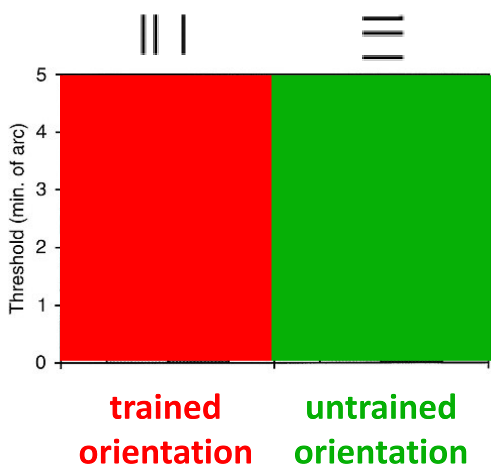
Boost in learning rates evokes full transfer in visual perception tasks

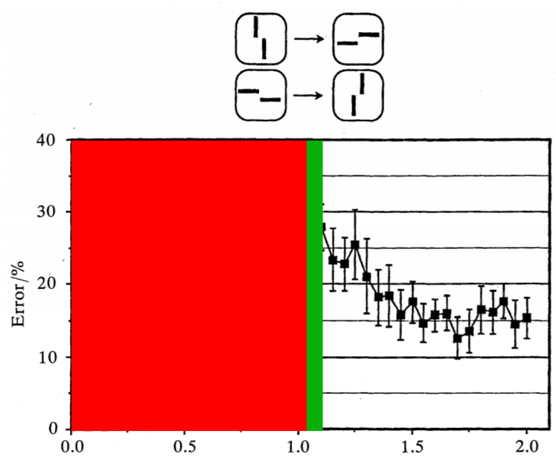
Too Bayesic

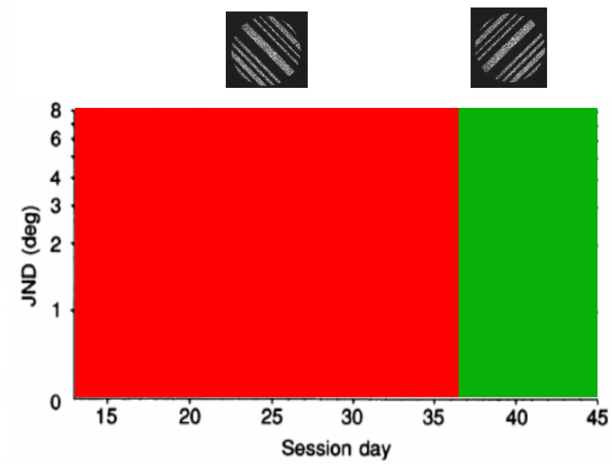
Pedro Cisneros-Velarde, Suraj Chakravarthi Raja, Wanying Jiang, Katrin Sutter

Phenomenon

- The improvement in stimulus discrimination is specific in perceptual learning (Crist et al, 1997; Fahle, 1994; Schoups et al, 1995; Werner et al, 2005).



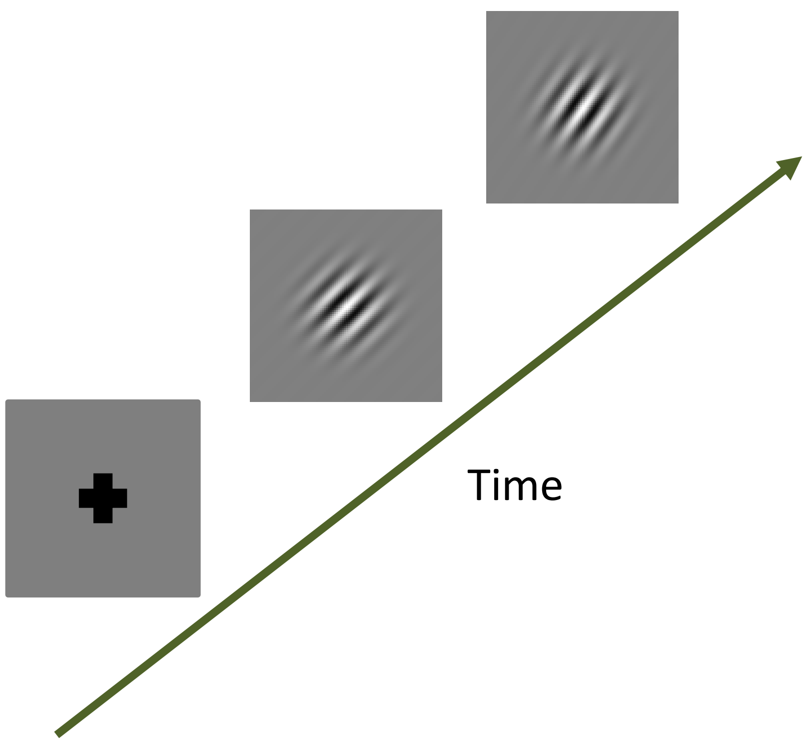




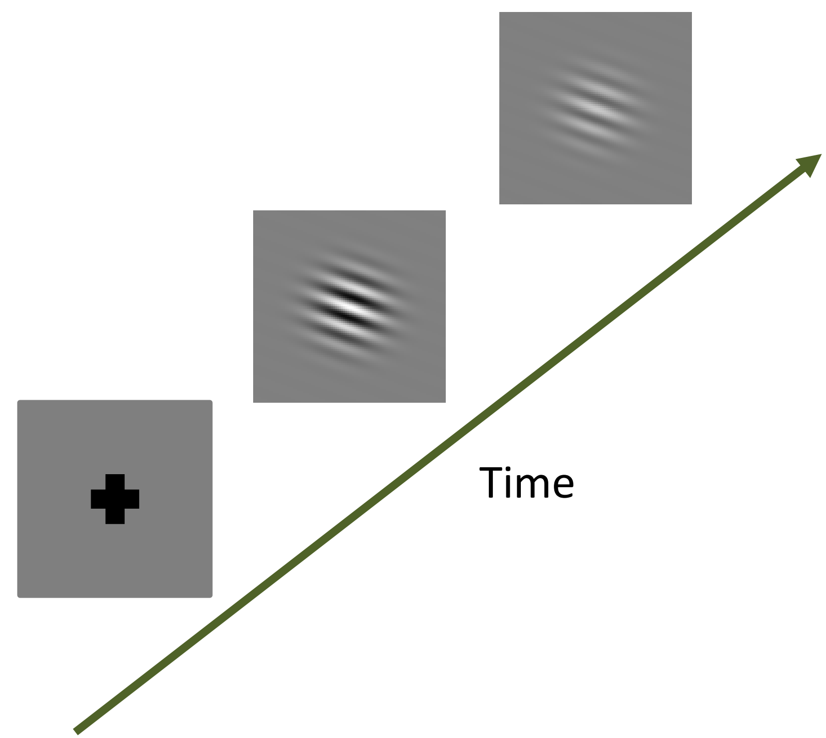
- **Double-training evokes transfer learning** (Xiao et al, 2008; Zhang et al, 2010).

Double-training

- **Training:** Orientation discrimination

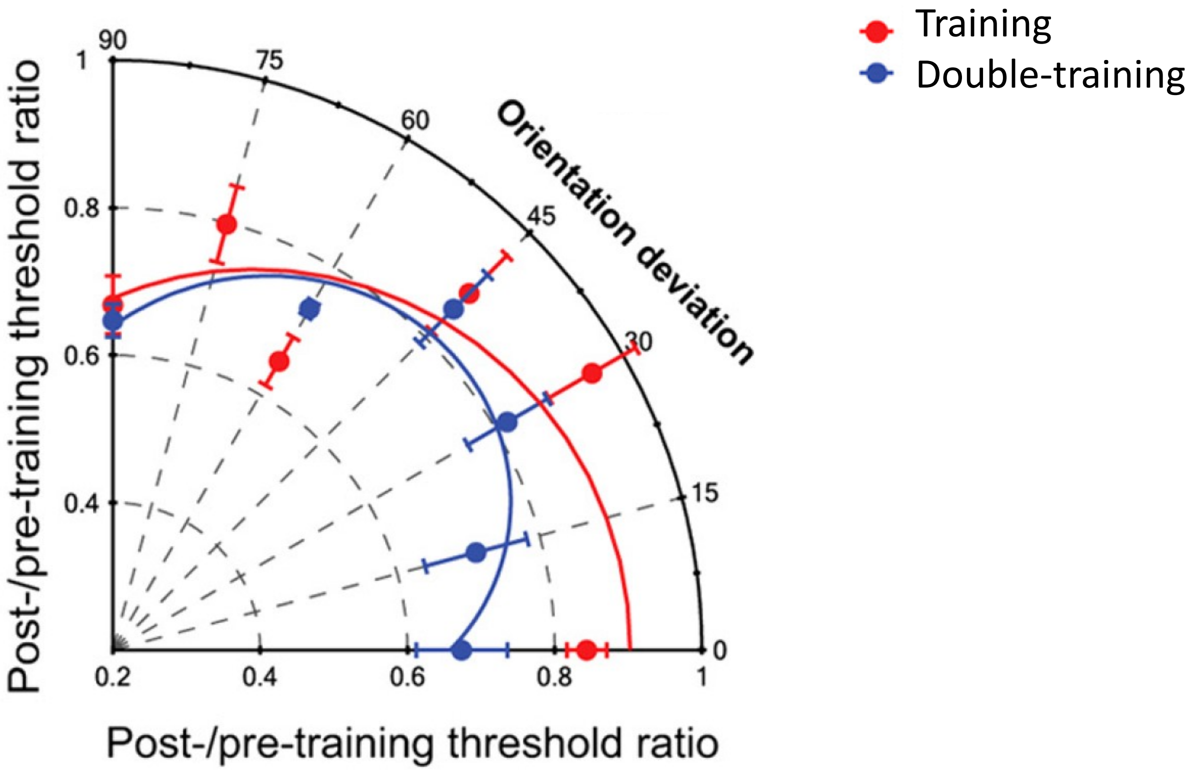


- **Exposure:** Contrast discrimination



Double-training

- Decreased orientation specificity due to double-learning

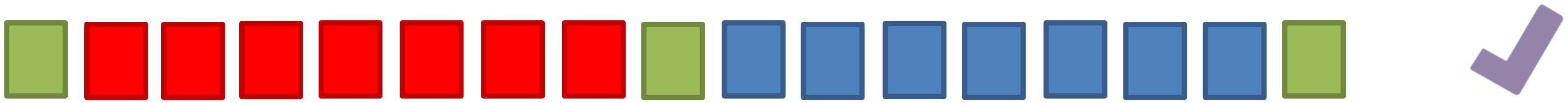


Zhang et al, 2010

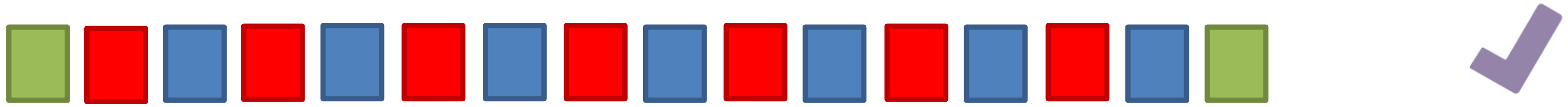
Double-training

Testing, Training, Exposure

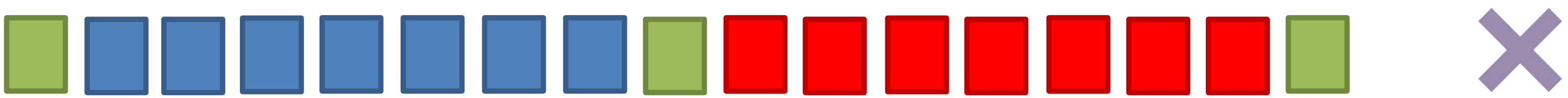
- ***Case 1:*** Training first, exposure later



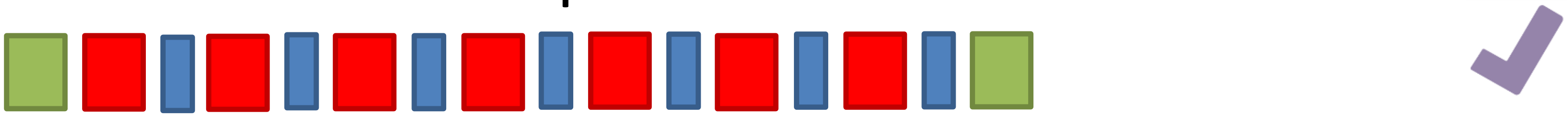
- ***Case 2:*** Interleaved training and exposure



- ***Case 3:*** Exposure first, training later



- ***Case 4:*** Reduced exposure



Double-training: Conclusions

1. Exposure phases must have mutual information with the training phases.

2. Evidence suggests that training phases boost the learning improvement in the following exposure phases, but not the other way around.

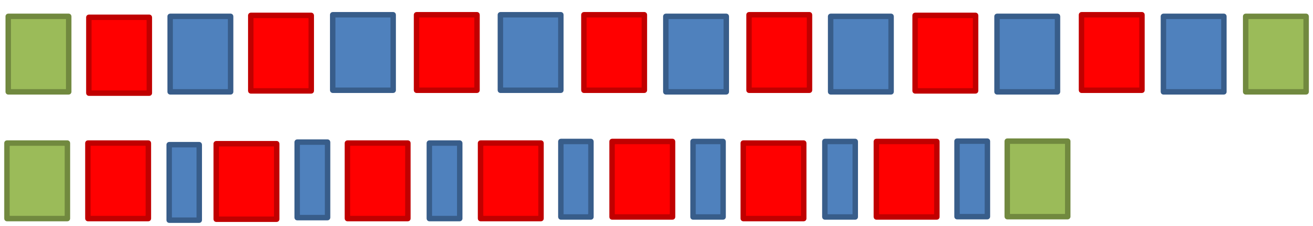
Question and Hypothesis

- **Question:**

How do the length and order of exposure phases affect the transfer of learning?

- **Hypothesis:**

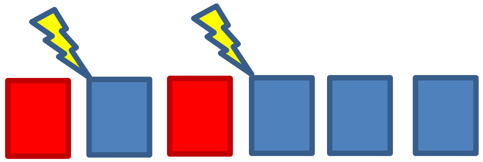
The boost mechanism in the learning rate allows a reduction in the length of exposure phases while achieving full transfer.



Assumptions

1. Learning is additive and its rate **decays exponentially** through time in both training and exposure phases.

2. A learning improvement or boost is **enabled in the transition** to the first exposure block following a training phase.



3. The **learning in training phases** is limited to a maximum learning value while the **learning in exposure phases** is limited to the accumulated learning of the training phases.

4. The model assumes **full retention** between any two blocks.

Model Variables

- **Inputs:** The sequence of training and exposure phases $u \in \{0,1\}^n$.

- **Outputs:** Current value of accumulated learning for training and transfer orientations $T_{tr}, T_{tf} \in [0,1]$.

- **Latent variables:** Decay rates for exposure phases $\alpha_{tf} \in (0,1)$.

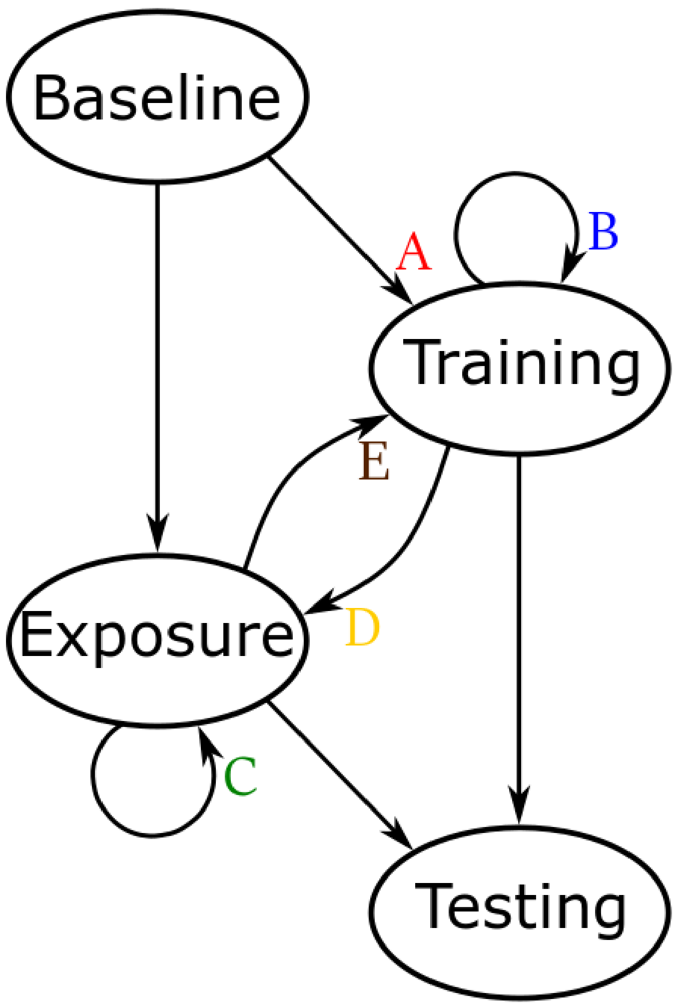
Model Parameters

- Decay rates for training $\alpha_{tr} \in (0,1)$.
- Learning improvement per unit time without decay $l \in (0,1)$.

Selected Toolkit

Finite State Machines (FSM)

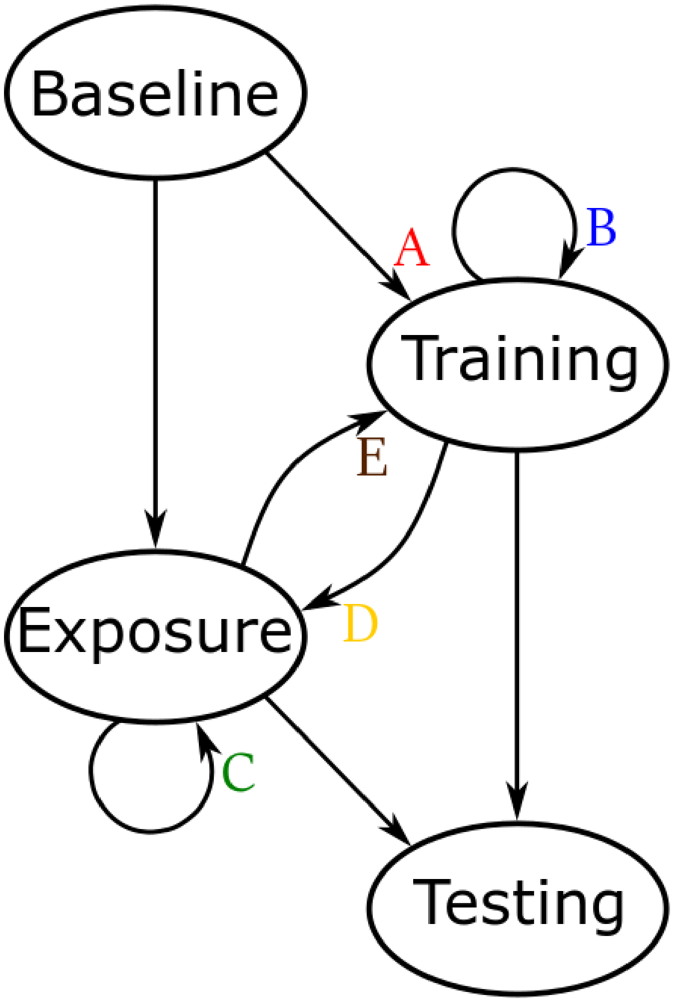
Model Schematics



Baseline

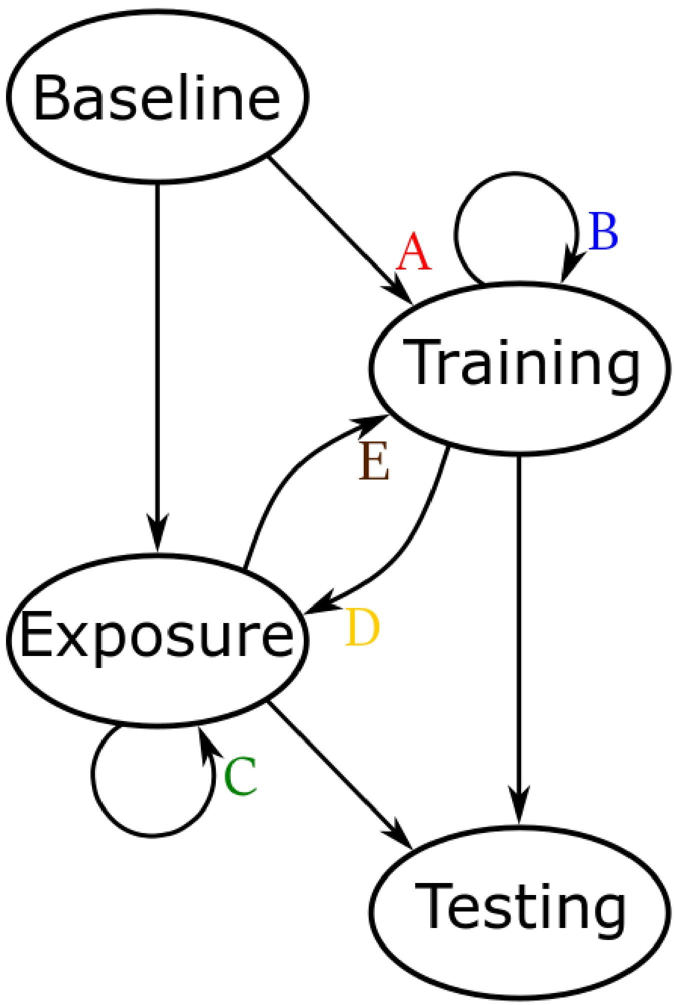
$$T_{tr} \leftarrow 0 \qquad C_{tr} \leftarrow 0$$
$$T_{tf} \leftarrow 0 \qquad C_{tf} \leftarrow 0$$
$$f_1 \leftarrow 0$$

Model Schematics



$$\text{B,A}\;\left|\;\begin{array}{l} C_{tr} \leftarrow C_{tr} + 1 \\ T_{tr} \leftarrow T_{tr} + \alpha_{tr}^{C_{tr}} \cdot l \end{array}\right.$$

Model Schematics



C

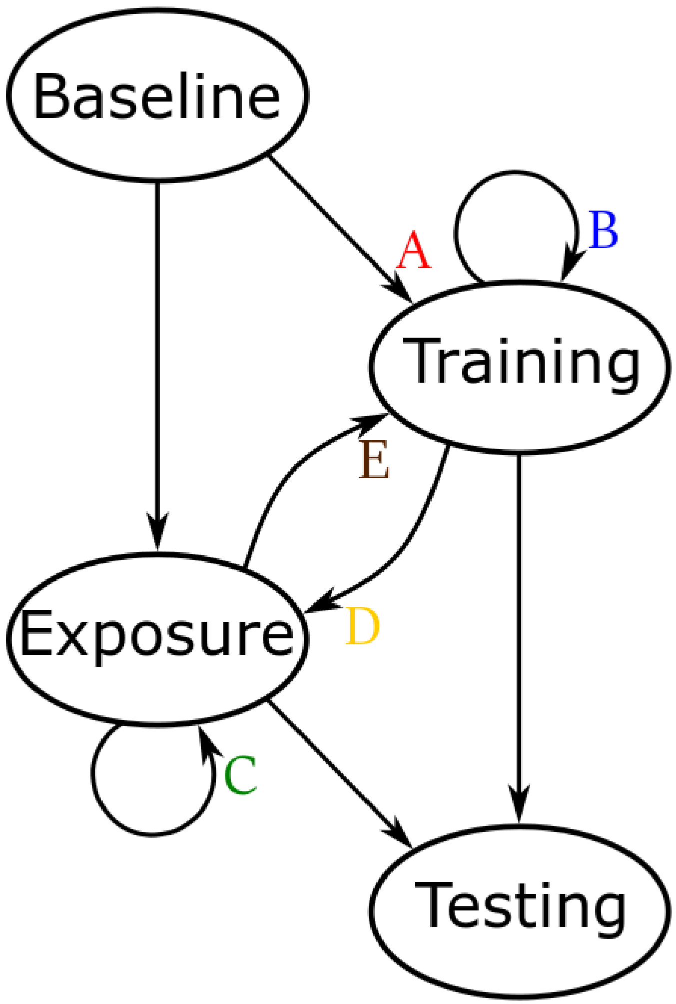
$$if(f_1 = 1)\quad C_{tf} \leftarrow C_{tf} + 1$$
$$if(f_1 = 1)\quad T_{tf} \leftarrow T_{tf} + \alpha_{tf}^{C_{tf}} \cdot l$$

D

$$f_1 \leftarrow 1 \quad C_{tf} \leftarrow 0$$
$$\alpha_{tf} \leftarrow 1 - e^{-\lambda\cdot(C_0+C_{tr}}$$

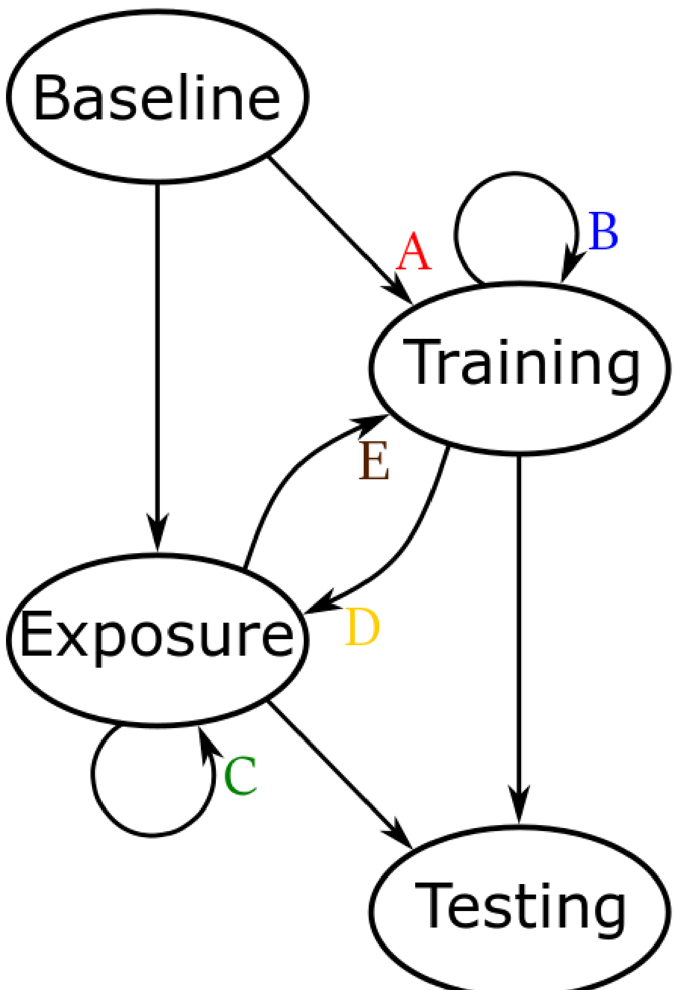
$$T_{tf} \leftarrow T_{tf} + \alpha_{tf} \cdot l$$

Model Schematics



$$\mathrm{E}\left|\begin{array}{l} C_{tr} \leftarrow C_{tr} + 1 \\ T_{tr} \leftarrow T_{tr} + \alpha_{tr}^{C_{tr}} \cdot l \end{array}\right.$$

Model Schematics



Baseline

$$T_{tr} \leftarrow 0 \qquad C_{tr} \leftarrow 0$$
$$T_{tf} \leftarrow 0 \qquad C_{tf} \leftarrow 0$$
$$f_1 \leftarrow 0$$

B,A

$$C_{tr} \leftarrow C_{tr} + 1$$
$$T_{tr} \leftarrow T_{tr} + \alpha_{tr}^{C_{tr}} \cdot l$$

C

$$if(f_1 = 1) \quad C_{tf} \leftarrow C_{tf} + 1$$
$$if(f_1 = 1) \quad T_{tf} \leftarrow T_{tf} + \alpha_{tf}^{C_{tf}} \cdot l$$

D

$$f_1 \leftarrow 1 \qquad C_{tf} \leftarrow 0$$
$$\alpha_{tf} \leftarrow 1 - e^{-\lambda \cdot (C_0 + C_{tr})}$$
$$T_{tf} \leftarrow T_{tf} + \alpha_{tf} \cdot l$$

E

$$C_{tr} \leftarrow C_{tr} + 1$$
$$T_{tr} \leftarrow T_{tr} + \alpha_{tr}^{C_{tr}} \cdot l$$

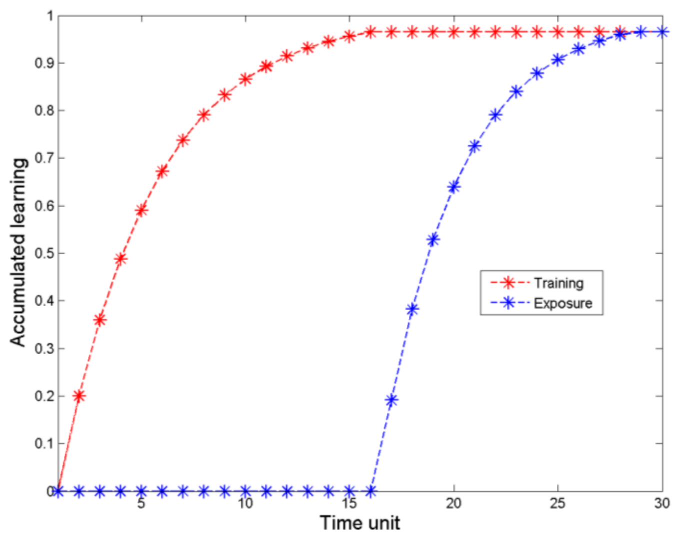
Simulations (I)

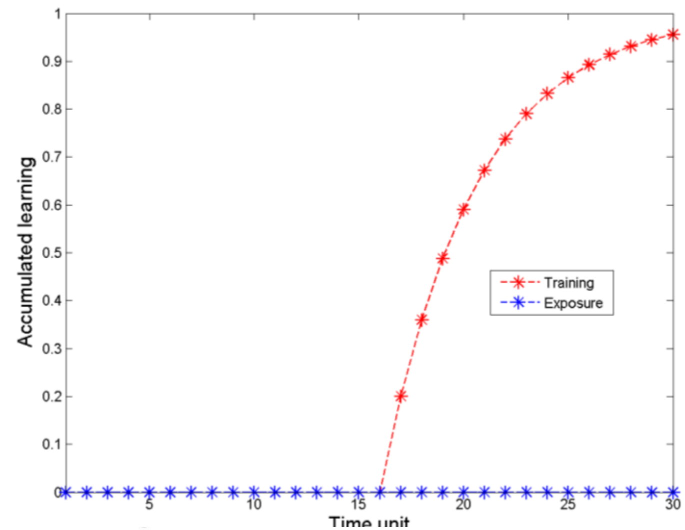
- There are 15 training blocks and at most 15 exposure blocks are inserted at different positions among the training blocks.
- All of these simulations replicate the qualitative behaviour of the data.

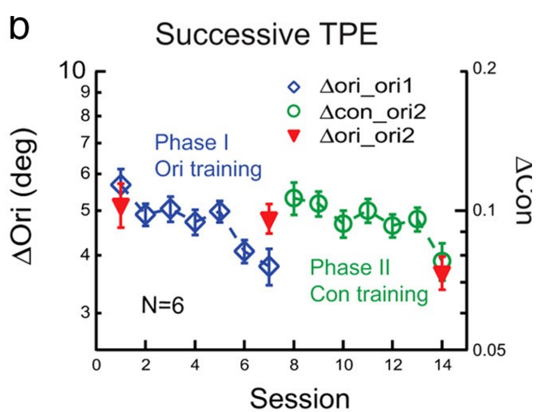
Simulations (I)

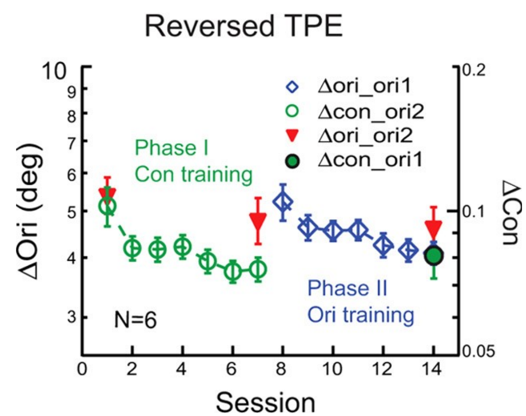
$u = 1 \times \{15\, Training, 15\, Exposure\}$

$u = 1 \times \{15\, Exposure, 15\, Training\}$



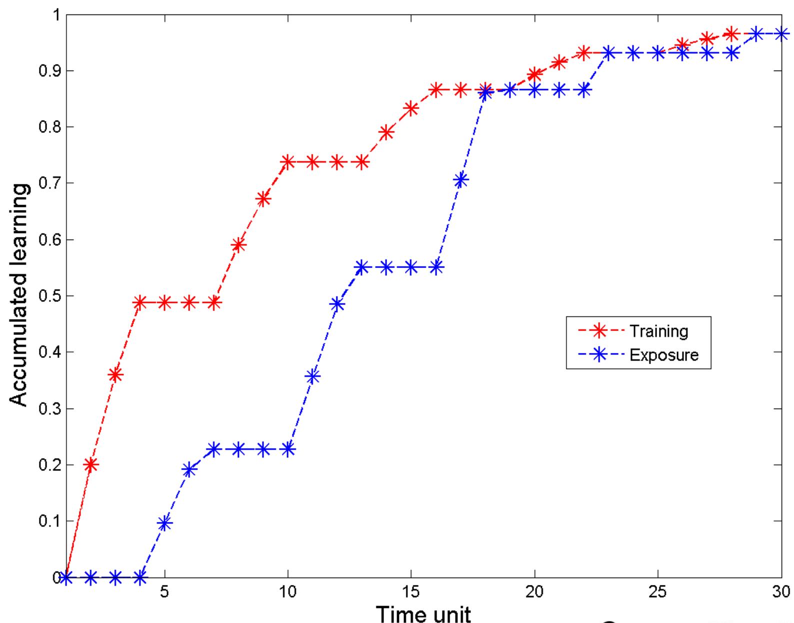




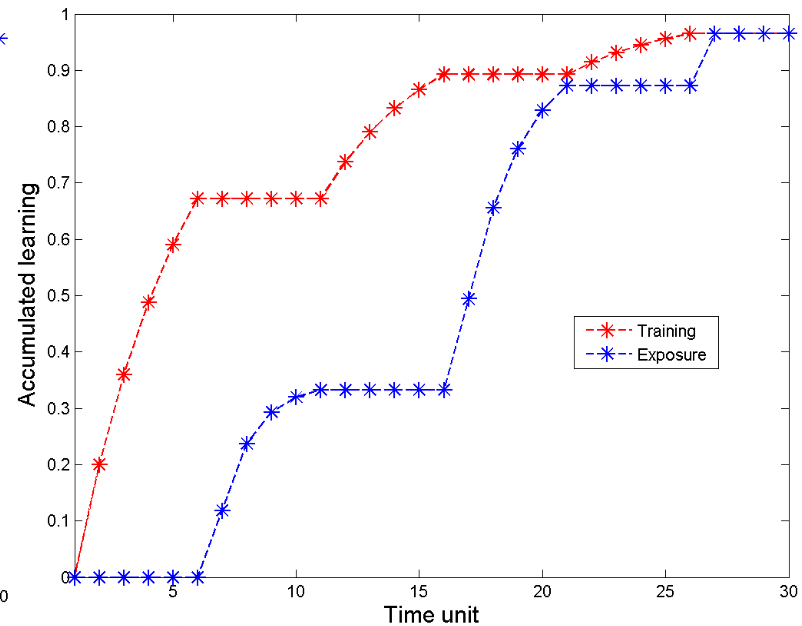


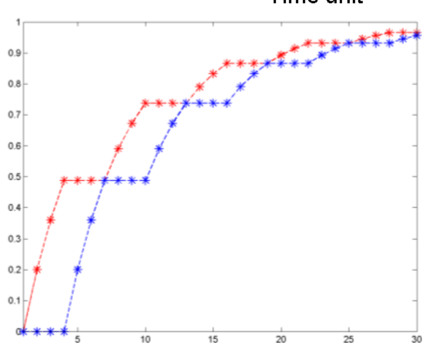
Simulations (I): Interleaved

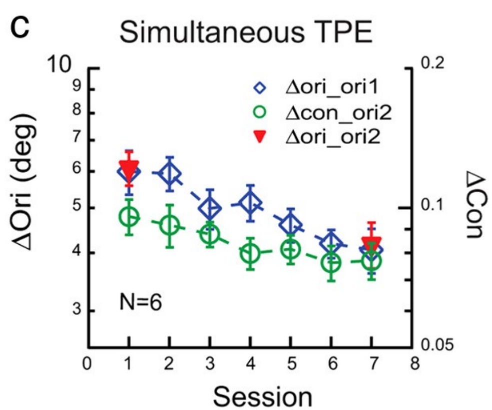
$u = 5 \times \{3\, Training, 3\, Exposure\}$

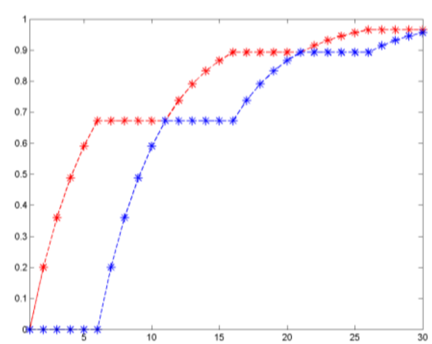


$u = 3 \times \{5\, Training, 5\, Exposure\}$



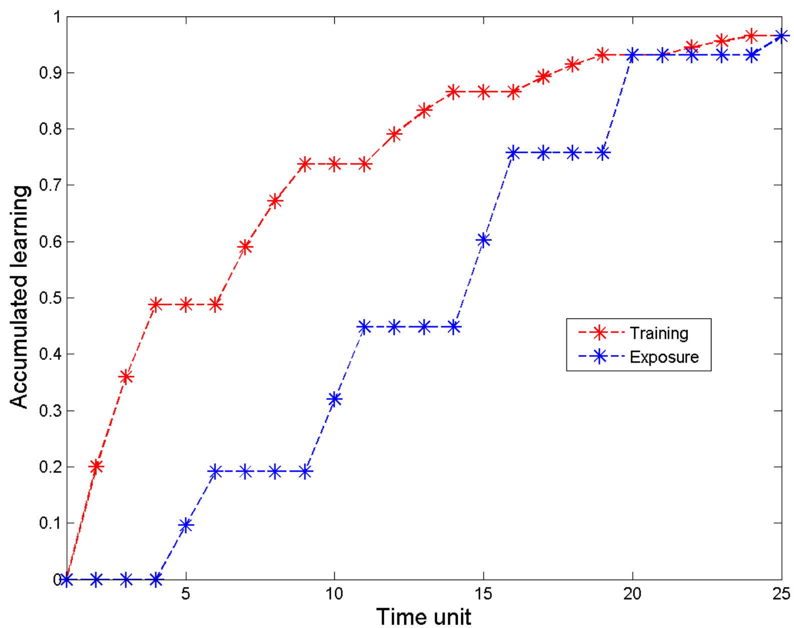


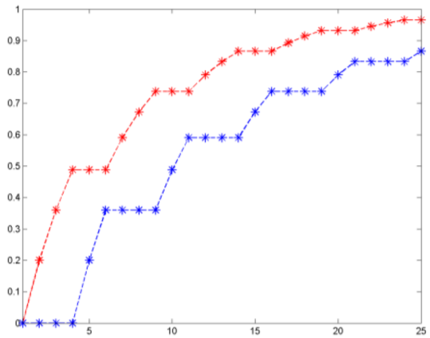




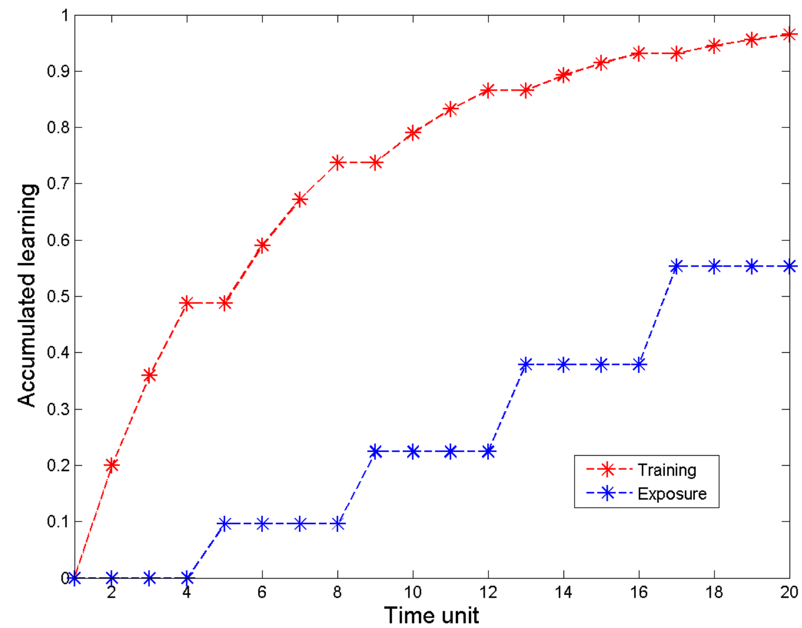
Simulations (I): Reduced Exposure

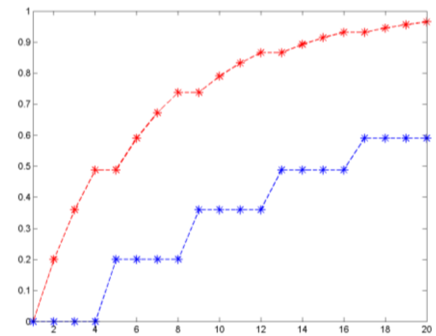
$u = 5 \times \{3\ Training, 2\ Exposure\}$





$u = 5 \times \{3\ Training, 1\ Exposure\}$



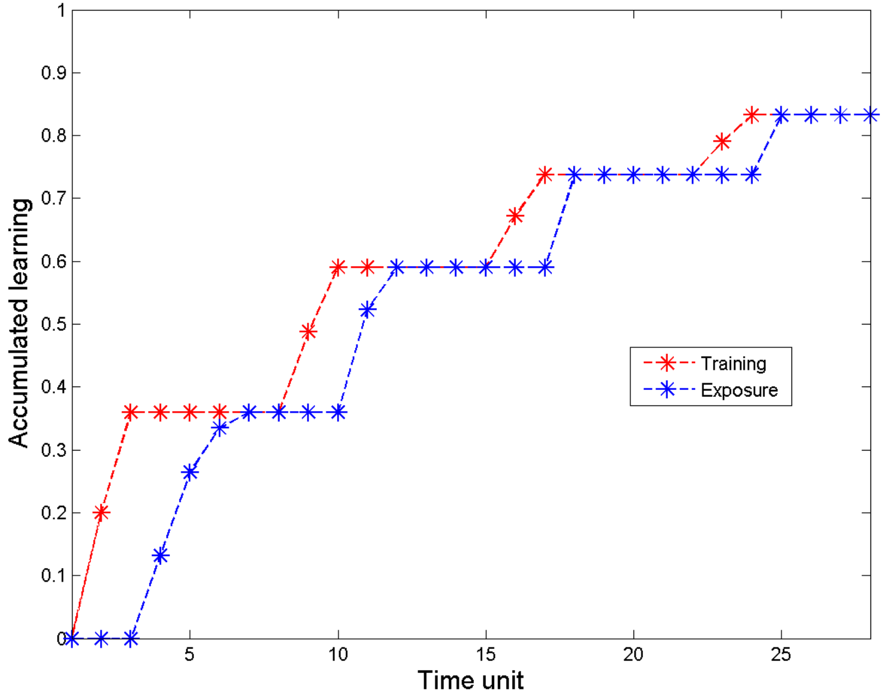


Simulations (II)

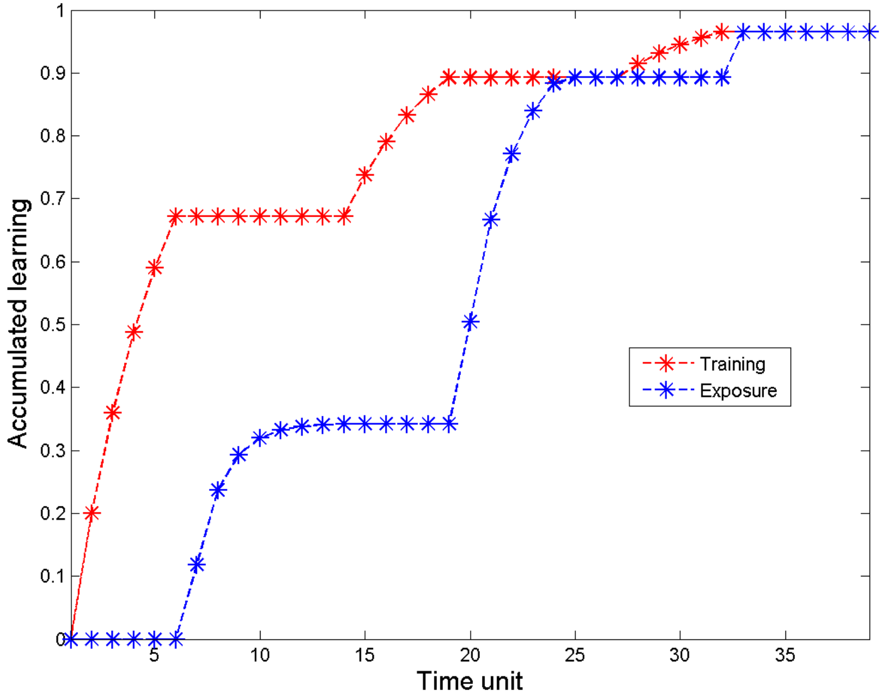
- It is assumed that 17 training blocks are needed for complete learning.
- All of these simulations are novel situations that have not been experimentally verified.

Simulations (II)

$u = 4 \times \{2\ Training, 5\ Exposure\}$

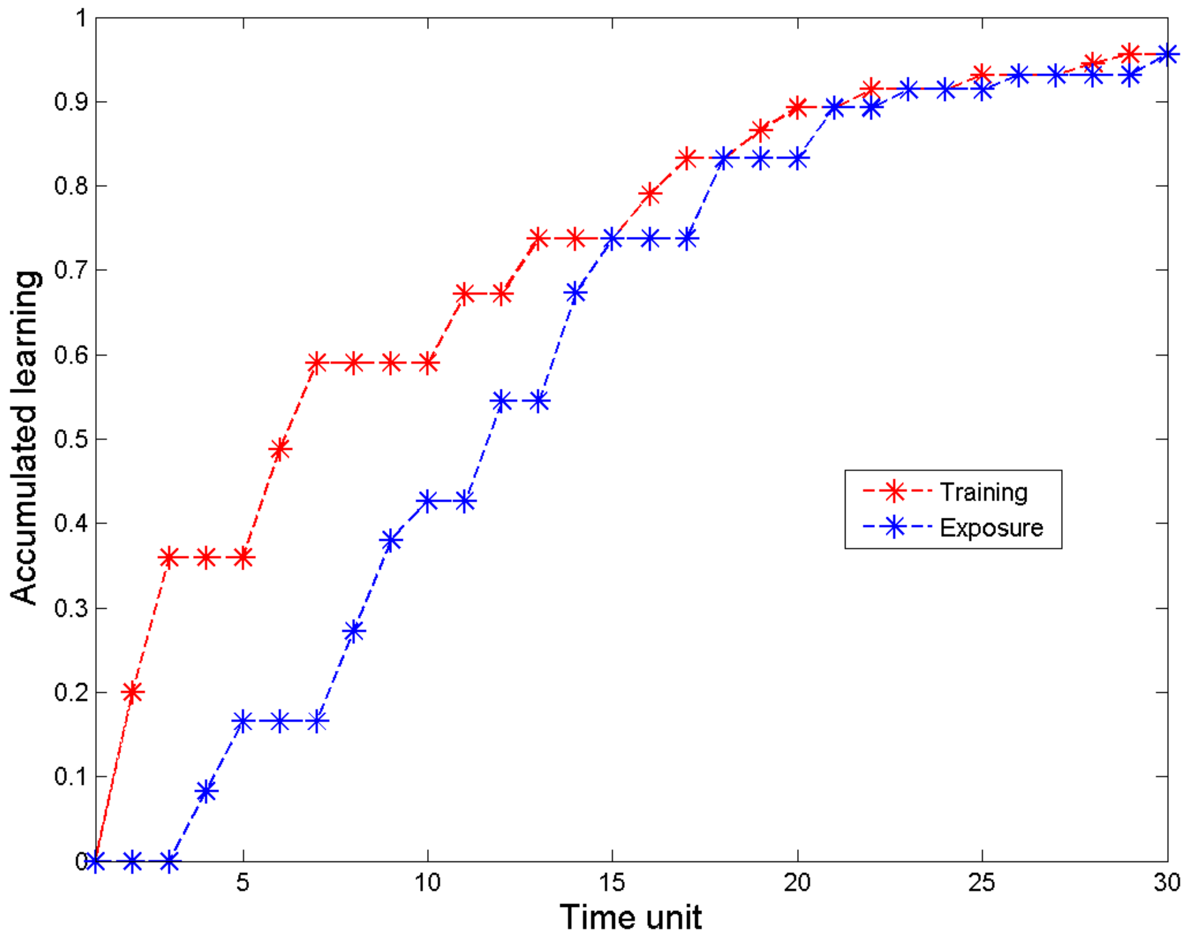


$u = 3 \times \{5\ Training, 7\ Exposure\}$



Simulations (II)

$$u = \{0,0,1,1,0,0,1,1,1,0,1,0,1,1,0,0,1,0,0,1,0,1,1,0,1,1,0,0,1,0\}$$



Simulation Conclusions

A “boosting” mechanism in the learning rate of the exposure phases gives an explanation for all the observed full transfer behavior in the data.

Critical model evaluation

- Our model can be generalized to other visual perceptual tasks (Zhang et al, 2010).

- Insufficient data to assess other combinations of training and exposure blocks, however, the model creates predictions for future experiments.
  - E.g., arbitrary increasing the length of the exposure phases does not lead to better performance at the testing angle.

- Careful hand-tuning is required.

- The model does not account for temporal decay of retention.

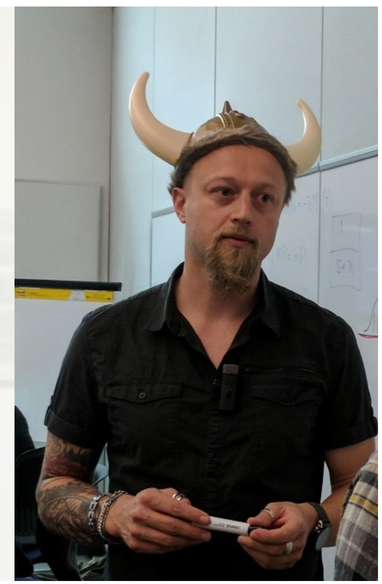
Summary & conclusions

- What have you learned?
  - Model parameters have a significant impact on fitting its outcomes to the qualitative behavior of the data
  - Additional testing of subjects will allow us to subject our model to more elaborate edge-case testing.

- Future work
  - We expect to extend our model to account for generalization in the double-training tasks.

Thank you!





References

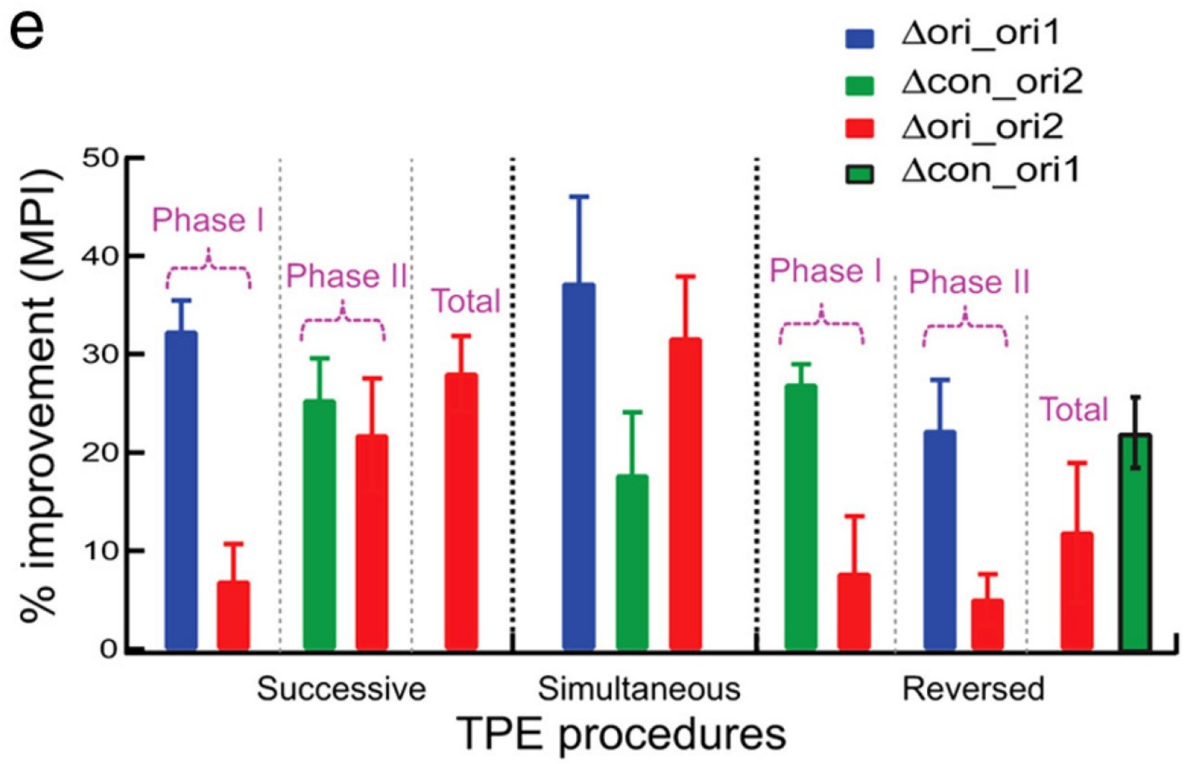
- Crist, R. E., Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (1997). Perceptual learning of spatial localization: specificity for orientation, position, and context. *Journal of neurophysiology*, *78*(6), 2889-2894.
- Fahle, M. (1994). Human pattern recognition: parallel processing and perceptual learning. *Perception*, *23*(4), 411-427.
- Zhang, J. Y., Zhang, G. L., Xiao, L. Q., Klein, S. A., Levi, D. M., & Yu, C. (2010). Rule-based learning explains visual perceptual learning and its specificity and transfer. *Journal of Neuroscience*, *30*(37), 12323-12328
- Schoups, A. A., Vogels, R., & Orban, G. A. (1995). Human perceptual learning in identifying the oblique orientation: retinotopy, orientation specificity and monocularity. *The Journal of physiology*, *483*(3), 797-810.
- Xiao, L. Q., Zhang, J. Y., Wang, R., Klein, S. A., Levi, D. M., & Yu, C. (2008). Complete transfer of perceptual learning across retinal locations enabled by double training. *Current Biology*, *18*(24), 1922-1926.

Additional slides

- For the transition Training to Exposure, after boosting (i.e., updating) the learning rate could have possibly be implemented by

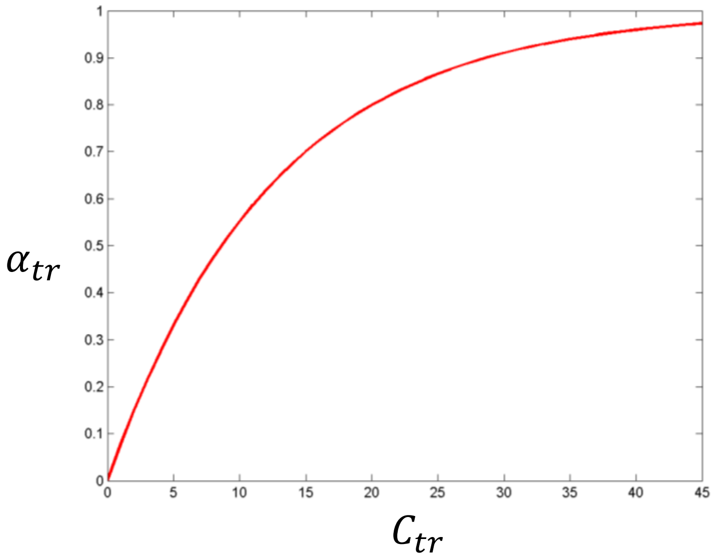
$$T_{tf} \leftarrow T_{tf} + \alpha_{tf}^{c_{tf}} l,\ f_1 \leftarrow 1$$

This approach does not work for any of the parameters.


e
Δori_ori1
Δcon_ori2
Δori_ori2
Δcon_ori1
% improvement (MPI)
50
40
30
20
10
0
Phase I
Phase II
Total
Phase I
Phase II
Total
Successive
Simultaneous
Reversed
TPE procedures

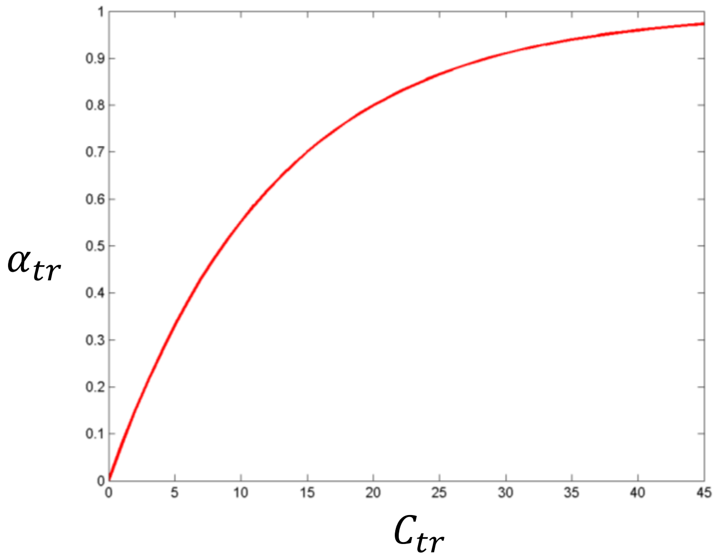
Equations

- Baseline
  - $T_{tr} \leftarrow 0, T_{tf} \leftarrow 0, f_1 \leftarrow 0, C_{tr} \leftarrow 0, C_{tf} \leftarrow 0$
- Baseline to Training, Training to Training
  - $C_{tr} \leftarrow C_{tr} + 1, T_{tr} \leftarrow T_{tr} + \alpha_{tr}^{C_{tr}} l$
- Training to Exposure
  - $\alpha_{tf} \leftarrow 1 - e^{-\lambda(C_0 + C_{tr})}$
  - $T_{tf} \leftarrow T_{tf} + \alpha_{tf} l, f_1 \leftarrow 1,$
- Exposure to Exposure
  - If $f_1 = 0$: $C_{tf} \leftarrow C_{tf} + 1, T_{tf} \leftarrow T_{tf} + \alpha_{tf}^{tf} l$
- Exposure to Training
  - $C_{tr} \leftarrow C_{tr} + 1, T_{tr} \leftarrow T_{tr} + \alpha_{tr}^{C_{tr}} l$



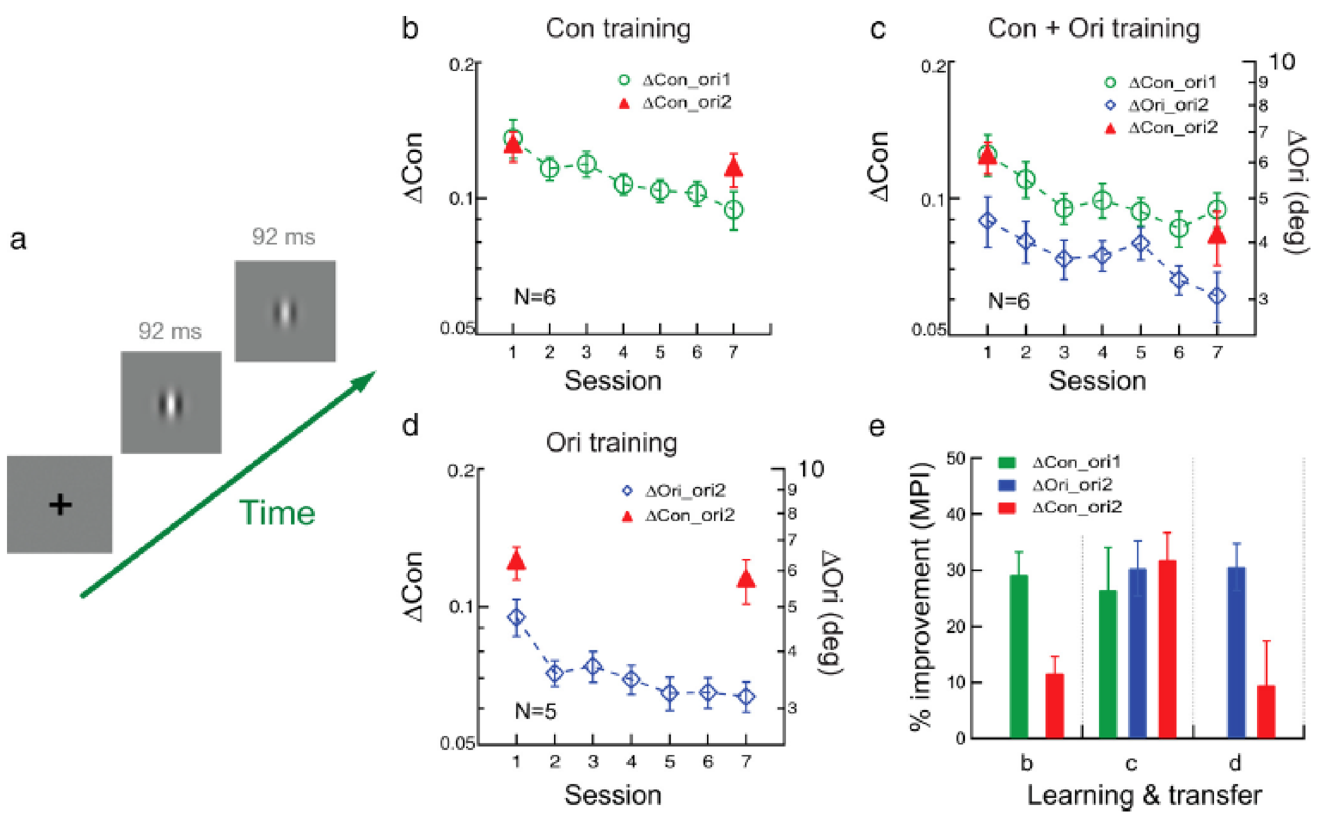
Exponential

```
delta = 0.973;
N = Ntraining*3;
lambda = -log(1-delta)/N;
alpha_SE = 1-exp(-lambda*3);
```



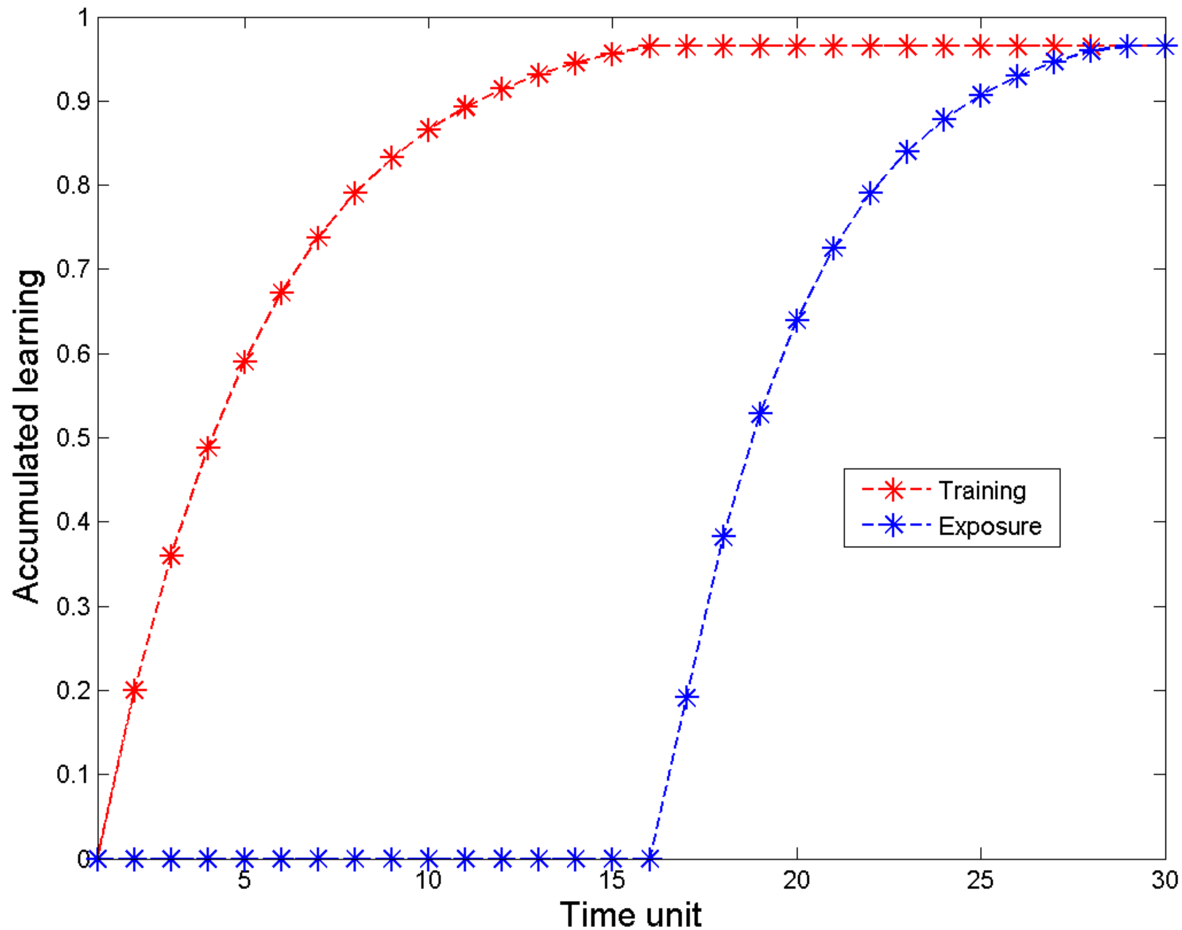
Double-training

- Learning transfer occurs also when the tasks are reversed (e.g. when contrast discrimination is being tested)



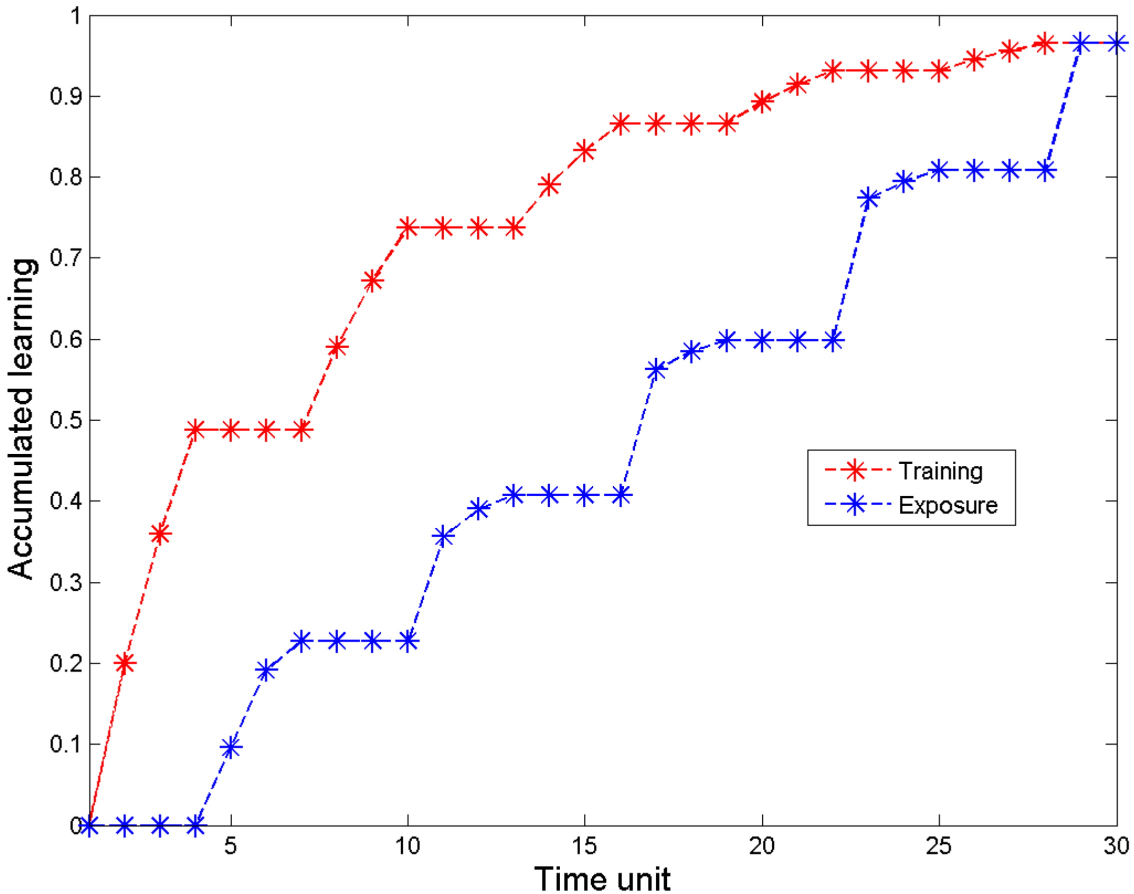
Simulations / results

- **Case 1:** $u = \{15 * Training, 15 * Exposure\}$



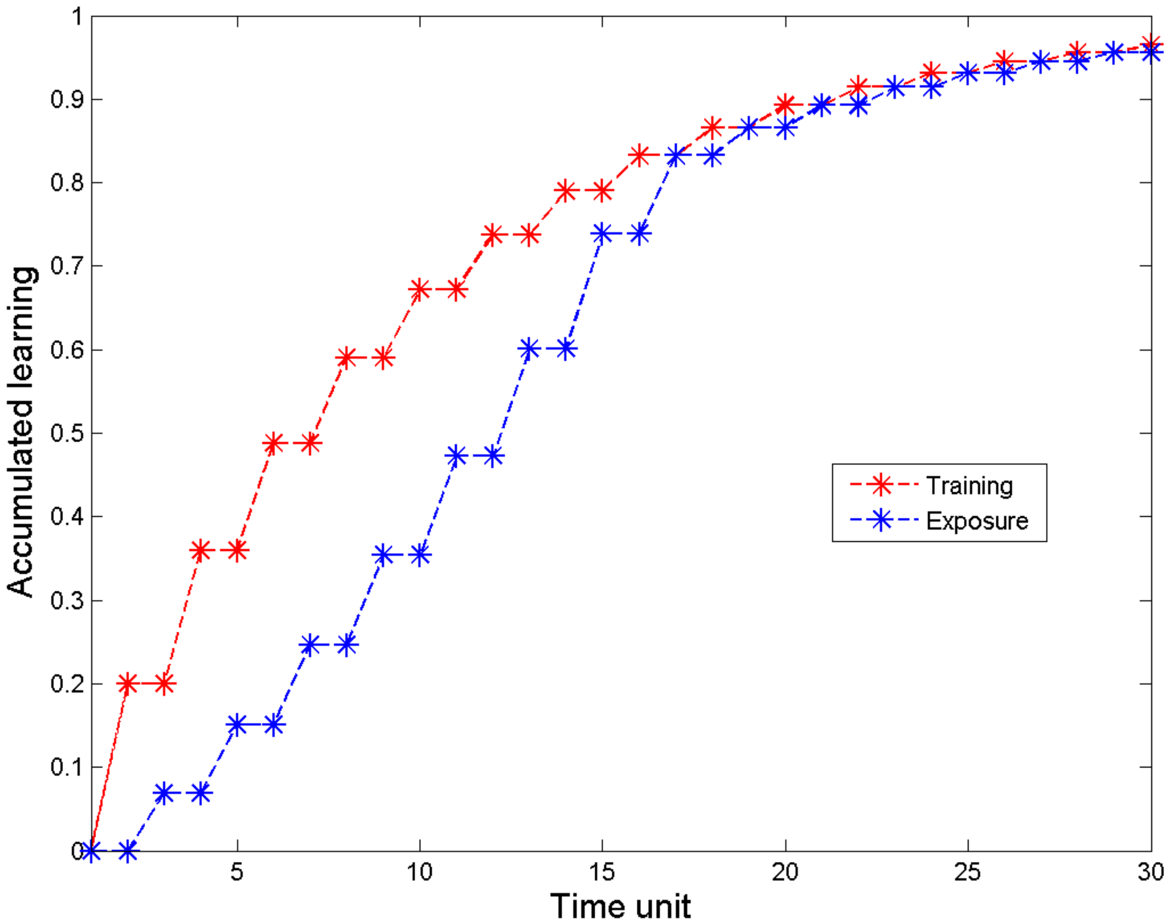
Simulations / results

- **Case 2:** $u = \{3 * Training, 3 * Exposure, 3 * Training, 3 * Exposure, \ldots, 3 * Training, 3 * Exposure\}$



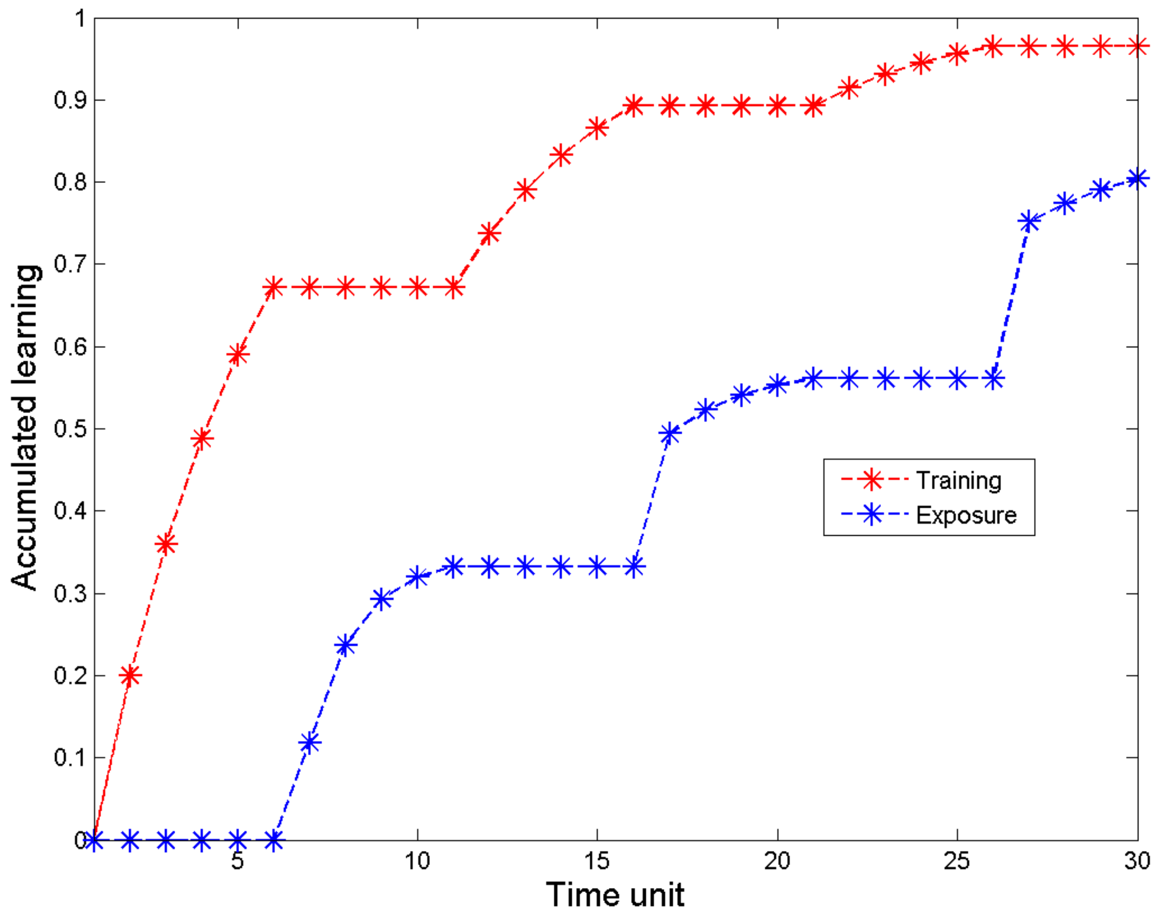
Simulations / results

- **Case 2:** $u = \{1 * Training, 1 * Exposure, 1 * Training, 1 * Exposure, \ldots, 1 * Training, 1 * Exposure\}$



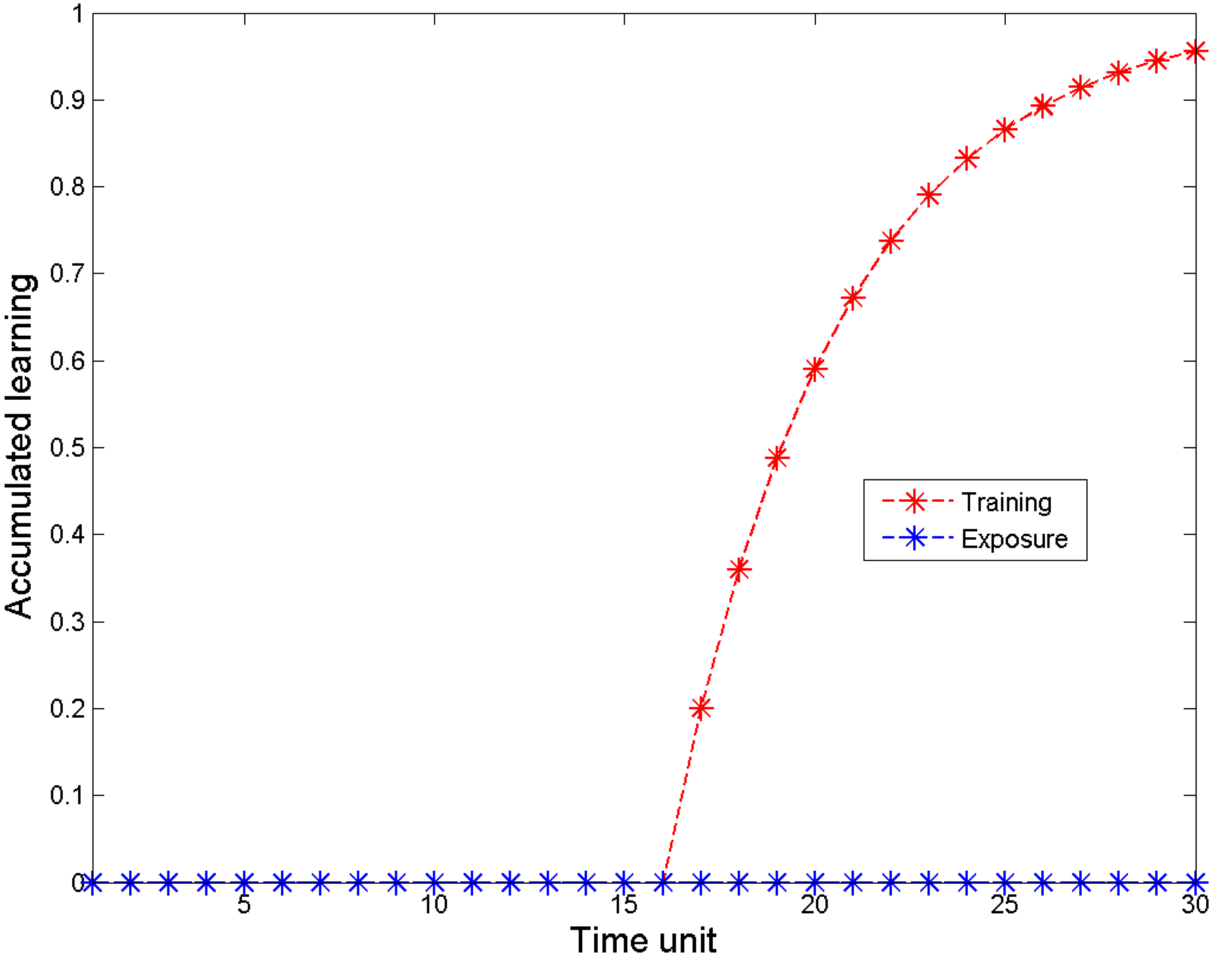
Simulations / results

- **Case 2:** $u = \{5 * Training, 5 * Exposure, 5 * Training, 5 * Exposure, \ldots, 5 * Training, 5 * Exposure\}$



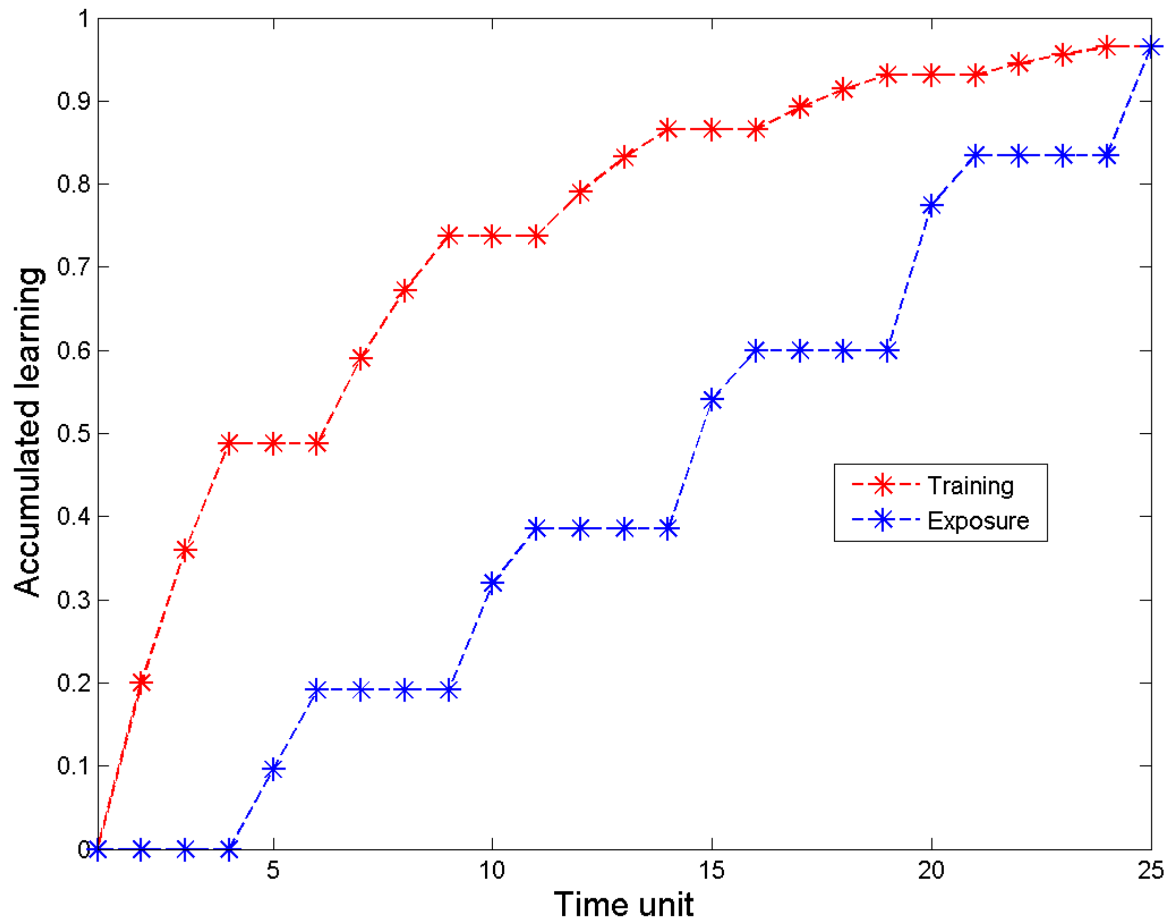
Simulations / results

- **Case 3:** $u = \{15 * Exposure, 15 * Training\}$



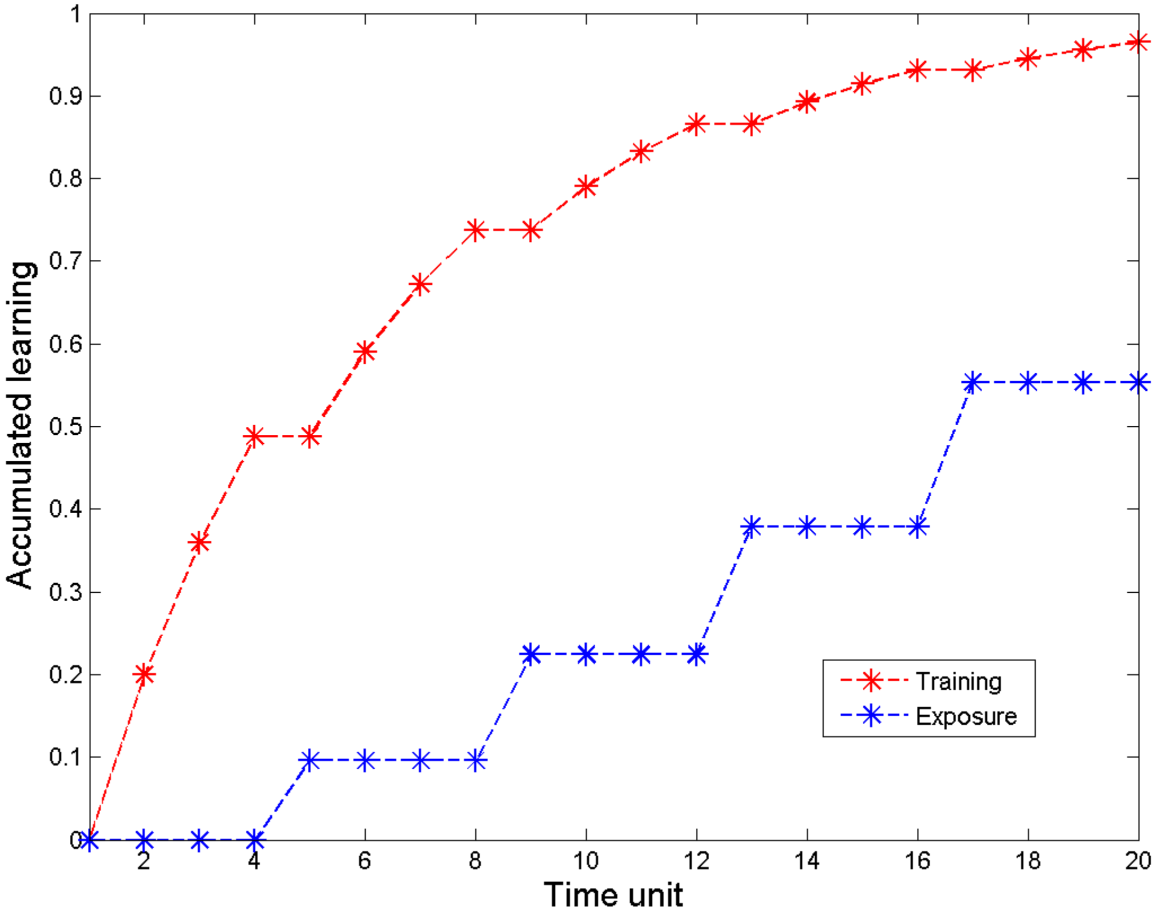
Simulations / results

- **Case 4:** $u = \{3 * Training, 2 * Exposure, 3 * Training, 2 * Exposure, \ldots, 3 * Training, 2 * Exposure\}$



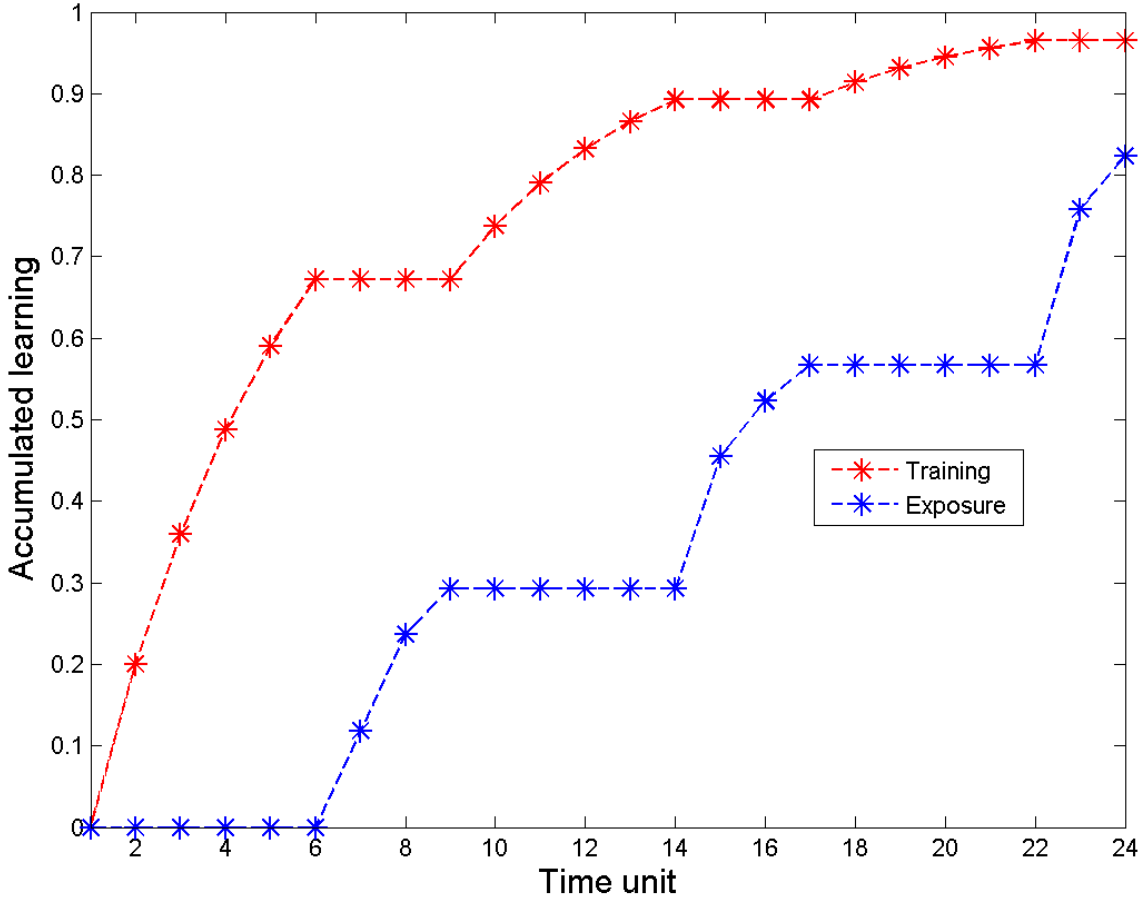
Simulations / results

- **New case:** $u = \{3 * Training, 1 * Exposure, 3 * Training, 1 * Exposure, \dots, 3 * Training, 1 * Exposure\}$



Simulations / results

- **Case 4:** $u = \{5 * Training, 3 * Exposure, 5 * Training, 3 * Exposure, \ldots, 5 * Training, 3 * Exposure\}$



Prevalent conjectures

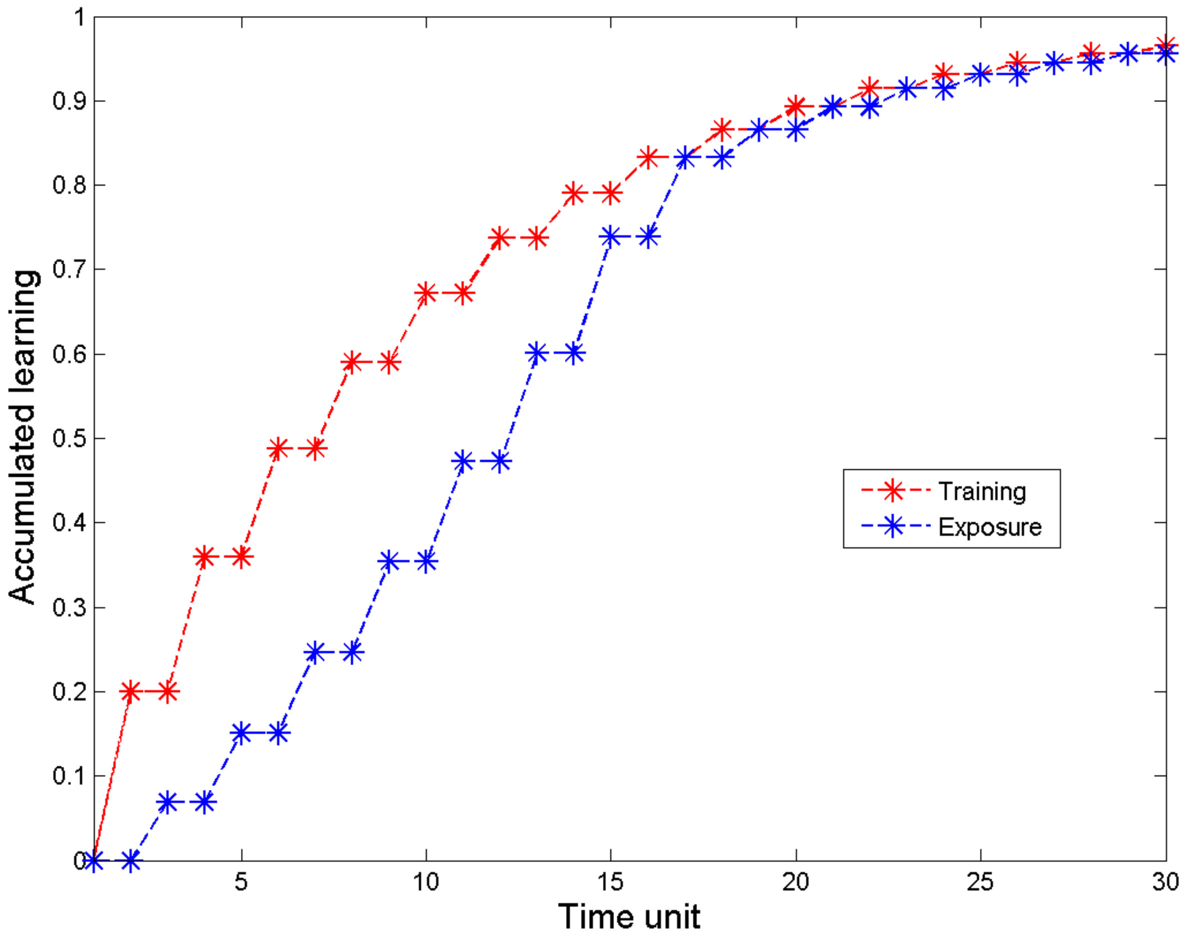
Rule based learning (Zhang et al, 2010)

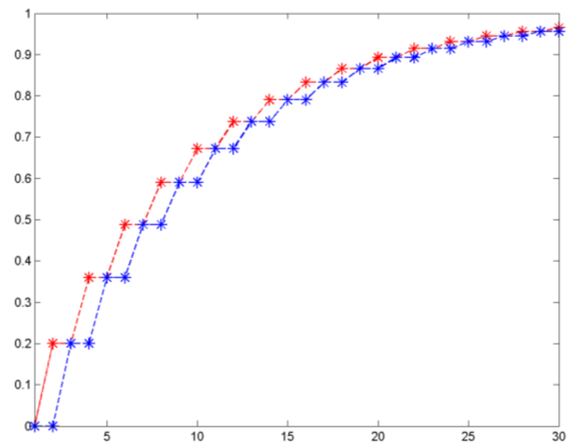
- V1 inputs representing untrained orientations or retinal locations are unattended and likely suppressed.
- Orientation exposure or location training establishes connections that enable learning transfer by reactivating the unattended or suppressed V1 inputs.

No computational model that explains how double-learning evokes learning transfer

Simulations

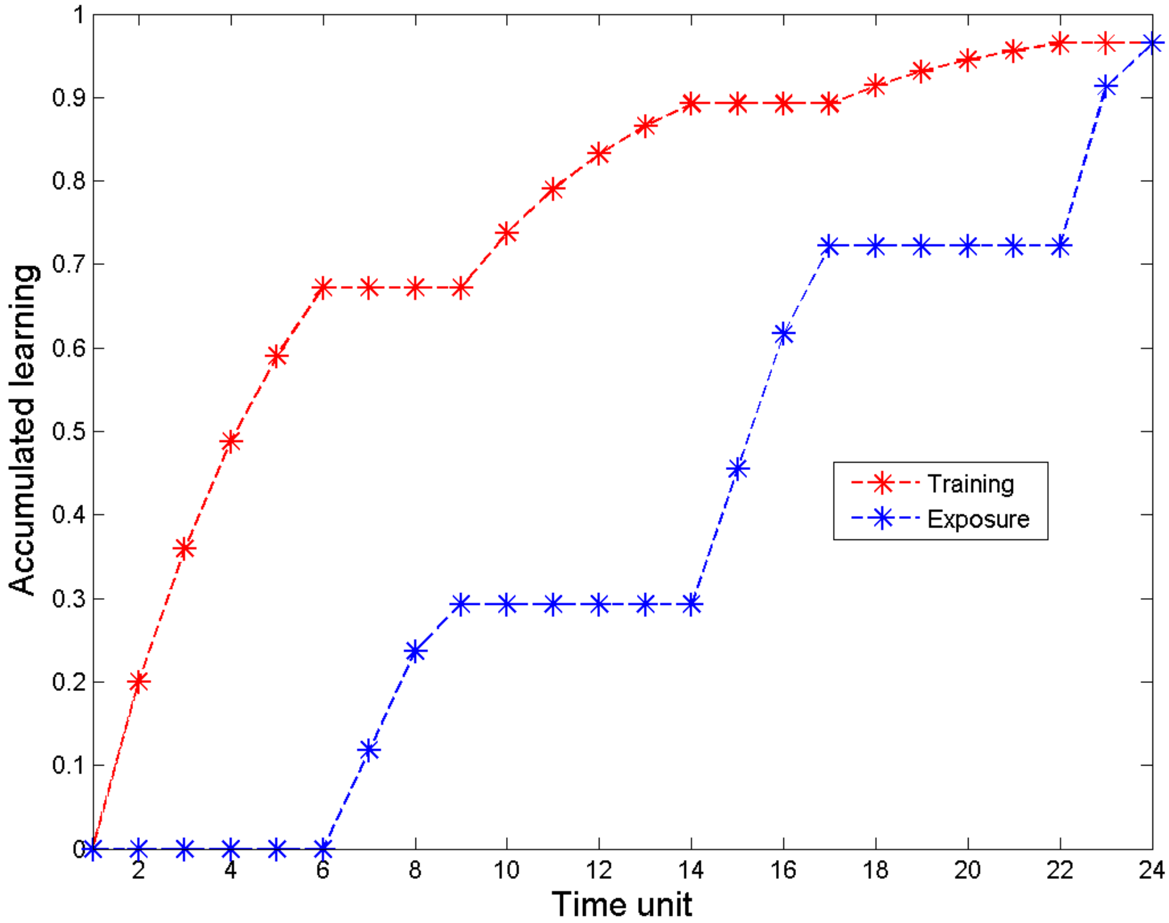
$u = 15 \times \{1\ Training, 1\ Exposure\}$ **(Case 2)**

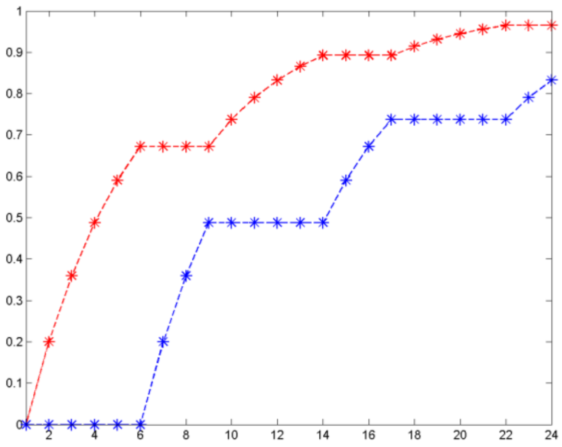




Simulations

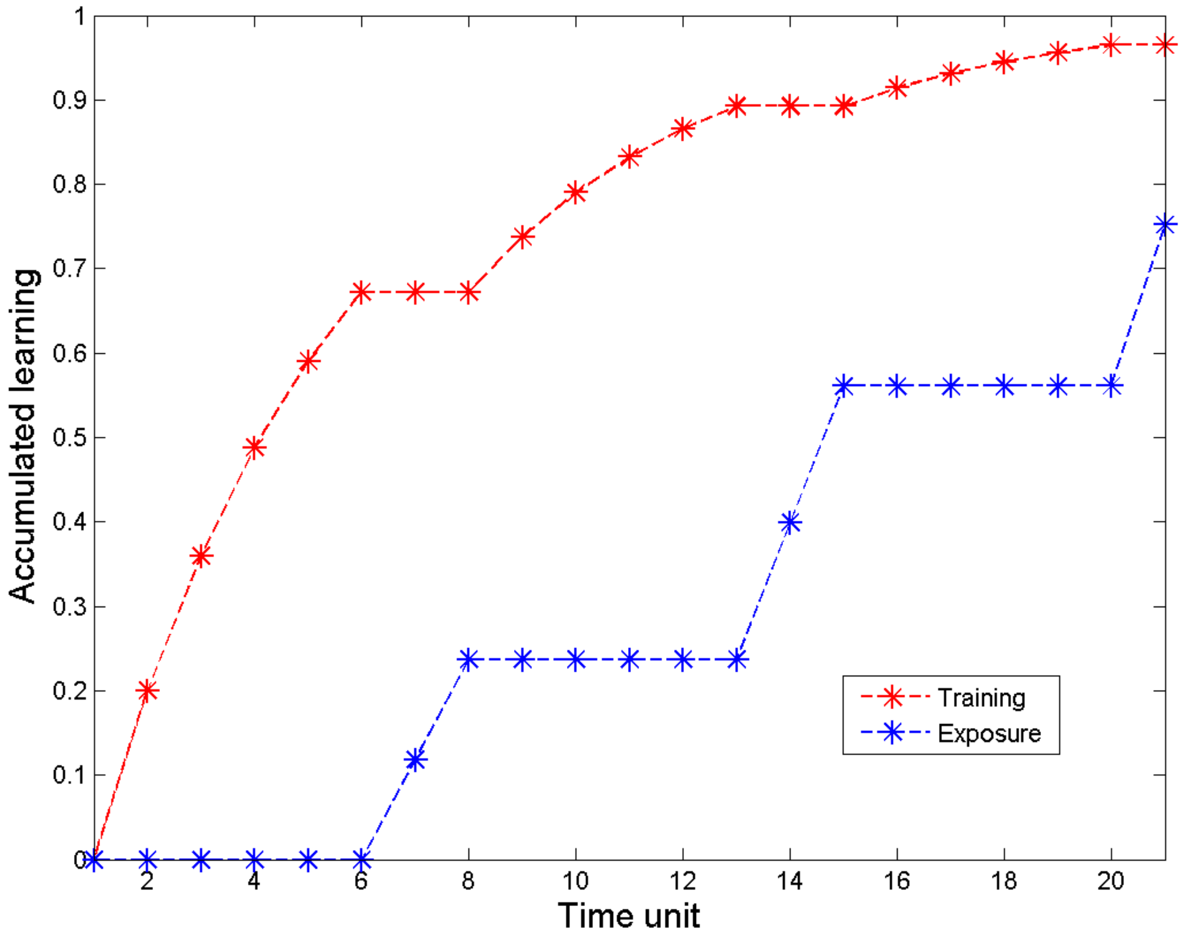
$u = 3 \times \{5\ Training, 3\ Exposure\}$ **(Case 4)**

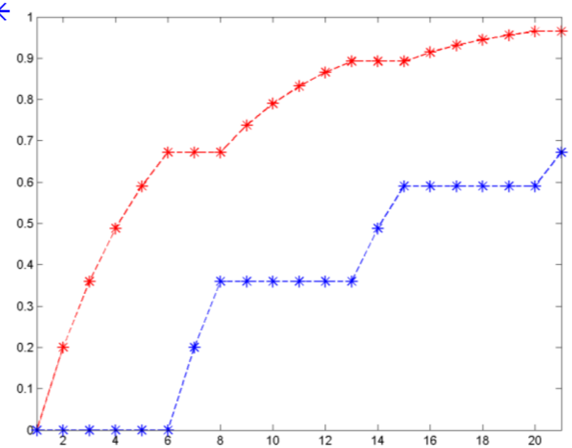




Simulations

$$u = 3 \times \{5\ Training, 2\ Exposure\}$$





Simulations

$$u = \{0,1,1,0,0,0,1,1,1,1,0,1,0,1,0,0,0,0,1,1,0,0,1,1,1,0,1,0,0,1\}$$