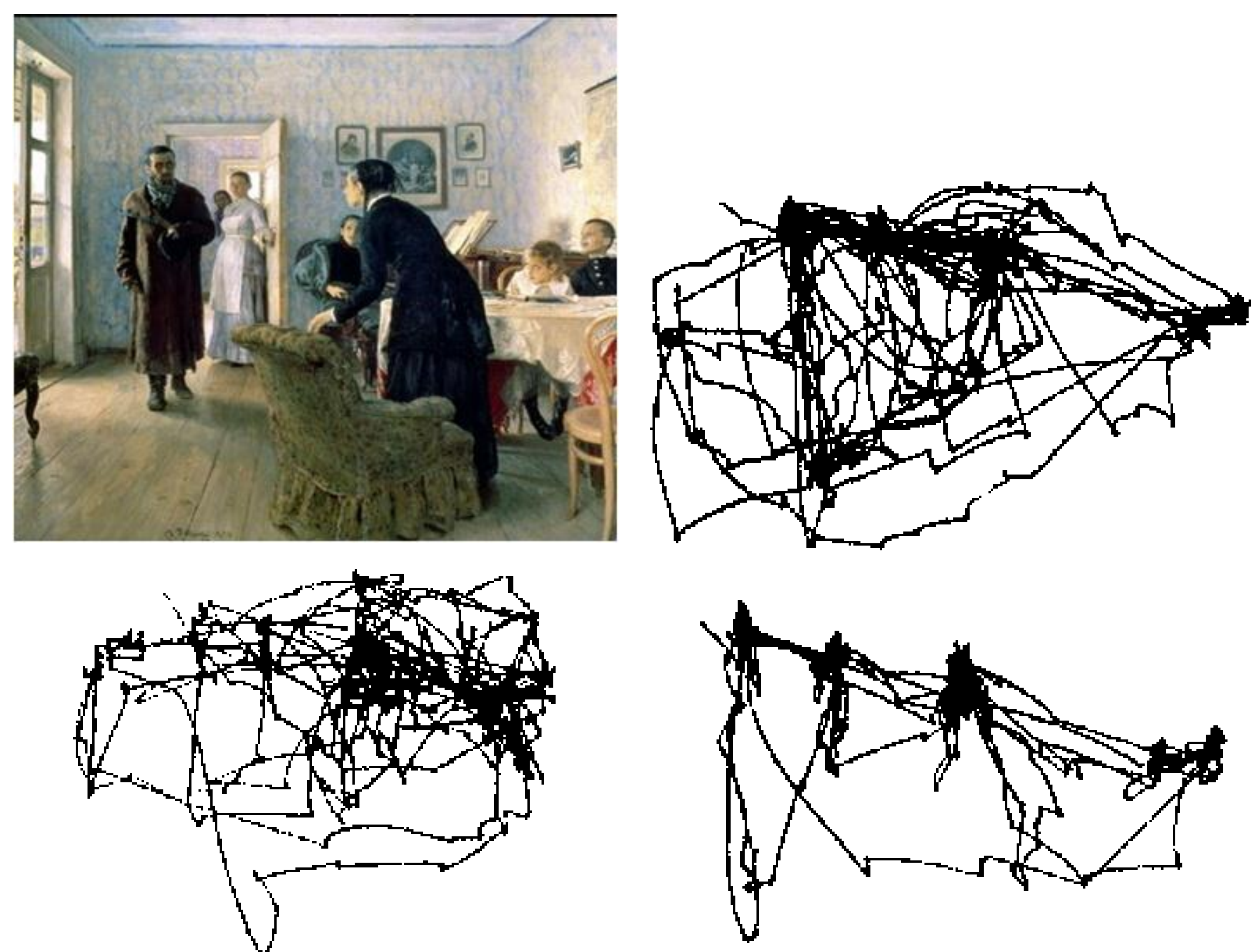


# Visual Task Inference in Conjunction Search Using Hidden Markov Models and Token Passing

Amin Haji-Abolhassani, James J. Clark  
Centre for Intelligent Machines, McGill University

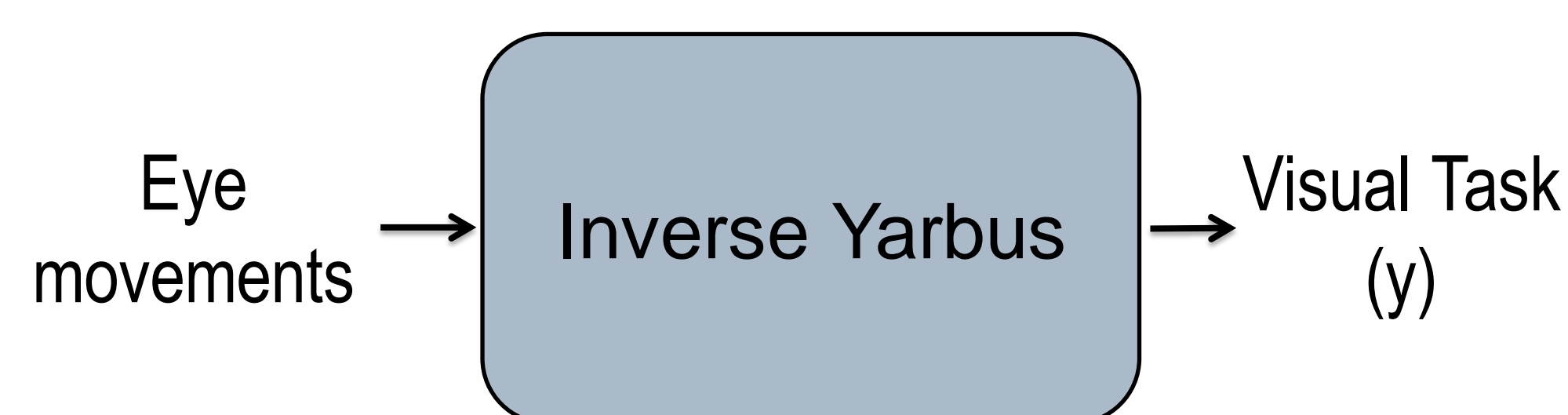
## PROBLEM DEFINITION

What we have: Yarbus process

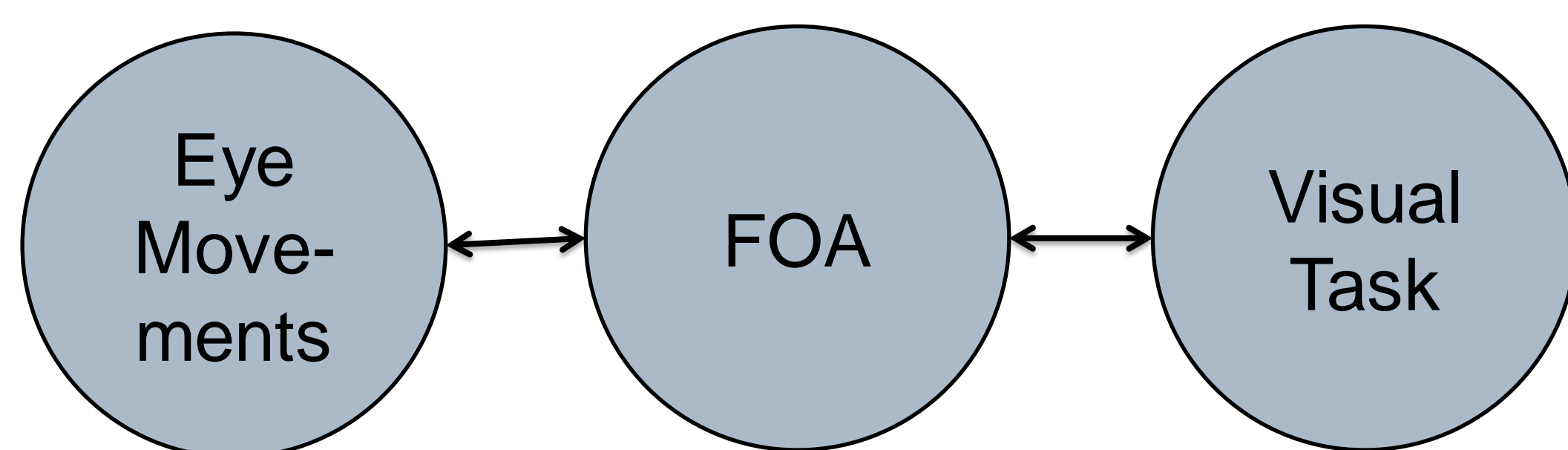


Eye movements measured by Yarbus [1] by viewers carrying out different tasks. Upper right - no specific task, lower left - estimate the wealth of the family, lower right - give the ages of the people in the painting.

What we need: Inverse Yarbus process



How we do that: Via tracking/predicting the Focus Of Attention (FOA)



## GENERATIVE LEARNING

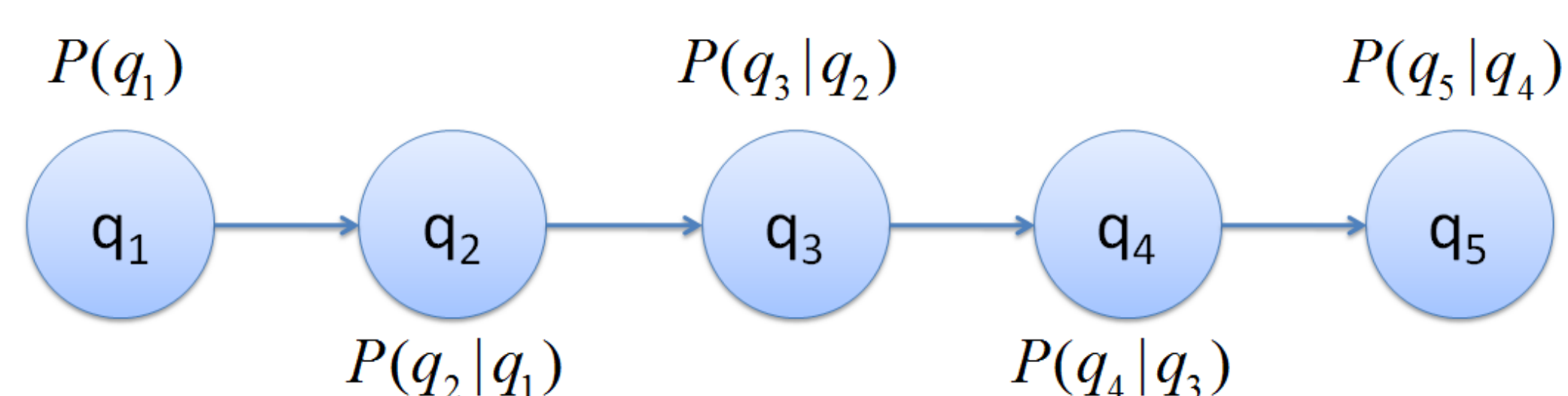
Bayesian Inference

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

$Q$  : Nearest states to fixation locations  $(q_1, q_2, \dots, q_T)$

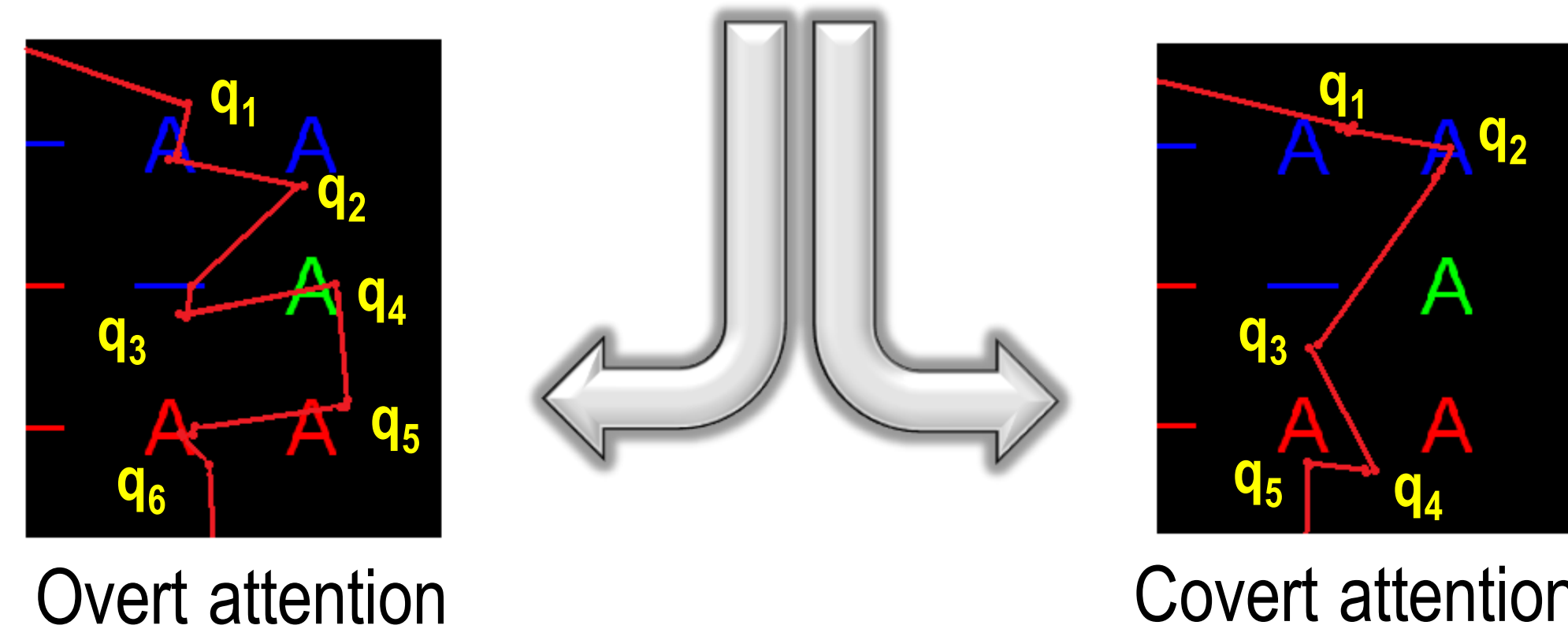
$y$  : Visual task ( $y \in Y$ )

Sequential Models: Markov Process



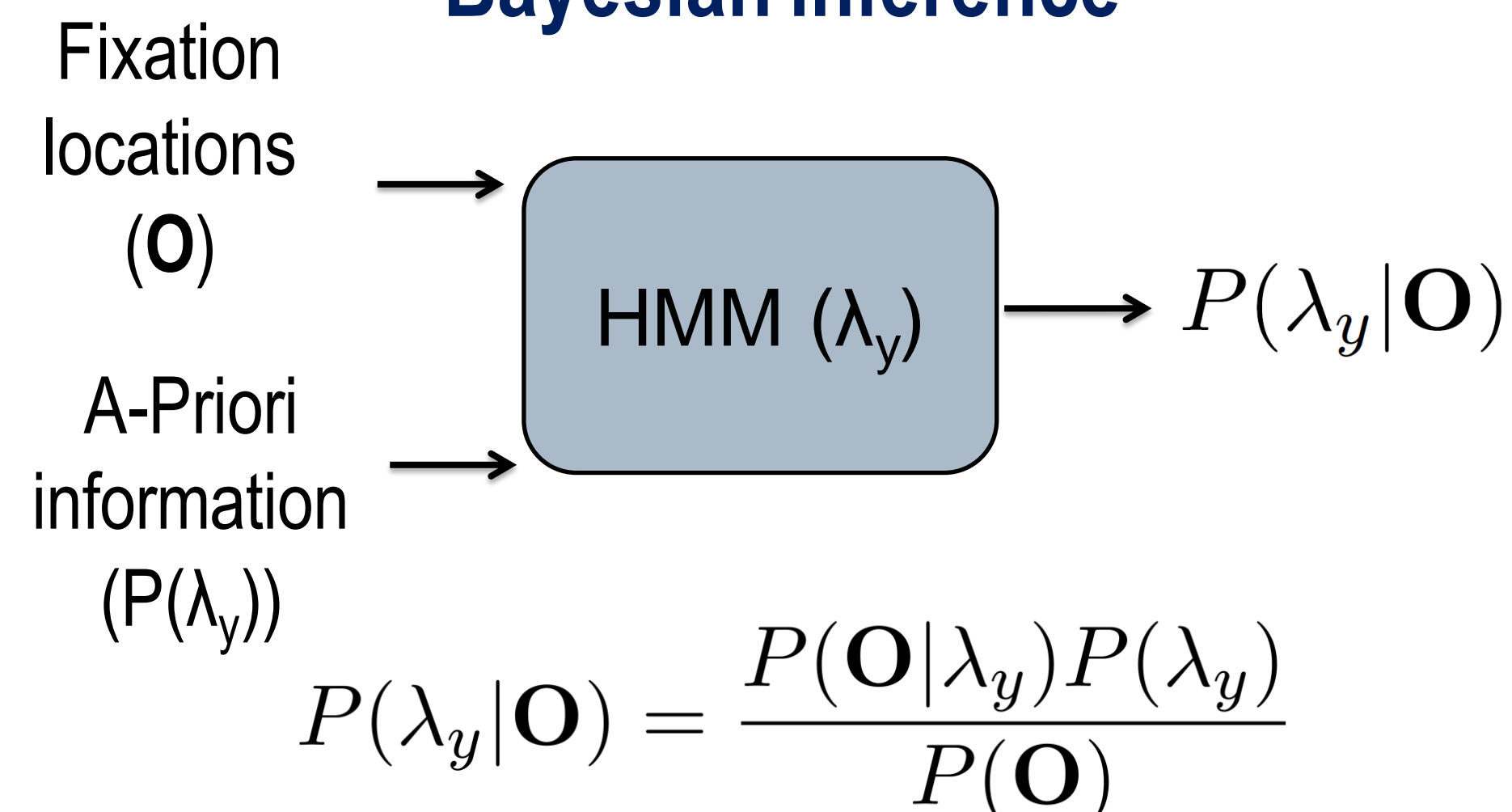
Problem with sequential models:

□ Attention is always assumed to be overt. However, it is sometimes covert, and not directly indicated by the eye position.



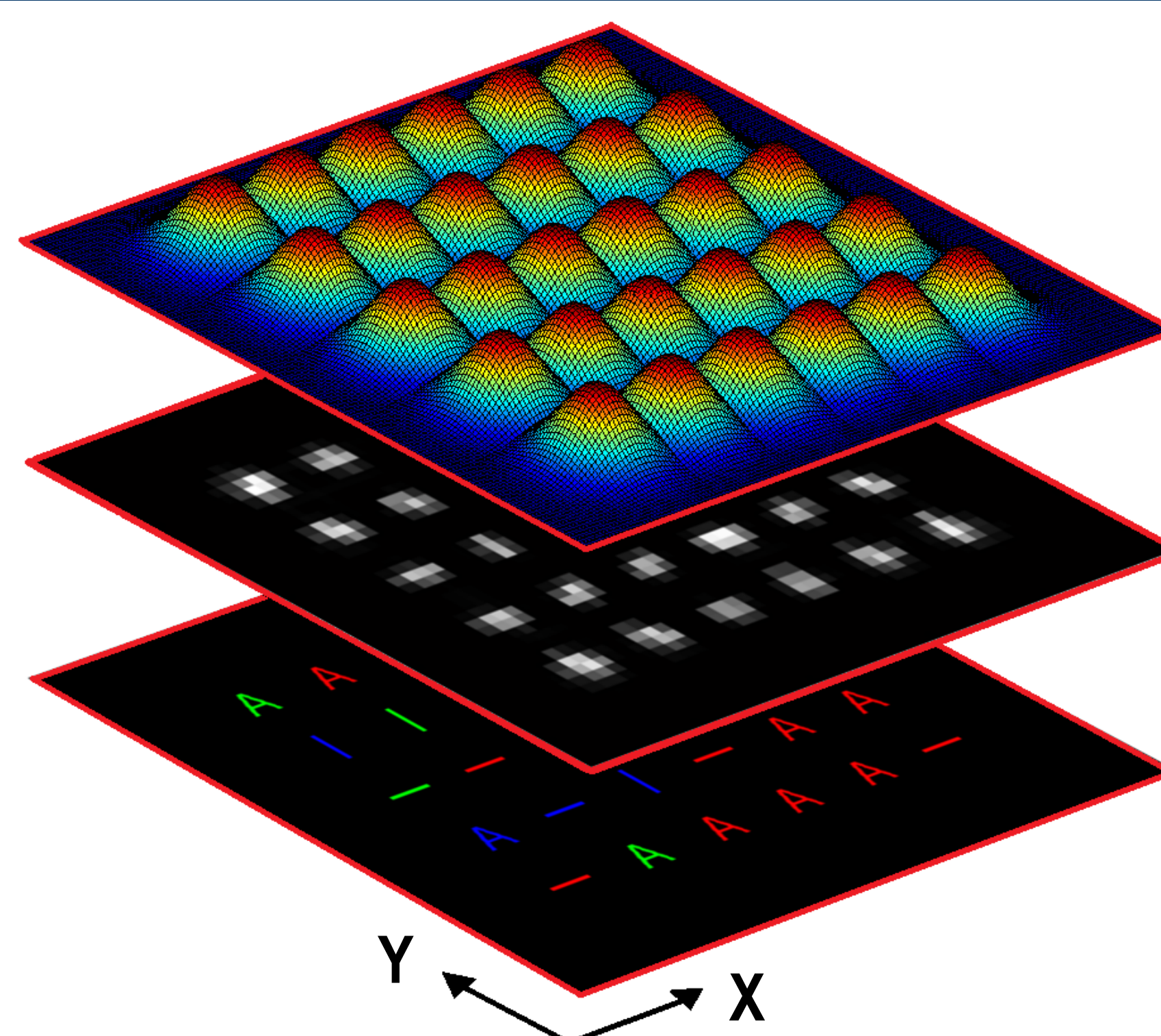
## HIDDEN MARKOV MODELS (HMM)

Bayesian inference



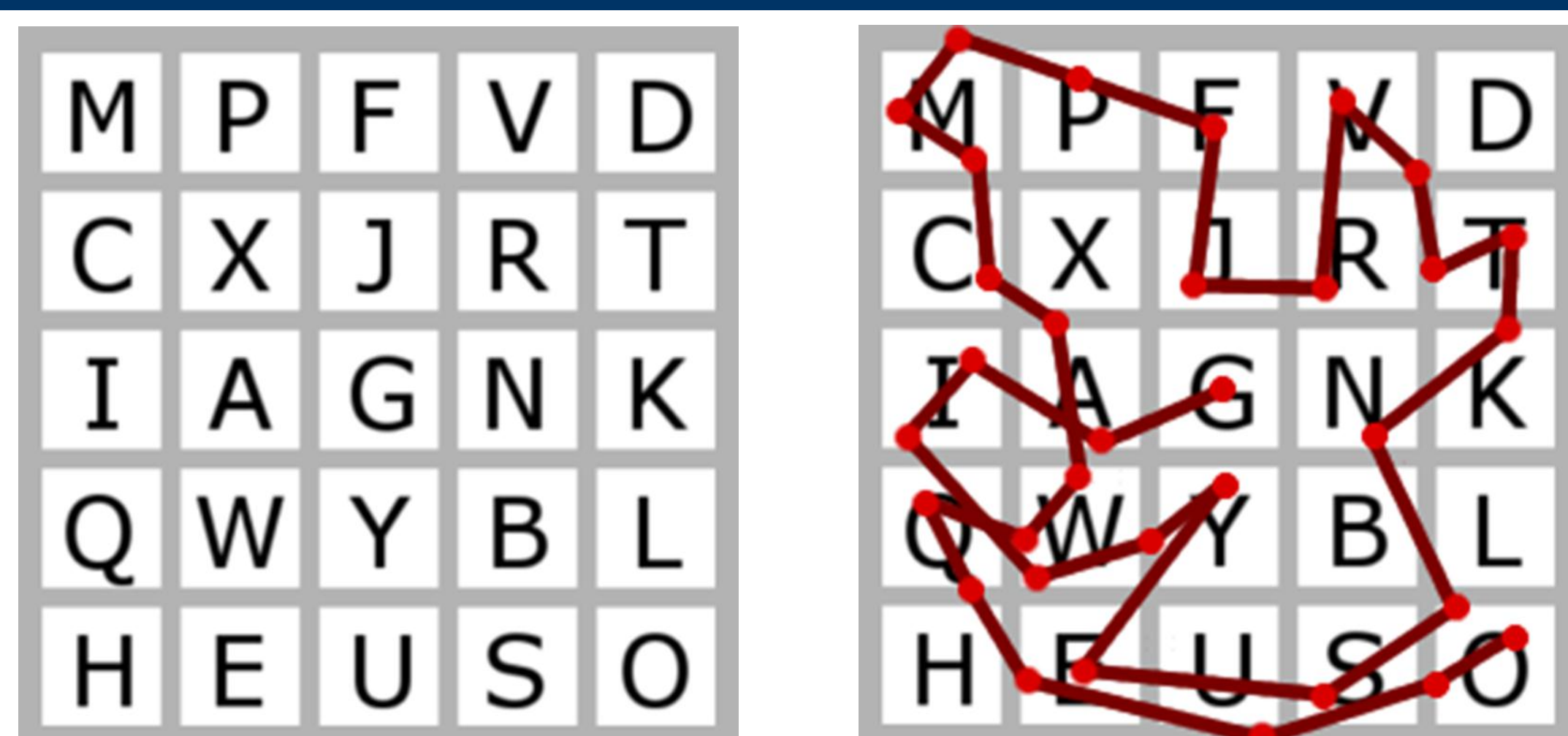
$\lambda_y$ : Task-specific HMM trained on eye movements of subjects performing task  $y$ . The hidden states represent the covert attention and the observations represent fixation location  $(O)$ .

## POP-OUT SEARCH: ERGODIC HMM (VSS 2011)



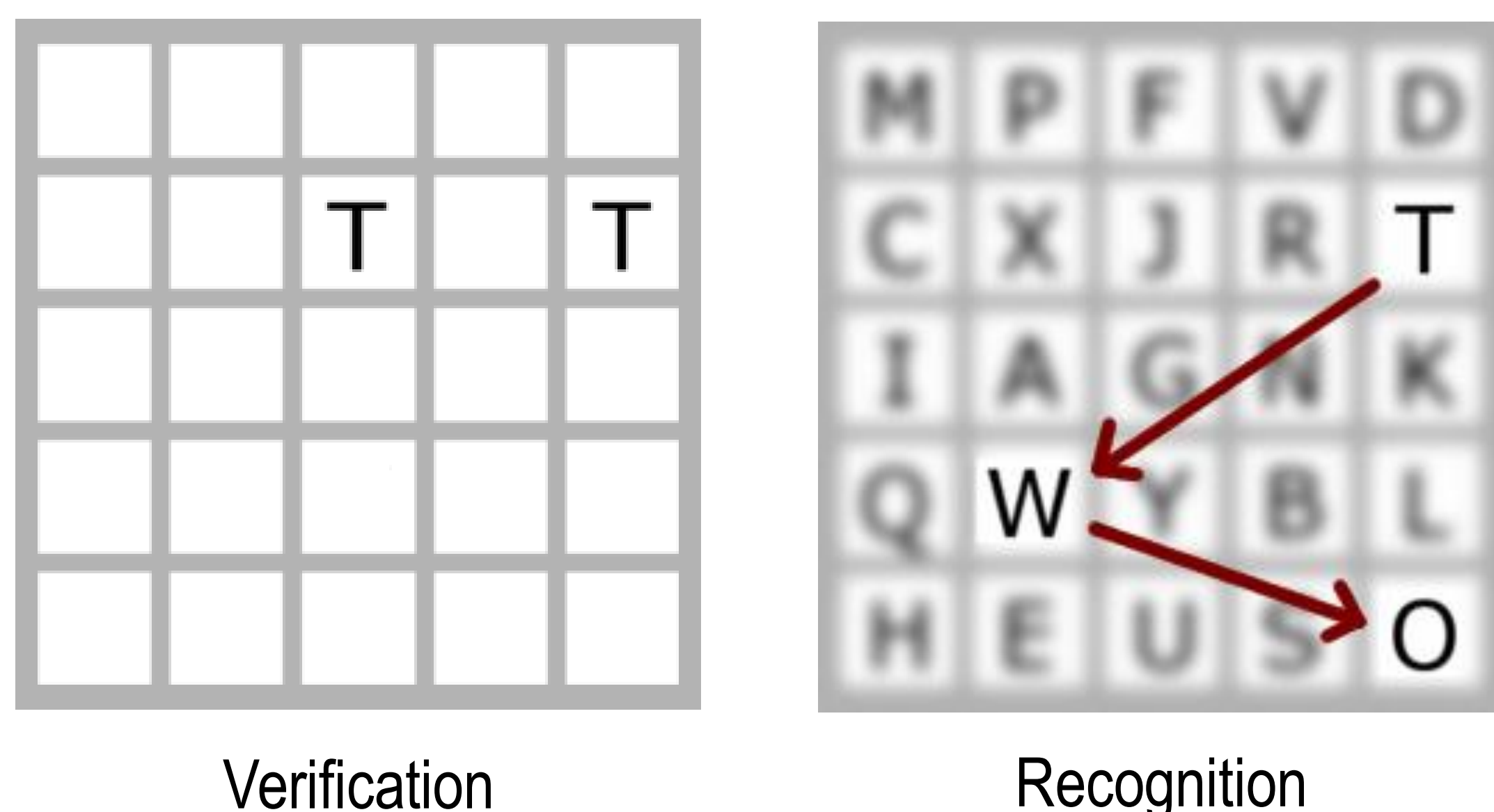
In [2] we proposed to use ergodic HMMs to infer the visual task in distractor-independent, pop-out visual search.

## CONJUNCTIVE SEARCH: TRI-STATE HMM



Keyboard Layout

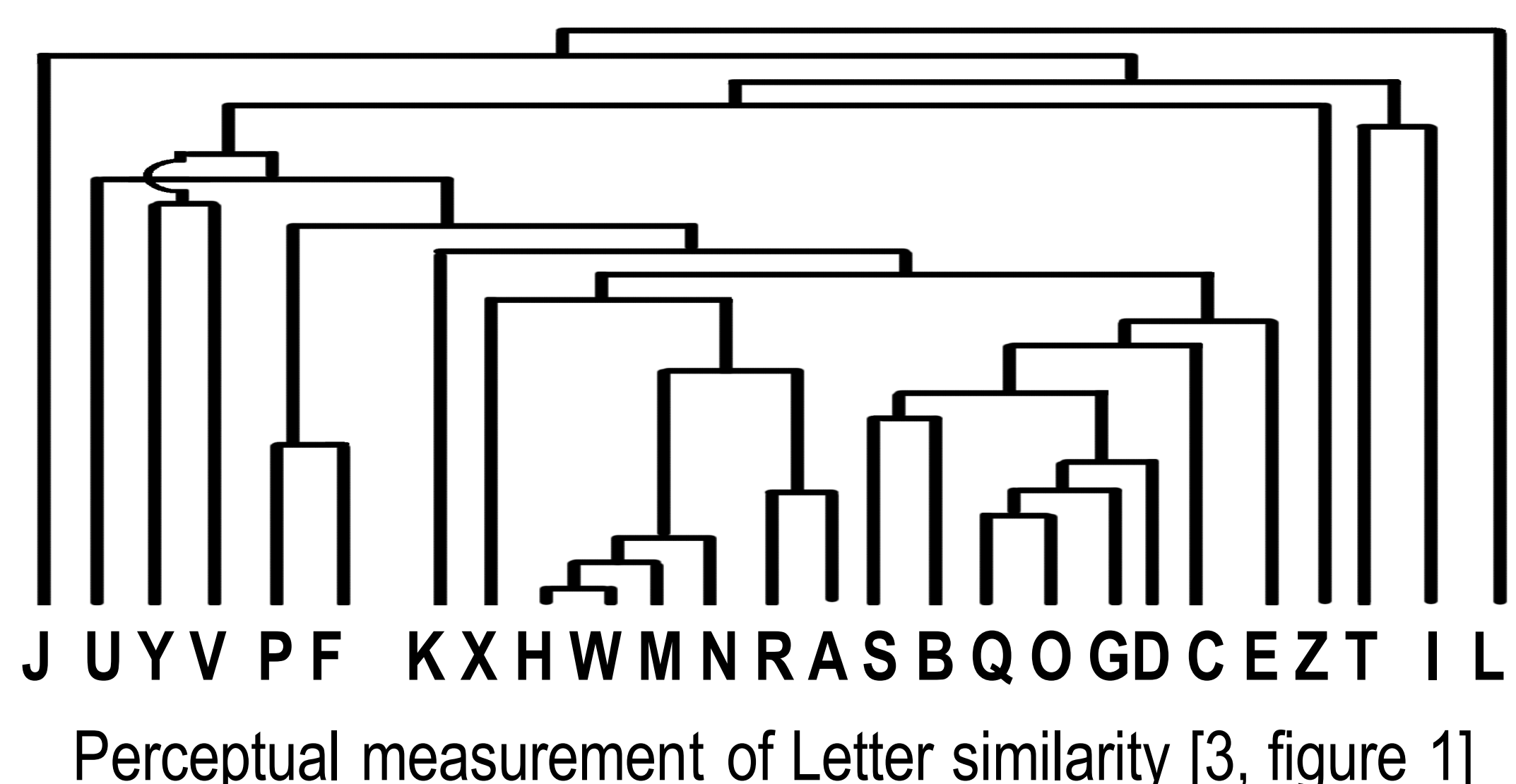
Eye-Movements while eye-typing the word "TWO"



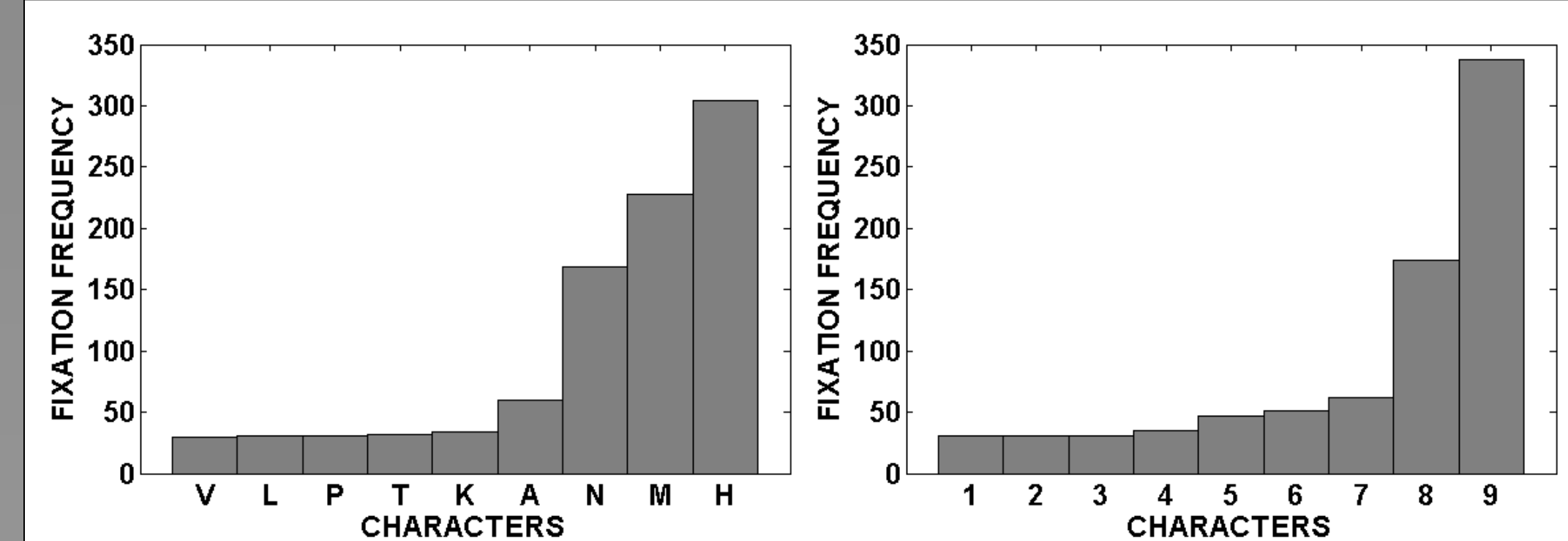
Verification

Recognition

## Analysis of Similarity

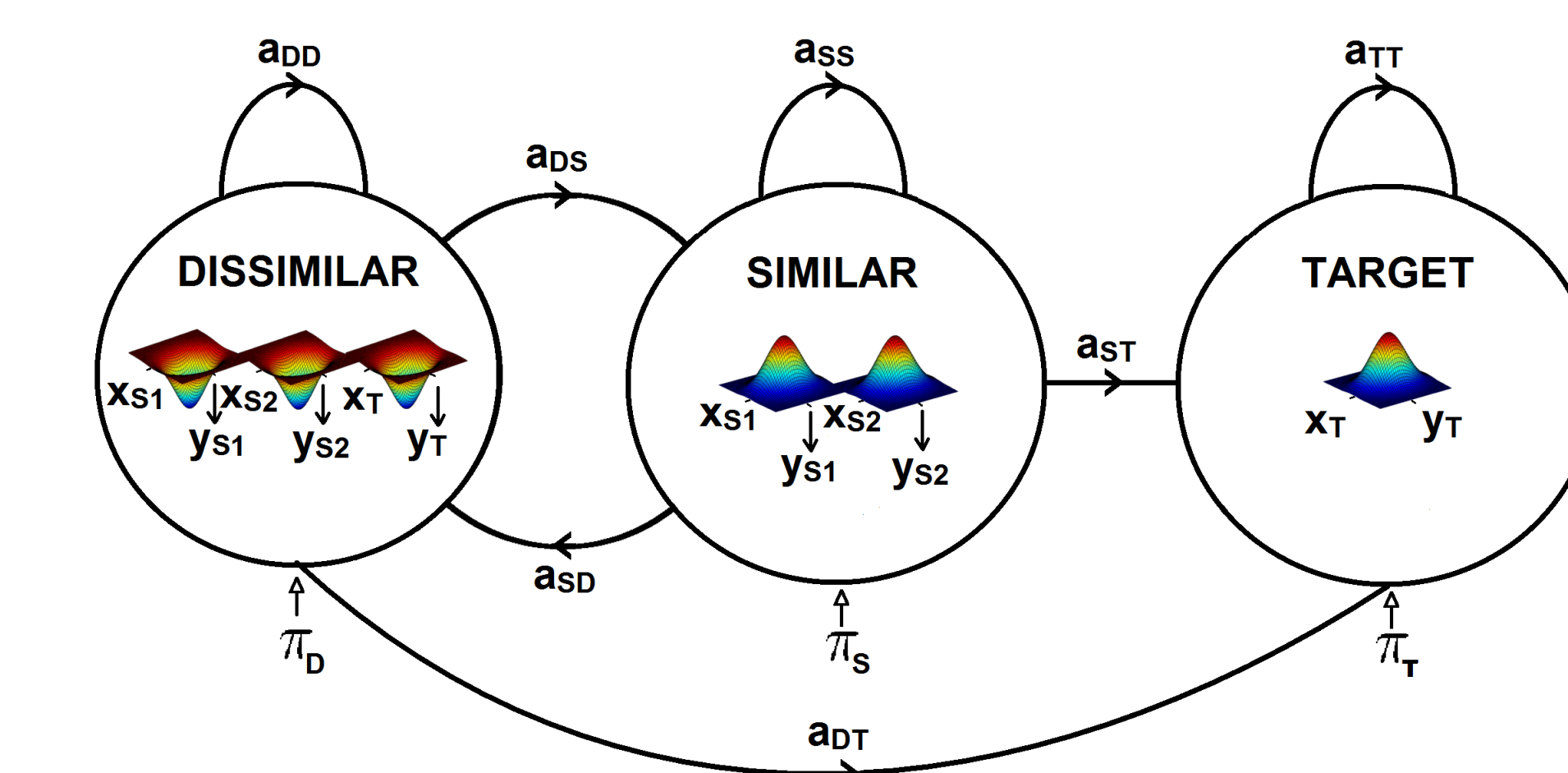


Perceptual measurement of Letter similarity [3, figure 1]

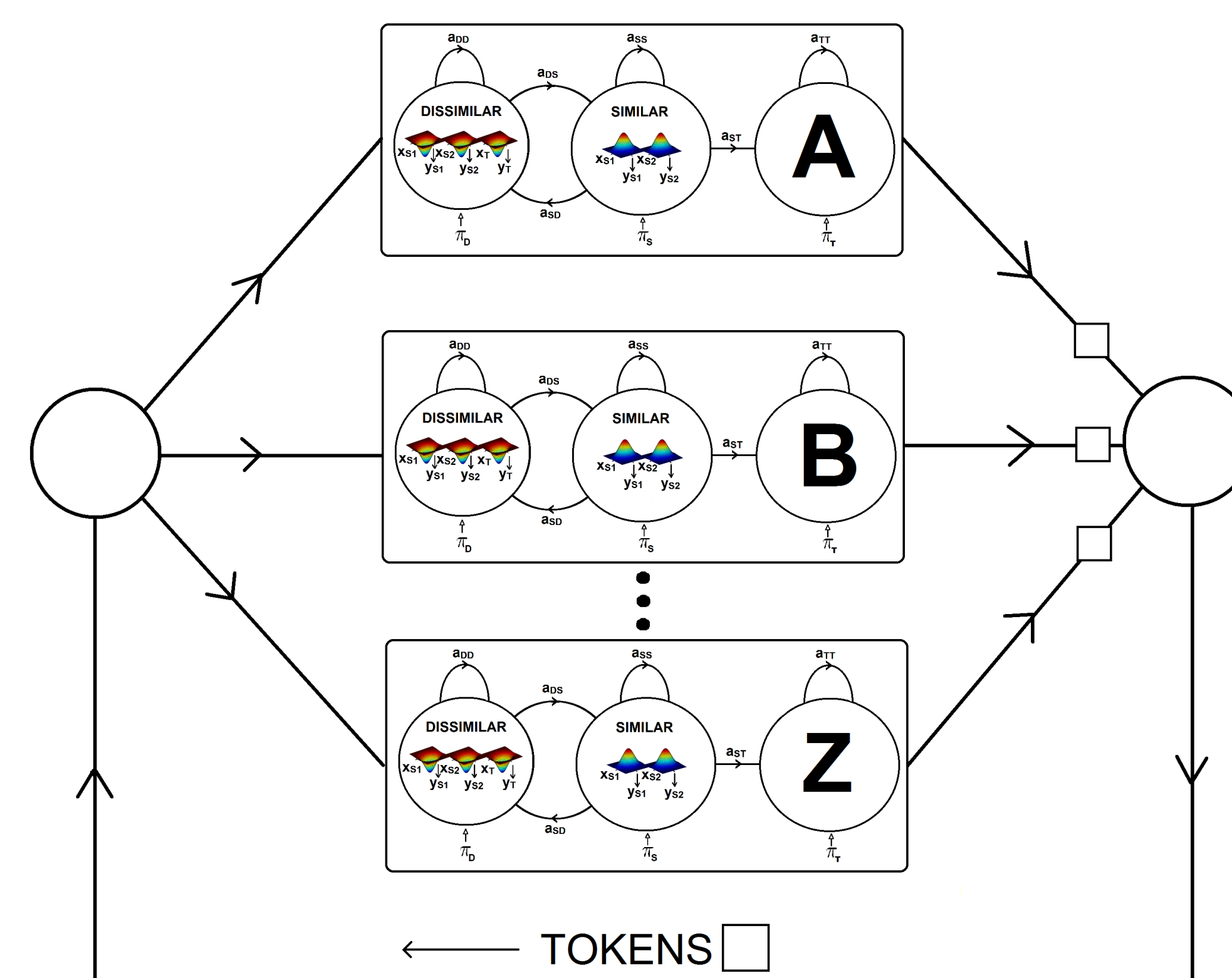


Looking for the letter "W"    Looking for different letters

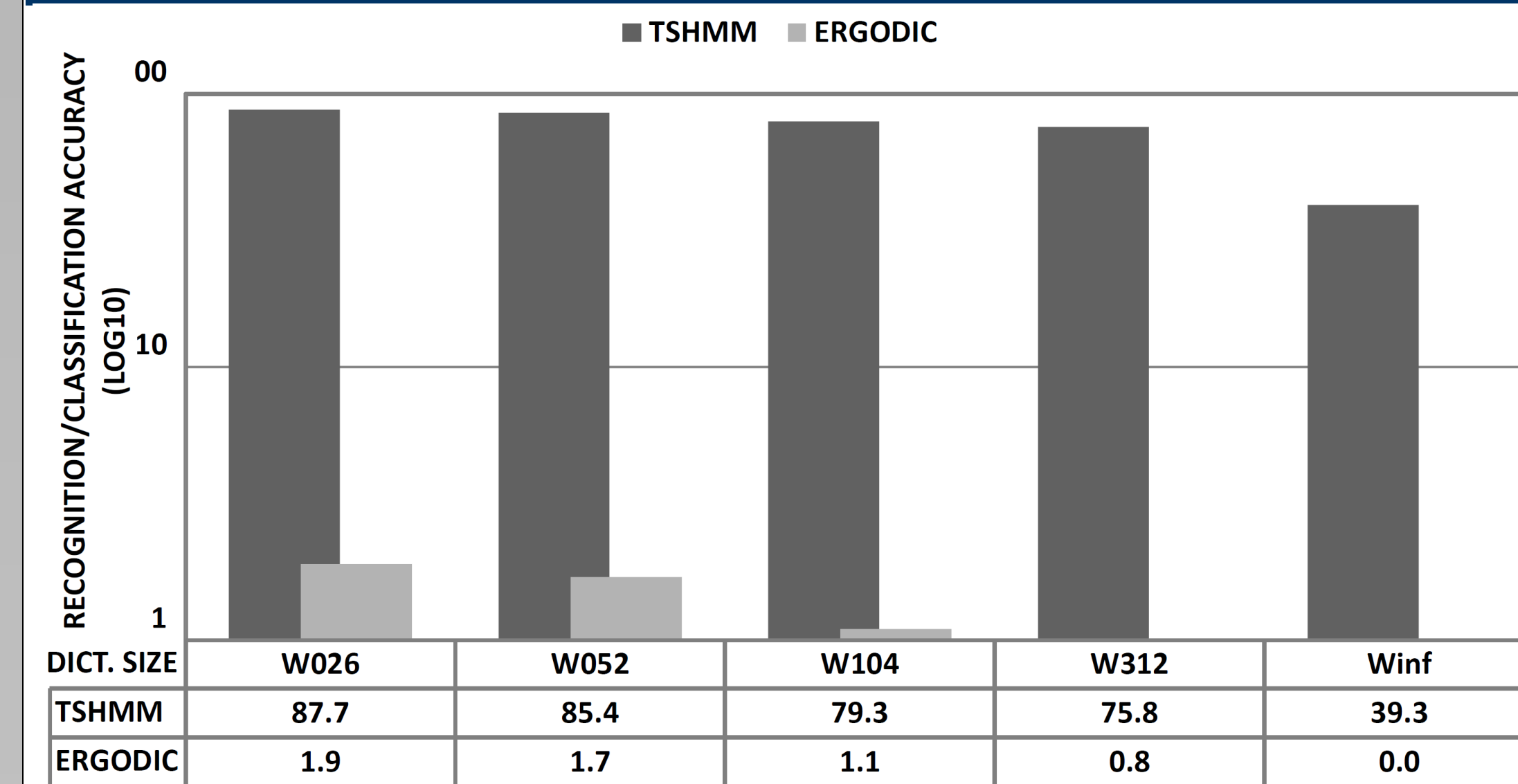
## Tri-State HMM (TSHMM)



## Decoding the Sequence Using Token Passing



## EXPERIMENTAL RESULTS



## CONCLUSIONS

- We had previously implemented a method in [2] to infer the ongoing visual task in pop-out search using ergodic HMMs.
- In order to infer the visual task in conjunction search we need to deal with the off-target fixations made on distractors.
- Off-target fixations give us information about the target due to the fact that similar-to-target objects are more likely to get fixated.
- We implemented a Tri-state HMM (TSHMM) for each letter that incorporates the perceptual similarity of the letters.
- We used a technique called *Token Passing* to decode the letter sequence that is eye-typed.

## REFERENCES

- [1] Yarbus, A., 1967. Eye movements during perception of complex objects. *Eye movements and vision*, 7: 171-196. Translated from the Russian edition by Haigh, B.
- [2] Haji-Abolhassani, A., Clark, J.J., 2011. Realization of an Inverse Yarbus Process via Hidden Markov Models for Visual-Task Inference. *Journal of Vision*, 11: 218-218.
- [3] Gilmore, G., Hersh, H., Caramazza, A., Grin, J., 1979. Multidimensional letter similarity derived from recognition errors. *Attention, Perception, & Psychophysics*, 25: 425-431.