Eye Tracking Experiment Design

Lecture 2

Antoni B. Chan
Dept. of Computer Science,
City University of Hong Kong

Janet H. Hsiao
Dept. of Psychology,
The University of Hong Kong

Outline

- Experiment Design
  - Face recognition
  - Web page usage
- Simulation Study
  - *How many trials are needed?*

- Develop Your Mini-Experiment

Aim of the study

Aging is associated with decline in..

Cognitive ability

- Impairment in various aspects such as memory, processing speed, attention, executive functions
  (Kail & Salthouse, 1994; Grady, 2002)

Face recognition performance

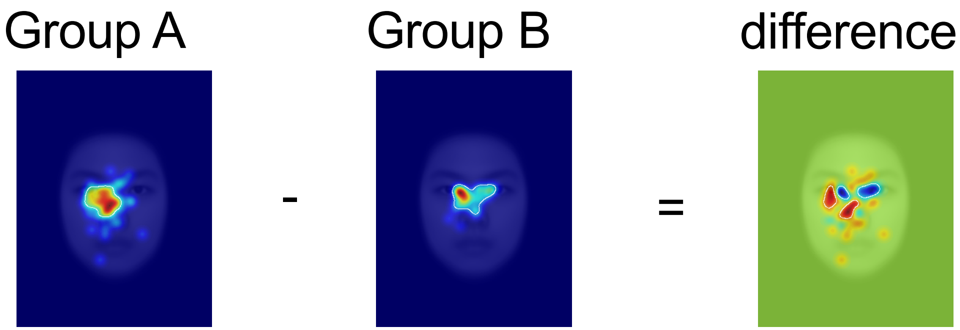
- Older adults recognize faces worse than young adults
  (Fulton & Bartlett, 1991; Grady, McIntosh, Horwitz, & Rapoport, 2000; Lamont, Stewart-Williams, & Podd, 2005)
- The declined performance is mainly due to high false alarm rates, instead of low hit rates, produced by older adults
  (Crook & Larrabee, 1992; Lamont et al., 2005).

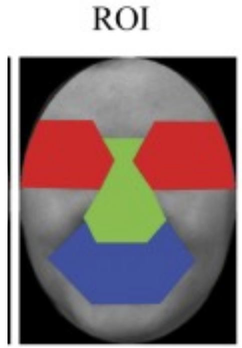
Aim of study: Could they be reflected in eye movement pattern?

Background

Can we use eye movements to understand cognitive processes related to aging?

- However, existing methods focused on group-level comparison (e.g. Heat map and ROI approaches)





- It would largely exclude individual differences:
  - cognitive or face recognition performance can be different among older people
  - Eye movement can also be different among older people
- Use HMM-based approach, which consider individual differences, to examine the link between cognitive ability and eye movement pattern

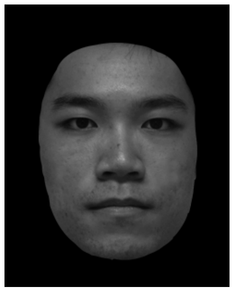
Research Questions

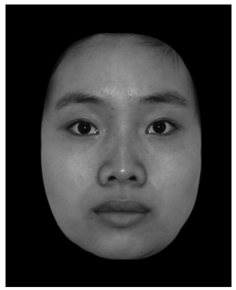
Investigate eye movement patterns of older adults in face recognition memory as compared with young adults via the HMM-based approach:

1. Do young and older adults have different eye movement patterns?
2. Is false alarm rate related to the eye movement pattern used?
3. Is cognitive ability of older adults associated with the eye movement pattern used?

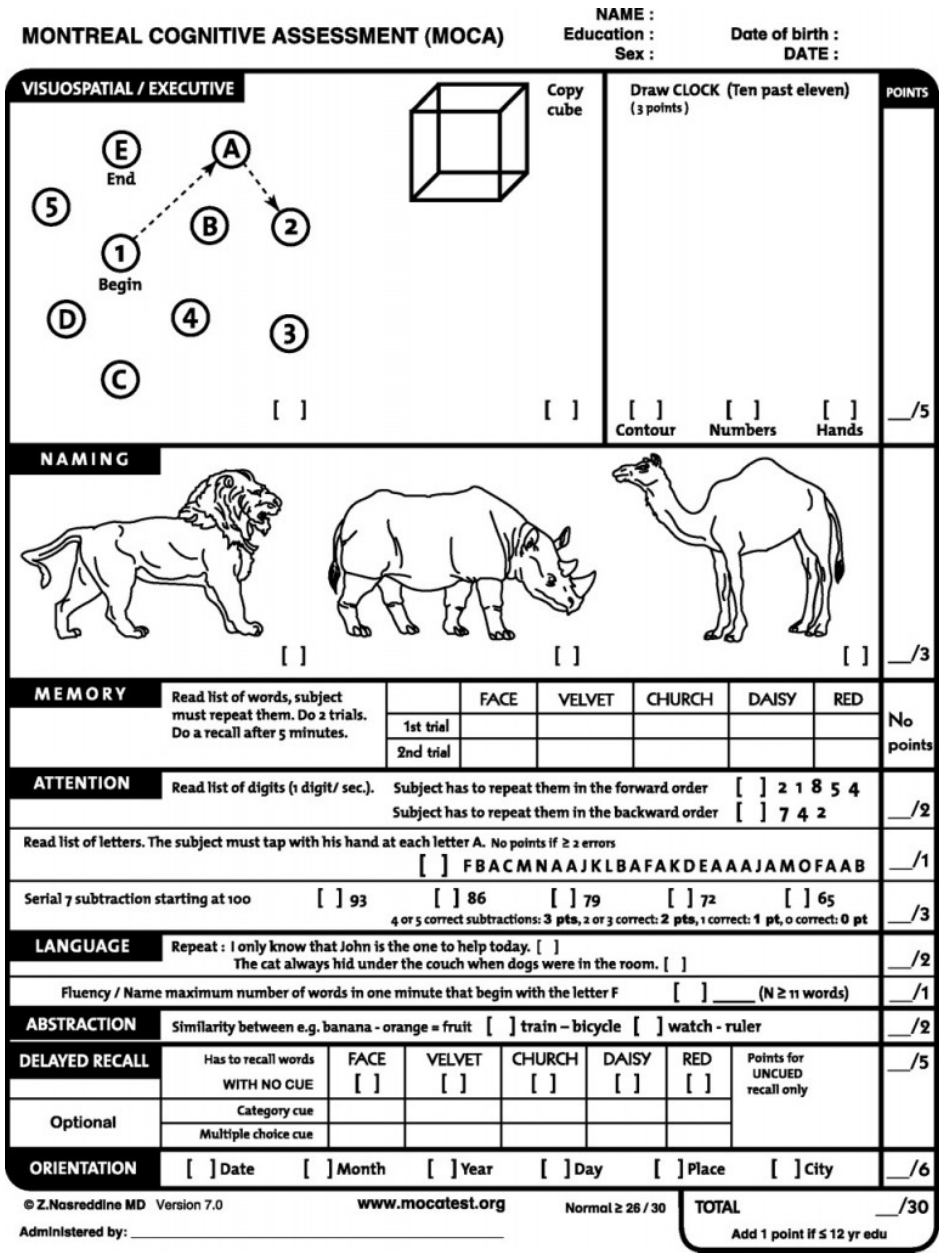
Methods

- **2 participant groups**
  - 34 Young adults (18-24, M = 20, SD = 2)
  - 34 Older adults (65-81, M = 69, SD = 8)
  - All are Chinese, right-handedness, with normal or corrected vision
  - All older adults are within normal range of cognitive ability
    - as assessed by the Montreal Cognitive Assessment (MoCA) Hong Kong version (Wong et al., 2009)
- **Materials:**
  - Young Chinese frontal view face images, gray-scale
- **Apparatus:**
  - Eyelink 1000 eye-tracking system





MoCA – Cognitive Assessment



MONTREAL COGNITIVE ASSESSMENT (MOCA)

NAME :
Education : Date of birth :
Sex : DATE :

		Points
VISUOSPATIAL / EXECUTIVE	(E) End, (A), (5), (B), (2), (1) Begin, (D), (4), (3), (C) [ ] ; Copy cube [ ] ; Draw CLOCK (Ten past eleven) (3 points) [ ] Contour [ ] Numbers [ ] Hands	__/5
NAMING	[ ] [ ] [ ]	__/3
MEMORY	Read list of words, subject must repeat them. Do 2 trials. Do a recall after 5 minutes.	No points
ATTENTION	Read list of digits (1 digit/ sec.). Subject has to repeat them in the forward order [ ] 2 1 8 5 4 ; Subject has to repeat them in the backward order [ ] 7 4 2	__/2
	Read list of letters. The subject must tap with his hand at each letter A. No points if ≥ 2 errors [ ] FBACMNAAJKLBAFAKDEAAAJAMOFAAB	__/1
	Serial 7 subtraction starting at 100 [ ] 93 [ ] 86 [ ] 79 [ ] 72 [ ] 65 — 4 or 5 correct subtractions: **3 pts**, 2 or 3 correct: **2 pts**, 1 correct: **1 pt**, 0 correct: **0 pt**	__/3
LANGUAGE	Repeat : I only know that John is the one to help today. [ ] The cat always hid under the couch when dogs were in the room. [ ]	__/2
	Fluency / Name maximum number of words in one minute that begin with the letter F [ ] ____ (N ≥ 11 words)	__/1
ABSTRACTION	Similarity between e.g. banana - orange = fruit [ ] train – bicycle [ ] watch - ruler	__/2
DELAYED RECALL	Has to recall words WITH NO CUE: FACE [ ] VELVET [ ] CHURCH [ ] DAISY [ ] RED [ ] — Points for UNCUED recall only	__/5
Optional	Category cue ; Multiple choice cue	
ORIENTATION	[ ] Date [ ] Month [ ] Year [ ] Day [ ] Place [ ] City	__/6

	FACE	VELVET	CHURCH	DAISY	RED
1st trial					
2nd trial					

© Z.Nasreddine MD Version 7.0 www.mocatest.org Normal ≥ 26 / 30 TOTAL __/30

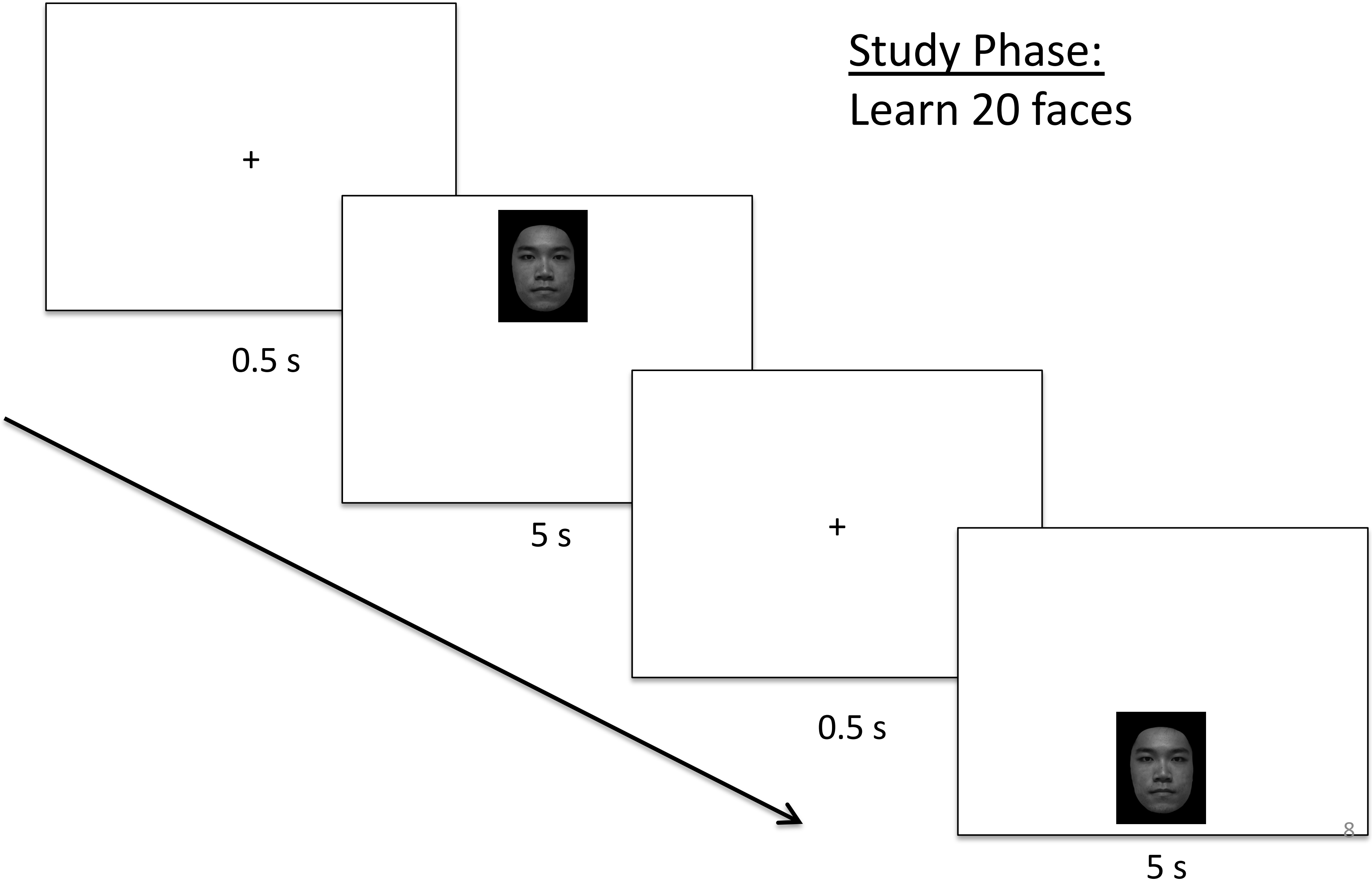
Add 1 point if ≤ 12 yr edu

Administered by: ____________________

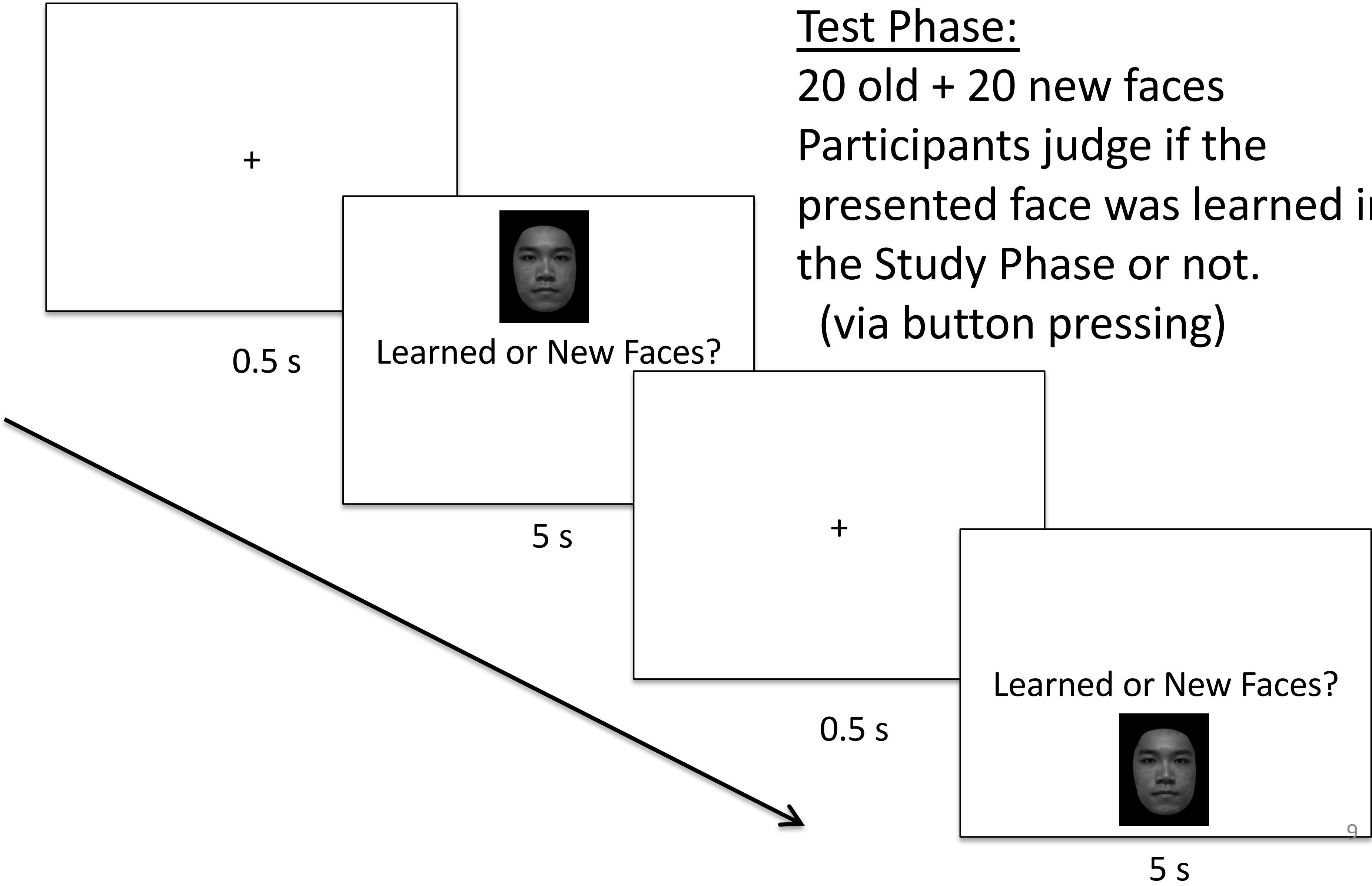
Methods: Face Recognition task

<u>Study Phase:</u>

Learn 20 faces



Methods: Face Recognition task



Test Phase:

20 old + 20 new faces

Participants judge if the presented face was learned in the Study Phase or not.

(via button pressing)

Notes on Experiment Design

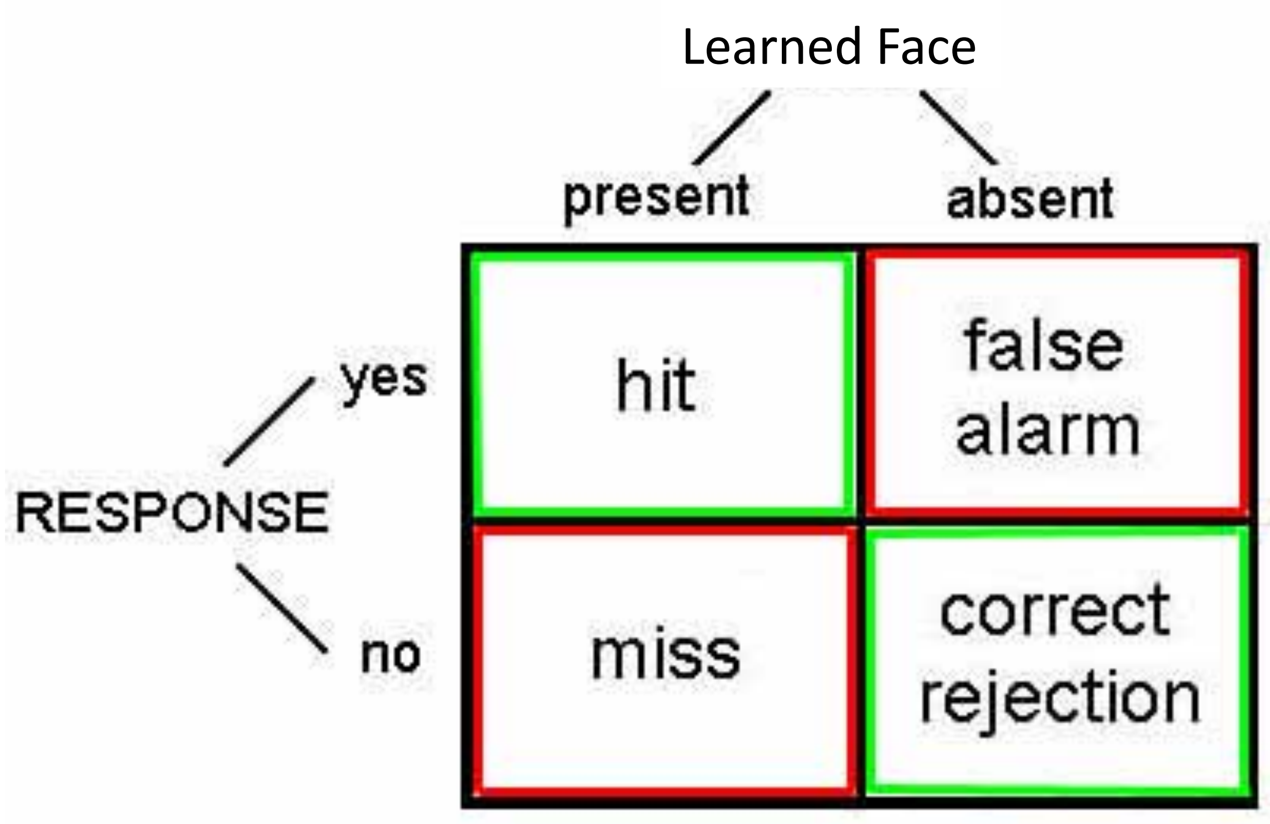
- Need to calibrate eye tracker for each subject
  - Fixation cross used for drift correction before each trial
- All stimuli are aligned by eyes-mouth triangle
  - Recorded eye gaze coordinates have the same semantic meaning over all stimuli.
- Counter-balancing experiment
  - Counter-balance or randomize the presentation of the faces on top and bottom halves. (or in the 4 quadrants)
  - Randomize order of faces.

Data Analysis

- Pre-processing
  - convert *raw* eye gaze to eye fixations
    - use eye tracker's software
  - convert *raw* fixation locations into locations on stimuli
    - subtract the coordinates of the top-left corner of the stimuli
- Analysis
  - Learn an HMM for each individual
  - Cluster individuals' HMMs into groups
  - Analyze differences in groups (pre-defined or data-driven)
    - Group membership (X^2 test)
    - Performance differences between groups (t-test)
    - Correlation between group similarity and behavioral data (linear regression)

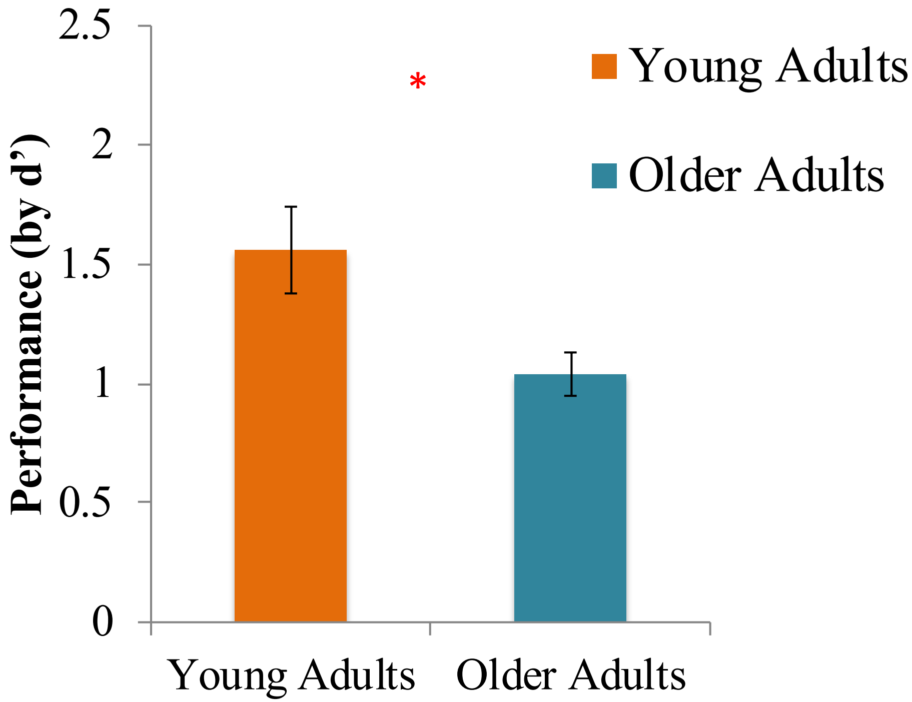
Face Recognition Performance

- Signal detection theory
- Measured by Discrimination sensitivity measure d’
  d’ = Z(hit) – Z(false alarm)

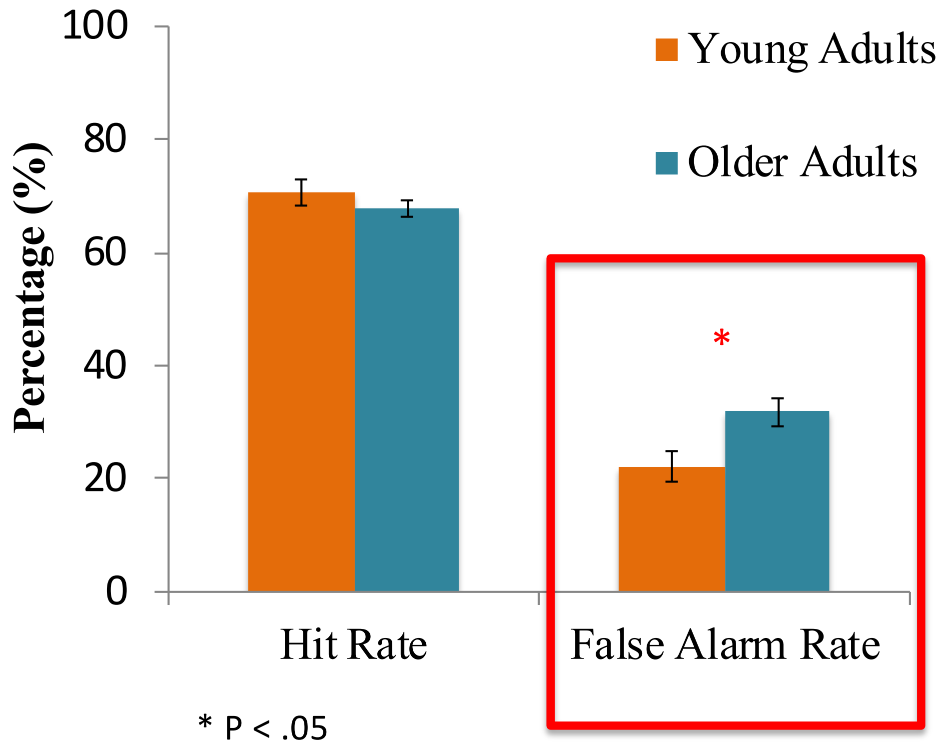


Face Recognition Performance

Group Analysis of Young/Older Adults



* P < .05

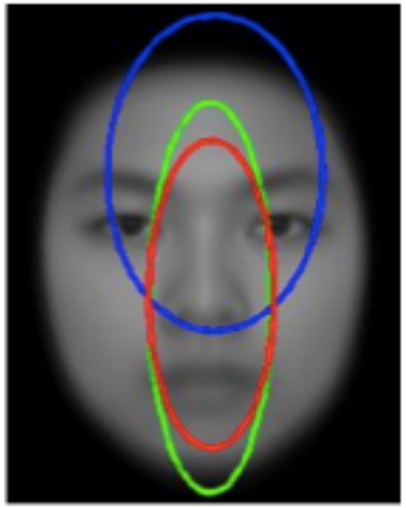


* P < .05

Young adults recognized faces better!

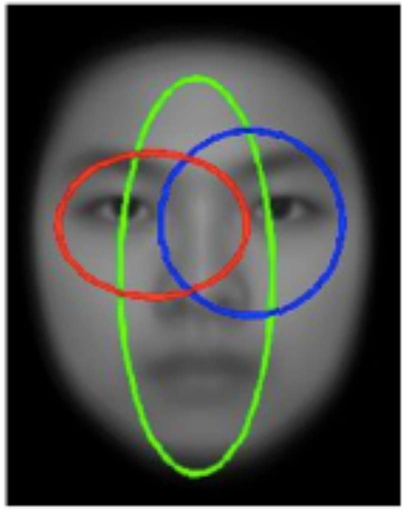
Older adults had significantly higher false alarm rates!

Clustered HMM – by clustering all participant HMMs into subgroups



Representative HMM of Holistic Pattern

Prior values	Red	Green	Blue
	.16	.80	.04
Transition probabilities	To Red	To Green	To Blue
From Red	.82	.12	.06
From Green	.93	.02	.05
From Blue	.07	.00	.93

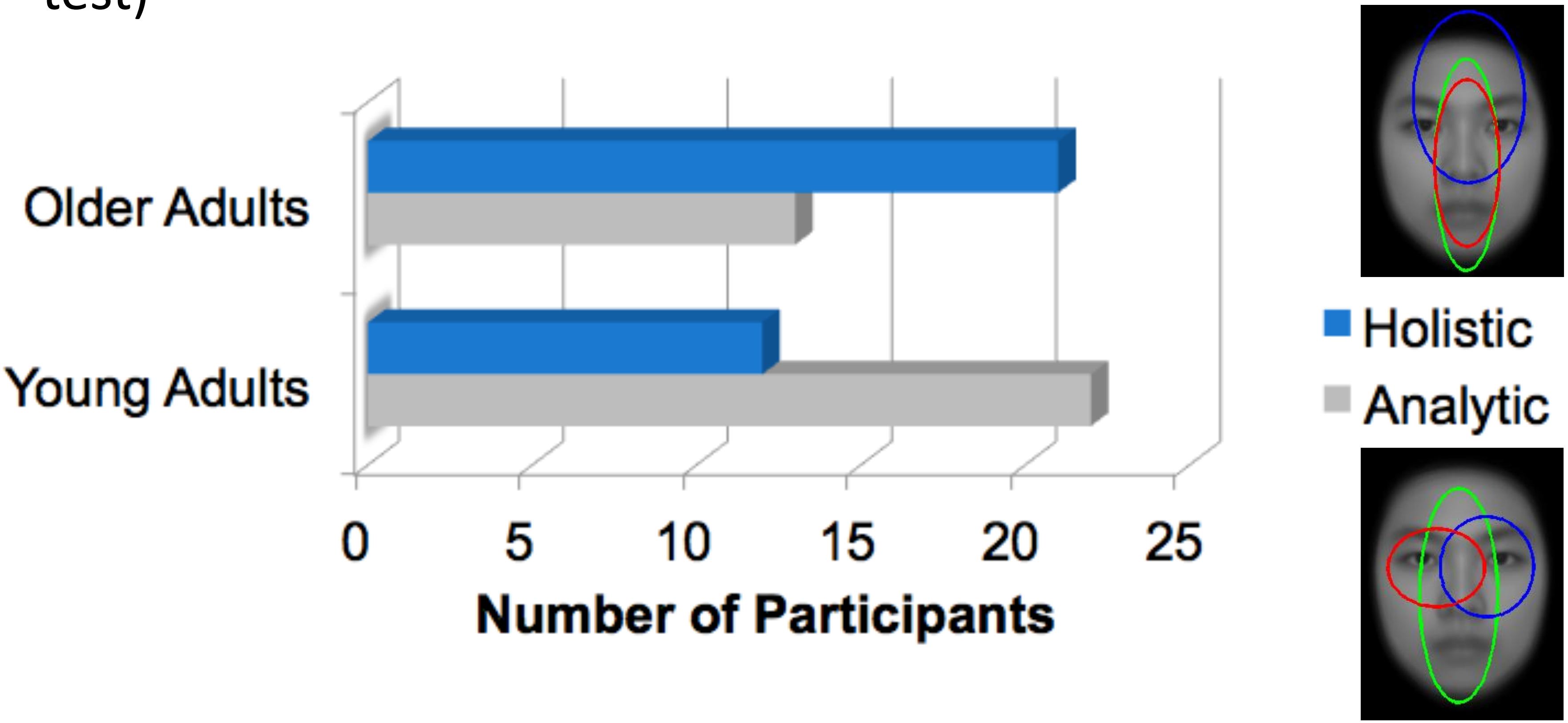


Representative HMM of Analytic Pattern

Prior values	Red	Green	Blue
	.03	.82	.15
Transition probabilities	To Red	To Green	To Blue
From Red	.21	.34	.45
From Green	.25	.36	.39
From Blue	.27	.26	.47

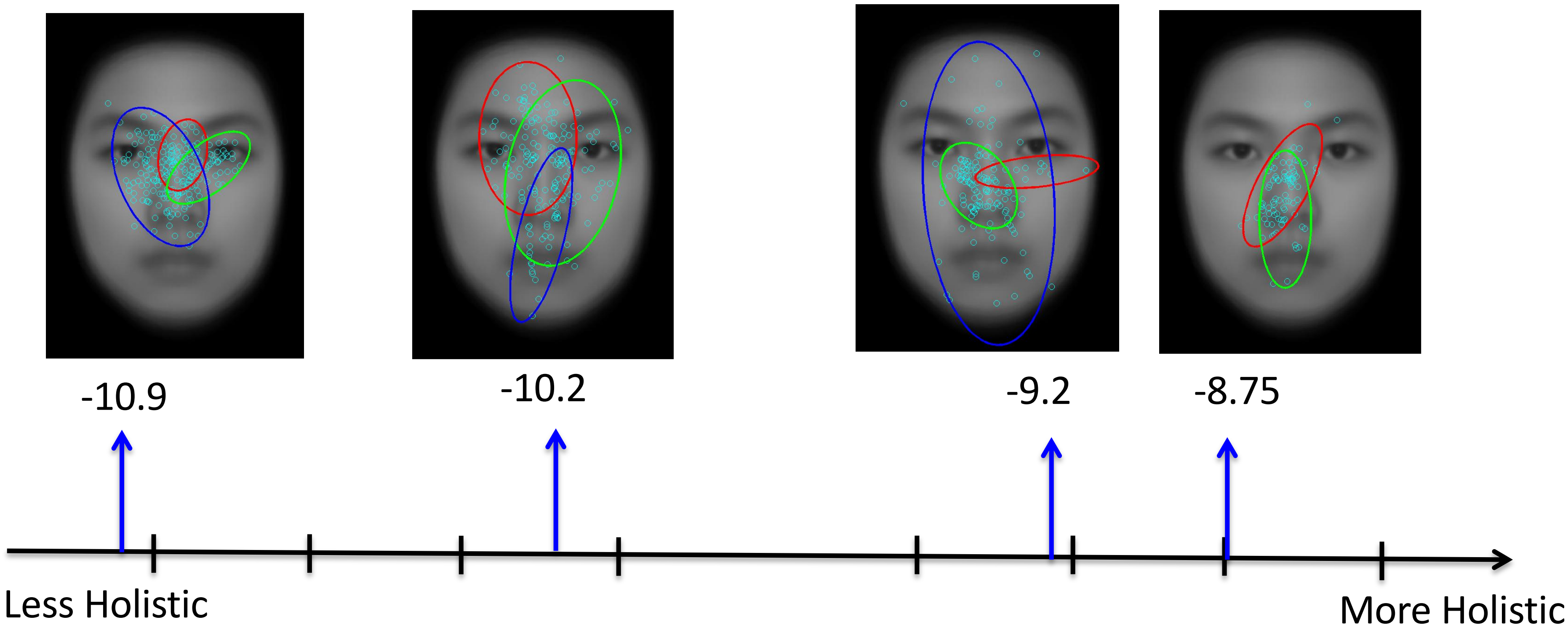
Result: Different Eye movement patterns

- Look at group membership (2x2 contingency table)
- More older adults adopted holistic patterns while more young adults adopted analytic patterns, X(2) = 4.77, p = .03 (Chi-square test)



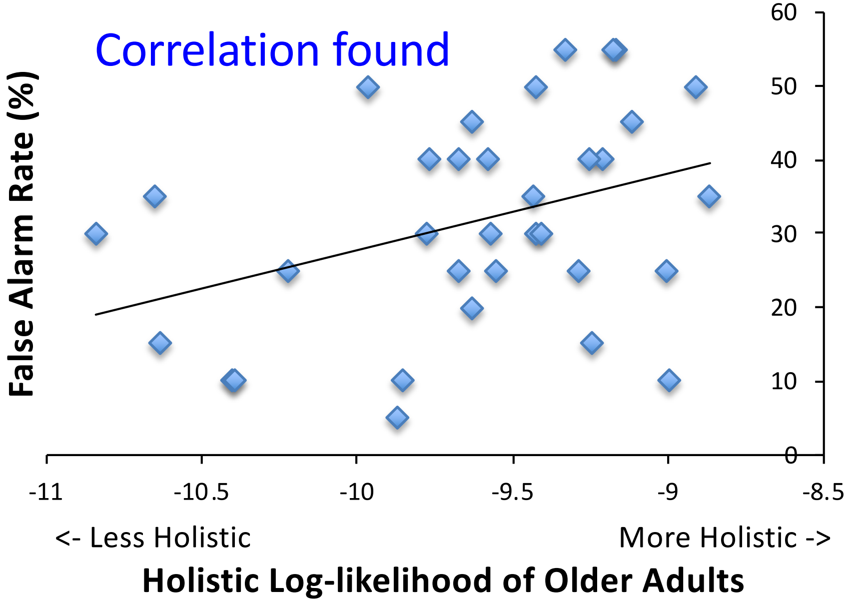
Results

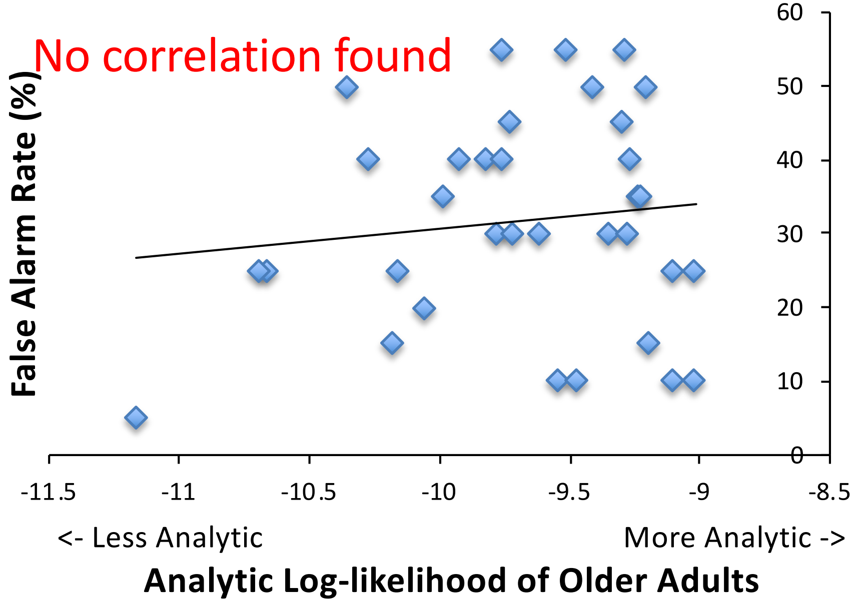
- Compute the likelihood of each participant's eye fixations to the holistic or analytic group HMM.
- **Holistic likelihood:** degree of being classified to holistic pattern
- **Analytic likelihood:** degree of being classified to analytic pattern
- For example:



Results: Relationship between eye movement pattern and performance

- Group analysis
  - Analytic group scored higher (d') than the holistic group:
    - lower false alarm rate
    - No difference in the hit rate.
- Linear regression analysis
  - Among older adults, the higher the log-likelihood of a holistic pattern, the higher the false alarm rate. This was not found in the analytic patterns.

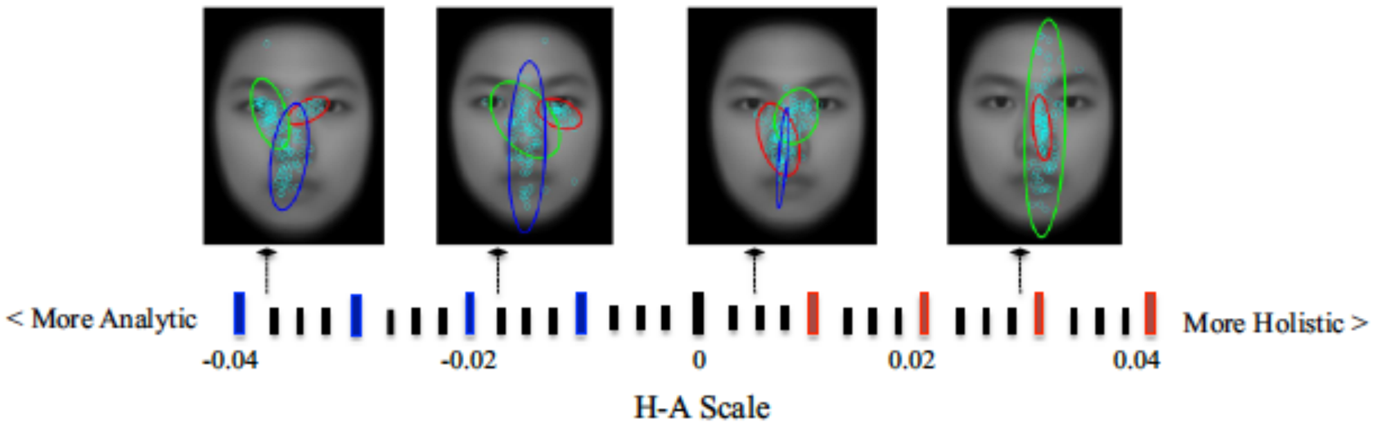




Relationship between eye movement patterns and cognitive ability in older adults

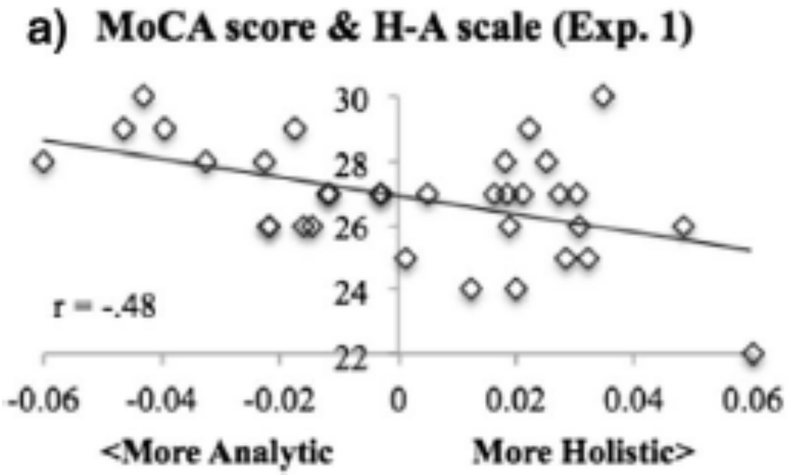
- Summarize similarity to the two groups using Holistic-Analytic (HA) scale:

$$\text{H−A } \mathit{Scale} = \frac{\mathit{Holistic}\ \log\!-\!\mathit{likelihood} - \mathit{Analytic}\ \log\!-\!\mathit{likelihood}}{|\mathit{Holistic}\ \log\!-\!\mathit{likelihood}| + |\mathit{Analytic}\ \log\!-\!\mathit{likelihood}|}$$



- Linear regression analysis
  - In older adults, the lower the cognitive ability (by MoCA), the higher the H-A scale:

R = -.48, P = 0.004



Eye movement & cognitive ability: Correlations

- **H-A scale was particularly correlated with executive and visual attention functions.**

	Correlation with H-A Scale	
	r	p
1. General Cognitive assessment (MoCA)	-.44	.010*
2. Executive Planning (TOL; Total moves)	.36	.043*
3. Visual Attention (TMT)	.37	.034*
4. Working Memory - Verbal 2-back	-.35	.062
- Spatial 2-back	-.35	.066
5. Verbal Memory (CAVLT)	-.17	.337
6. Verbal Fluency (CVFT)	-.15	.394

  – Could also use a hierarchical regression analysis to determine the most important predictors

Chan, C. Y. H., Chan, A. B., Lee, T. M. C., & Hsiao, J. H. (2018). Eye movement patterns in face recognition are associated with cognitive decline in older adults. *Psychonomic Bulletin & Review, 25(6), 2200-2207.*

Experiment Design for Web Pages

- Subjects were given a self-evaluation survey on their familiarity with technology.
  - Used to form groups of *experienced* / *inexperienced*
- Subjects were asked to perform a task on LinkedIn
  - "Ask your contact "Christian Maier" for a letter of recommendation."
- Some given incentive to perform quickly.
  - Used to form groups of *pressure* / *no-pressure* to perform
- Recorded eye movements while performing the task.
  - discover the ROIs from data using GMMs / HMMs.

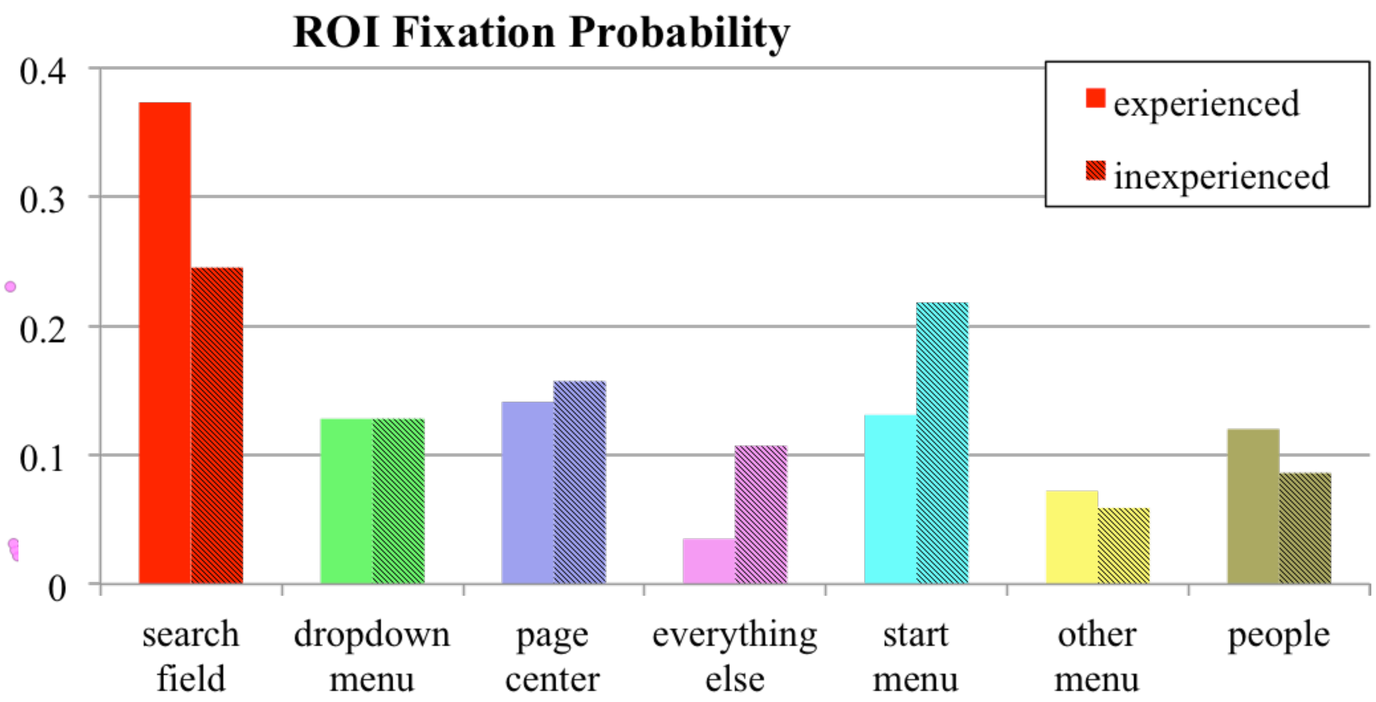
A. Eckhardt, C. Maier, JJ P.-A. Hsieh, T. Chuk, A.B. Chan, J.H. Hsiao, and R. Buettner (2013). "Objective Measures of IS Usage Behavior Under Conditions of Experience and Pressure Using Eye Fixation Data." In: *International Conference on Information Systems.*

Data Analysis and Webpage ROIs

- Learn one HMM for each group.
  - share ROIs among groups.
- Look at differences in eye gaze strategy among groups.

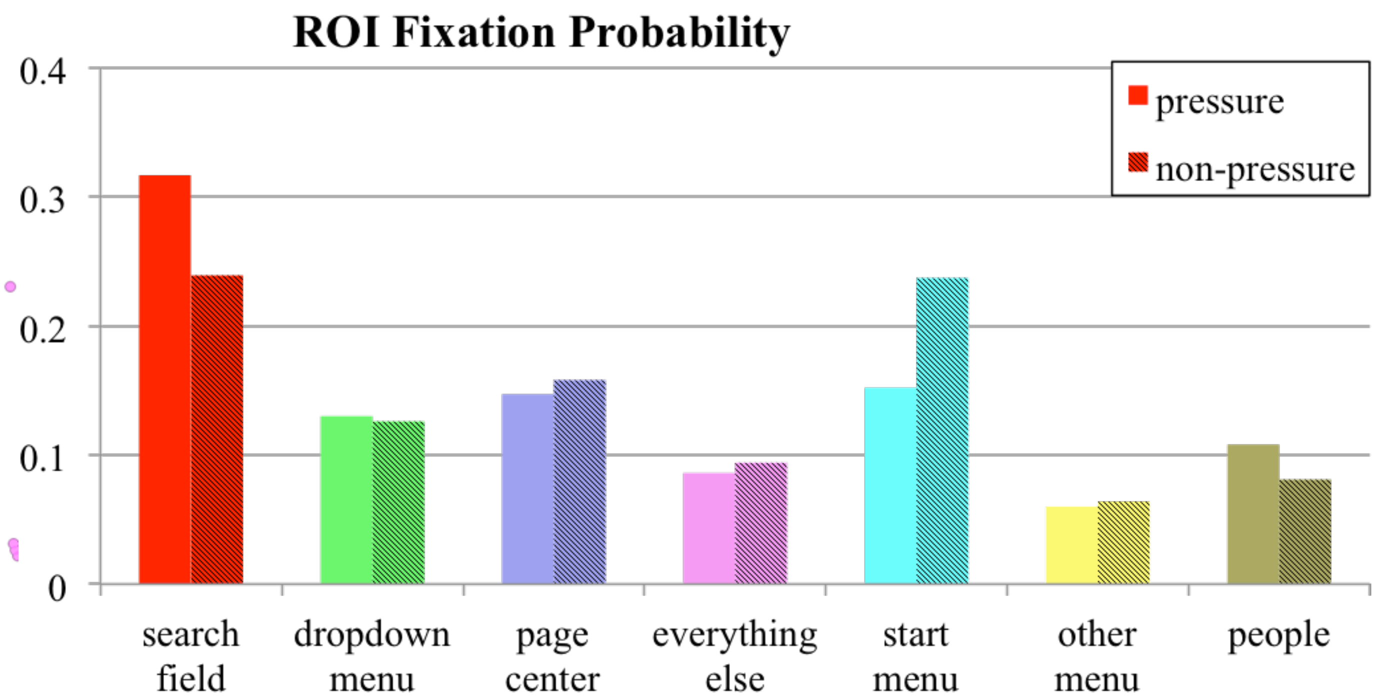
A. Eckhardt, C. Maier, JJ P.-A. Hsieh, T. Chuk, A.B. Chan, J.H. Hsiao, and R. Buettner (2013). “Objective Measures of IS Usage Behavior Under Conditions of Experience and Pressure Using Eye Fixation Data.” In: *International Conference on Information Systems.*

Behavioral IS usage patterns of experienced and inexperienced users



- Experienced users focus on search area.
- Inexperienced users use the start menu.
  - also get distracted by “everything else”

Behavioral IS usage patterns with(out) pressure to perform



- Pressured users focus on search area.
- Non-pressured users use the start menu.

Different Stimuli Layouts

- So far the stimuli have the same feature layouts
  - e.g., all faces are aligned by eyes and mouth
- **How to summarize a general eye movement strategy in tasks where stimuli's feature layouts differ significantly** (e.g., scene perception, visual search, reading etc.)?
  - **EMHMM with co-clustering**: 1 stimulus 1 model; use co-clustering to discover participants using the same pattern across stimuli.

More details in next lecture

Outline

- Experiment Design
  - Face recognition
  - Web page usage
- Simulation Study
  - *How many trials are needed?*
- Develop Your Mini-Experiment
  - *PsychToolbox and Eyelink*

Simulation Study

How many trials do we need?

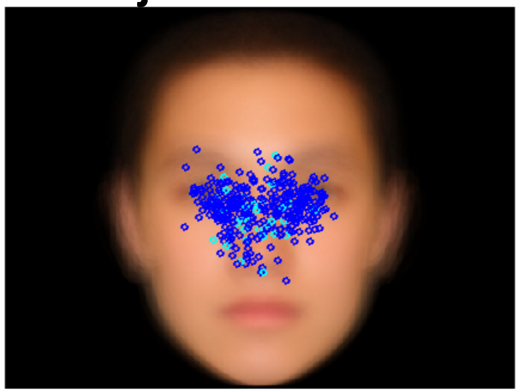
Antoni B. Chan
Dept. of Computer Science,
City University of Hong Kong

Janet H. Hsiao
Dept. of Psychology,
The University of Hong Kong

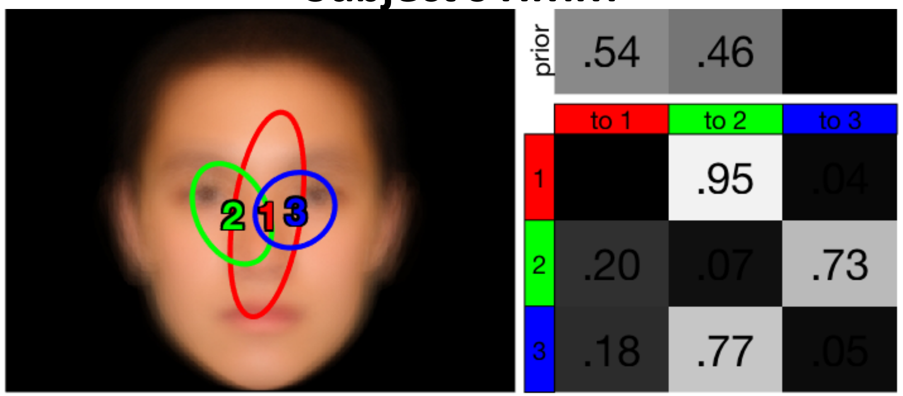
Eye Tracking Experiment

- For each subject:
  - Collect eye fixations on set of stimuli (e.g., faces) while performing a task (e.g., recognition).
    - Typically 30-50 trials → 30-50 fixation sequences.
    - Depending on the task, 2-10 fixations per trial.
  - Estimate an individual HMM for the subject.
  - Interpret the estimated HMM as the subject's eye gaze strategy for this task.

Subject's fixations



Subject's HMM

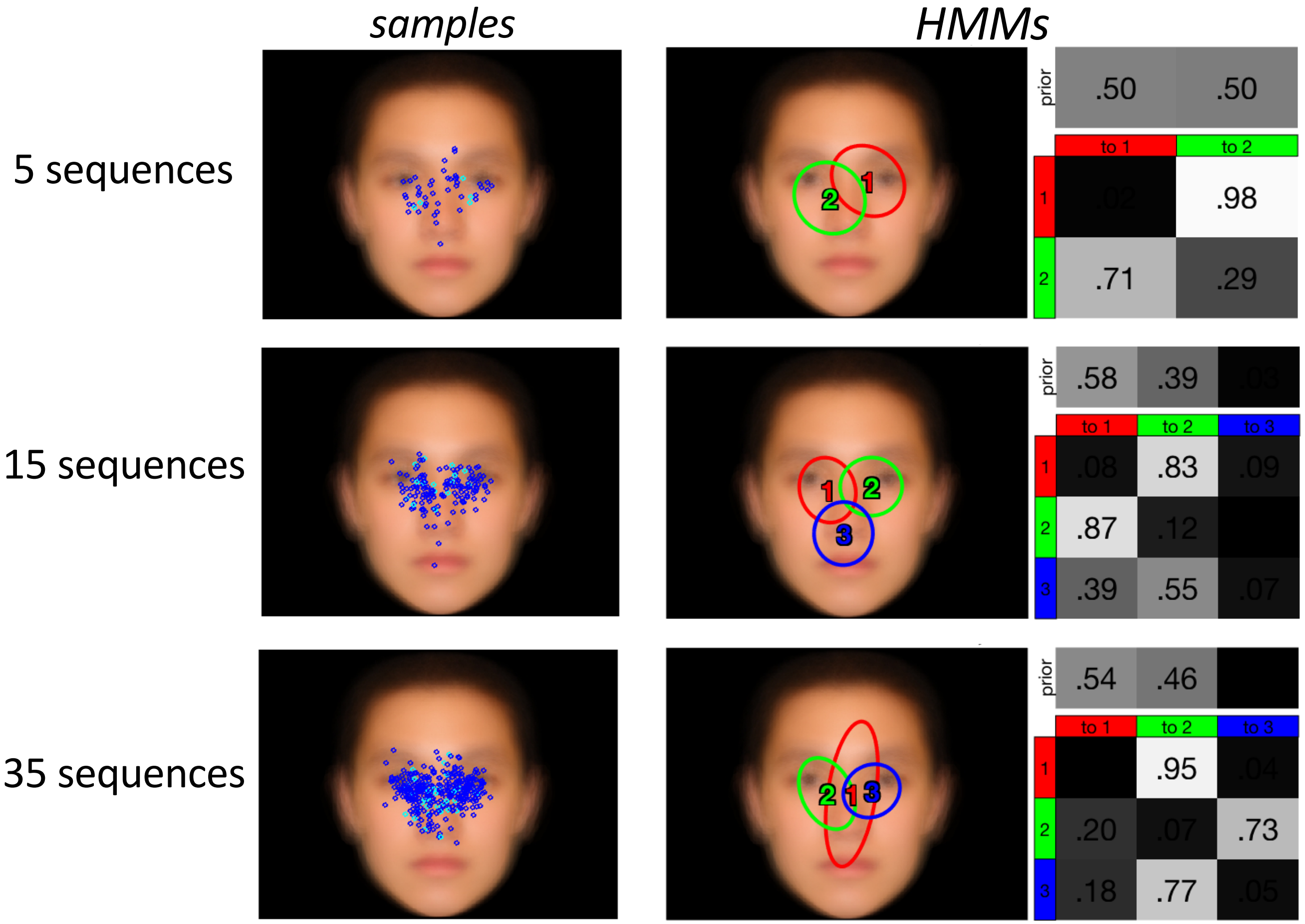


Subject's Strategy

Looks between eyes (analytic strategy)

How much data?

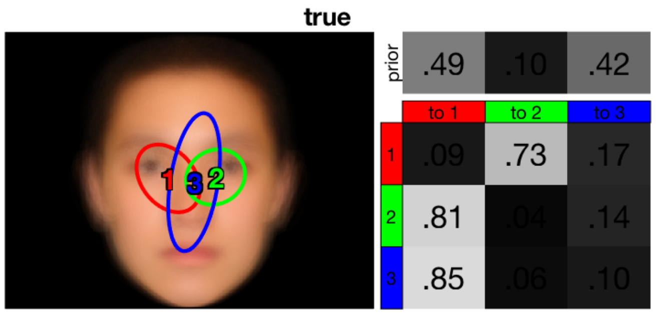
- **Question**: How much data do we need to ensure that the estimated HMM can well represent the subject’s overall eye gaze strategy?



How much data?

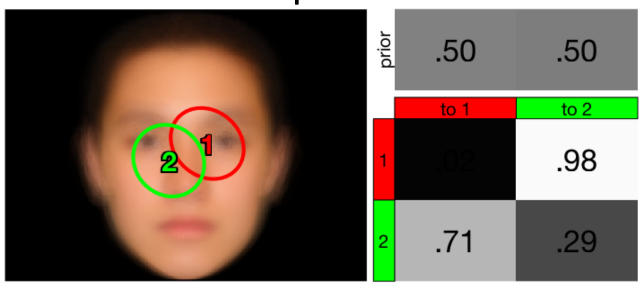
- Accuracy of the estimated HMM parameters, relative to the ***true*** underlying strategy (*true HMM*), depends on the amount of data used.
  - Analogously, the accuracy of a sample mean relative to the true mean depends on the sample size.

True HMM:

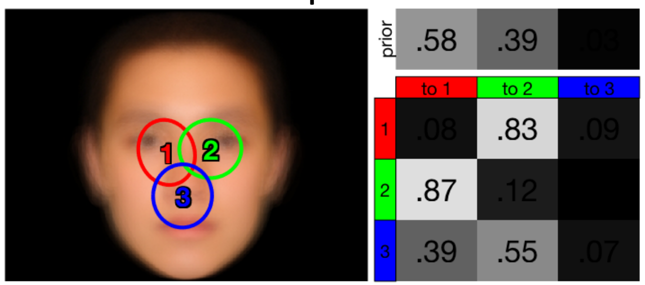


HMMs estimated with…

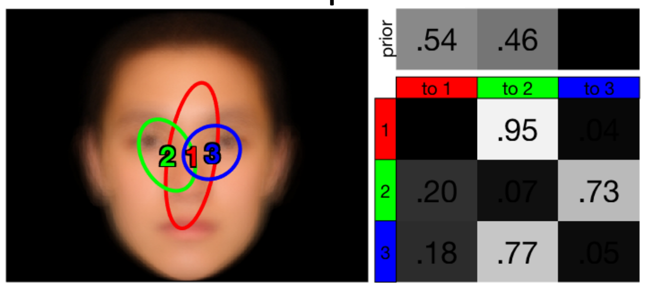
5 sequences



15 sequences



35 sequences



How much data?

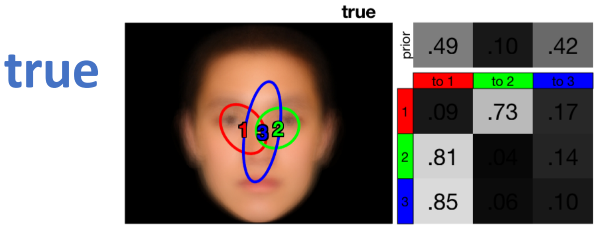
- **Question**: How much data do we need to ensure that the estimated HMM can well represent the subject's overall eye gaze strategy?
  - *i.e., Given samples from a true HMM, how much data is needed to estimate the HMM within some error tolerance?*
- Can claim that the estimated HMM represents the subject's ***overall*** eye gaze pattern on that type of stimuli.
- *Note*:
  - Using less data, then the estimated HMM only represents the eye gaze pattern over the ***particular*** stimuli.

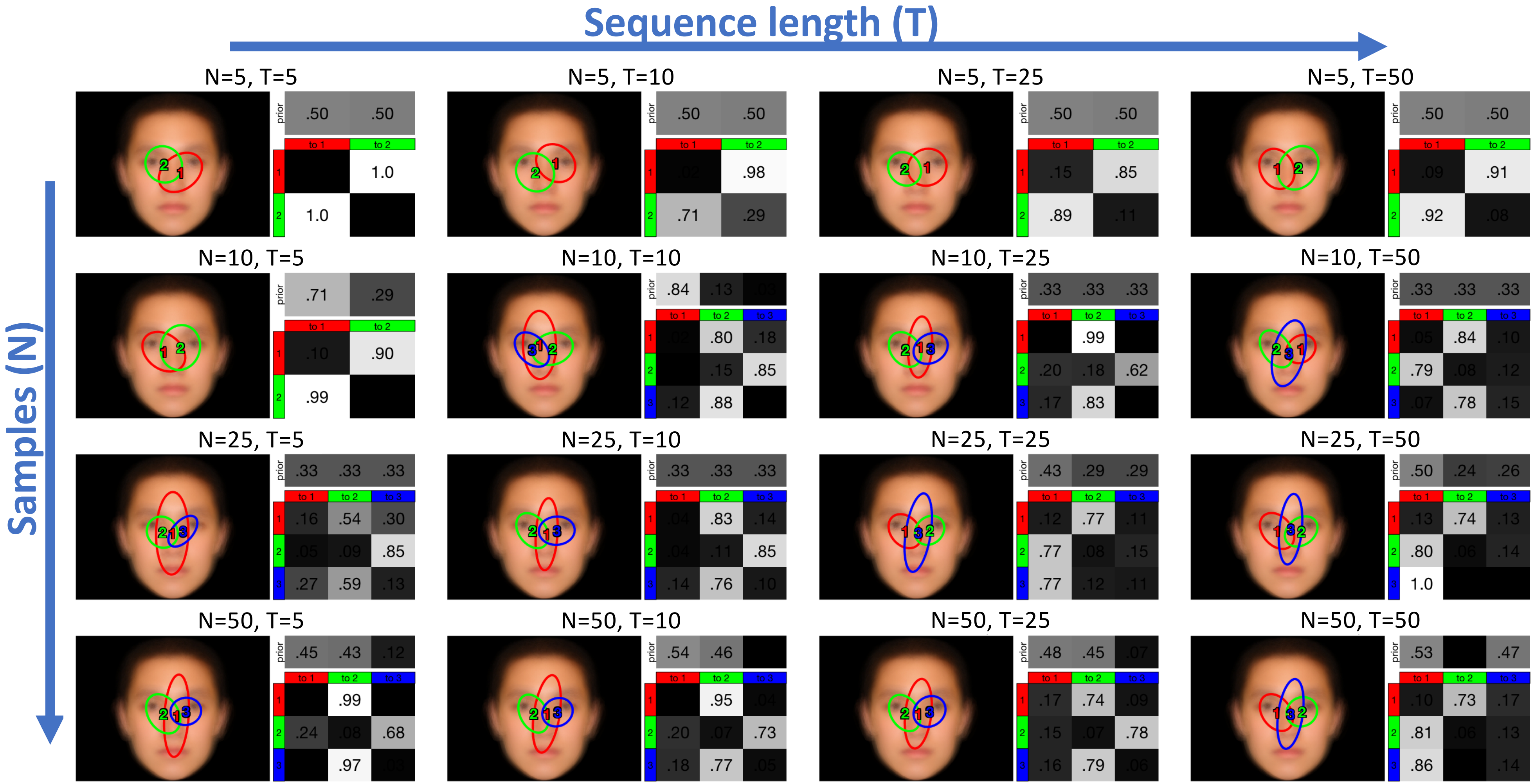
Simulation Study

- **To quantify the error when estimating HMMs from limited samples**
- 10 ground-truth HMMs from 10 subjects.
  - Image size is 512x384 and face region is 300x350.
- Estimation quality depends on the number of samples
  - Consider different combinations of number of sequence samples *N* and sequence lengths *T.*
- For each combination (*N*, *T*) we run 500 trials:
  - Select a random ground-truth HMM.
  - Sample *N* sequences of length *T* from the true HMM.
  - Estimate an HMM from the samples using VBHMM with automatic hyperparameter selection.
  - Compute *error measure* between the estimated and true HMMs.
- Average error over trials.

Example

- Estimated HMMs for different T, N





more fixations → better quality

Estimation Error for whole HMMs

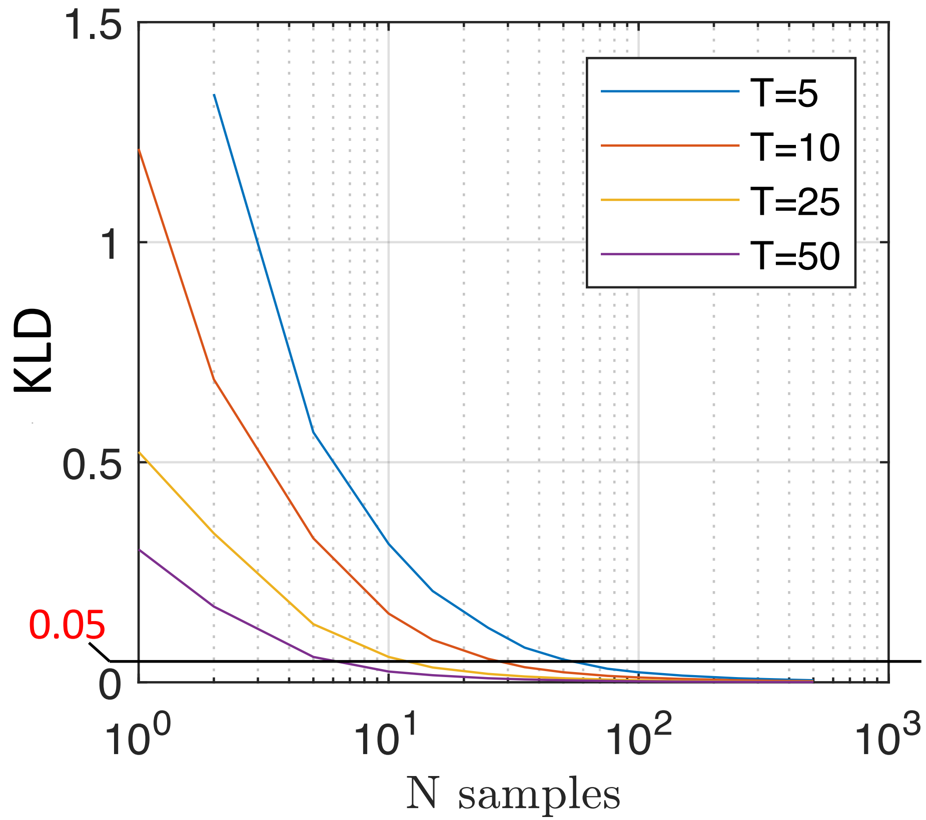
- Kullback-Leibler divergence (KLD) rate between probability distributions of the true and estimated HMM distributions.

$$\phi(p(y), \hat{p}(y)) = \int_y p(y) \log \frac{p(y)}{\hat{p}(y)} dy$$

  - Zero when the two HMMs are the same.
  - Positive values indicate more dissimilarity.
- Intuitively, KLD is the amount of “information” that needs to be added to turn the estimated distribution into the true distribution.

Estimation Error for whole HMMs

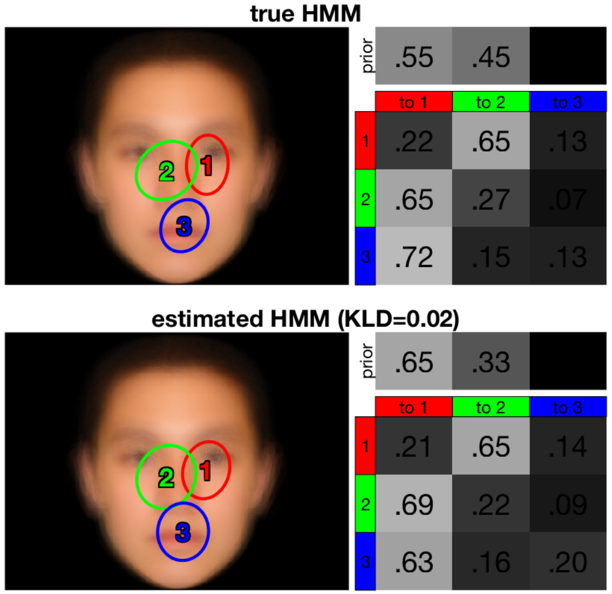
- KLD decreases to zero as number of samples increases or length of sequences increases.
- To obtain low KLD of 0.05 requires ~250 fixations:
  - *N*=52, *T*=5
  - *N*=26, *T*=10
  - *N*=11, *T*=25
  - *N*=6, *T*=50



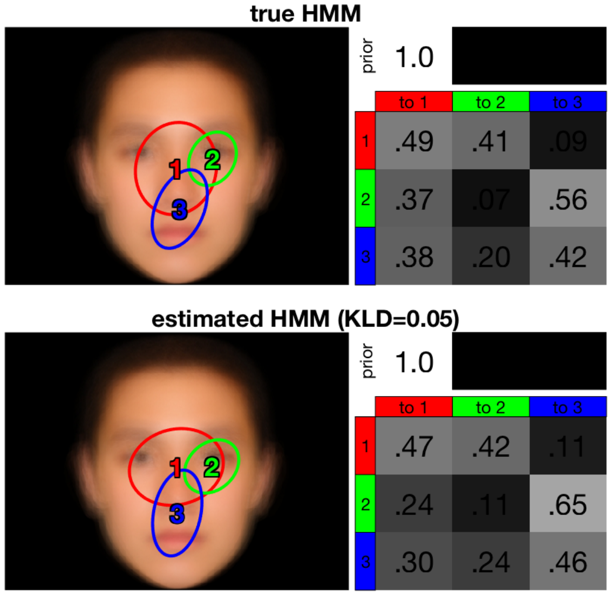
HMM Examples

- True and estimated HMMs

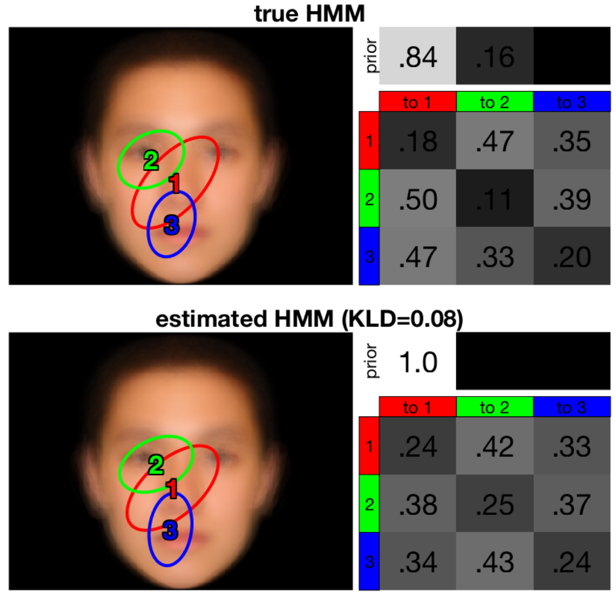
KLD=0.02



KLD=0.05

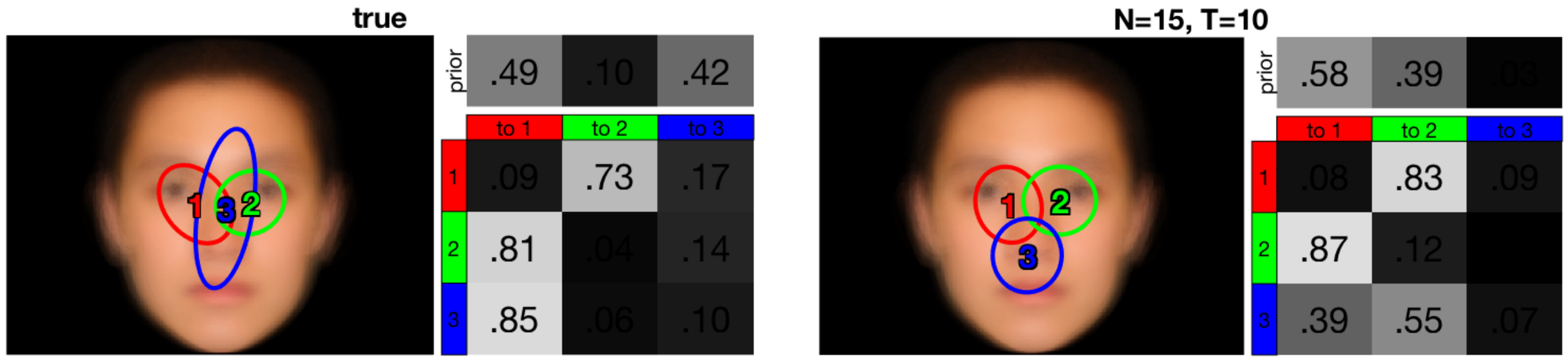


KLD=0.08



Estimation Error of Components

- KLD measures the **overall** error between GT and estimated HMMs.
- We can also measure error of the estimated model components
  - ROIs
  - transition matrix
  - initial probabilities

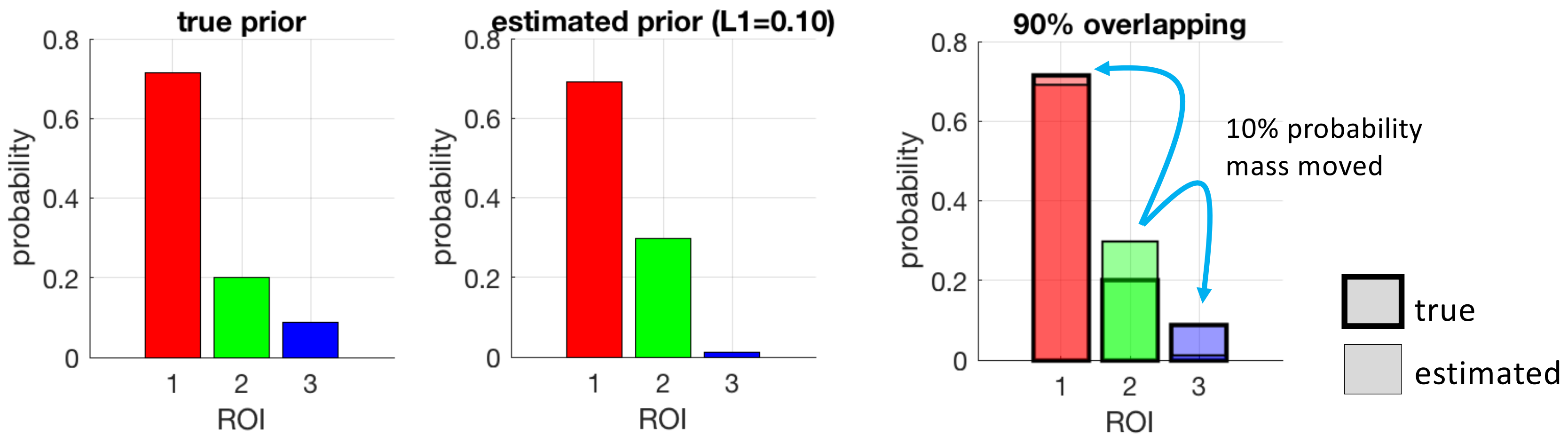


L1-norm error

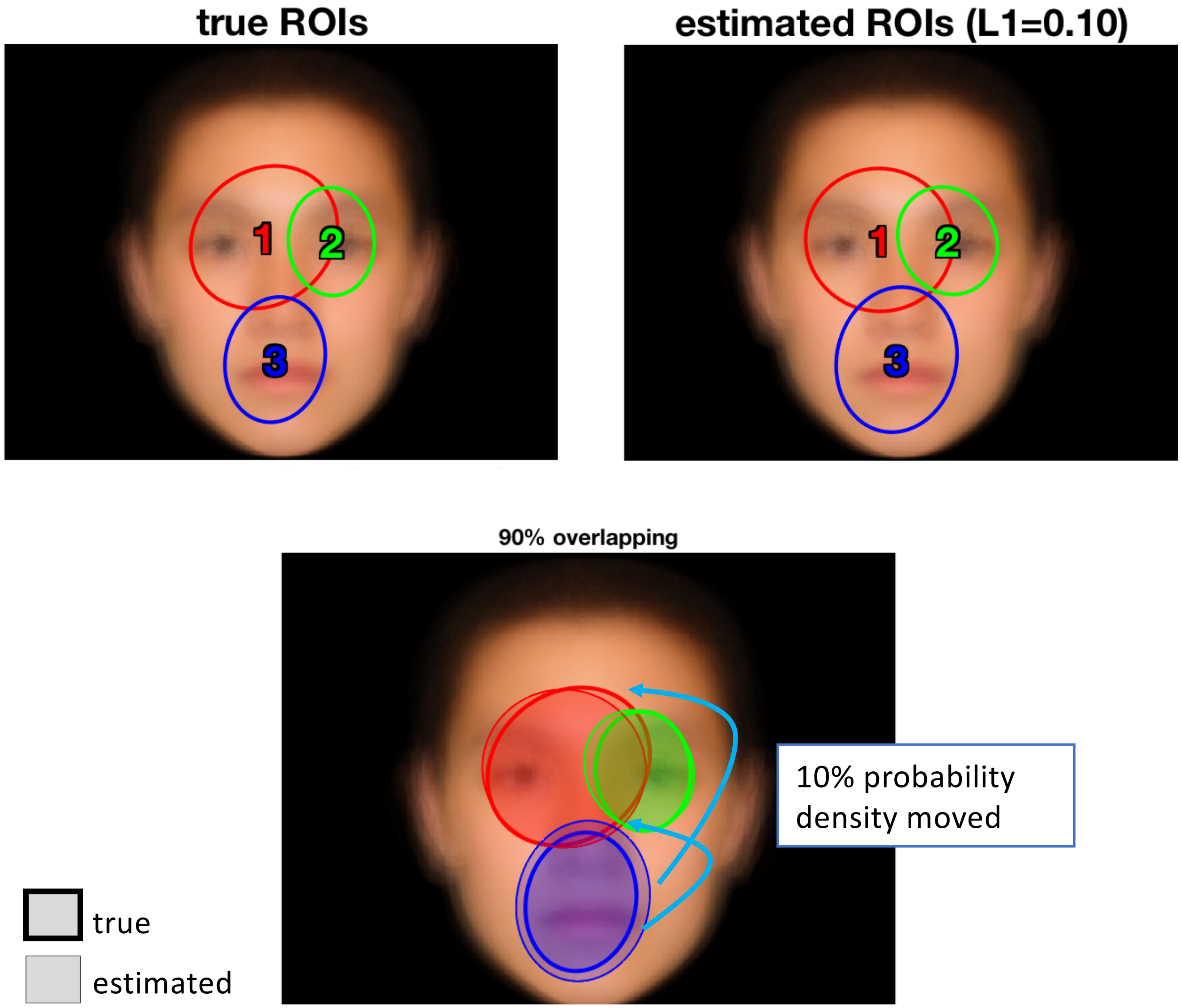
- L1-norm measures the absolute difference between two probability distributions.

$$\psi(p(y), \hat{p}(y)) = \frac{1}{2}\int_y |p(y) - \hat{p}(y)| dy$$

  - Intuitively, it is the amount of probability that needs to be moved to turn one distribution into the other.
  - Amount of non-overlapping between distributions.

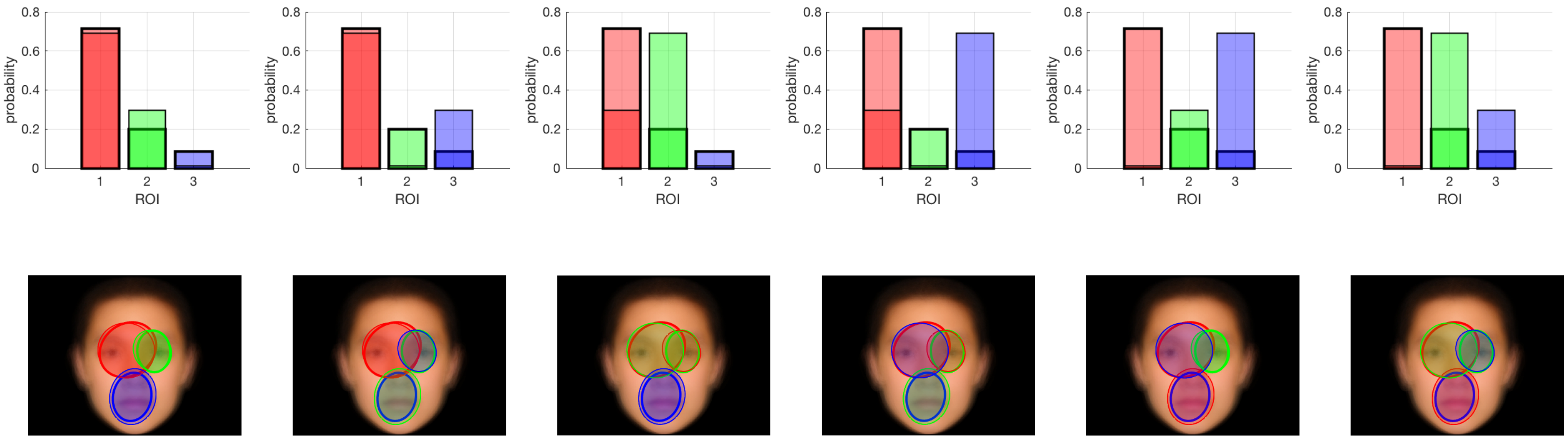


ROI Example


true ROIs
1
2
3
estimated ROIs (L1=0.10)
1
2
3
90% overlapping
10% probability density moved
true
estimated

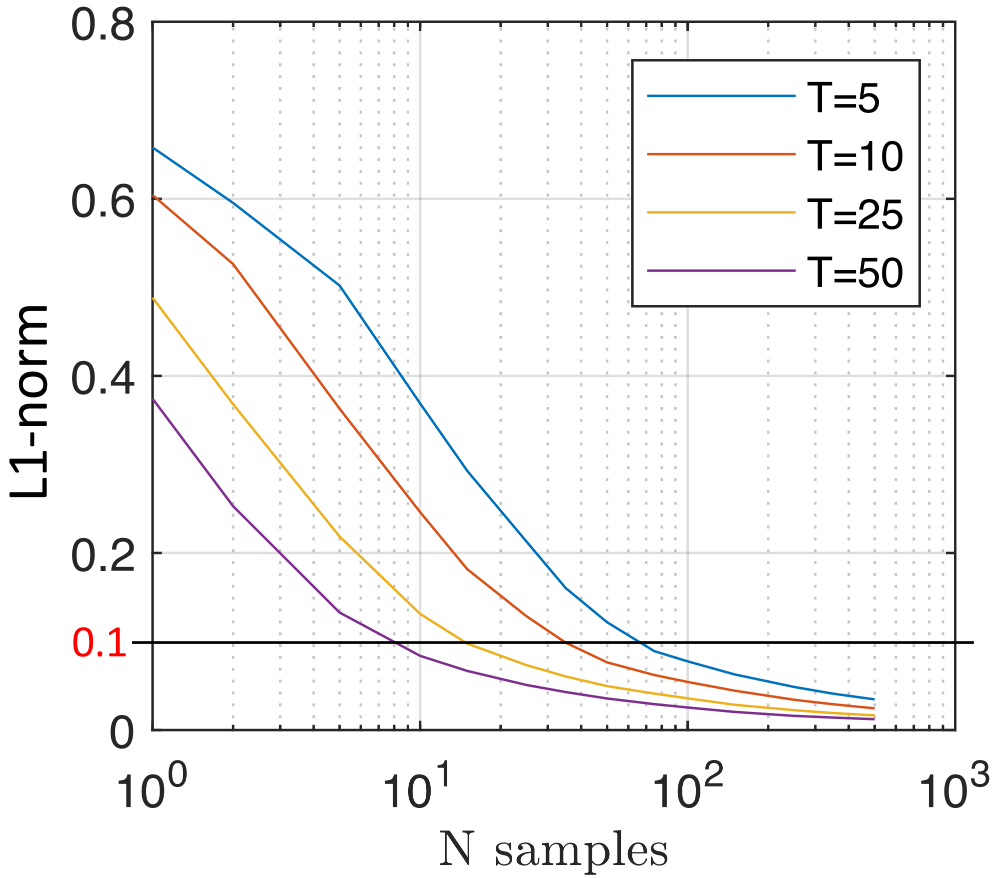
Matched L1-norm error

- **Problem**:
  - the order of the states in the GT and estimated HMMs may be different.
- **Solution:** align the states
  - Matched L1-norm: the smallest L1-norm over all permutations of states.



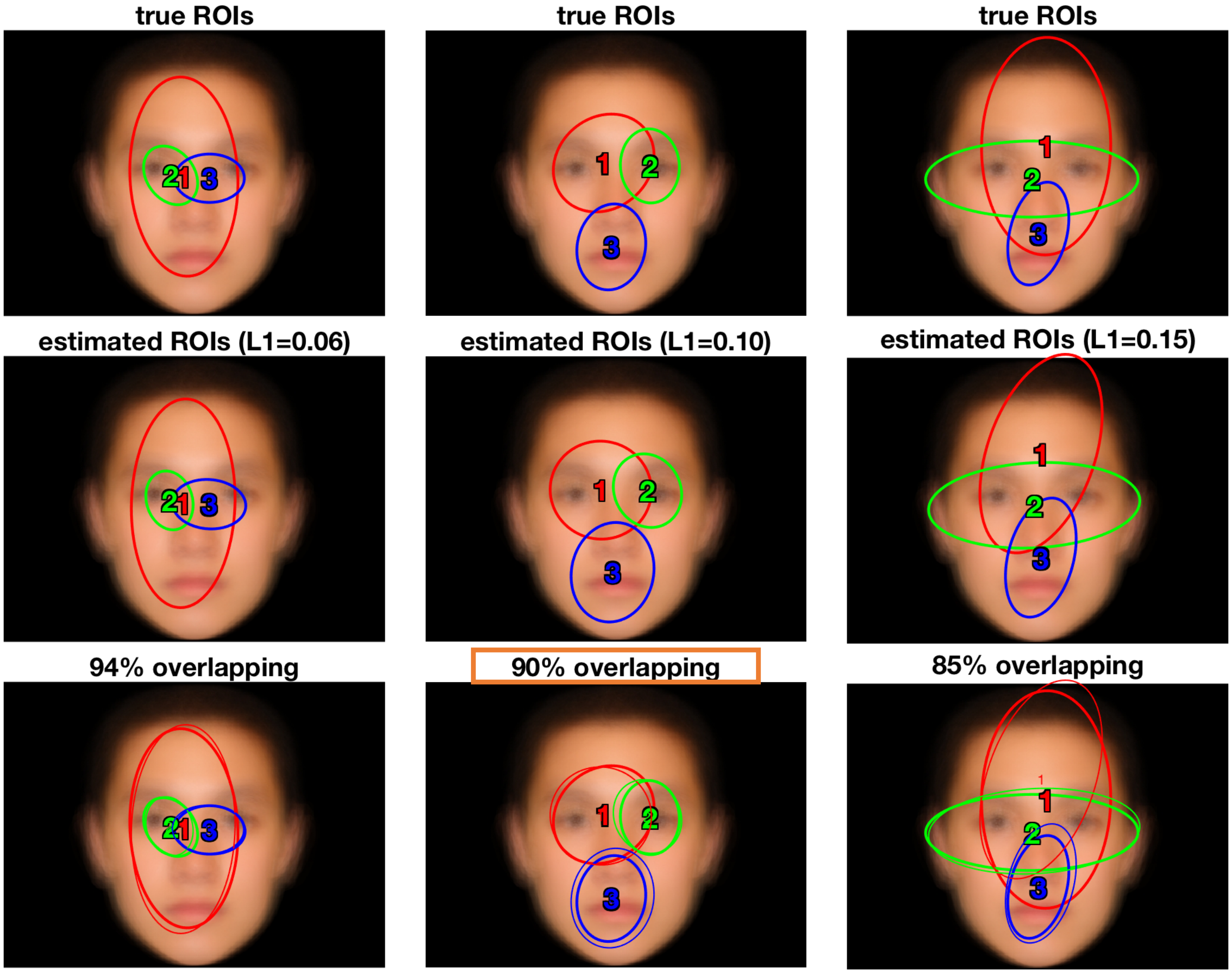
Estimation Error for ROIs

- ROI overlap increases with number of fixations.
- For 90% overlap (L1-norm = 0.1) requires 350 total fixations:
    - *N*=66, *T*=5
    - *N*=35, *T*=10
    - *N*=15, *T*=25
    - *N*=8, *T*=50



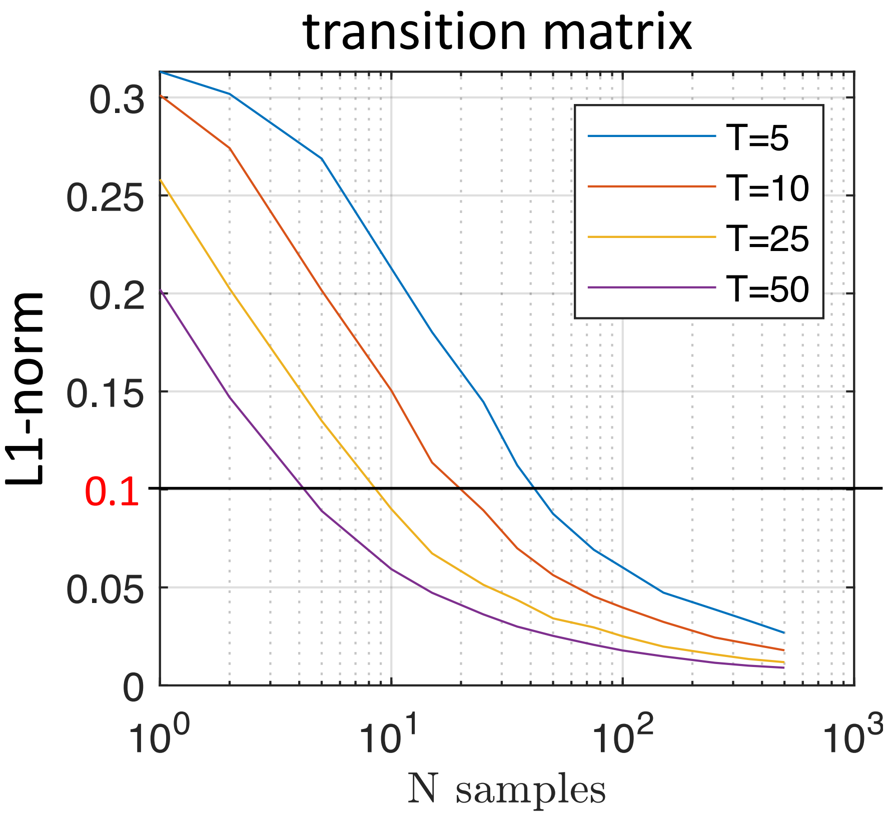
ROI Examples

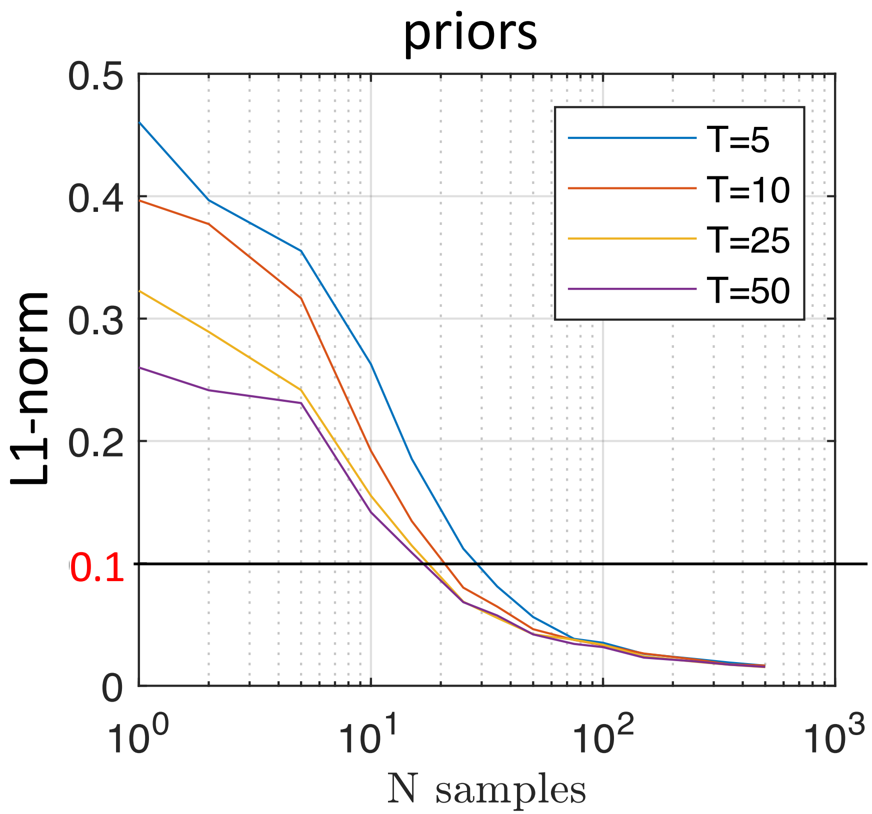
- 90% overlap of ROIs is equivalent to:
  - 4.7 pixel error in the ROI means
  - 4.5% change in the ROI sizes



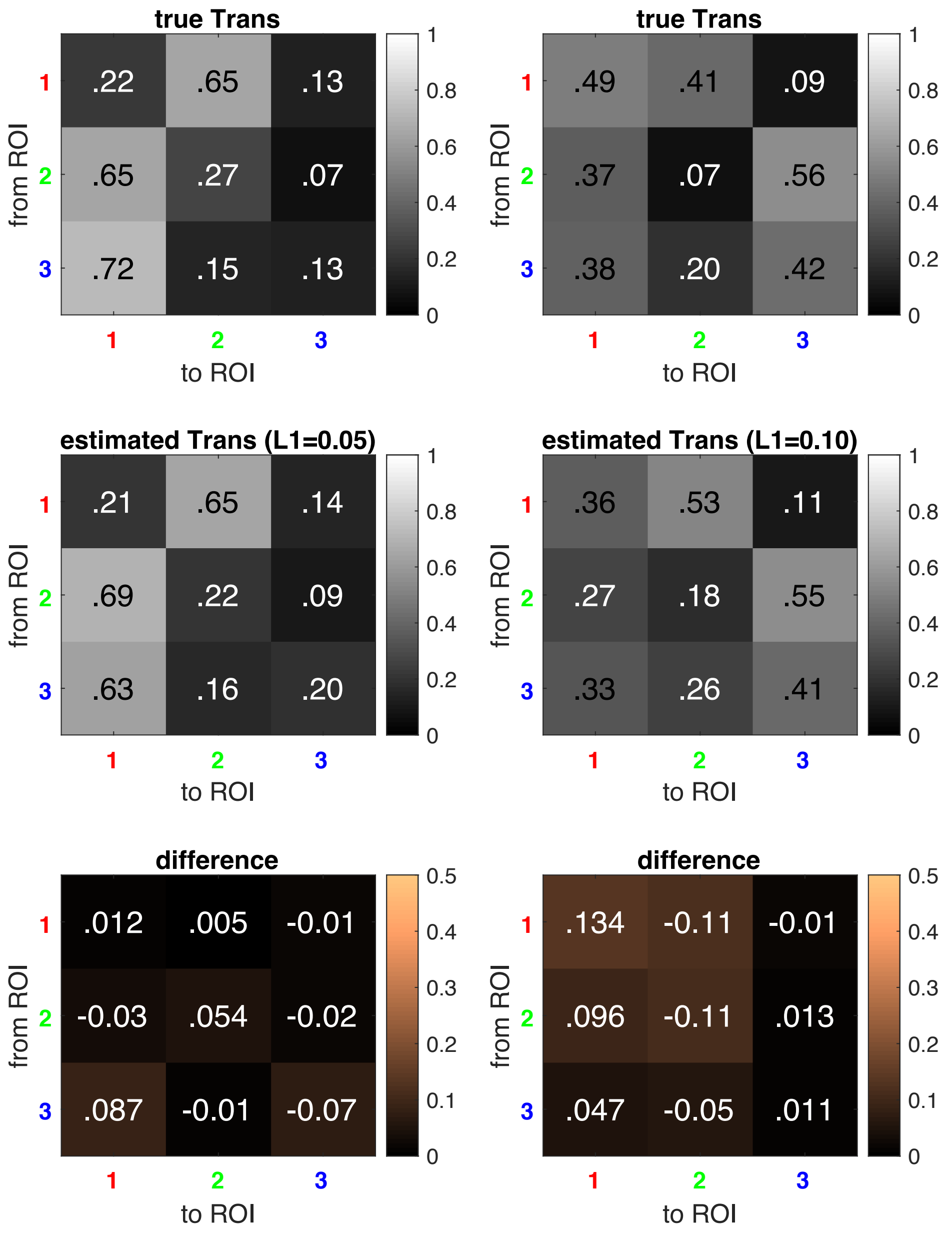
Error for Transition/Prior

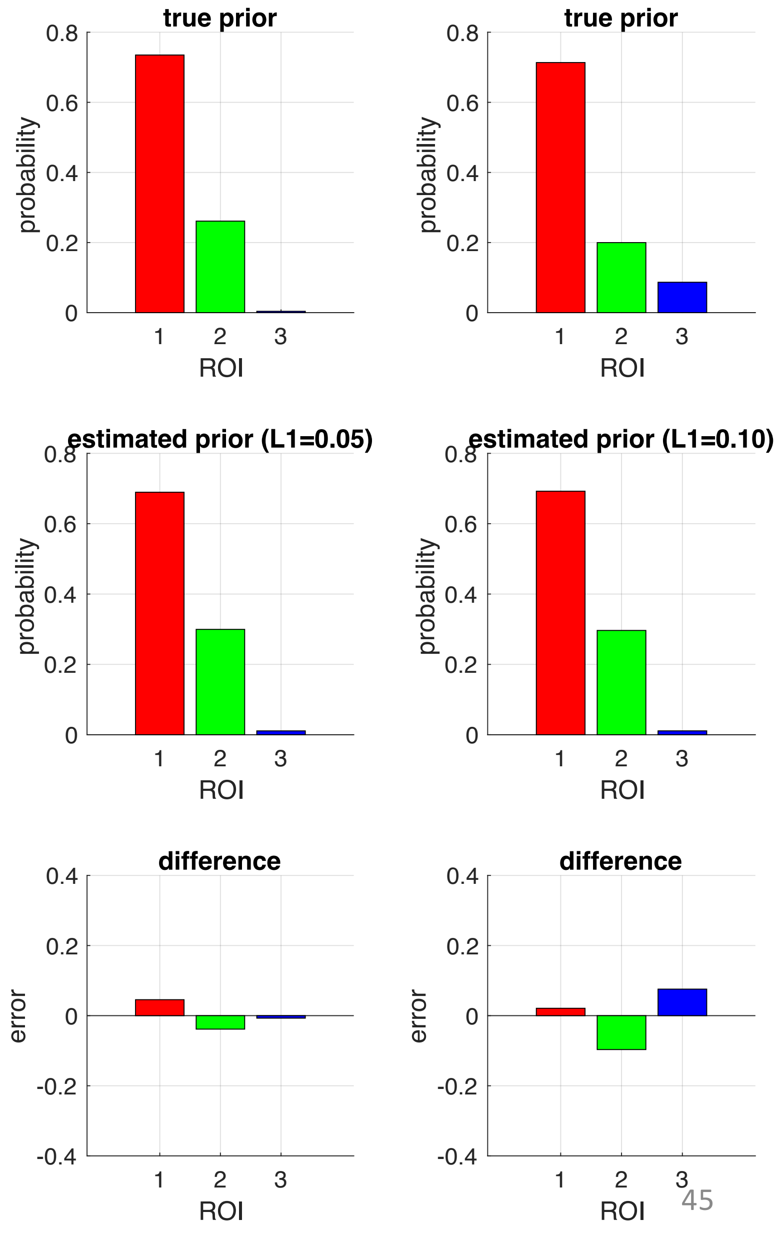
- 215 fixations for 90% overlap of transition matrices.
- ~25 sequences for 90% overlap of priors.





Transition/Prior Examples





Summary

- **Recommendations:**
  - To estimate an HMM with 90% overlap with the *true* HMM:
    - at least 250-350 total fixations, at least 25 fixation sequences.
      - E.g., 25 fixation sequences of length 10
      - E.g., 50 fixation sequences of length 5
  - For 80% overlap:
    - at least 150 total fixations, 10 fixation sequences
  - Estimated HMM is close to the true underlying strategy of the subject.
    - For interpreting *overall* eye gaze strategy of the person.
- **Note:**
  - For less data, still okay to use estimated HMMs.
    - Estimated HMM represents eye gaze pattern on the *particular stimuli*, not the subject's underlying strategy.
    - Still valid for analysis, e.g., a classification/regression task.

Questions?

- EMHMM toolbox:
  - http://visal.cs.cityu.edu.hk/research/emhmm/

- Acknowledgements:

  - This research was supported by the Research Grant Council of Hong Kong SAR: #17402814 and HKU 745210H for J.H. Hsiao; CityU 110513 and G-CityU109/14 for A.B. Chan. We also thank the HKU Seed Funding Programme for Basic Research (Project number 201311159131 to J.H. Hsiao.

Outline

- Experiment Design
  - Face recognition
  - Web page usage
- Simulation Study
  - *How many trials are needed?*

- Develop Your Mini-Experiment

Mini-Experiment and Analysis

- 1) collect your own data
  - Psych Toolbox and EyeLink
  - Try to find a sample script that collects eye gaze data for similar tasks
- 2) use data available online (published papers)
- 3) use data you already collected
- 4) use data we collected
  - Scene perception: Use co-clustering to summarize eye movement patterns for animal and vehicle images separately, and see whether participants use consistent patterns across the two image types.
  - Website viewing: co-clustering across webpages; analyze other tasks.