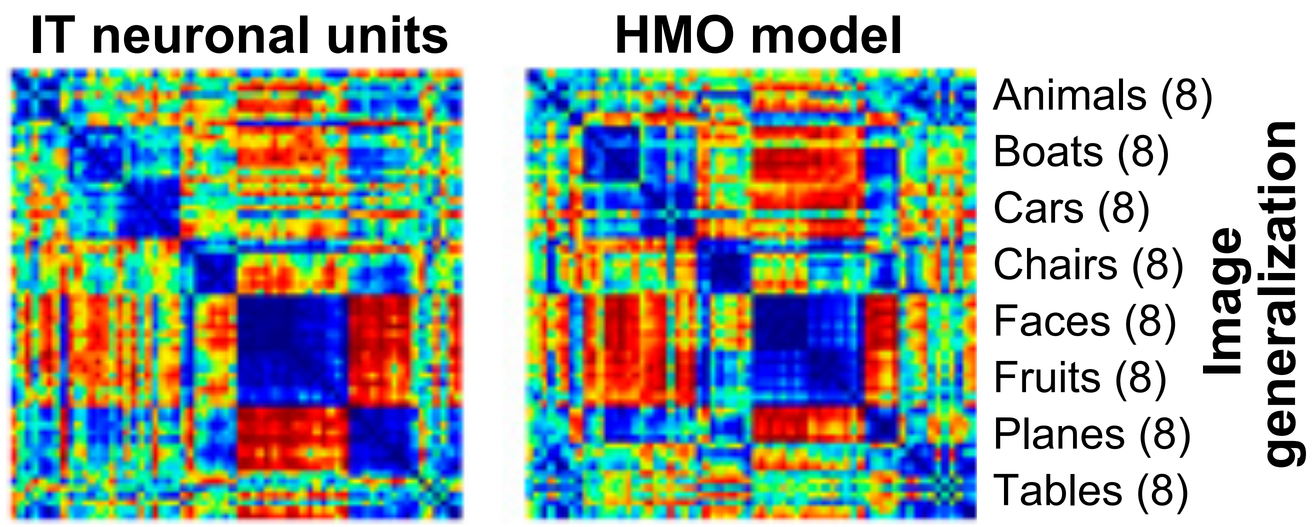




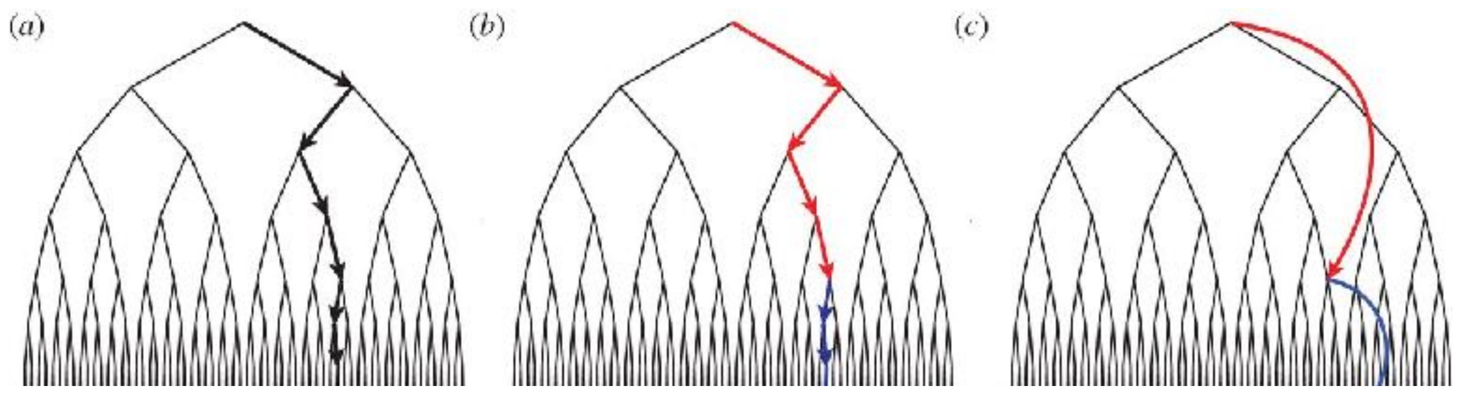


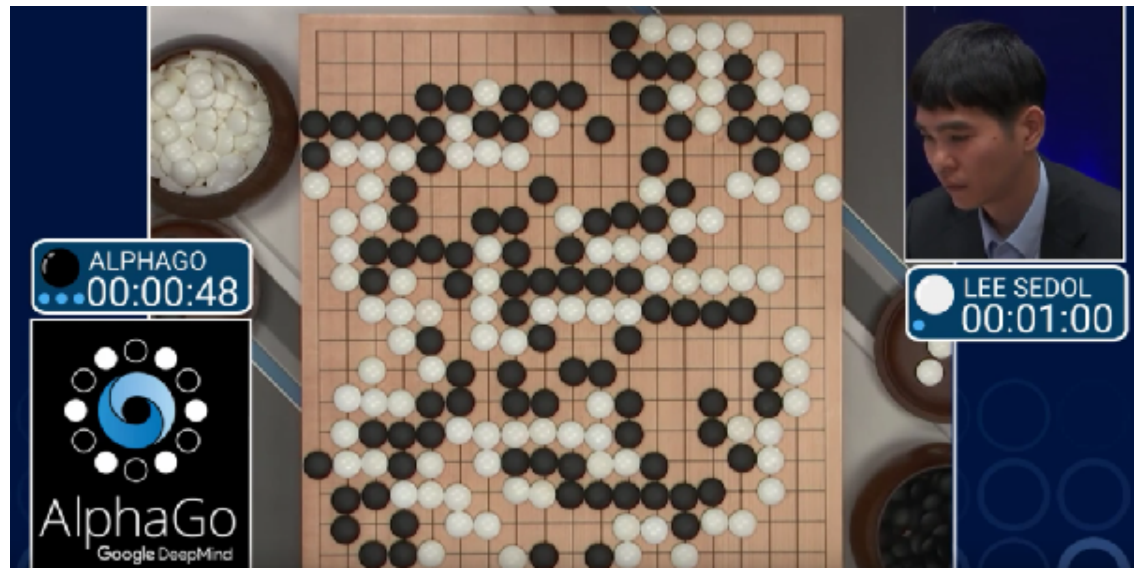


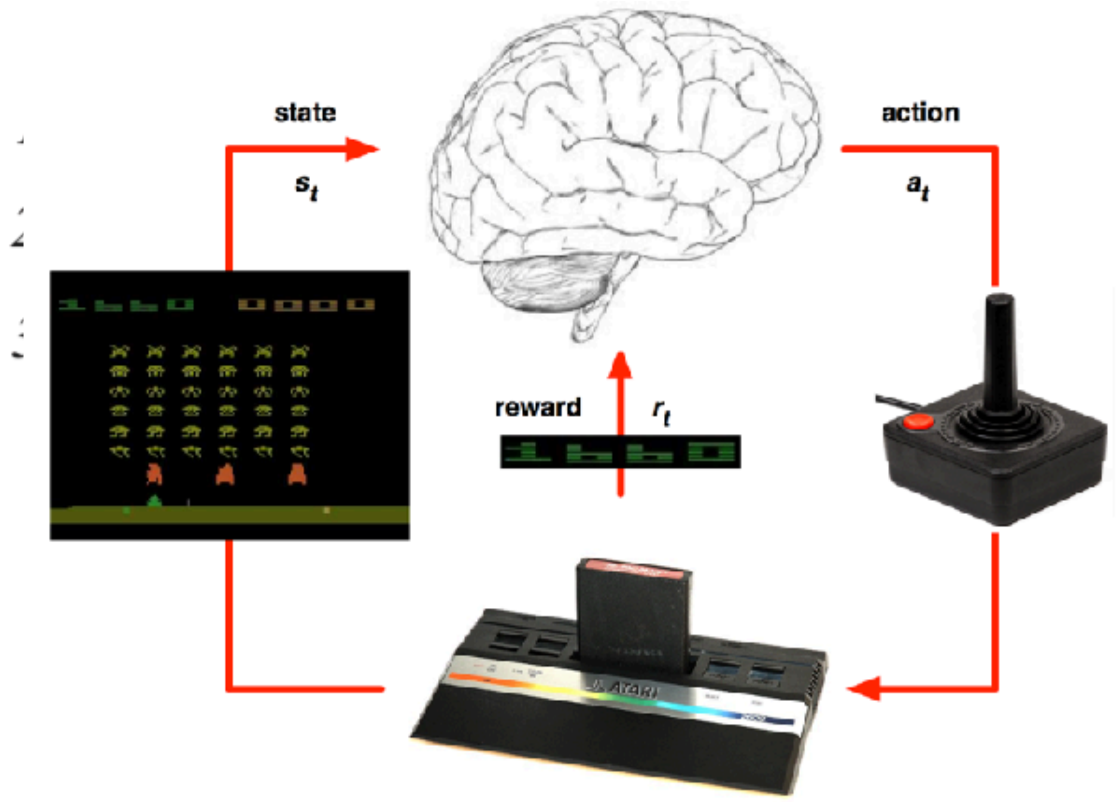


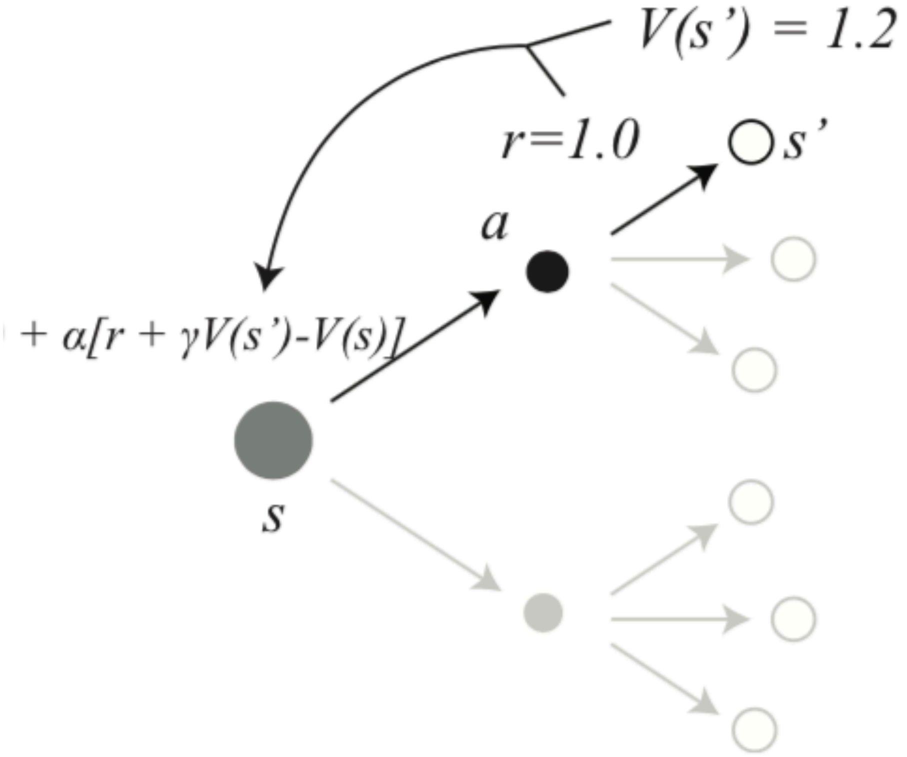


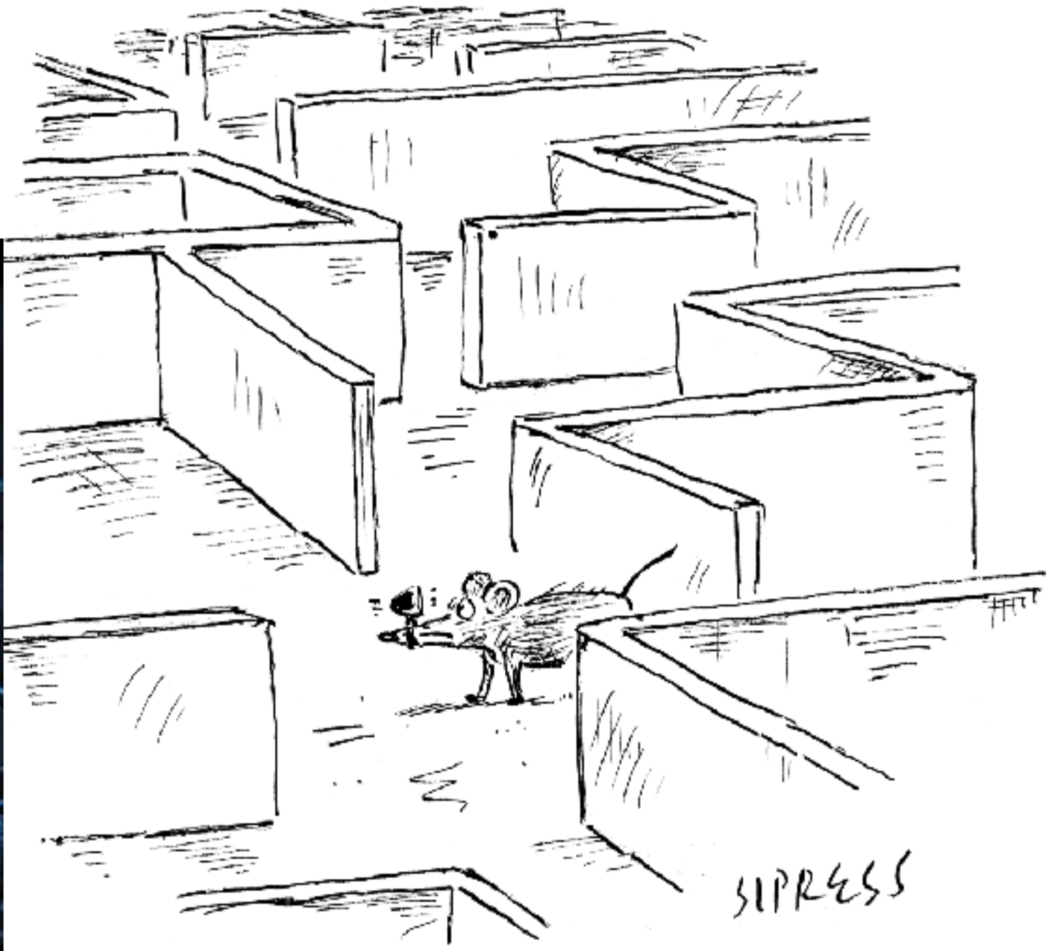
Reinforcement learning



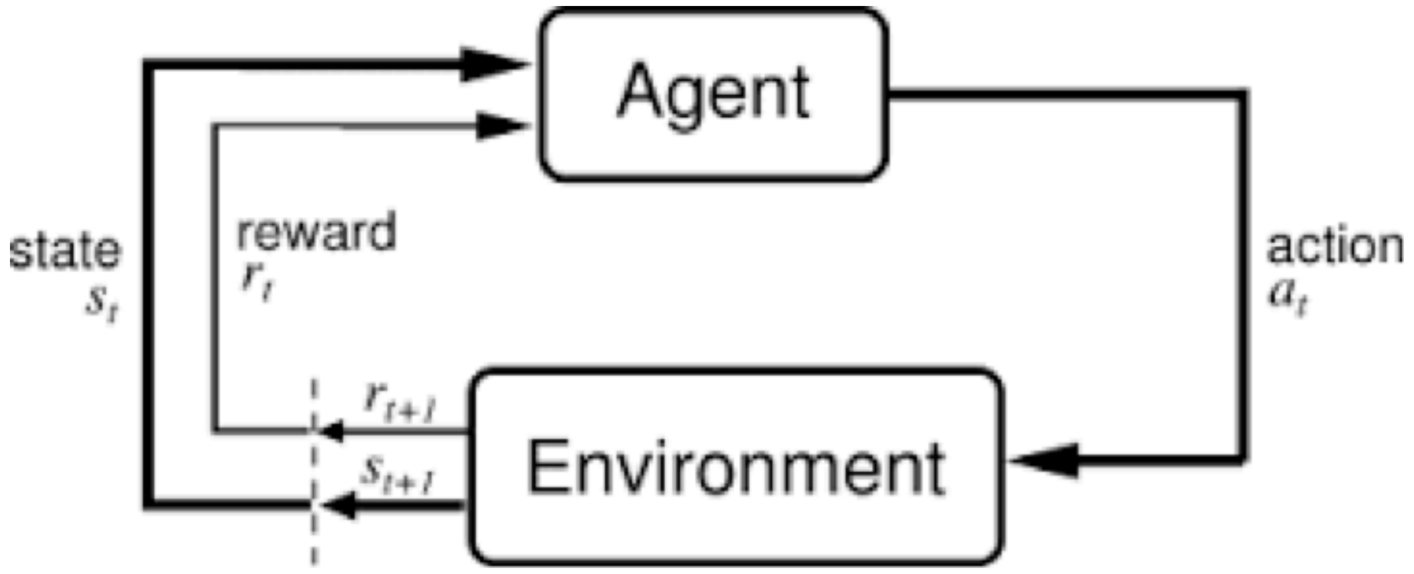


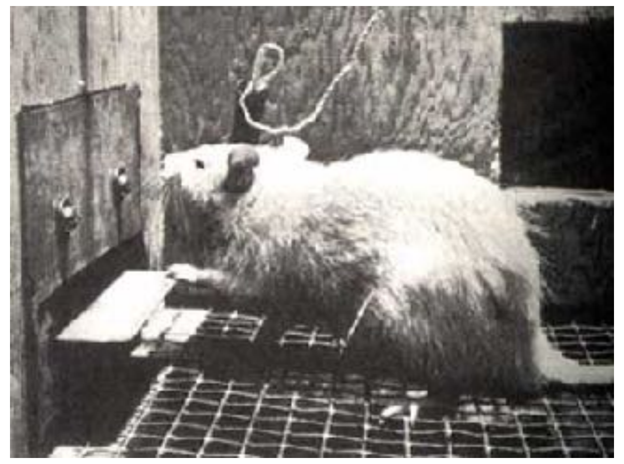






"Recalculating ... recalculating ..."



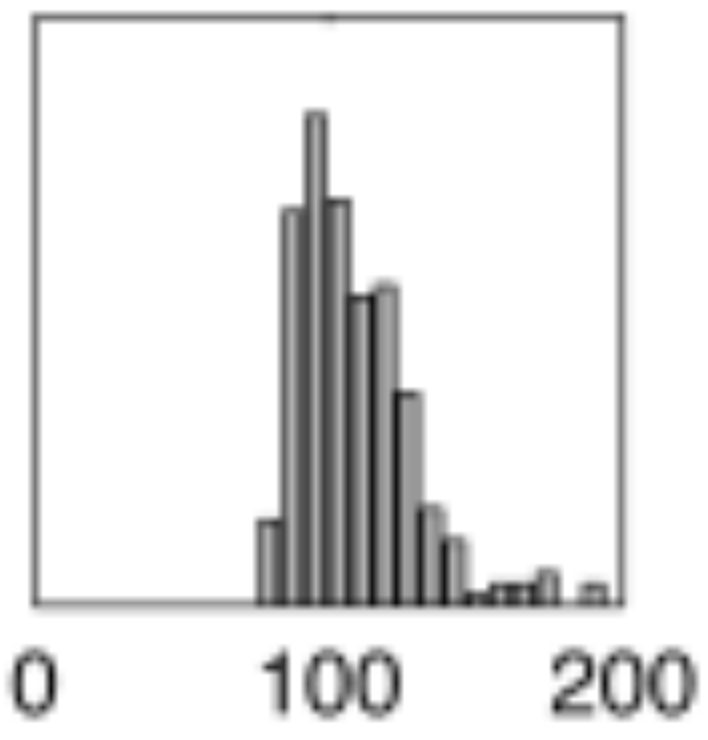


Bayesian modeling

$$P(h|D) = \frac{P(h)P(D|h)}{\sum_{h_i} P(h_i)P(D|h_i)}$$

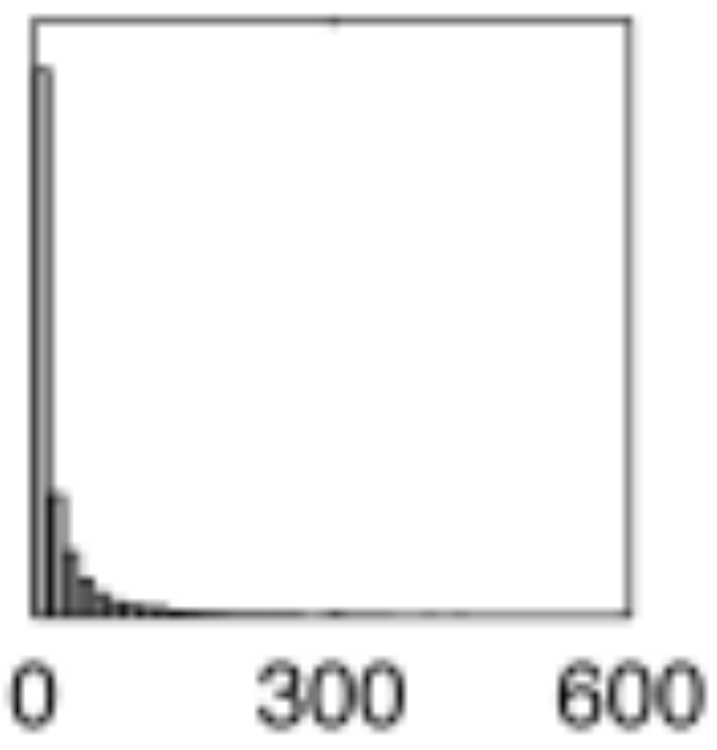
h : hypothesis $\quad D$: data

Movie runtimes (Gaussian)

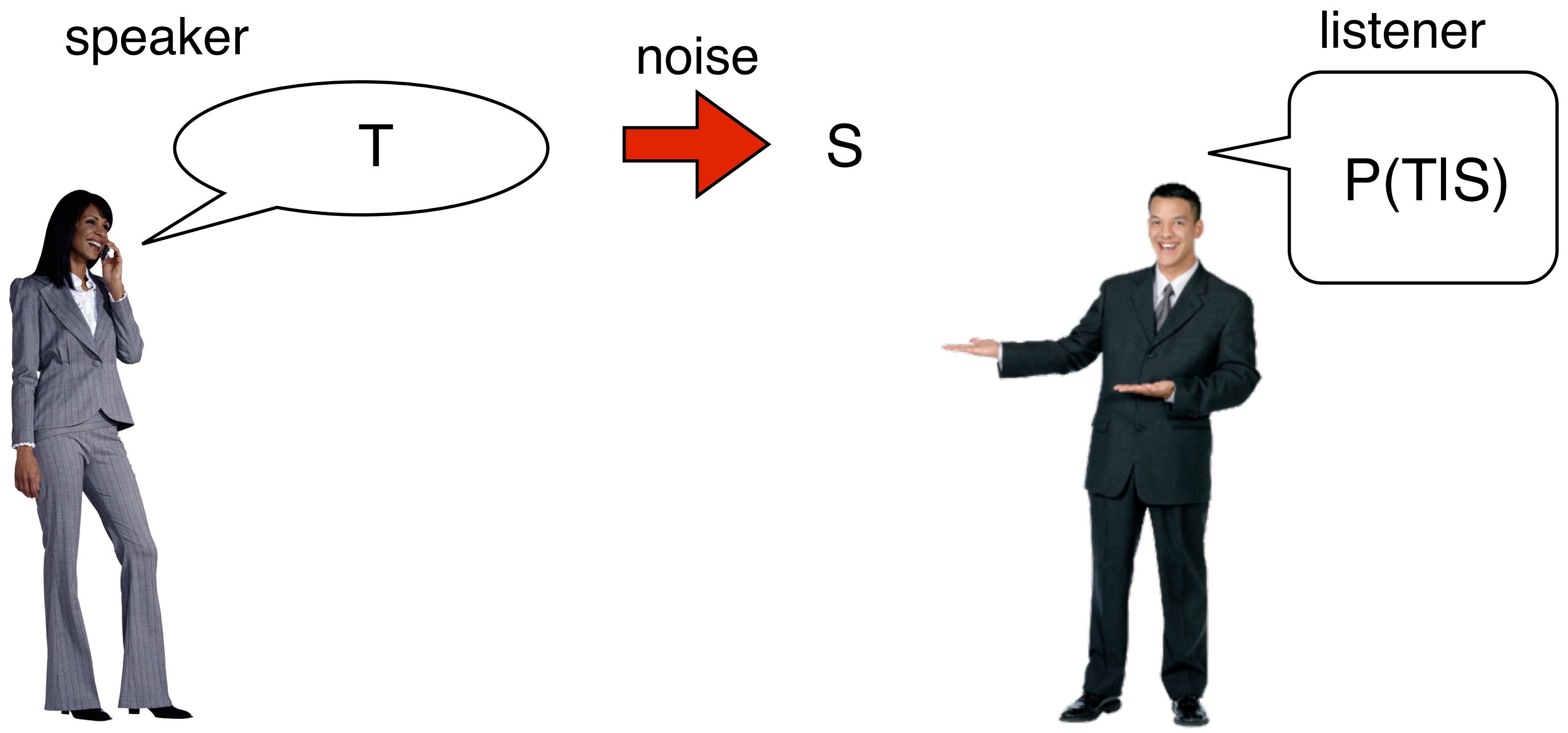


t_{total}

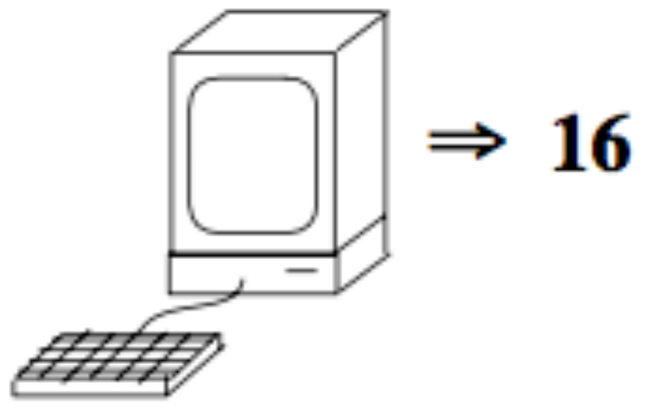
Movie grosses (Power-law)



t_{total}



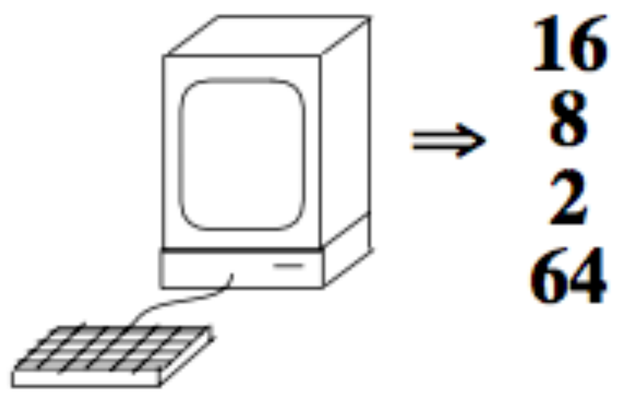
1 random "yes" example:



Which numbers will be accepted by the same computer program?

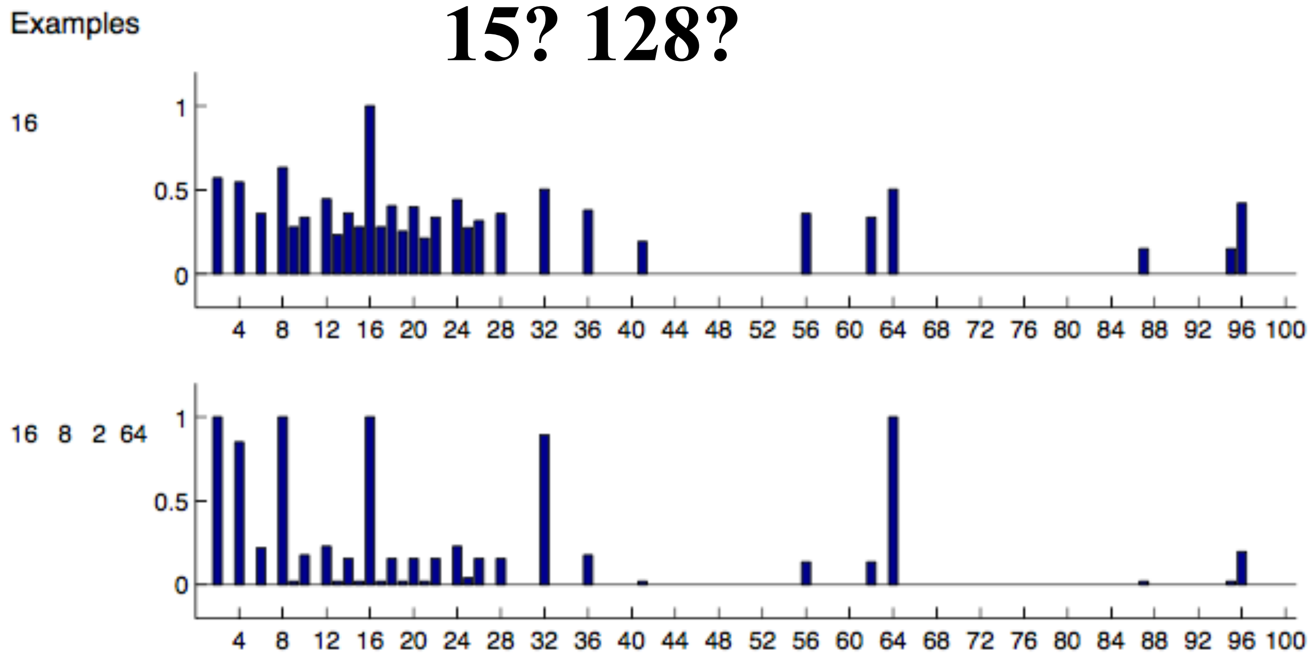
15? 128?

4 random "yes" examples:



Which numbers will be accepted by the same computer program?

15? 128?



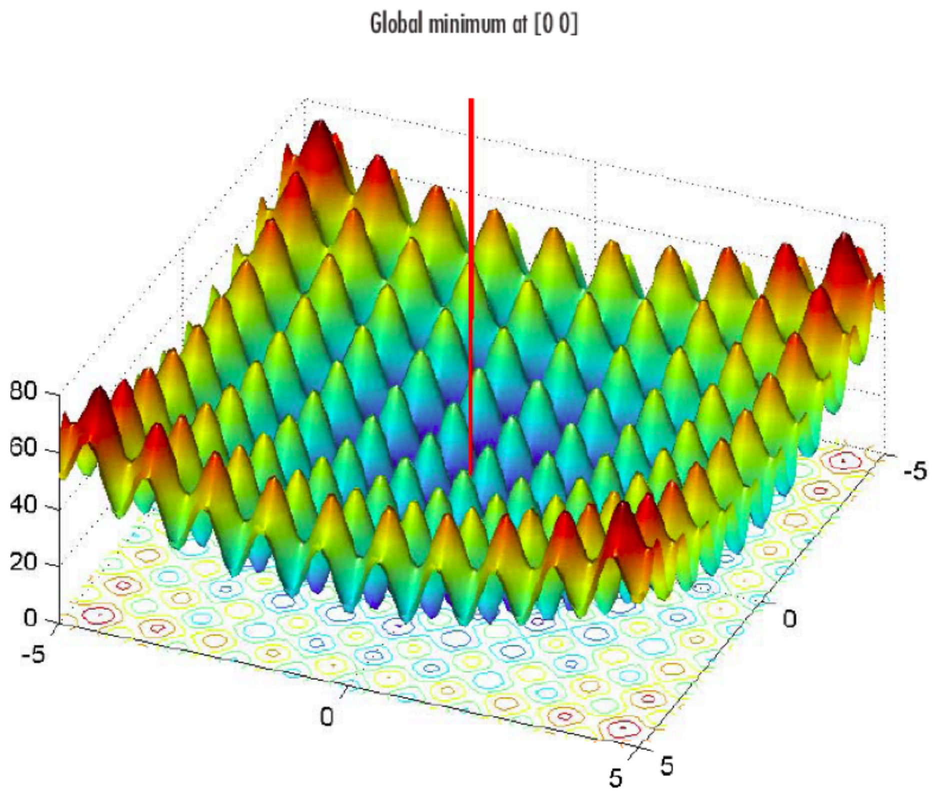
Model fitting, evaluation, and comparison

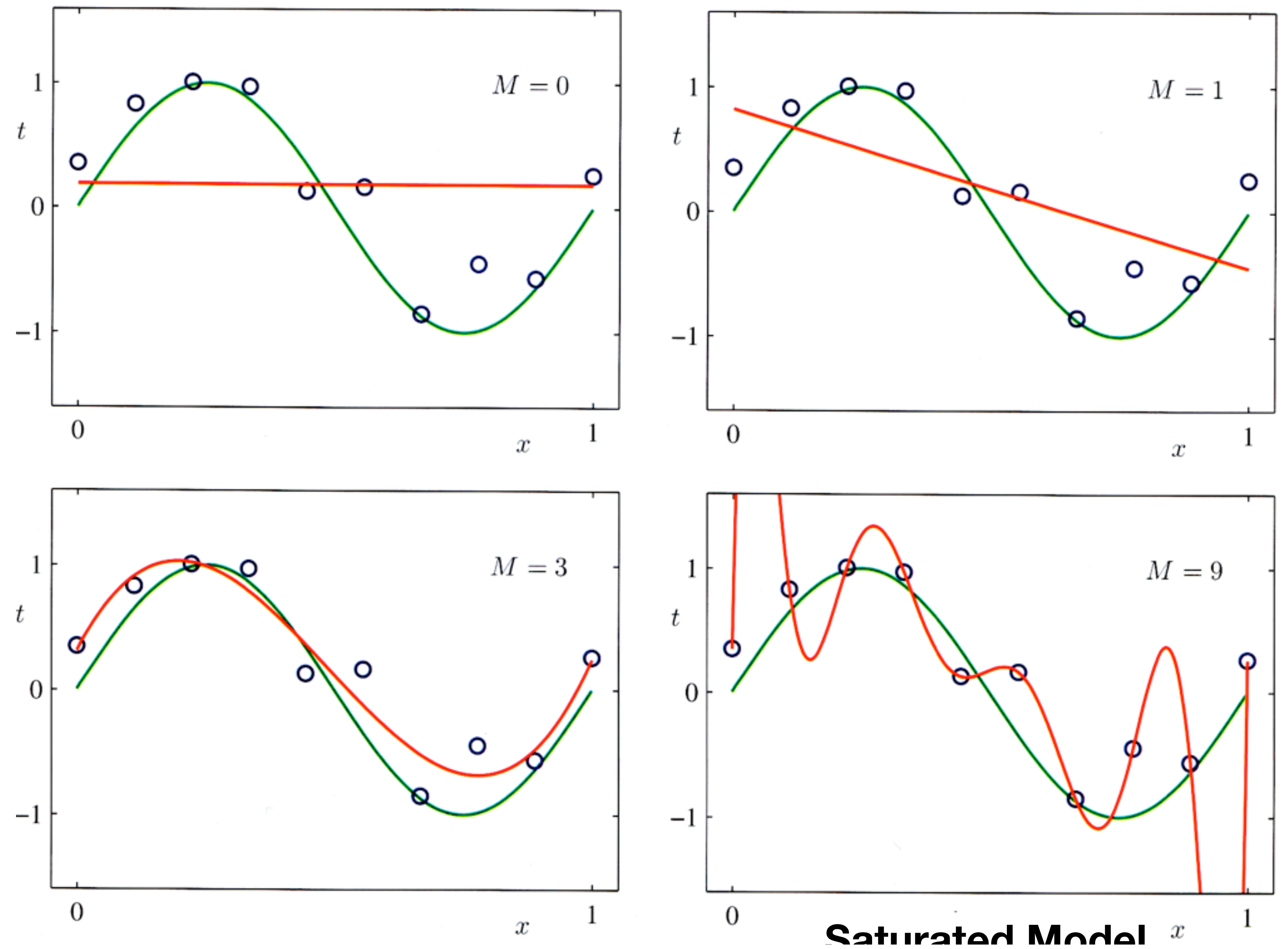
- Akaike's Information Criterion (AIC)

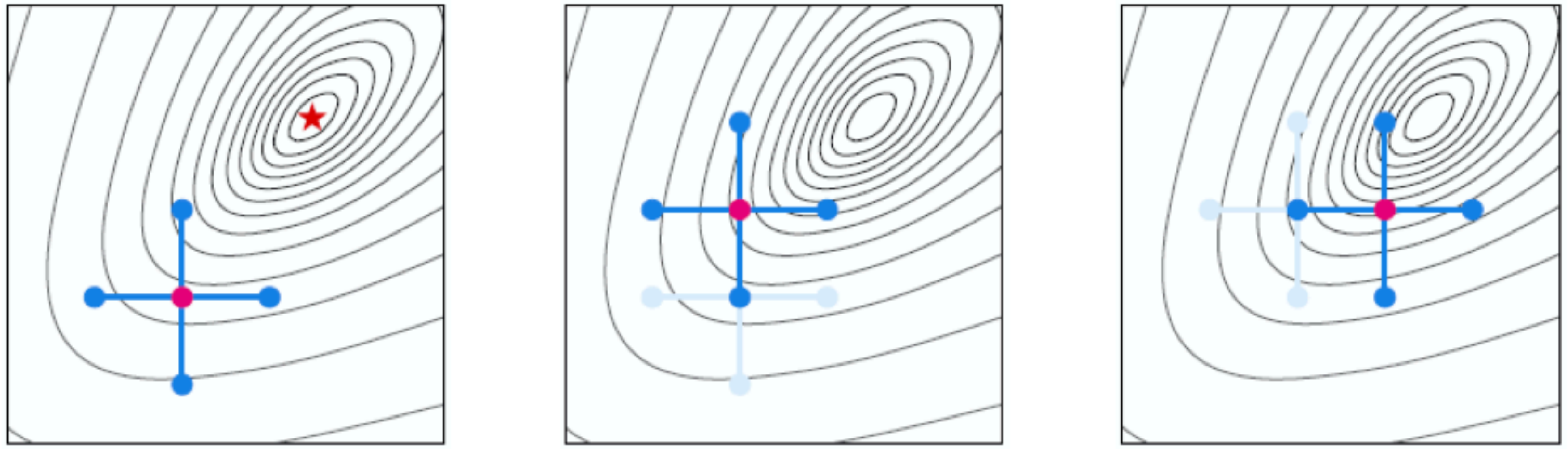
$$AIC = -2lnL(\theta|u, M) + 2K$$

- Bayesian Information Criterion (BIC)

$$BIC = -2lnL(\theta|u, M) + KlnN$$







(a) Initial pattern (b) Move North (c) Move West

classification and category learning

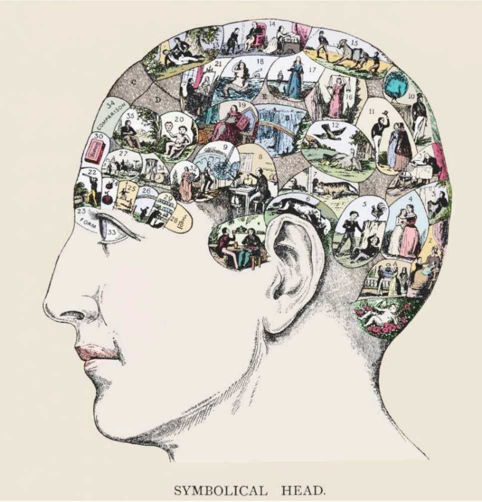
the human cognition framework

What is the function $y = f(\mathbf{x})$ that best characterizes how people make categorization decisions?

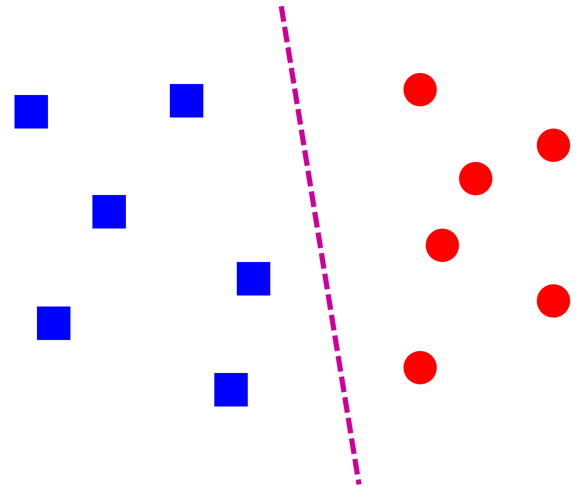
$$y = f(\mathbf{x})$$

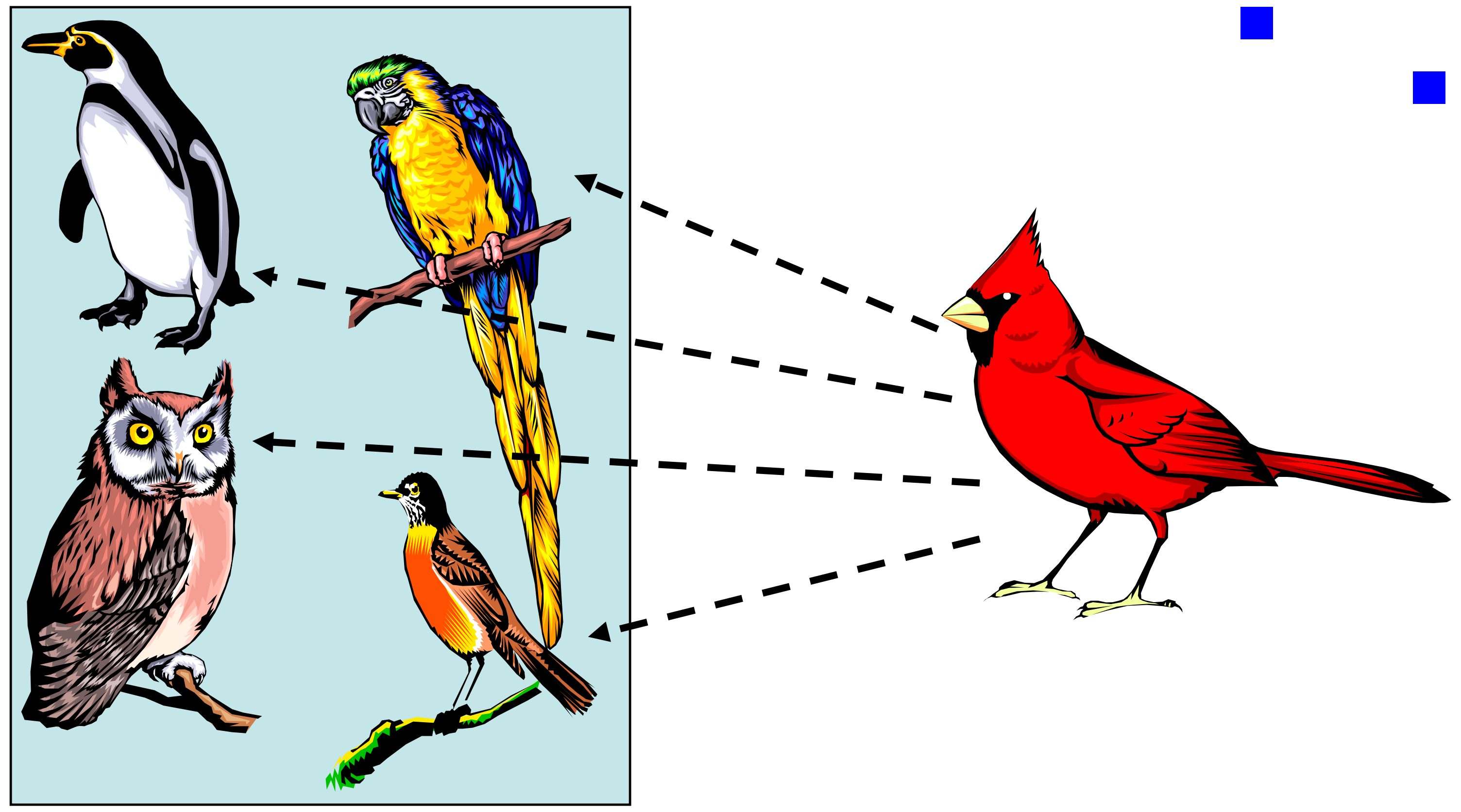
output prediction function Image feature

x_0	x_2	x_3	x_4	x_5	x_6	x_7
0	0	0	1	0	0	0
0	1	0	0	1	0	1
1	0	0	1	0	0	0
1	1	1	1	0	1	1



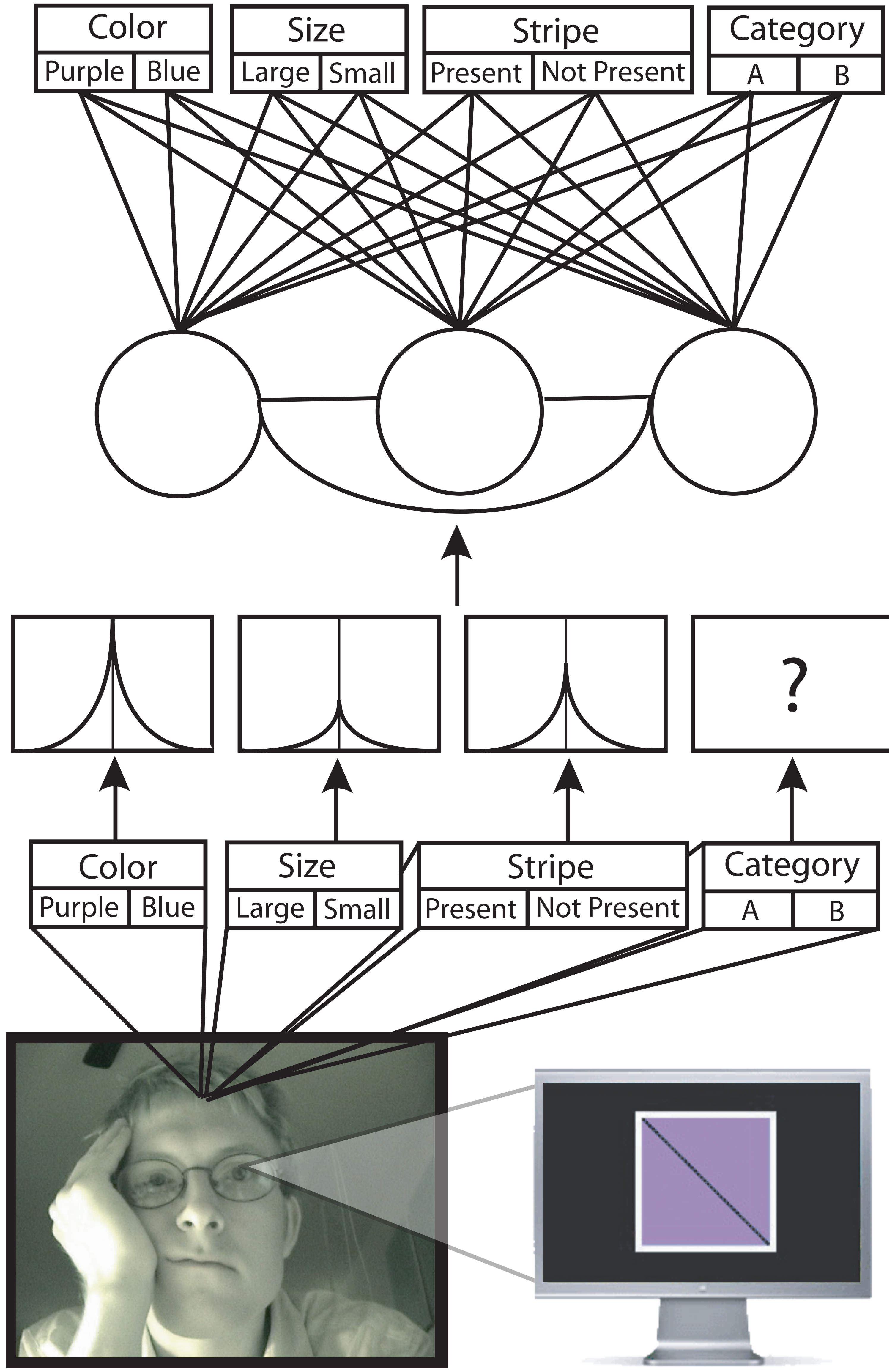
y
0
0
1
1



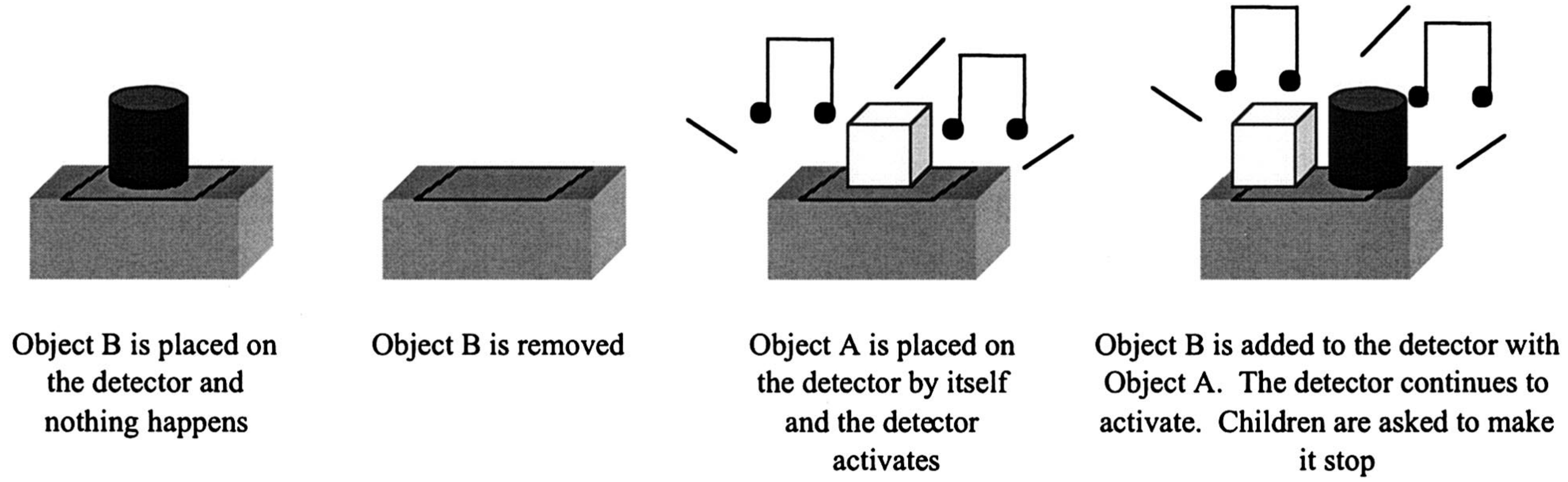


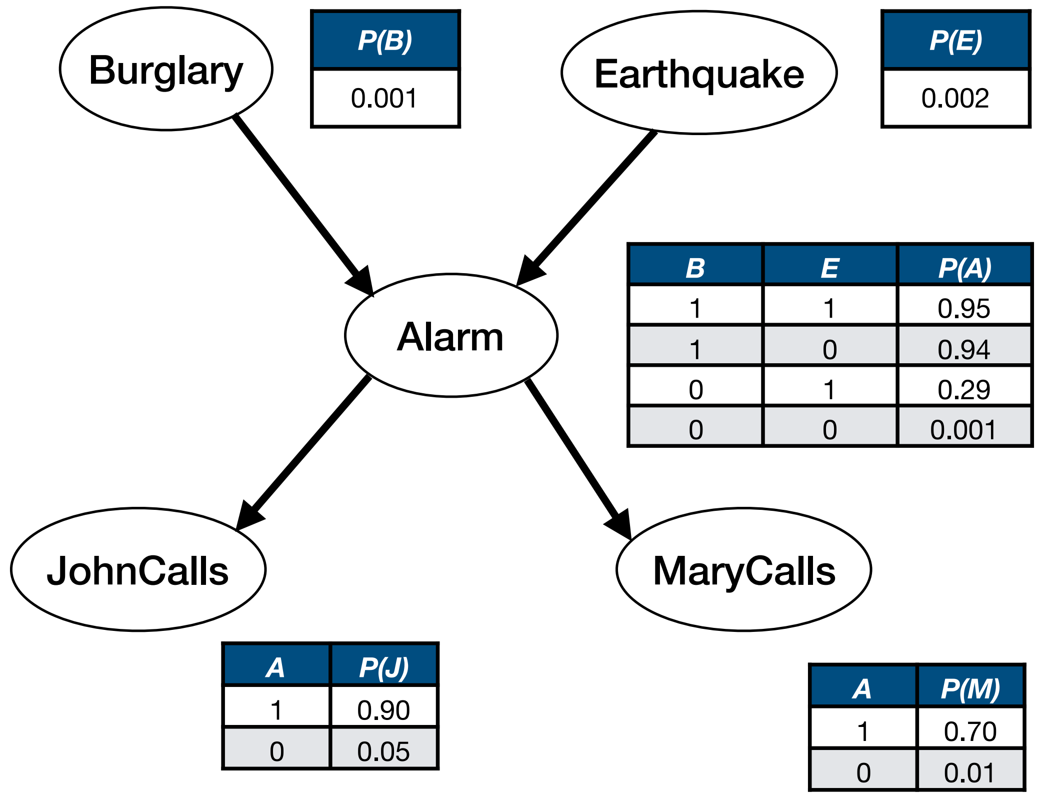
Birds You've Seen

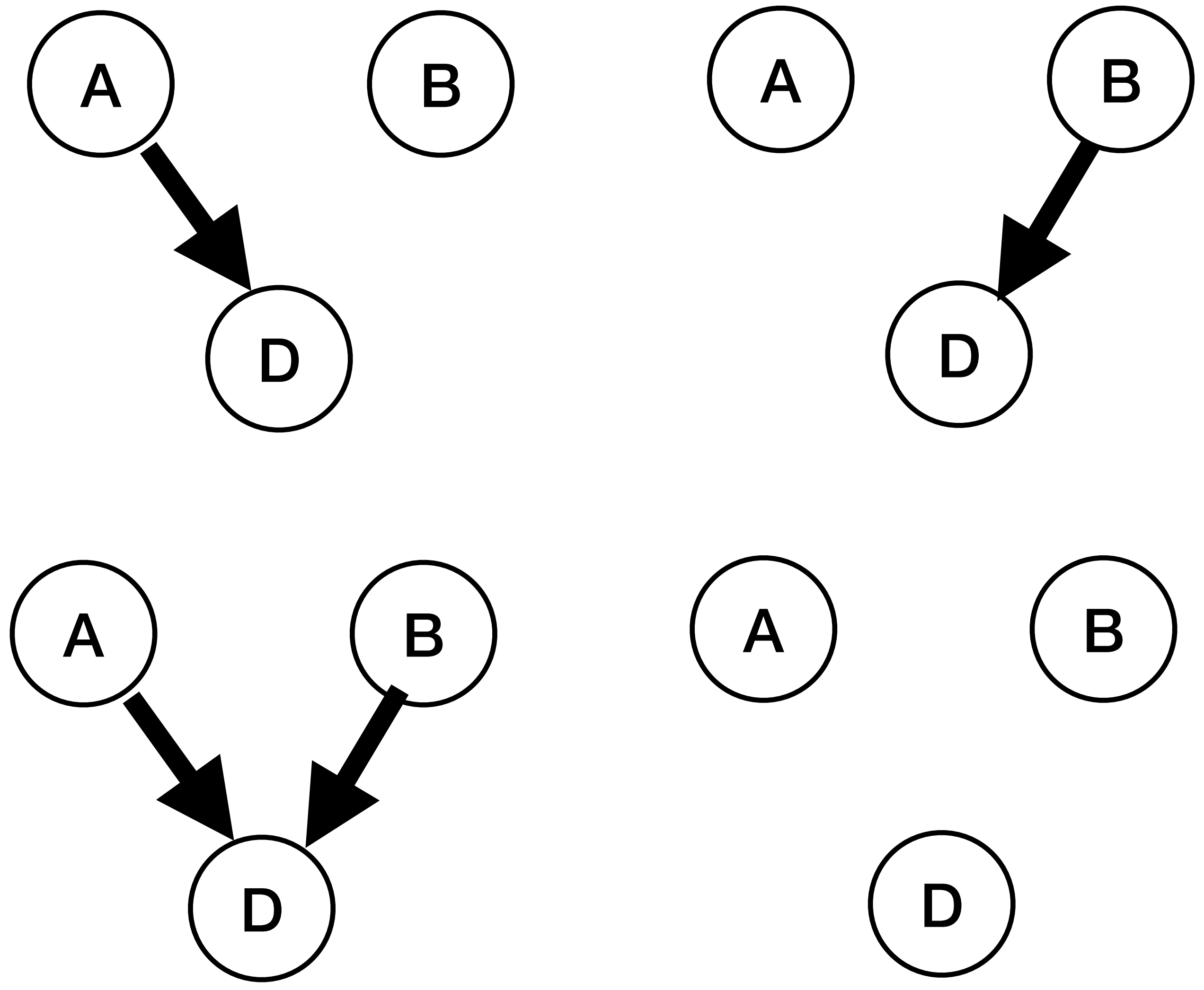
Bird?

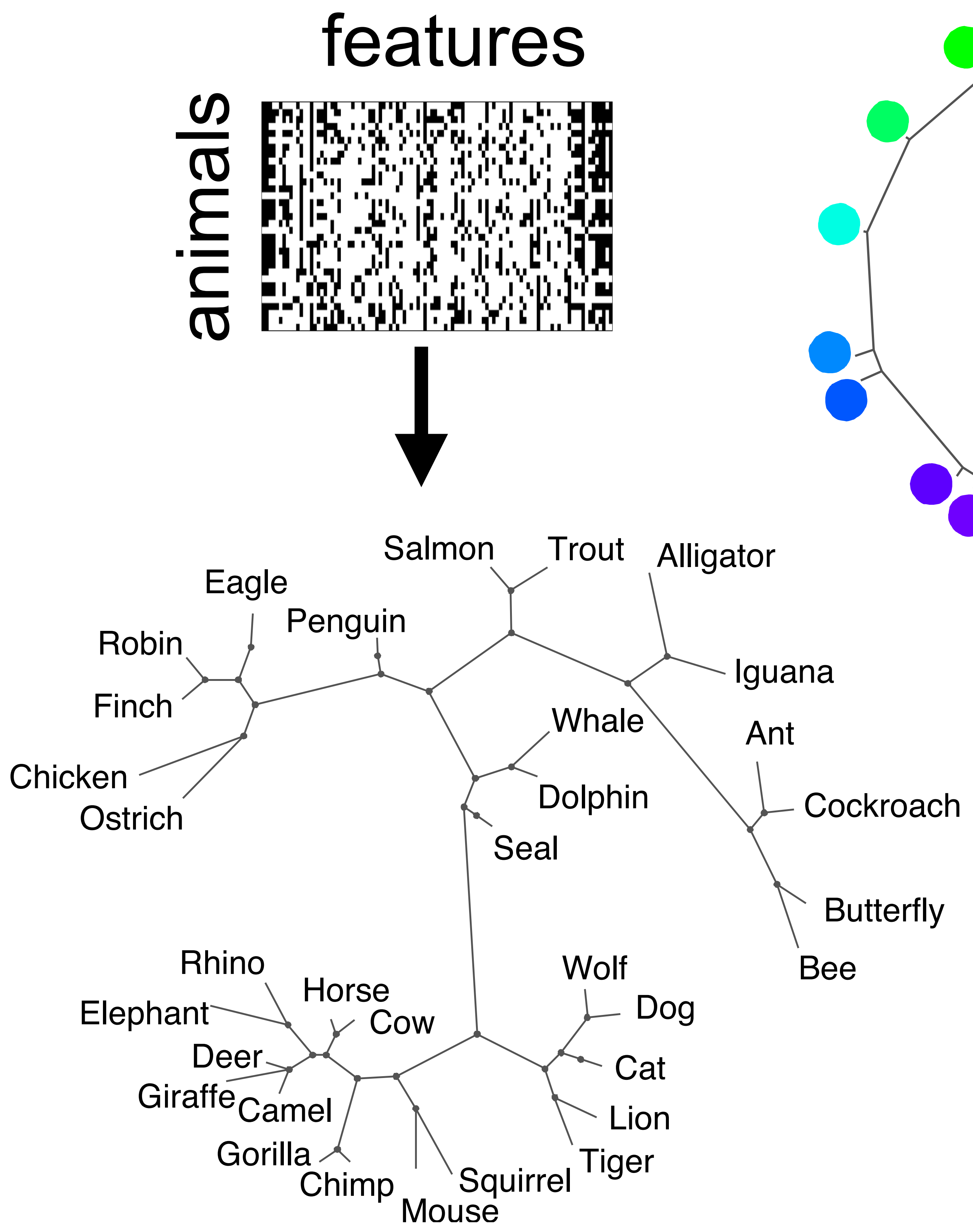


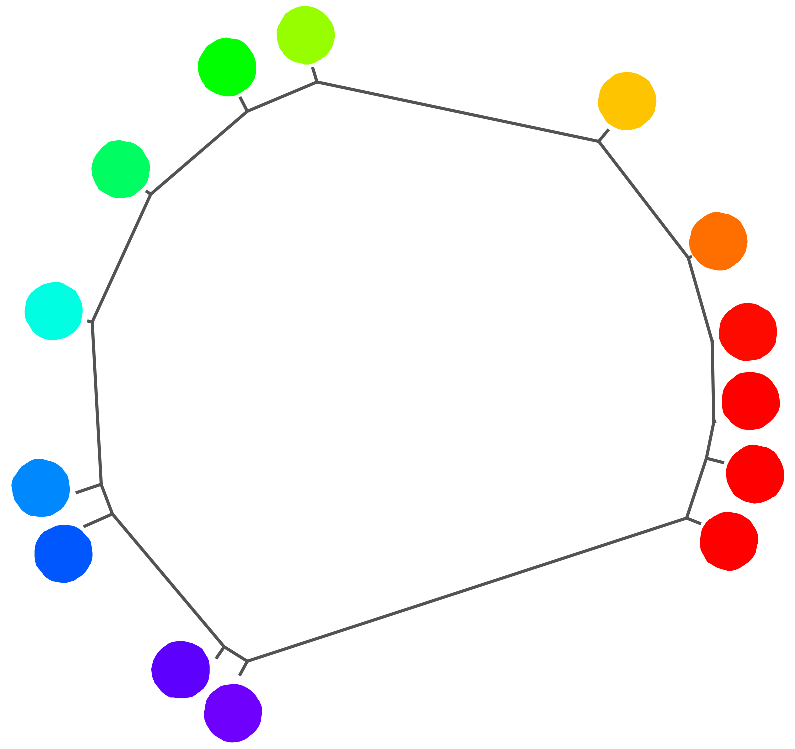
Probabilistic graphical models

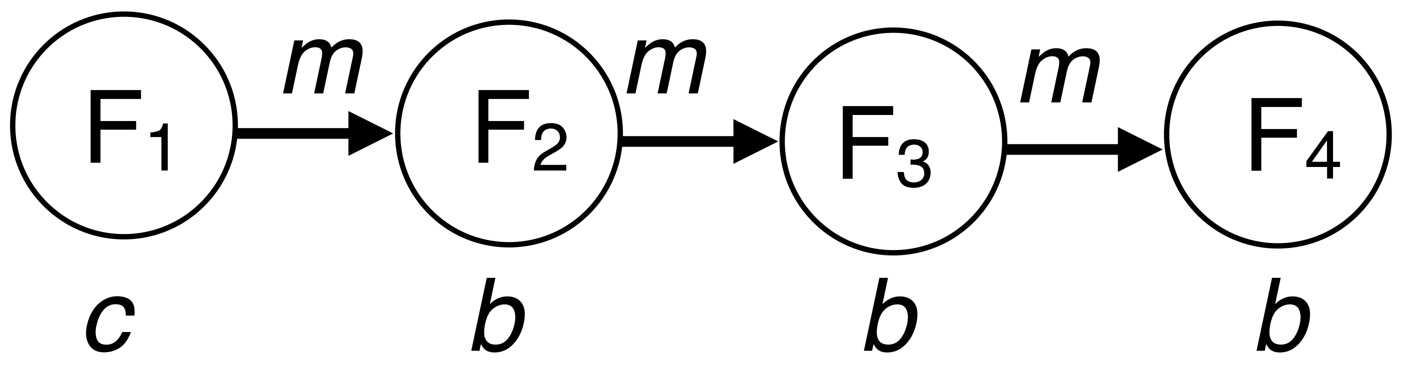




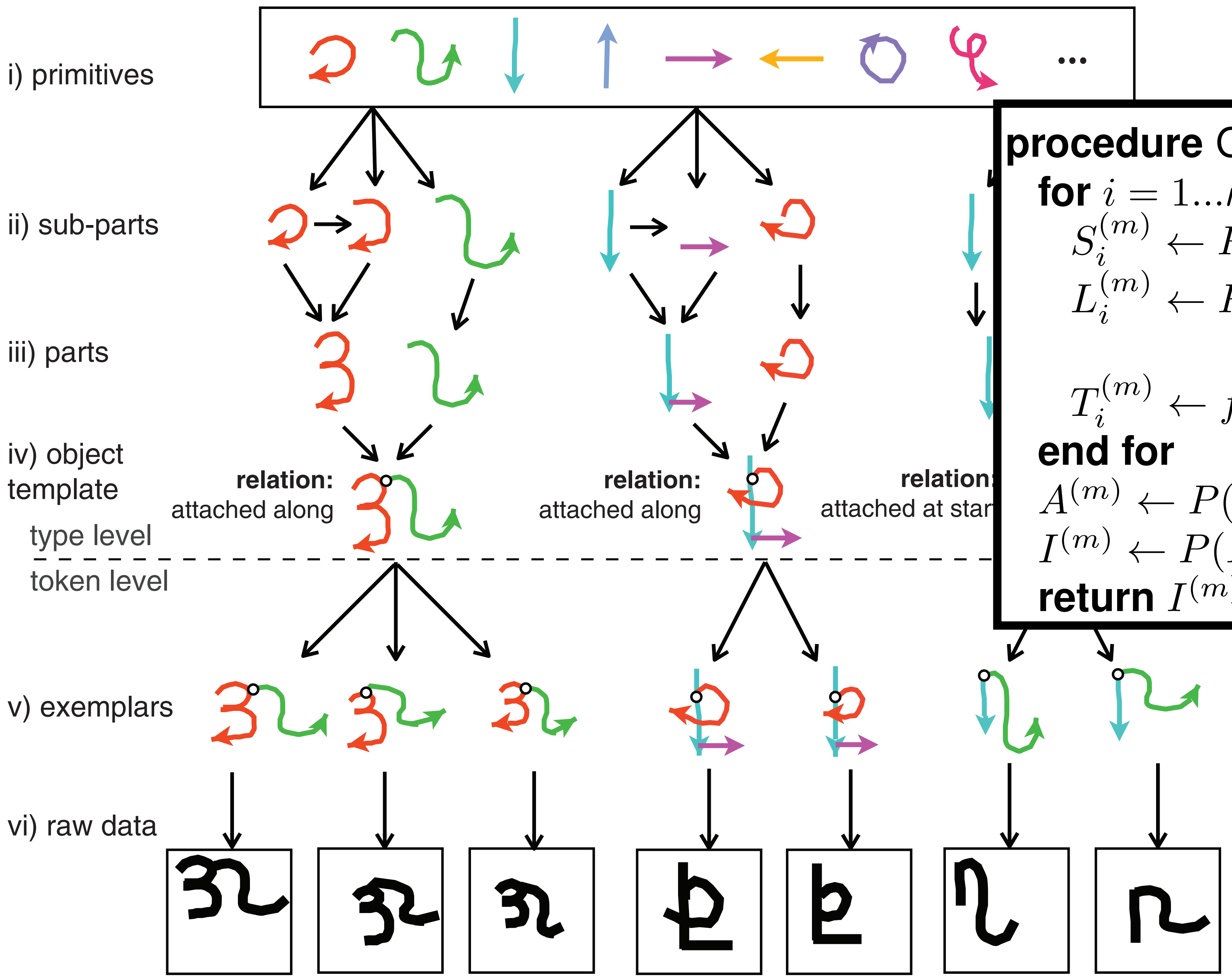




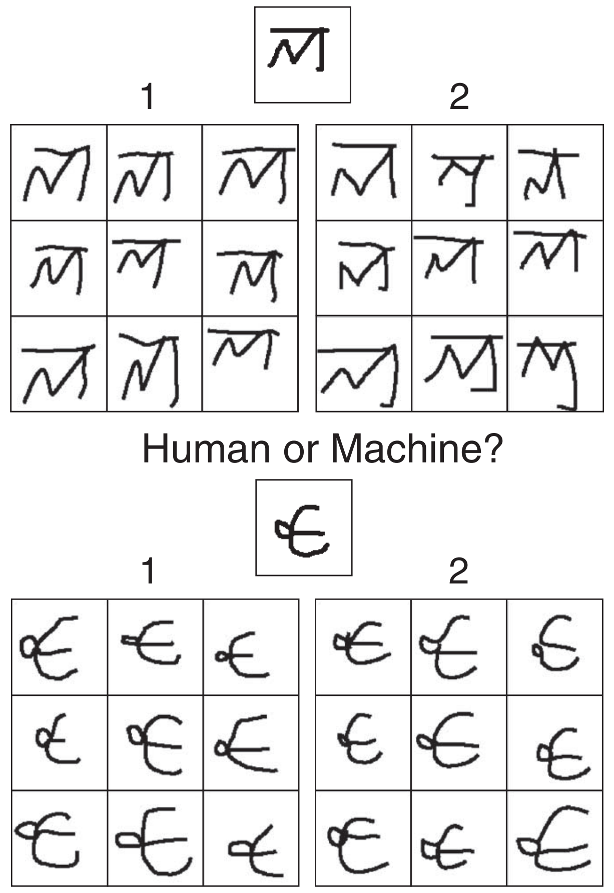




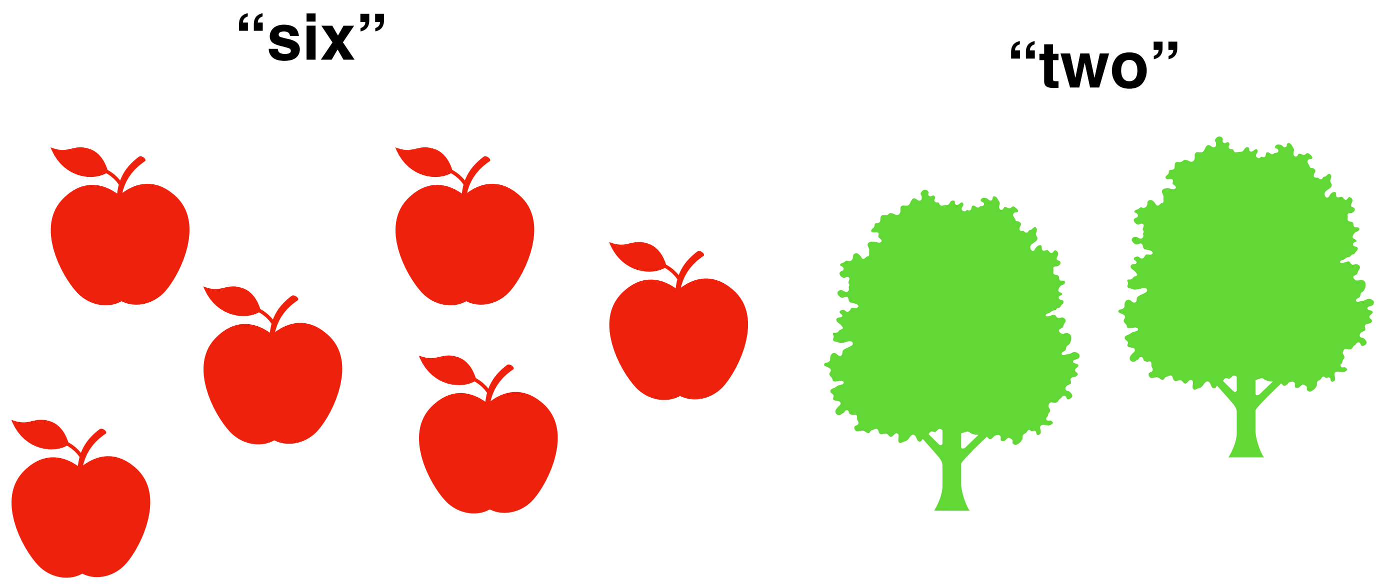
Probilistic programs and program induction

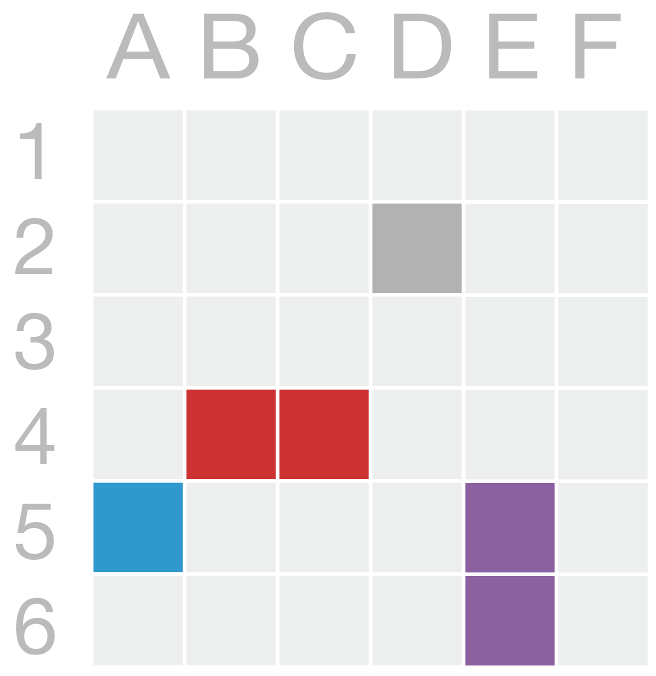


procedure GENERATETOKEN(ψ)
  for $i = 1...\kappa$ **do**
    $S_i^{(m)} \leftarrow P(S_i^{(m)}|S_i)$ ▷ Add motor variance
    $L_i^{(m)} \leftarrow P(L_i^{(m)}|R_i, T_1^{(m)}, ..., T_{i-1}^{(m)})$ ▷ Sample part's start location
    $T_i^{(m)} \leftarrow f(L_i^{(m)}, S_i^{(m)})$ ▷ Compose a part's trajectory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



λ S . (if (singleton? S)
"one"
(if (doubleton? S)
"two"
undef))





What is the top left of all the ship tiles?

```
(topleft (setDifference (set 1A ... 6F) (coloredTiles Water)))
```

Are all the ships horizontal?

```
(all (map (lambda x (== H (orient x))) (set Blue Red Purple)))
```

Are blue and purple ships touching and red and purple not touching (or vice versa)?

```
(== (touch Blue Purple) (not (touch Red Purple)))
```

Open questions

- How does a computationally limited, time constrained, noisy/wet/squish brain perform sophisticated (probabilistic) inferences?

- How do these noisy/wet/squishy neurons hook up in neural networks and maintain stability and function, even under damage or disease?

- How does the mind and brain learn, represent and reason with rich structural representations (graphs, trees, programs, etc.)? These representations sometimes seems as antithetical to brain processes (e.g., neural networks) but we are on verge of seeing massive convergence in approaches.

More open questions

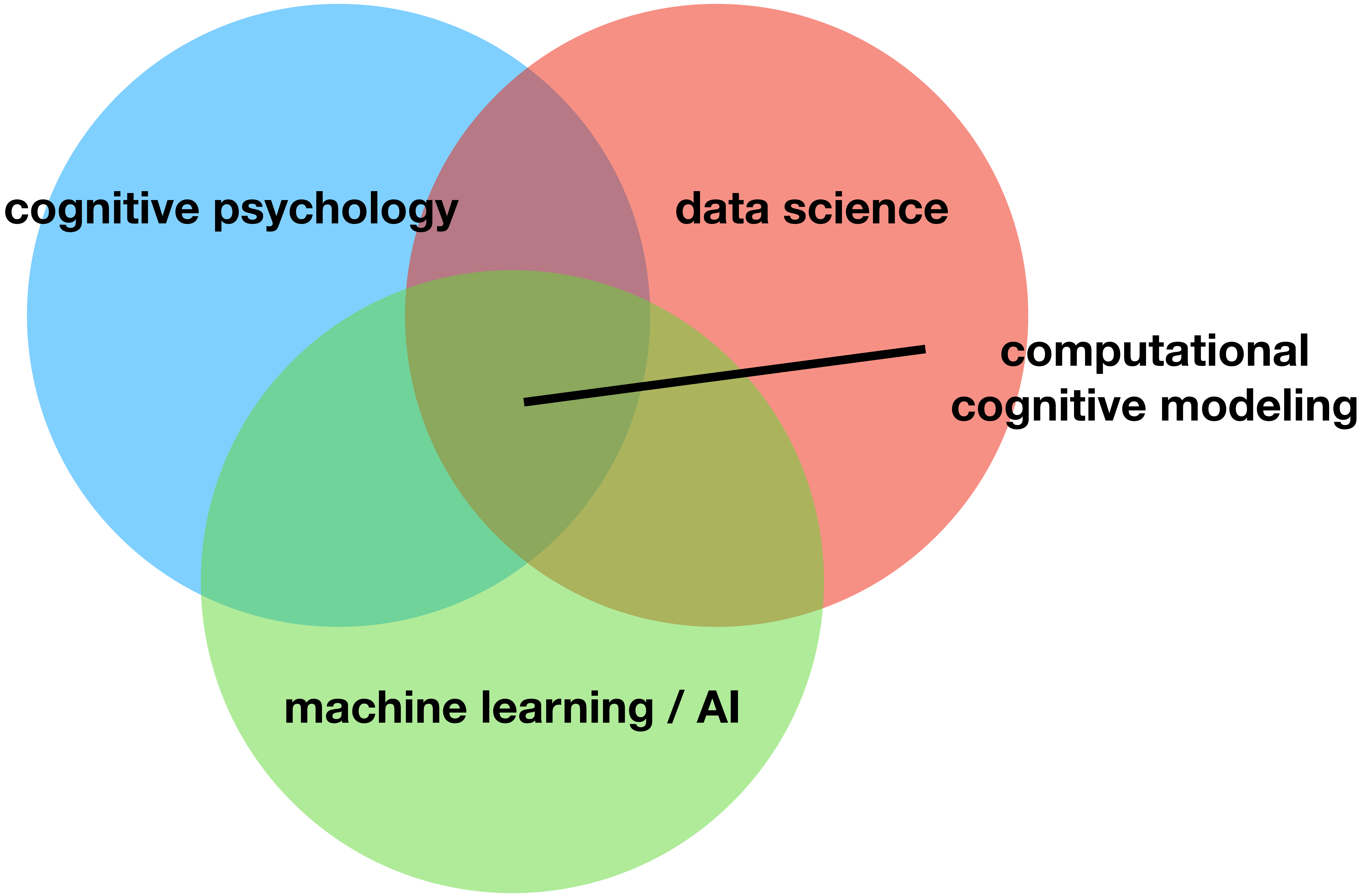
- How can recent advances in AI best advance computational cognitive modeling? How can recent advances in computational cognitive modeling best advance AI?

- Many human abilities lack compelling computational models:

- scene understanding
- language understanding
- creativity
- general purpose problem solving
- learning new video games
- commonsense reasoning, etc.

- How do deep learning, reinforcement learning, Bayesian modeling, graphical models, and probabilistic programming fit together? Is there are unifying computational framework for understanding human intelligence?

- How can understanding the structure of the cognitive system (e.g., the algorithmic or computational level) help us interpret the function and organization of the human brain?


cognitive psychology
data science
computational
cognitive modeling
machine learning / AI

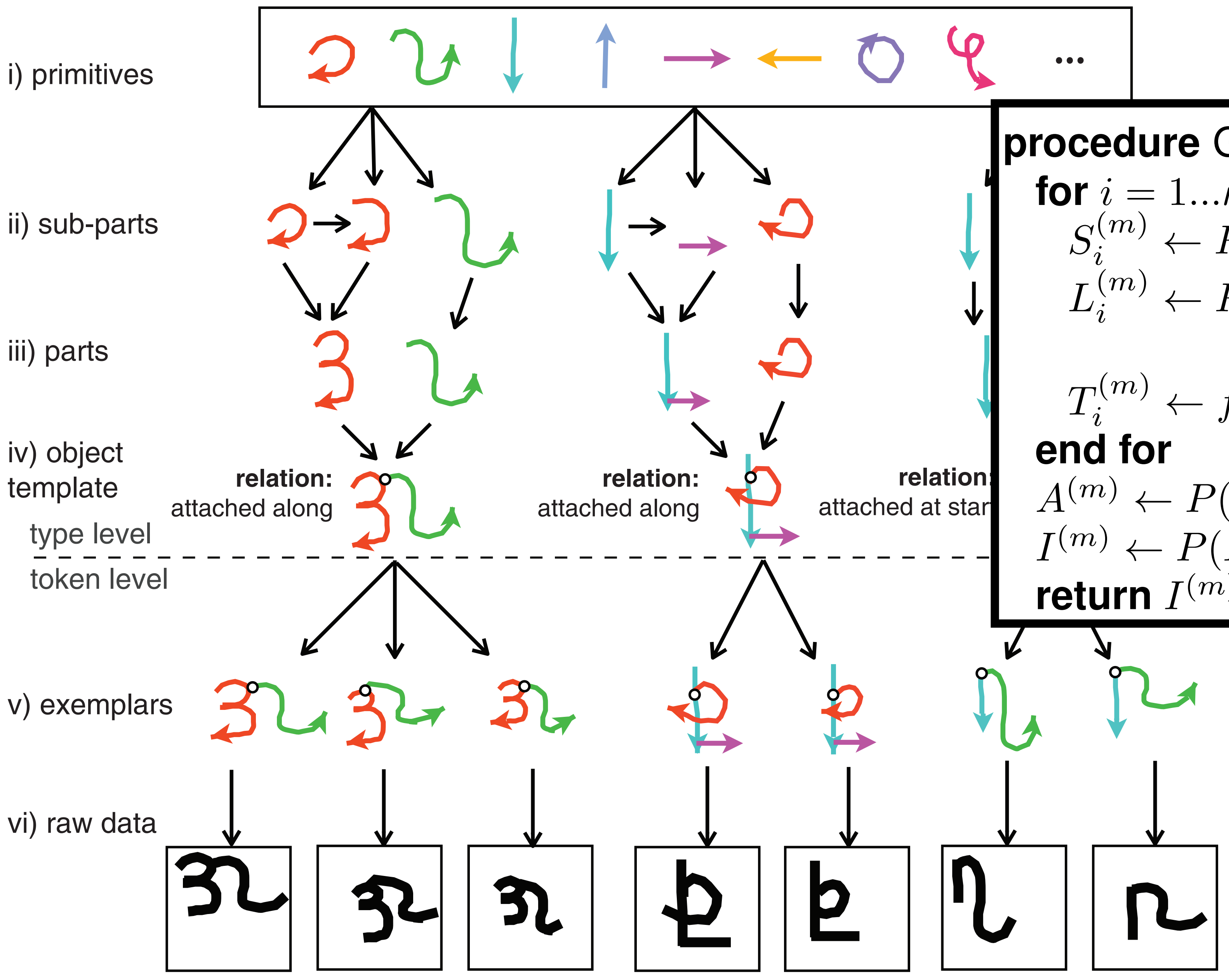
Open Questions in Computational Cognitive Modeling

How do deep learning, reinforcement learning, Bayesian modeling, graphical models, and probabilistic programming fit together? Is there are unifying computational framework for understanding human intelligence?

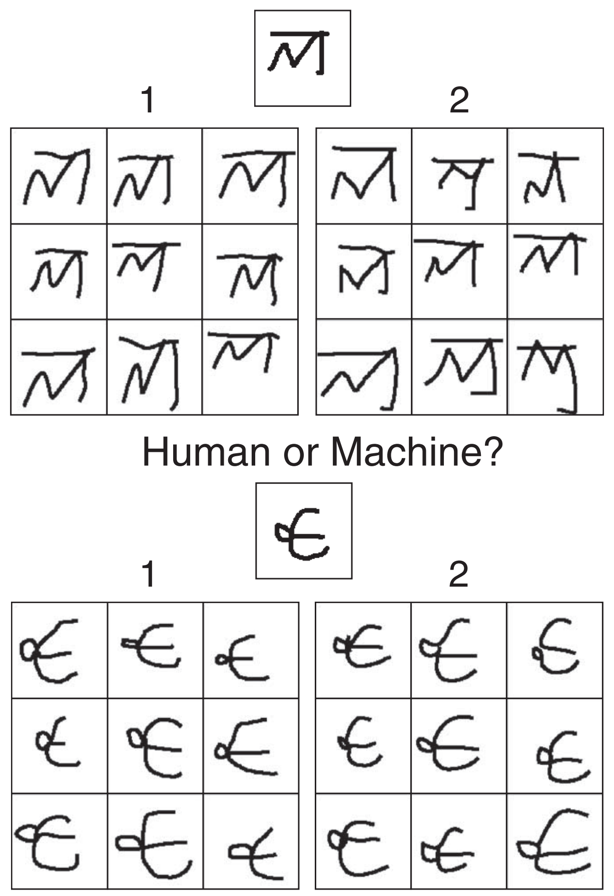
How do deep learning, reinforcement learning, Bayesian modeling, graphical models, and probabilistic programming fit together? Is there are unifying computational framework for understanding human intelligence?

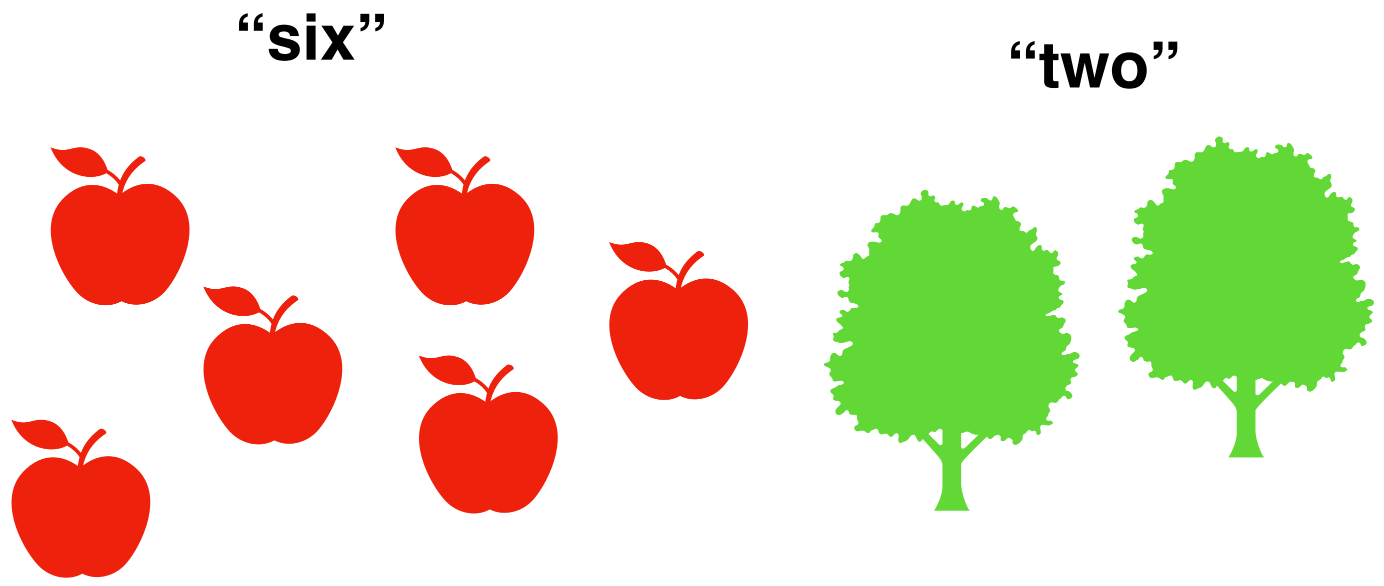
deep learning + reinforcement learning = deep RL
deep learning + Bayesian modeling = Bayesian deep learning
deep learning + symbolic modeling = neuro-symbolic modeling

Probabilistic programs and program induction

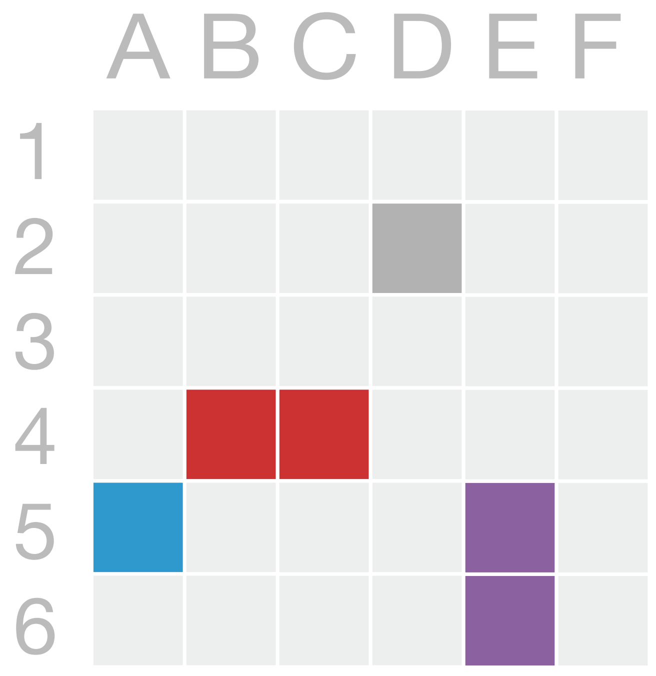


procedure GENERATETOKEN(ψ)
  for $i = 1...\kappa$ **do**
    $S_i^{(m)} \leftarrow P(S_i^{(m)}|S_i)$ ▷ Add motor variance
    $L_i^{(m)} \leftarrow P(L_i^{(m)}|R_i, T_1^{(m)}, ..., T_{i-1}^{(m)})$ ▷ Sample part's start location
    $T_i^{(m)} \leftarrow f(L_i^{(m)}, S_i^{(m)})$ ▷ Compose a part's trajectory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





λ *S* . (*if* (*singleton? S*)
  "one"
  (*if* (*doubleton? S*)
    "two"
    undef))

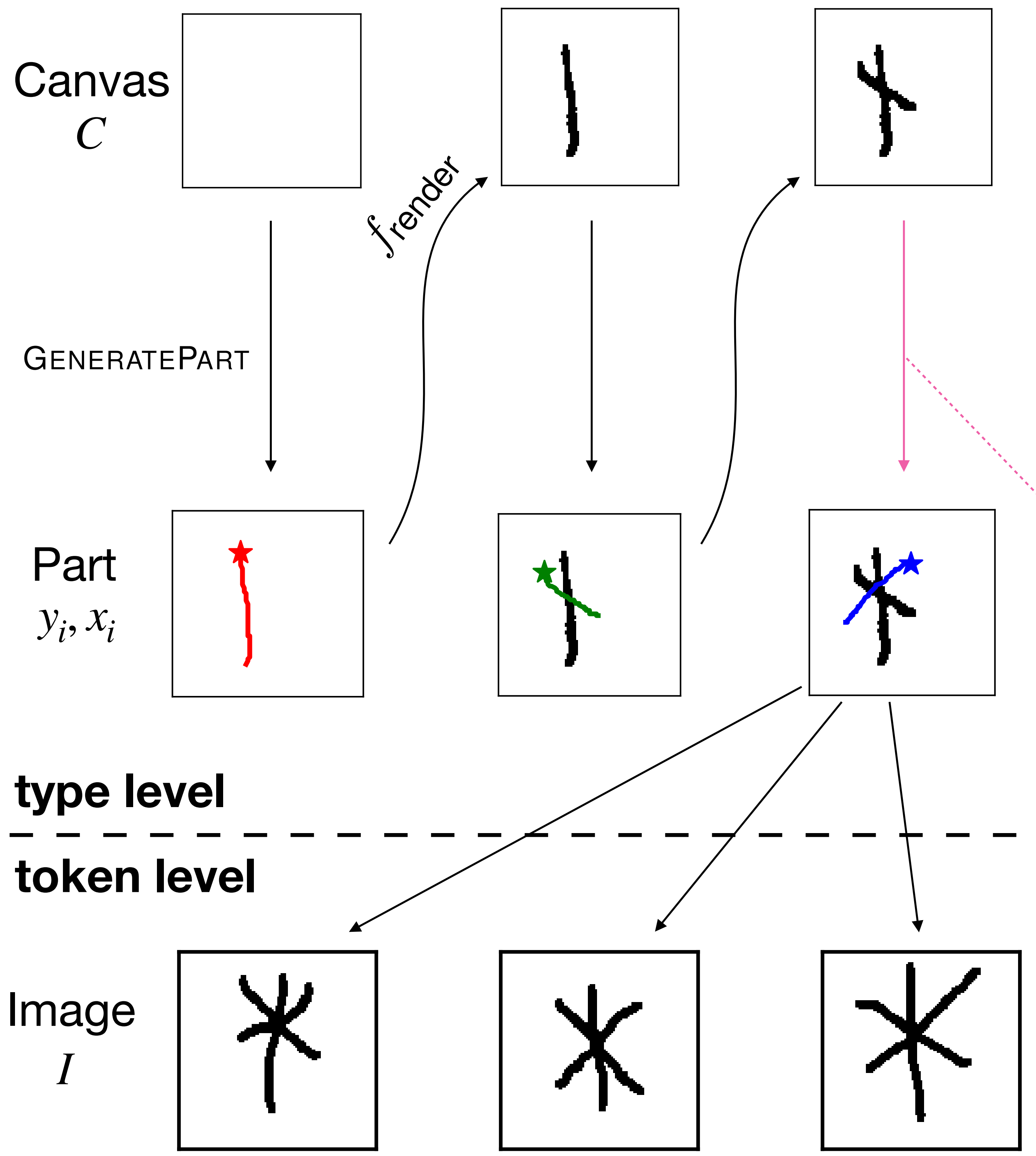


What is the top left of all the ship tiles?
```
(topleft (setDifference (set 1A ... 6F) (coloredTiles Water)))
```

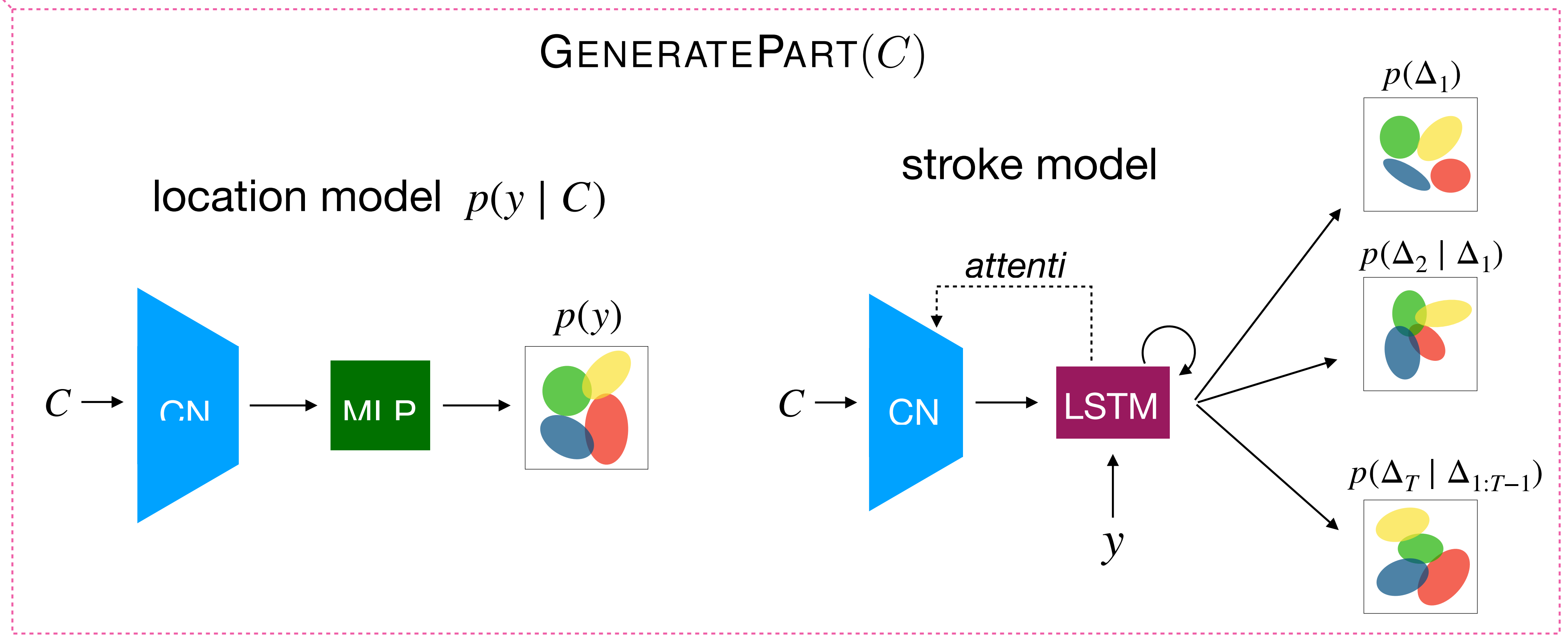
Are all the ships horizontal?
```
(all (map (lambda x (== H (orient x))) (set Blue Red Purple)))
```

Are blue and purple ships touching and red and purple not touching (or vice versa)?
```
(== (touch Blue Purple) (not (touch Red Purple)))
```

Generative neuro-symbolic modeling



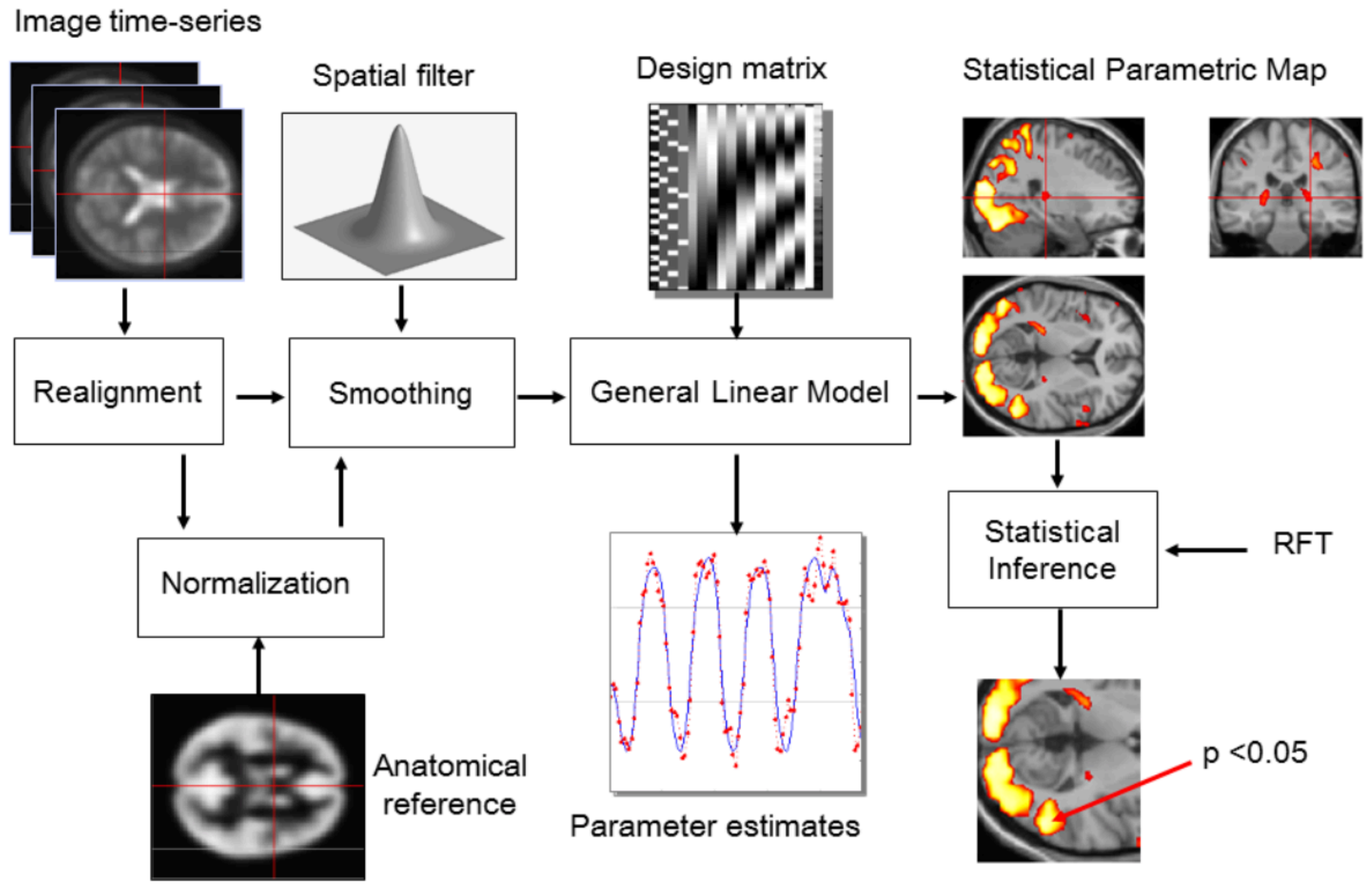
procedure GenerateType
$C \leftarrow 0$ ▷ Initialize blank canvas
for $i = 1, \ldots, \infty$ **do**
$x_i \leftarrow$ GeneratePart(C) ▷ Sample part from neural net
$C \leftarrow$ Render (y_i, x_i, C) ▷ Update canvas
$v_i \leftarrow$ Terminate?(C) ▷ Sample termination indicator
if v_i **then**
break ▷ Terminate sample
$\psi \leftarrow \{i, x_{1:i}, y_{1:i},\}$
return GenerateToken(ψ) ▷ Return concept type

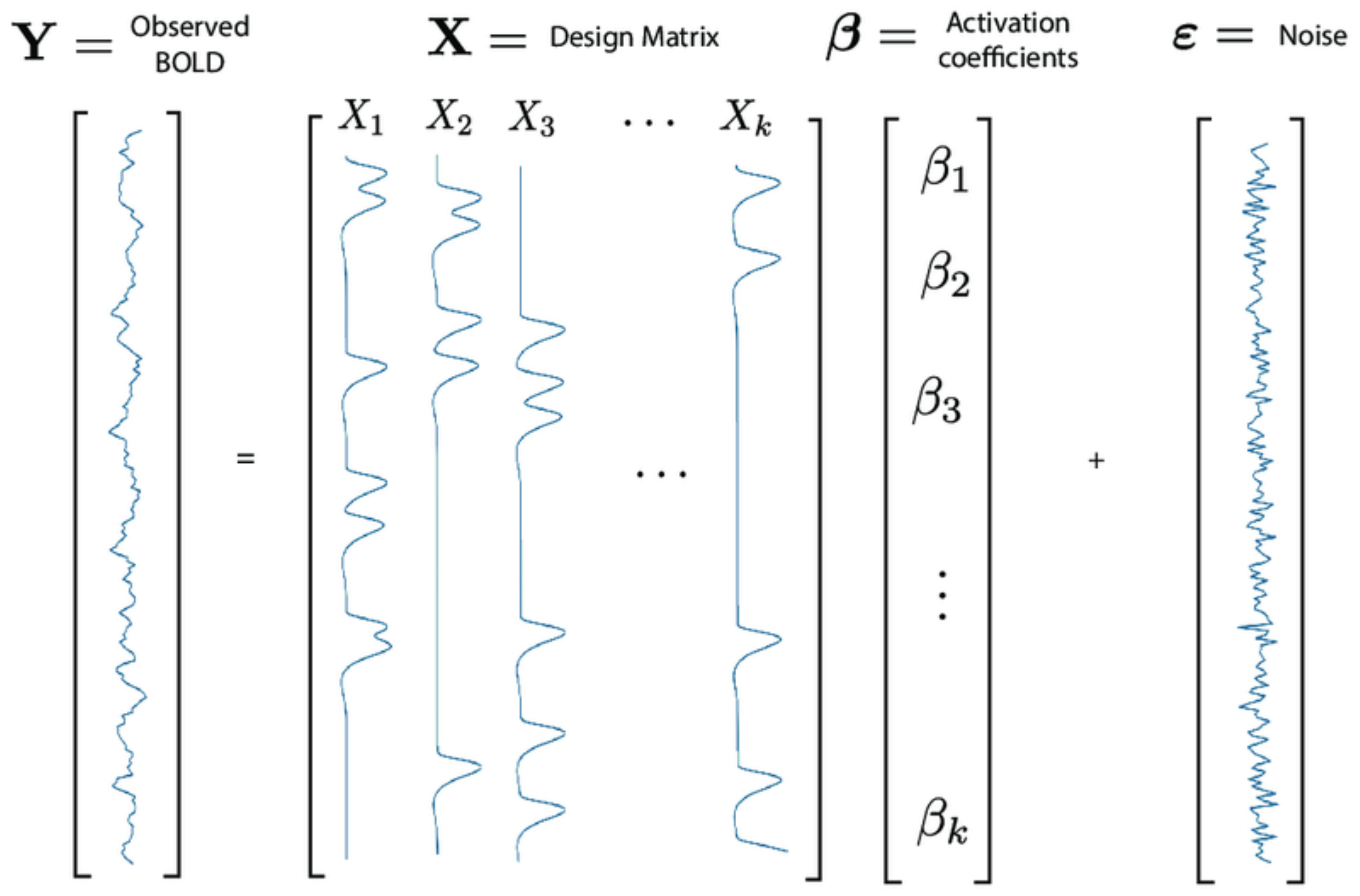


How can understanding the structure of the cognitive system (e.g., the algorithmic or computational level) help us interpret the function and organization of the human brain?

Classic approach

Directly relate operational definitions of psychological constructs to the brain:





Model-based fMRI

Novelty and Reward

- Novelty bonuses (Kakade & Dayan, 2002)
- Daw Novelty study

Striatal Activity Underlies Novelty-Based Choice in Humans

Bianca C. Wittmann,[1,3,*] Nathaniel D. Daw,[2,3,4] Ben Seymour,[1] and Raymond J. Dolan[1]
[1]Wellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, London WC1N 3BG, UK
[2]Gatsby Computational Neuroscience Unit, University College London, 17 Queen Square, London WC1N 3AR, UK
[3]These authors contributed equally to this work
[4]Present address: Center for Neural Science and Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA
*Correspondence: b.wittmann@ucl.ac.uk
DOI 10.1016/j.neuron.2008.04.027

Table 1. Parameter Estimates for the Behavioral Model, Shown as Mean (Over Subjects) ± 1 SE

Learning rate ν	0.23 ± 0.038
Softmax inv. temperature β	8.5 ± 1.2
Initial value, familiarized Q_f	0.37 ± 0.071
Initial value, novel Q_n	0.41 ± 0.076

Due to poor identification of β and ν, one subject is omitted from these averages.

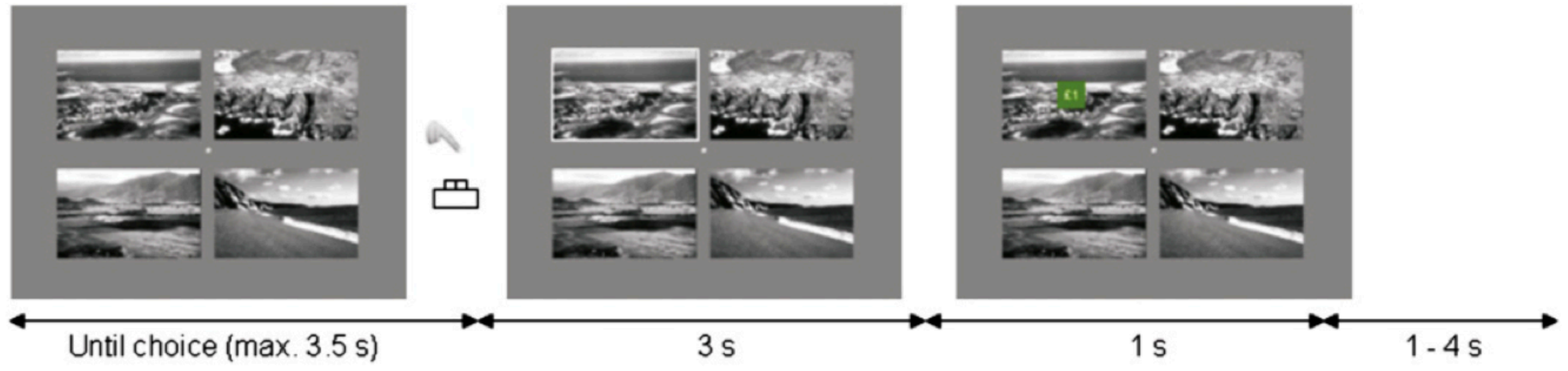


Figure 1. Experimental Design
Following a familiarization phase, participants were shown four pictures on each trial and asked to choose one. Both familiarized and novel pictures were presented at randomized locations that changed on each trial. Each picture was repeated for an average of 20 trials and then replaced. Participants were informed that each picture had been assigned a unique probability of winning £1 that would not change as long as that picture was repeated. They were given feedback at the end of each trial indicating whether they had won or received nothing.

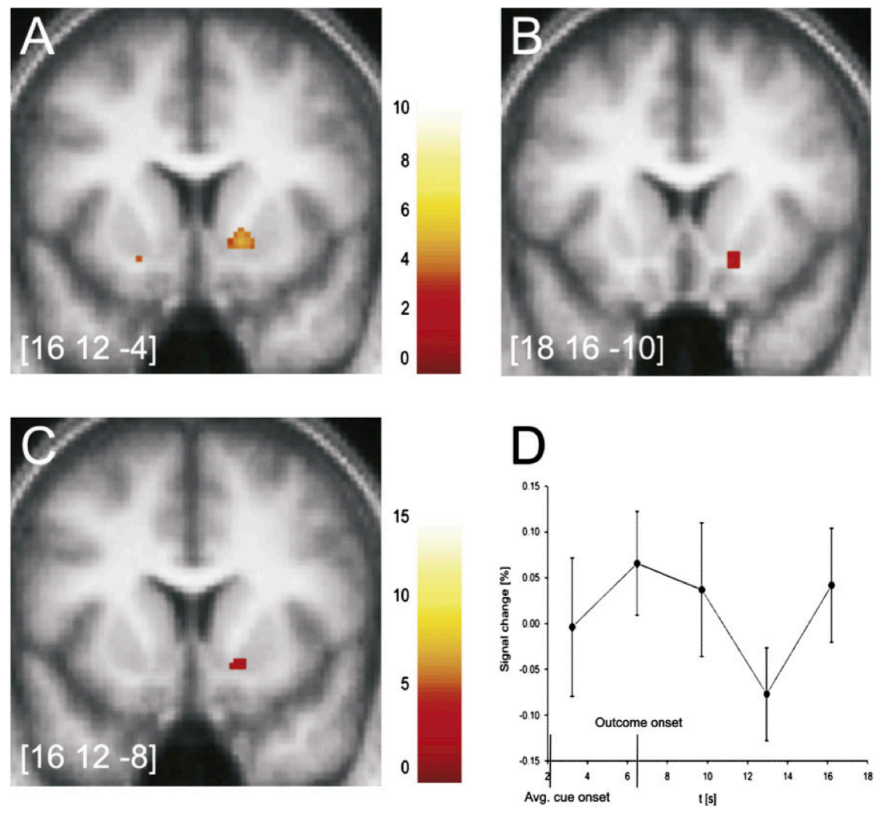
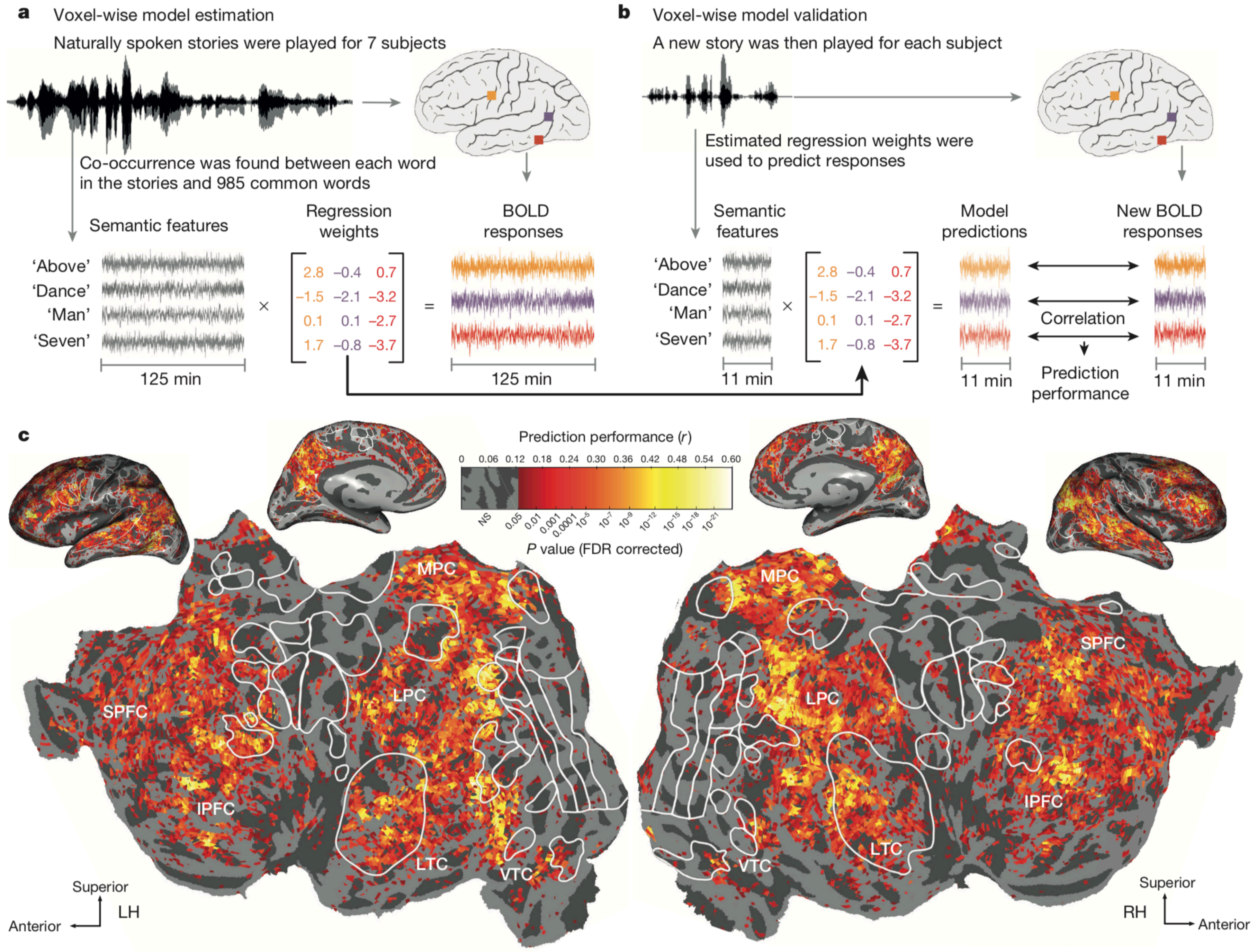
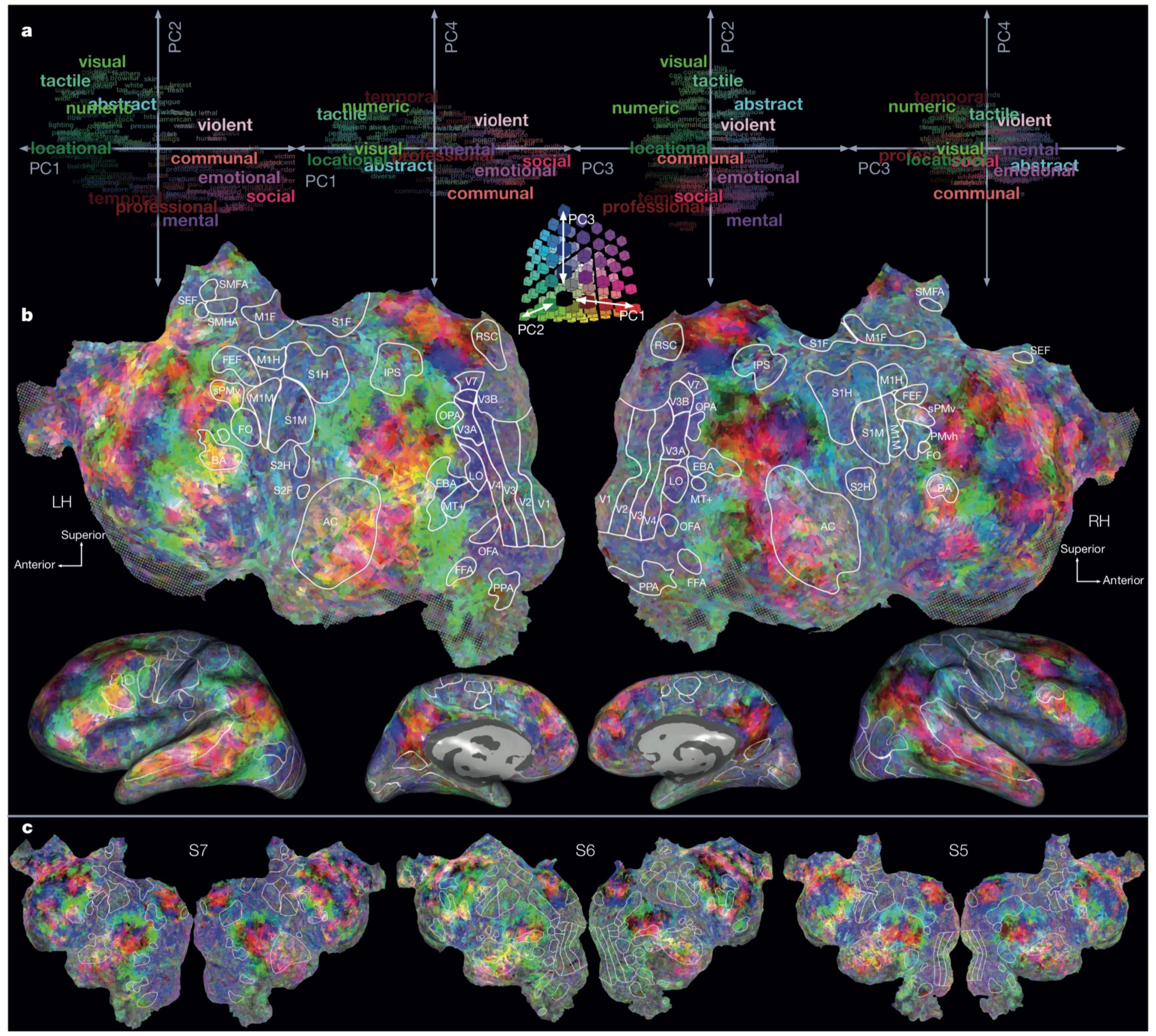


Figure 2. Ventral Striatal Response to Prediction Error and Novelty
Peak coordinates are given in MNI space on all images. Color bars indicate T values.
(A) Activation in right ventral striatum correlated significantly with reward prediction errors generated by the standard TD model ($p < 0.001$ uncorrected, $p < 0.05$ SVC, cluster > 5 voxels).
(B) Activation in right ventral striatum correlated significantly with additional prediction error due to inclusion of a novelty bonus ($p < 0.001$ uncorrected, $p < 0.05$ SVC, cluster > 5 voxels).
(C) Significant overlap between activation in right ventral striatum for the novelty bonus (see [B]) and activation obtained for standard model (see [A]) derived by inclusively masking (B) with (A) ($p < 0.005$, uncorrected, for both contrasts, cluster > 5 voxels).
(D) Striatal activation time courses calculated for the first two trials a novel stimulus is chosen minus the first two choices of familiar stimuli, shown for the peak voxel correlating with the novelty bonus (MNI coordinates: 14, 20, −10). Trials are aligned by the time of reward outcome at 6.5 s; the average stimulus onset time is also indicated. Error bars indicate SEM.

Brain mapping of semantic space





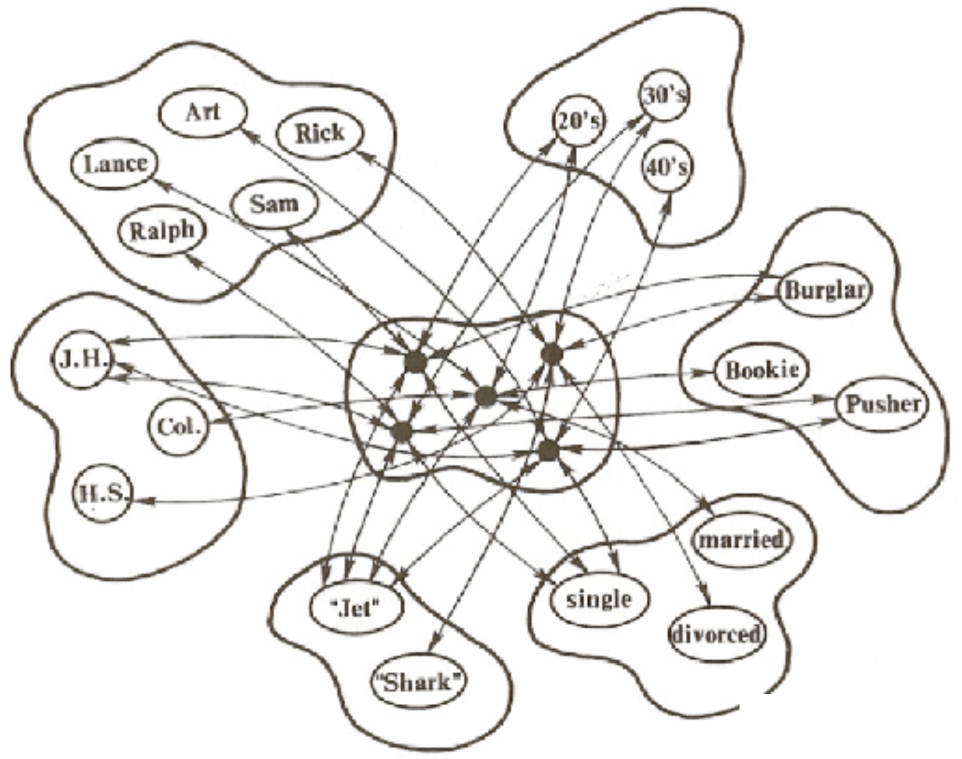
Natural speech reveals the semantic maps that tile human cerebral cortex

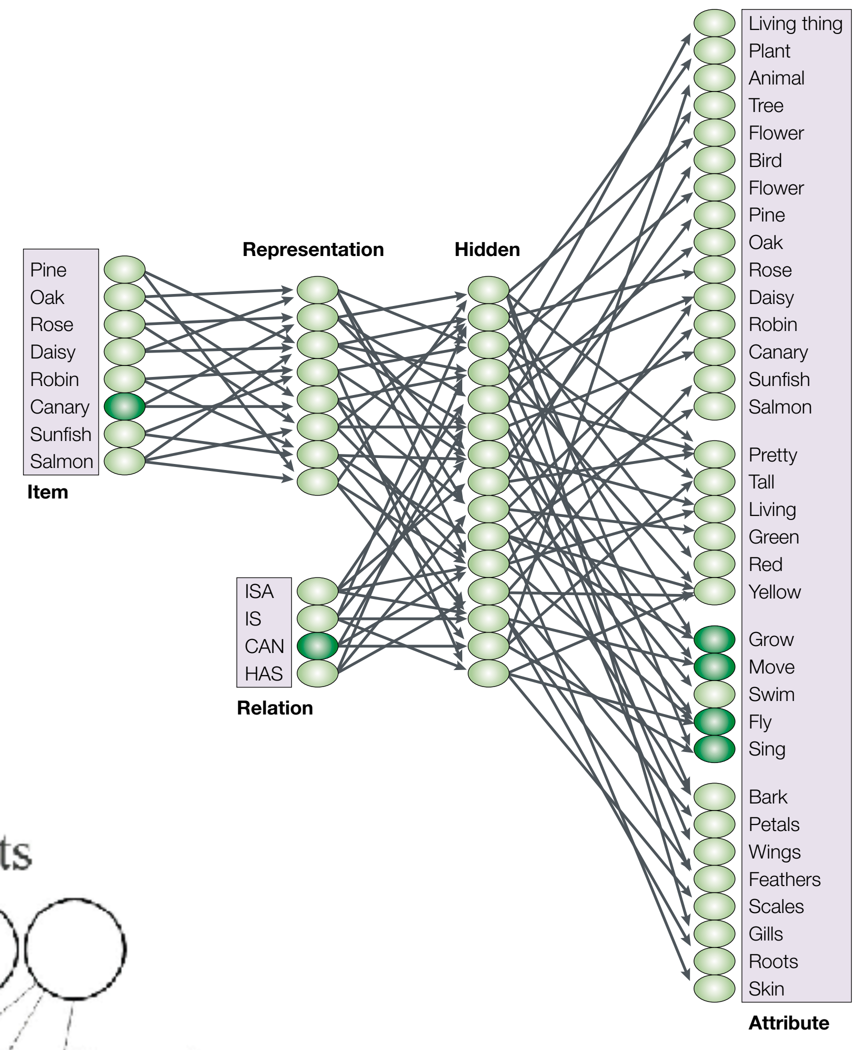
Alexander G. Huth[1], Wendy A. de Heer[2], Thomas L. Griffiths[1,2], Frédéric E. Theunissen[1,2] & Jack L. Gallant[1,2]

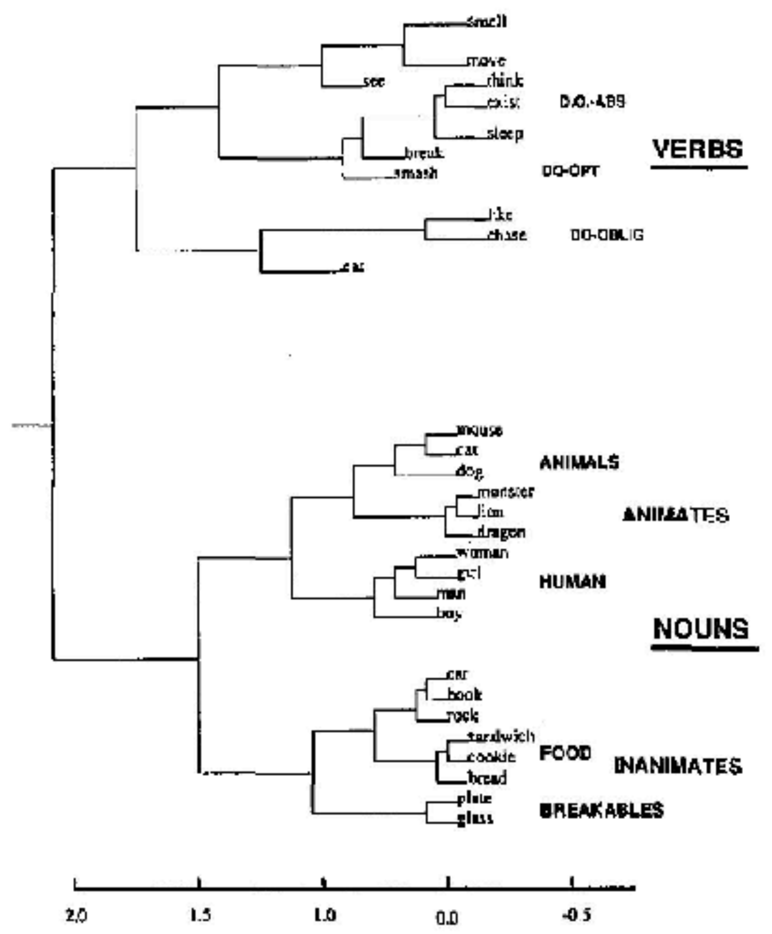
How can recent advances in AI best advance computational cognitive modeling?

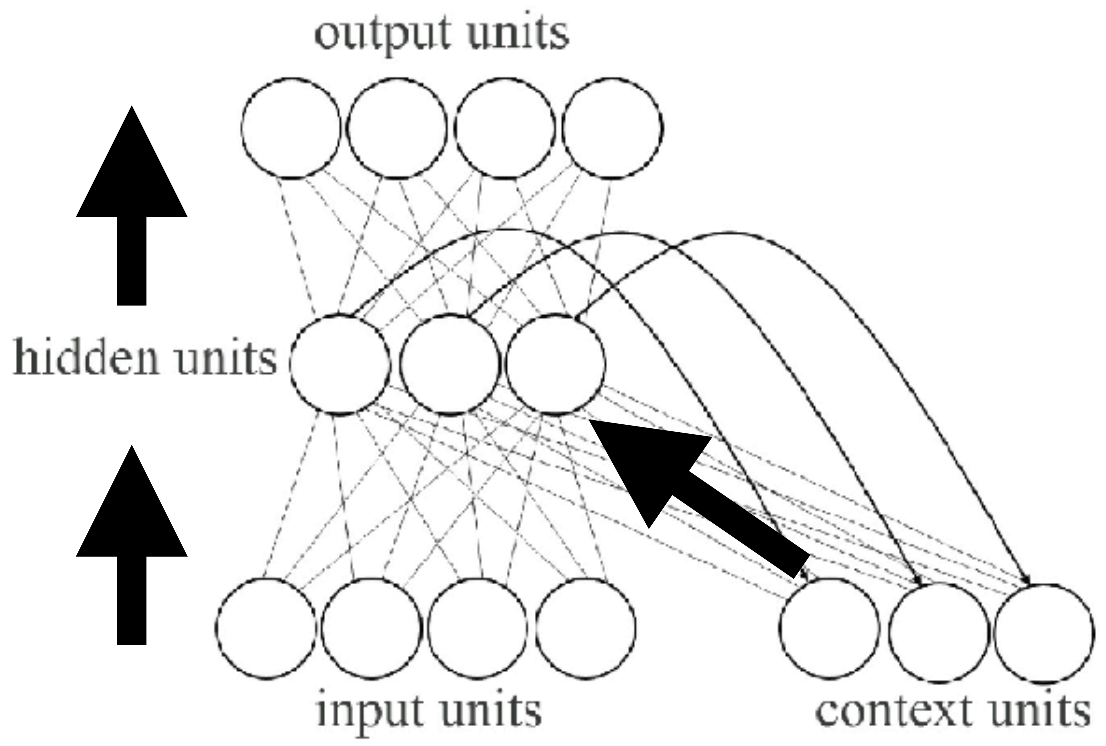
How can recent advances in computational cognitive modeling best advance AI?

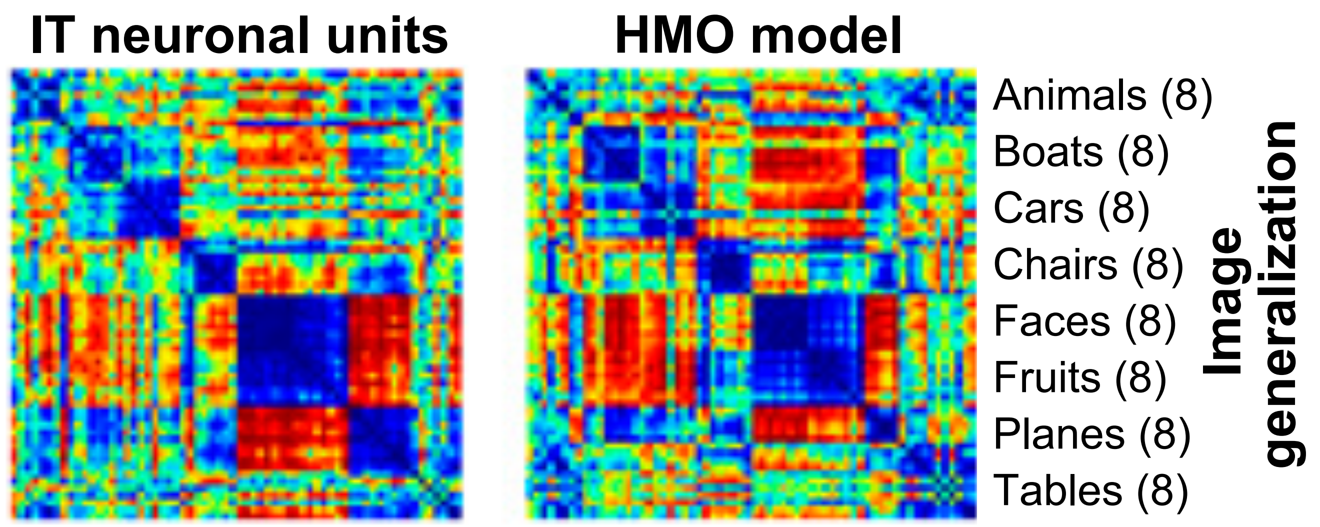
How can recent advances in AI best advance computational cognitive modeling?

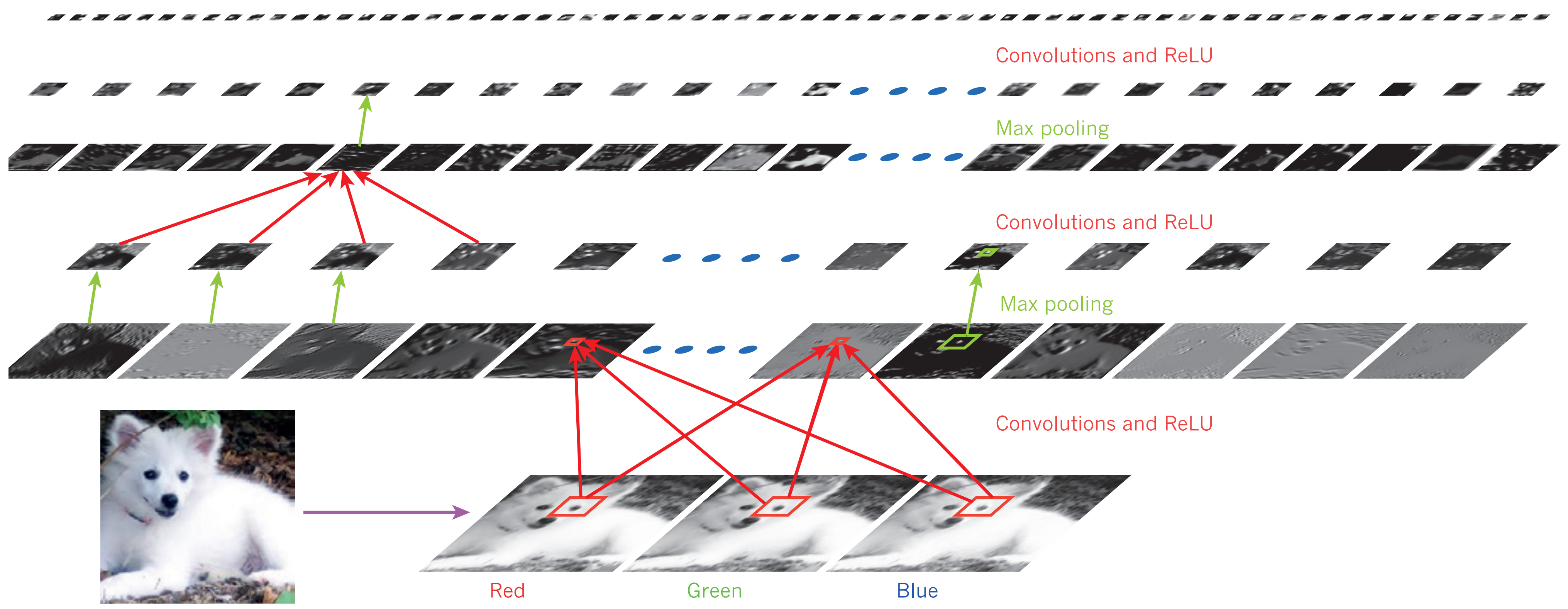




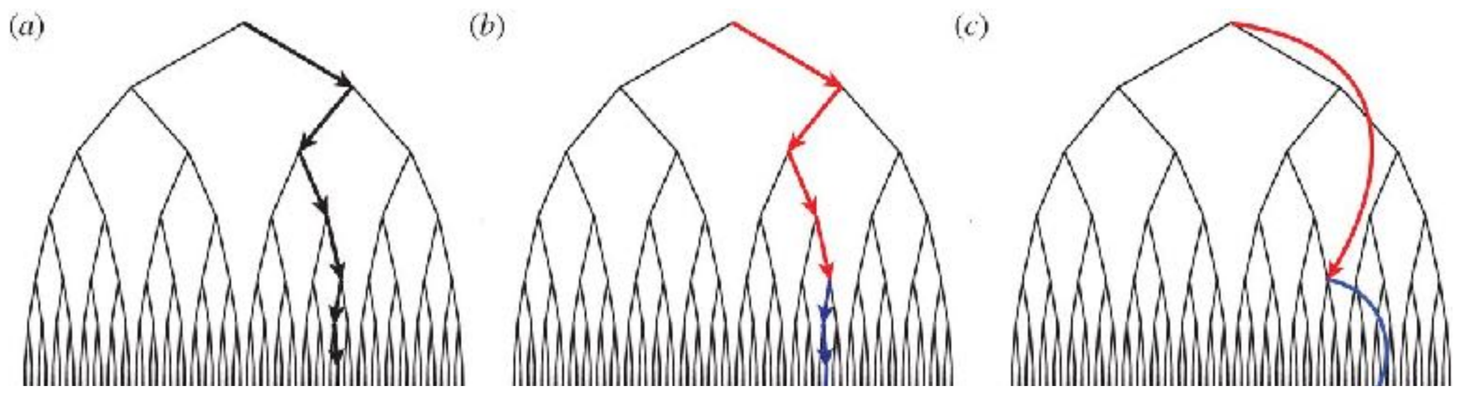


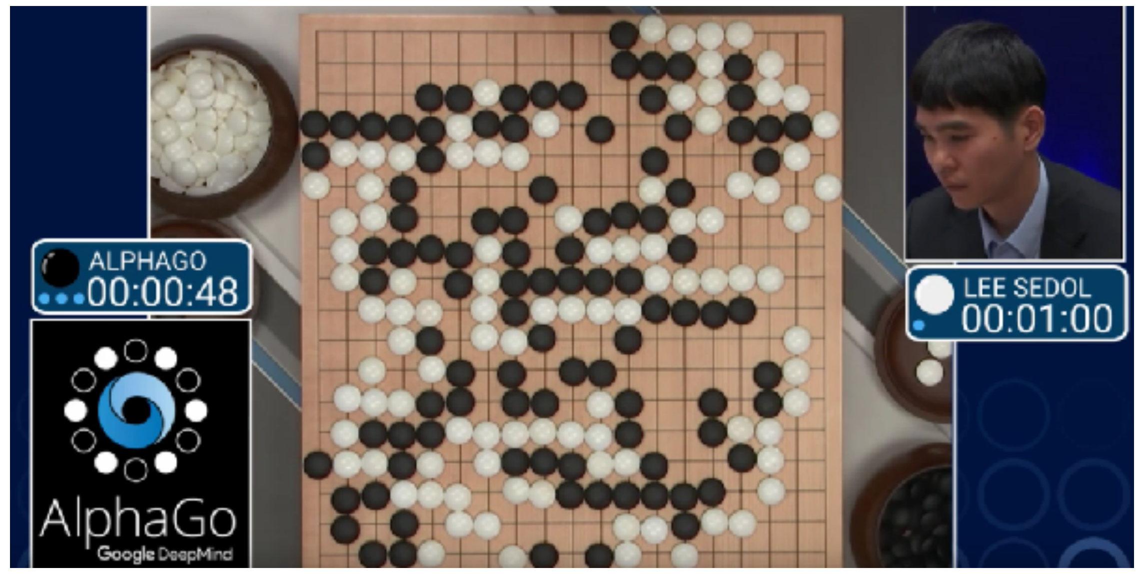


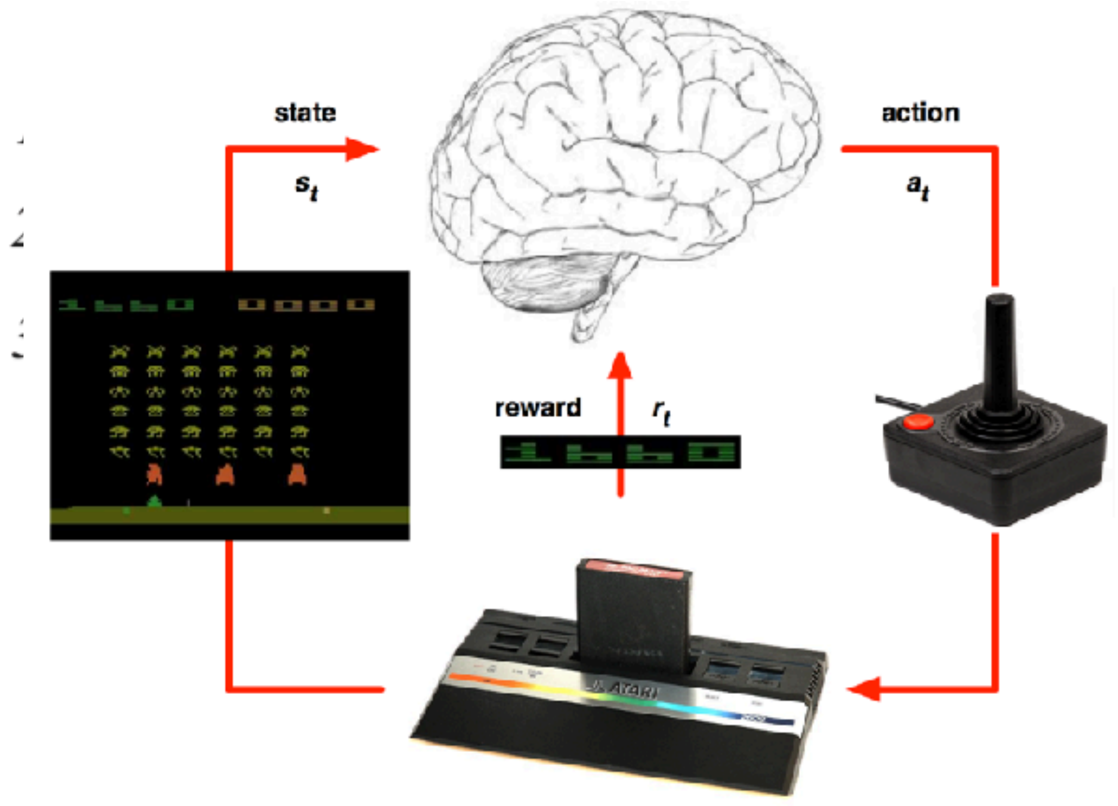


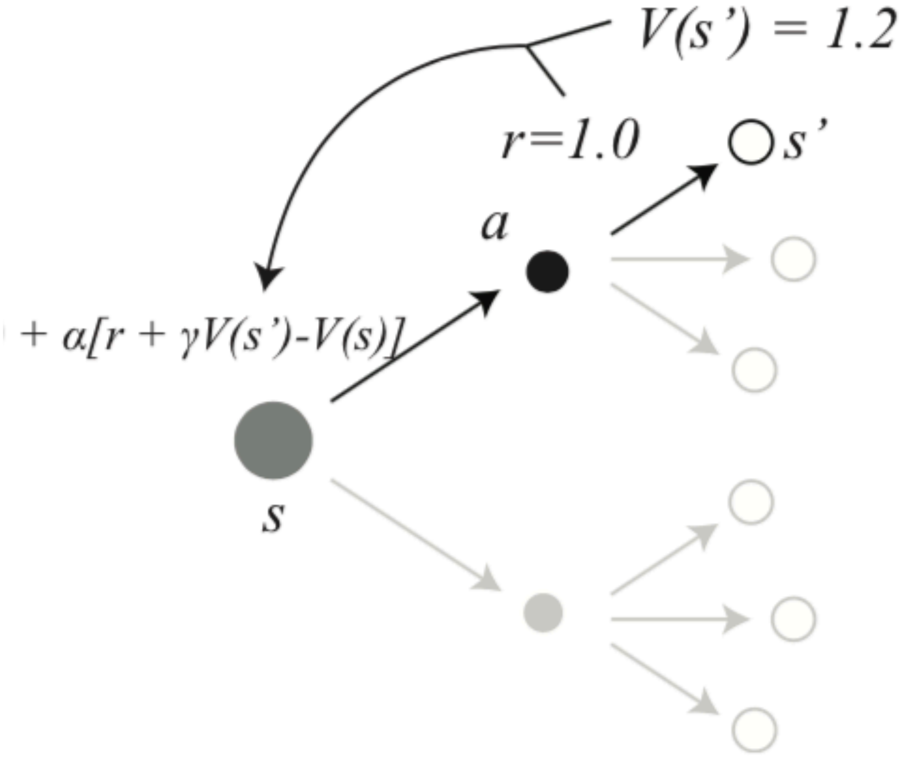


How can recent advances in AI best advance computational cognitive modeling?

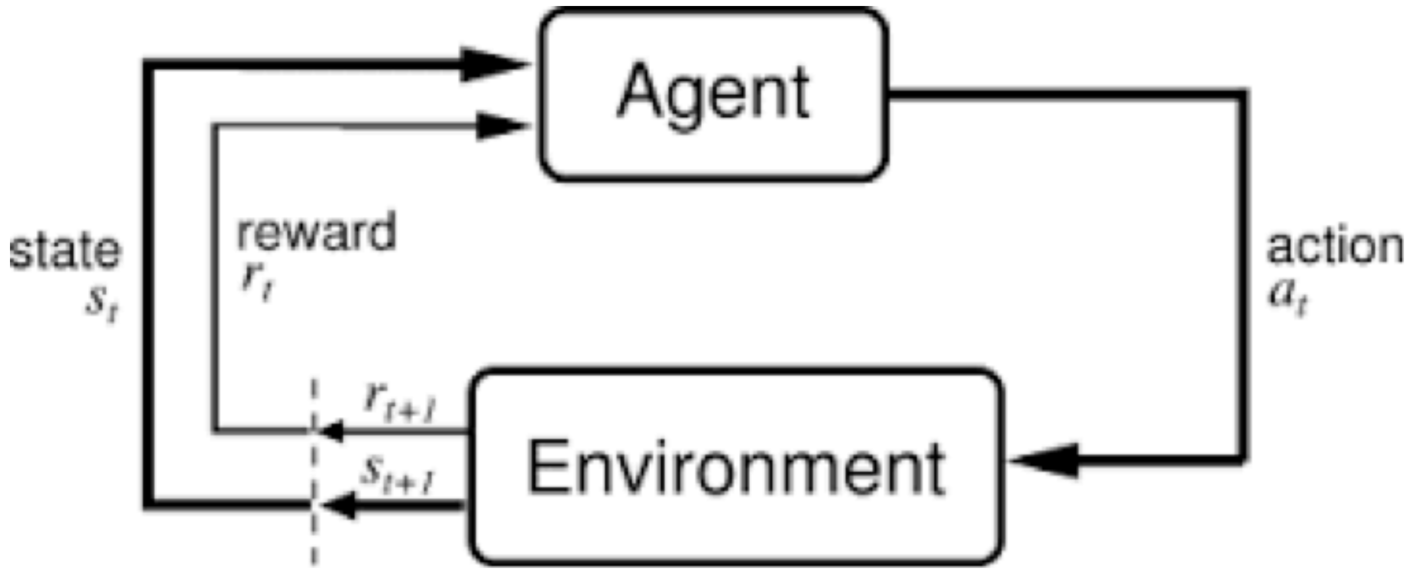


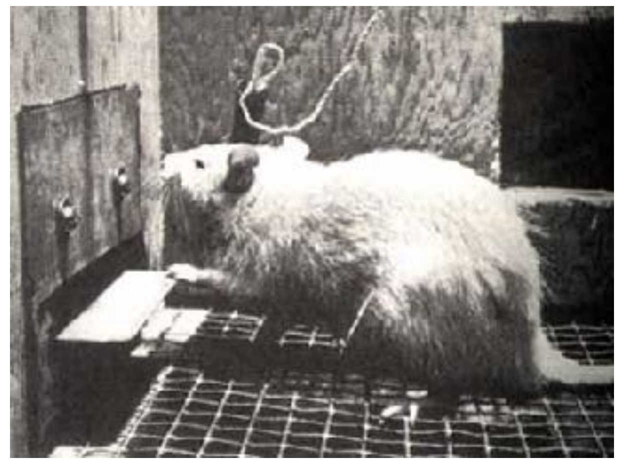






"Recalculating ... recalculating ..."





How can recent advances in computational cognitive modeling best advance AI?

Many human abilities lack compelling computational models:

- scene understanding
- language understanding
- creativity
- general purpose problem solving
- learning new video games
- commonsense reasoning, etc.