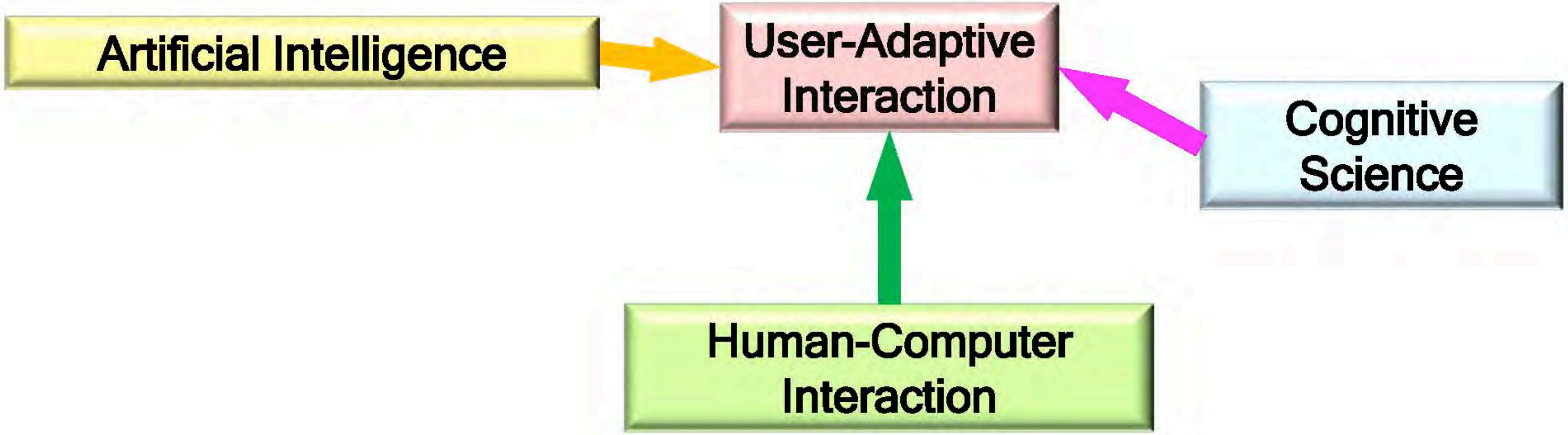
The Eyes Are the Windows to the Mind: Implications for Intelligent User Interfaces

Cristina Conati

University of British Columbia

Research Context (1)

- User-Adaptive Interaction (UAI)



- Create intelligent user interfaces that support personalized interaction by
  - capturing a user's needs, states, abilities
  - adapting the interaction accordingly

Research Context (cont'd)

- User-adaptive techniques are being investigated for a variety of exciting applications, e.g.
  - Recommender systems
  - Digital assistants
  - Intelligent Tutoring Systems
  - Entertainment (e.g., AI in Games)
  - Assistive technologies
  - Smart homes

Just to name a few

Adaptation Cycle

Forms of adaptation

- *Give help on task*
- *Adapt interface layout/functions*
- *Tailor Information Presentation*

....

User states/traits modeled

- *Goals*
- *Knowledge*
- *Behavioral Regularities*
- *Preferences*
- *Emotions*
- *Cognitive Load*
- *.........*

Input sources

- *Interface Actions*
- *Facial/verbal expression*
- *Movements*
- *Speech*

Adaptation Cycle

Forms of adaptation

- *Give help on task*
- *Adapt interface layout/functions*
- *Tailor Information Presentation*

....

User states/traits modeled

- *Goals*
- *Knowledge*
- *Behavioral Regularities*
- *Preferences*
- *Emotions*
- *Cognitive Load*
- *.........*

Input sources

- *Interface Actions*
- *Facial/verbal expression*
- *Movements*
- *Speech*
- *EYE TRACKING*

Why Eye-Tracking (ET)

- Not many other options in tasks that are mainly perceptual in nature
  - E.g., processing visualizations
- But information on gaze patterns is useful in most interactions
- ET also gives additional data (pupil dilation, head distance from the screen)
  - Helpful in predicting cognitive and affective states

First Eye-Tracker I used

Getting much better now

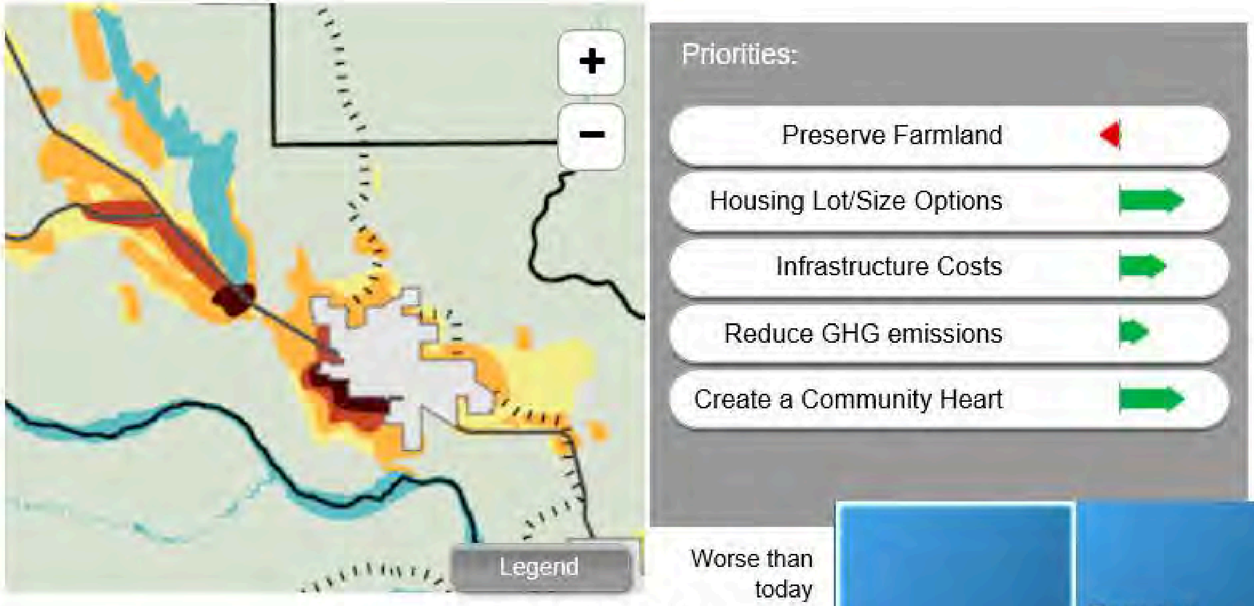
And even better

In This Talk

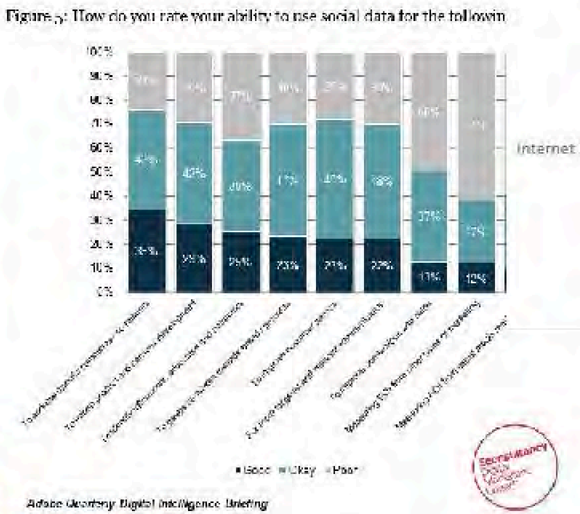
- Example applications of ET in user modeling for
  - User-Adaptive Visualizations
    - relevant cognitive abilities
    - confusion
  - Intelligent Tutoring Systems (ITS)

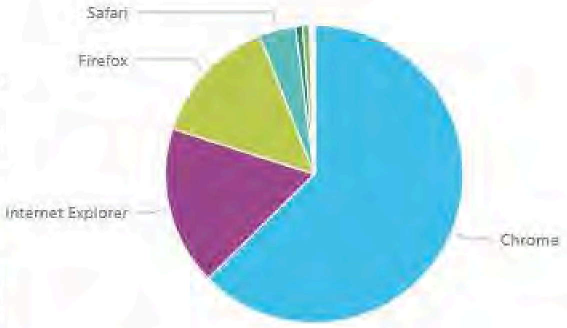
User-Adaptive Visualizations

- Information Visualization (InfoVis) more and more prominent in our life
  [e.g., Roberts et al. 2014; Huang et al. 2015; Blumenstein et al. 2016]



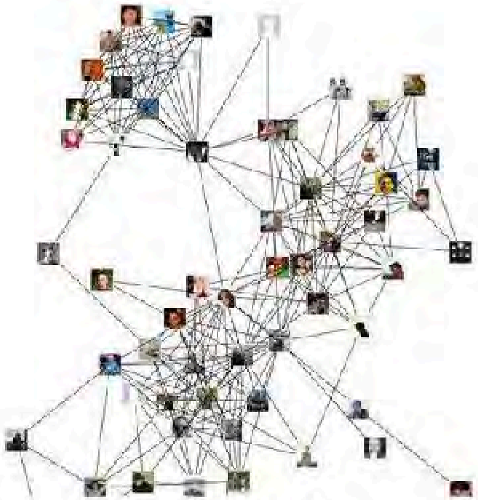
MetroQuest





Google Analytics

DiskUsage (Android)



Facebook

User-Adaptive Visualizations

- Traditional visualizations: one-size-fits all

- Evidence that user cognitive abilities can impact visualization effectiveness
  [e.g., Velez et al. 2005; Conati et al. 2012; Toker et al. 2014; Ottley et al. 2016]

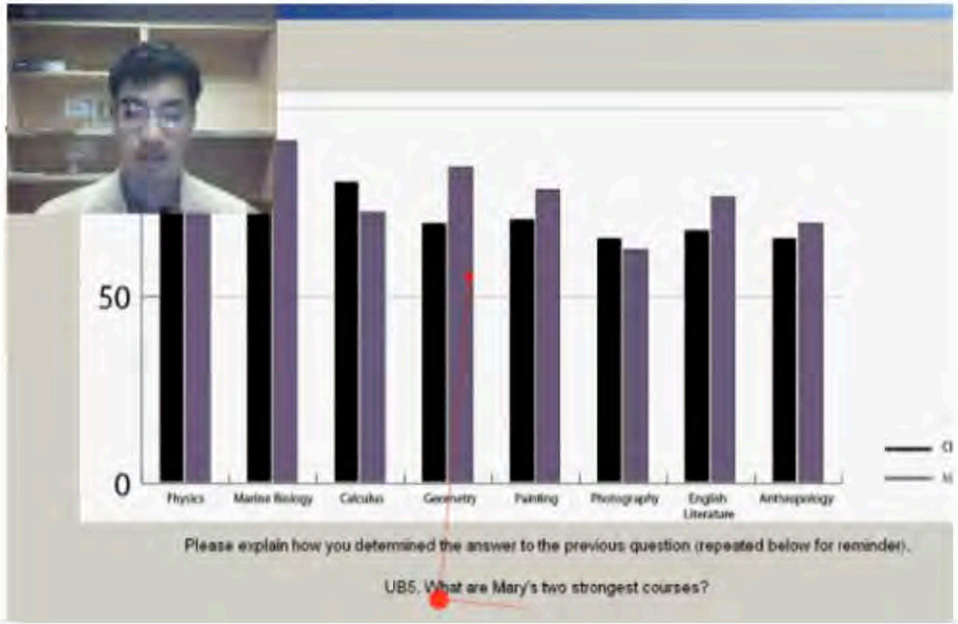
- Evidence that eye tracking can predict some of these abilities in real time *[e.g., Ooms et al. 2014; Steichen et al. 2014; Gingerich et al. 2015]*

Potential for intelligent user-adaptive visualizations that can:

- Detect relevant user abilities
- Provide support personalized to these abilities, to facilitate visualization processing

Before: we used ET to predict cognitive abilities when processing

- Bar-graph based visualizations
- Simple question-answering tasks



Here: generalize these findings

- to other visualizations: maps and deviation charts
- with more realistic tasks
  - use MetroQuest, a commercial visualization-based decision support platform

MetroQuest

- Version to compare three transit scenarios to our campus

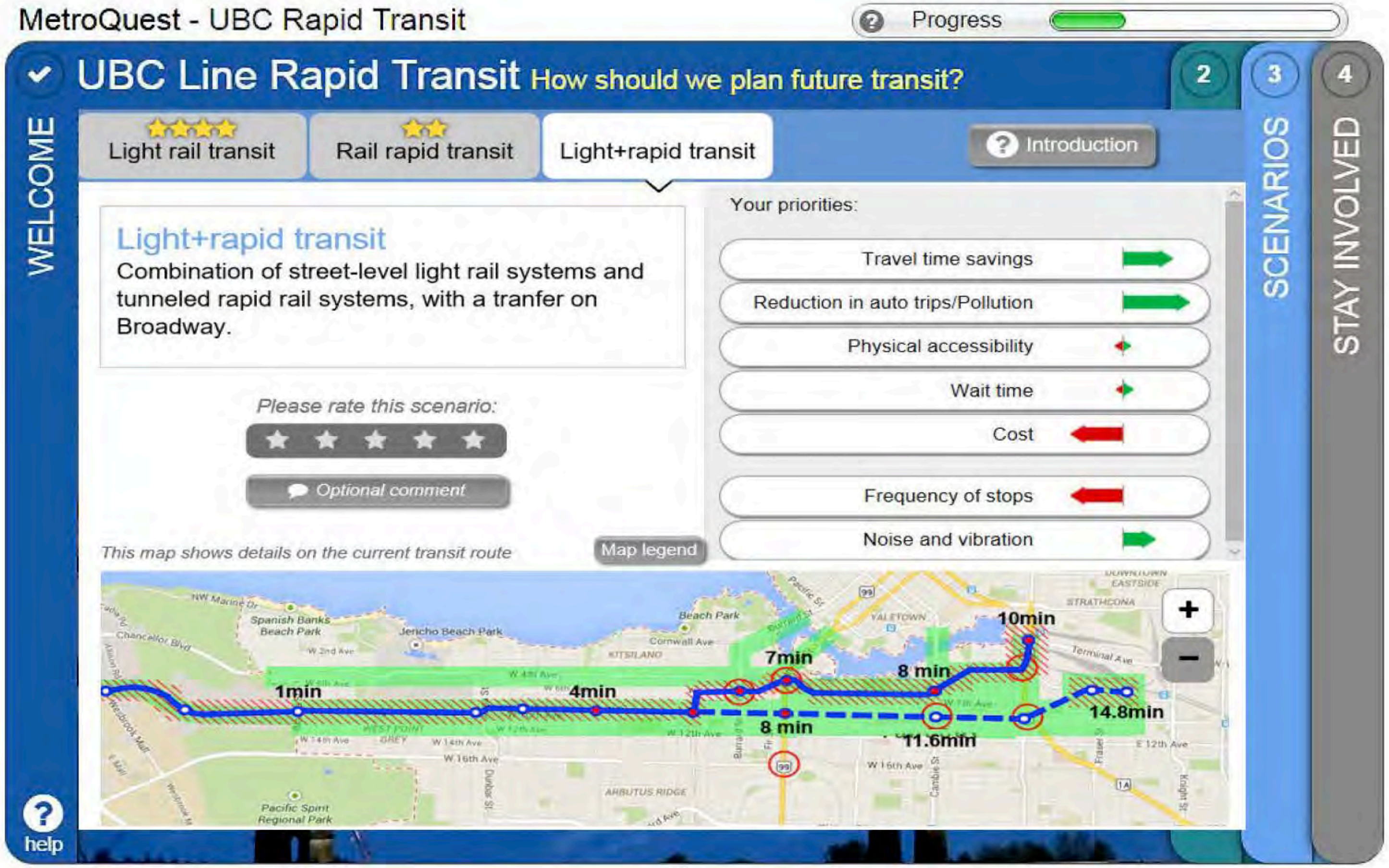
Real project of the city of Vancouver, regularly covered by the news

Used real data published by the City

Need for Personalization

Designing effective and engaging visualizations for MQ is challenging

- large diversity of users
- rich visualizations may satisfy some users, but overwhelm others.



How to Personalize?

User study (166 users) to investigate

- 10 user characteristics that may impact user performance with Metroquest visualizations
  - **WHAT TO ADAPT TO?**

- User models that use eye-tracking data to predict the relevant user characteristics
  - **WHEN TO ADAPT?**

Study details

Participants

- From the community directly affected by the transit project
  - UBC campus + residential area
- Gaze tracked with Tobii T120 eye tracker
- Completed a questionnaire on their experience with MetroQuest
  - usefulness of each visualization
  - visualization preference
  - decision confidence

Collected user characteristics

- **6 cognitive measures:**

1. Perceptual Speed (PS)
   Measure of speed when performing simple perceptual tasks.
2. Visual Working Memory (Visual WM)
   Storage and manipulation capacity of *visual* information (e.g. shape, color)
3. Verbal Working Memory (Verbal WM)
   Storage and manipulation capacity of *verbal* information.
4. Spatial Memory (SPM)
   Storage and manipulation capacity of spatial arrangement of objects
5. Visual Scanning
   Speed when locating objects in surroundings
6. Visualization literacy
   Ability to use well established Infovis.

- **2 personality traits:**

1. Locus of Control
   Tendency to see oneself as controlled by or in control of external events
2. Need for Cognition
   Tendency to like effortful mental tasks

- **2 measures of expertise:**

1. Task expertise
   How often the participant make preferential choice
2. Visualization expertise
   How often the participants to visualizations to make preferential choice

See http://www.cs.ubc.ca/cs-research/lci/research-groups/intelligent-user-interfaces/userchar.html

Impact of User Cognitive Abilities

These five significantly impacted user experience with MetroQuest [Lalle et al. 2017]

- **Perceptual Speed (PS)**

  Measure of speed when performing simple perceptual tasks.
- **Visual Working Memory (VisWM)**

  Storage and manipulation capacity of *visual* information (e.g. shape, color)
- **Verbal Working Memory (VerWM)**

  Storage and manipulation capacity of *verbal* information.
- **Spatial Memory (SpM)**

  Storage and manipulation capacity of spatial arrangement of objects
- **Visual Scanning (VisScan)**

  Speed when locating objects in surroundings

e.g., →

Low SpM users found charts less useful than high SP users

HIgh VisWM users preferred chart over map

↓

Personalized support based on these findings

See http://www.cs.ubc.ca/cs-research/lci/research-groups/intelligent-user-interfaces/userchar.html

Examples of Personalization

- For low spatial memory users:
  - Help them process the charts
    - » e.g., display fewer factors, add visual cues
  - Allow for hiding the charts

- For high Vis WM users:
  - Help them process the maps
    - e.g., show simpler maps
  - Allow for hiding the maps

Not always possible to test users for these user characteristics

↓

Can they be predicted during interaction?

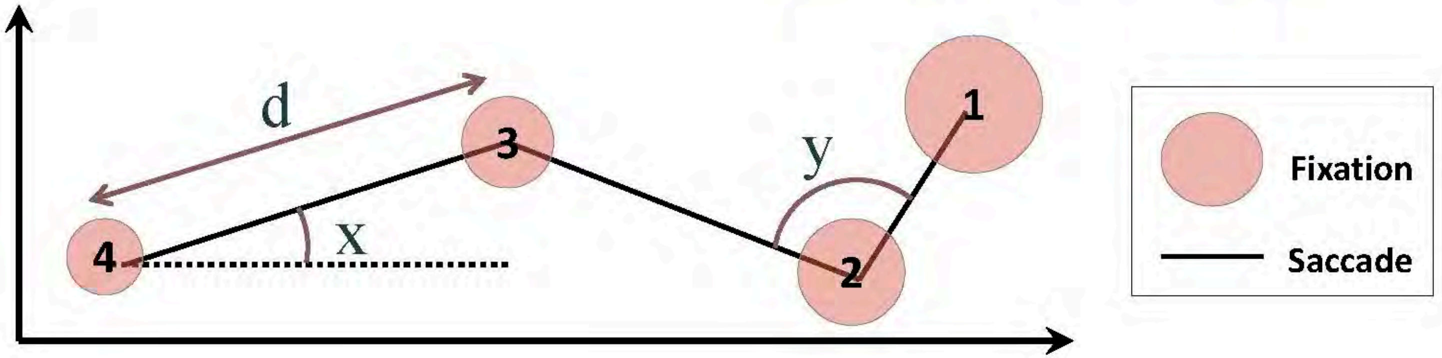
Predicting User Abilities from Eye-tracking Data

[Conati et al. IJCAI 2017]

Use machine learning to build on-line classifiers

- For binary classification of PerSpeed, VisWM, VerWM, SpM, VisScan
  - Predict high / low levels of these abilities (median split)
- From eye-tracking data:
  - gaze, pupil, and head distance information

Sample Gaze Features



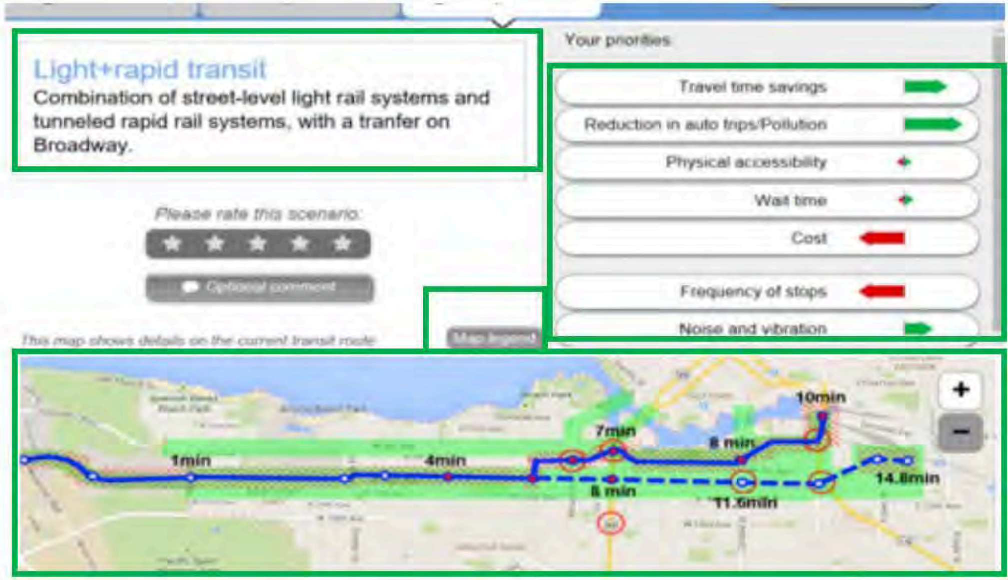
General Measures

- Number of Fixations
- Fixation rate
- Fixation Duration
- Saccade Length (d)
- Relative Saccade Angles (y)
- ....

Measures specific to Areas of Interest (AOI)

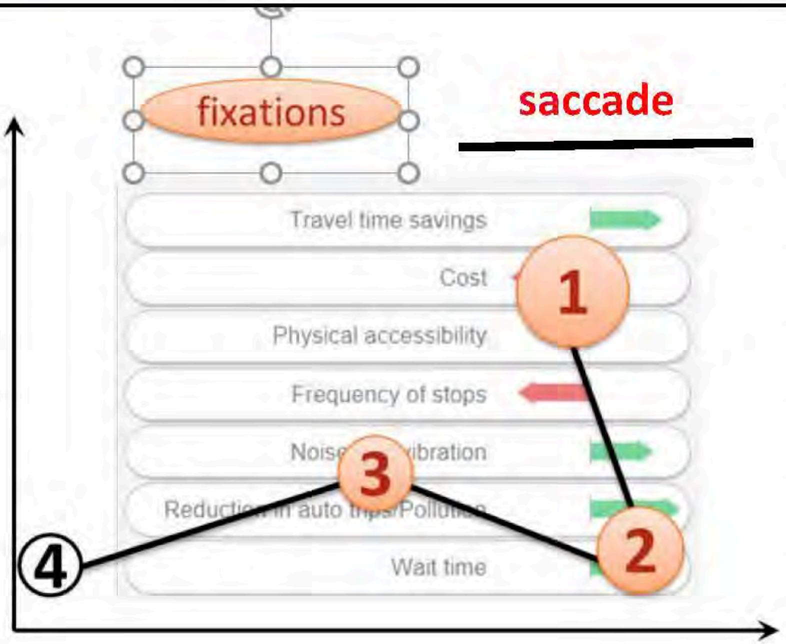
- Proportional number of fixations
- Proportional fixation time
- Time to first fixation
- Transitions between AOIs

....



Features based on sum, mean and st. dev.

Features for Classification



Gaze Features (68)

Overall Gaze Features (12):

- Fixation rate
- Mean & Std. deviation of fixation durations
- Mean & Std. deviation of saccade length
- Mean, Rate & Std. deviation of relative saccade angles
- Mean, Rate & Std. deviation of absolute saccade angles
- Mean saccade velocity

AOI Gaze Features for each AOI (56):

- Fixation rate in AOI
- Longest fixation in AOI, Time to first & last fixation in AOI
- Proportion of time, Proportion of fixations in AOI
- Number & Prop. of transitions from this AOI to every AOI

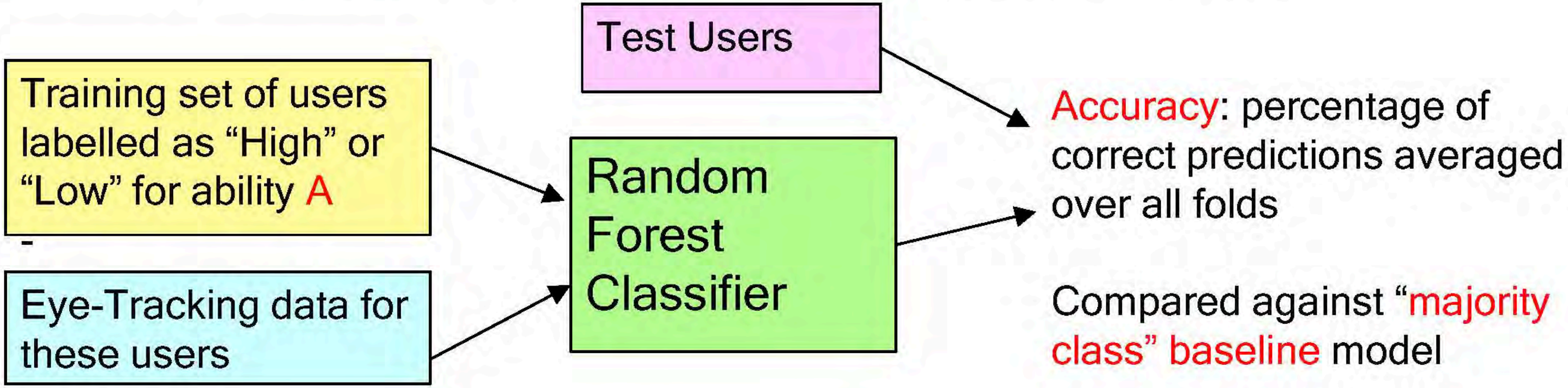
Pupil Features (6) and **Head Distance Features (6)**

- Mean, Std. deviation, Max., Min. of pupil width/head distance
- Pupil width/head distance at the *first* and *last fixation* in the data window

Model Training and Evaluation

For each cognitive ability A, trained a Random Forest classifier:

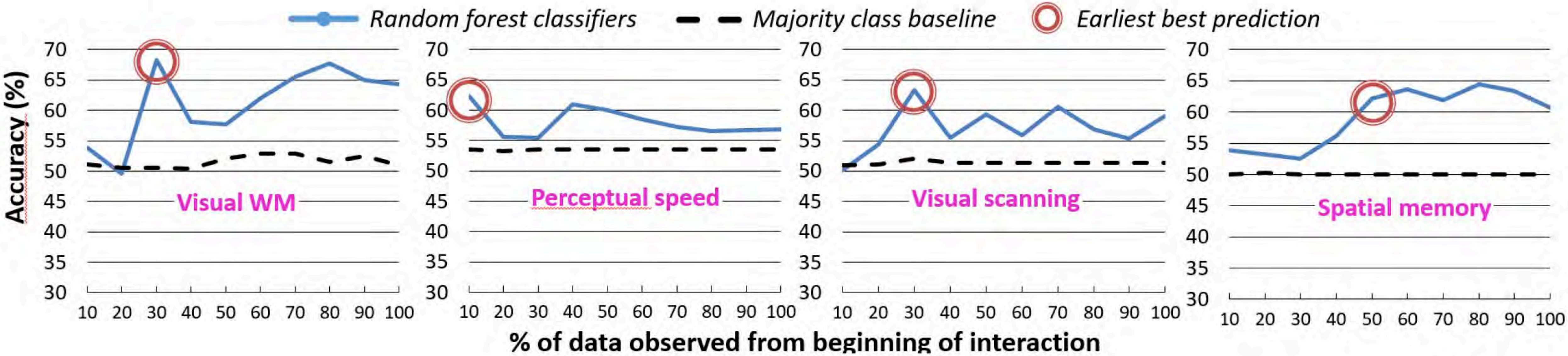
- 10-fold-cross-validation over users
  - Train model on a subset of users in the dataset, and test on rest
  - Repeat 10 times on different partitions (folds) of training and testing users



- Process repeated for increasing percentages of data
  - Gives insights on how accuracy evolves overtime during interaction

Results Summary (1)

- Random Forest outperforms baseline in 4 out of the 5 cognitive abilities



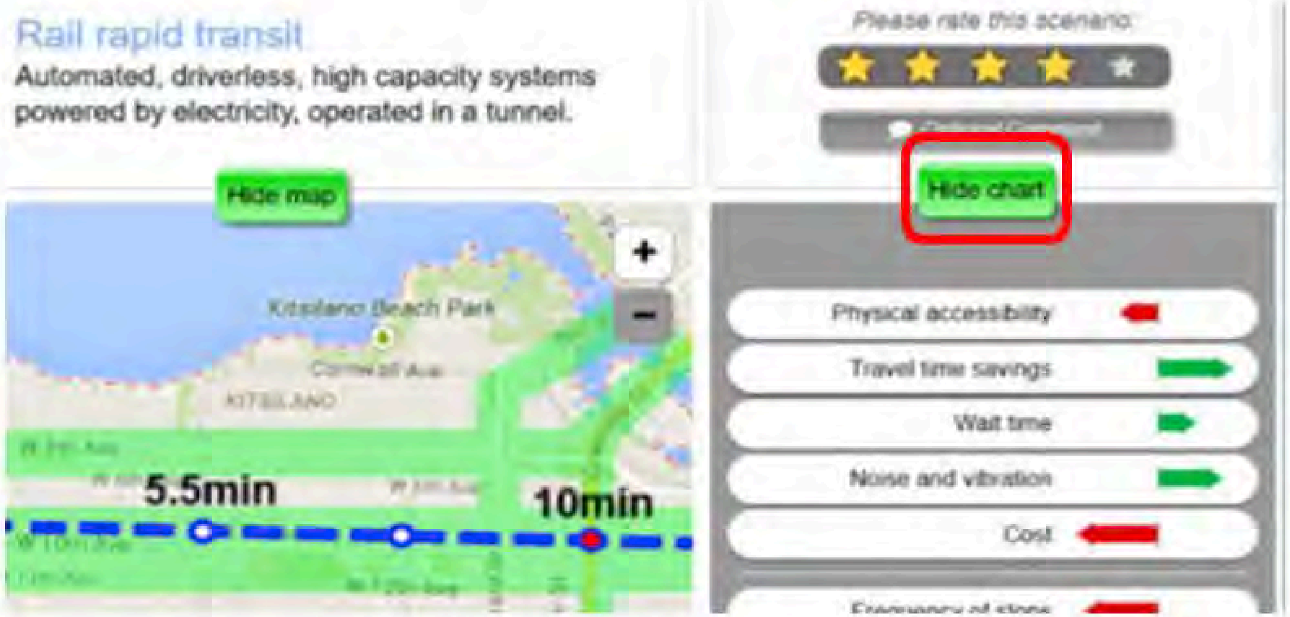
- Best accuracy achieved after observing the first 10% to 50% of interaction data
    - Important for providing prompt adaptation

What's next?

- Showed that cognitive abilities relevant for visualization processing can be predicted from eye-tracking data
  - This helps with the question of when to provide adaptation

  ⬇

  - look for lower levels of the relevant cognitive abilities and provide suitable interventions



What's next?

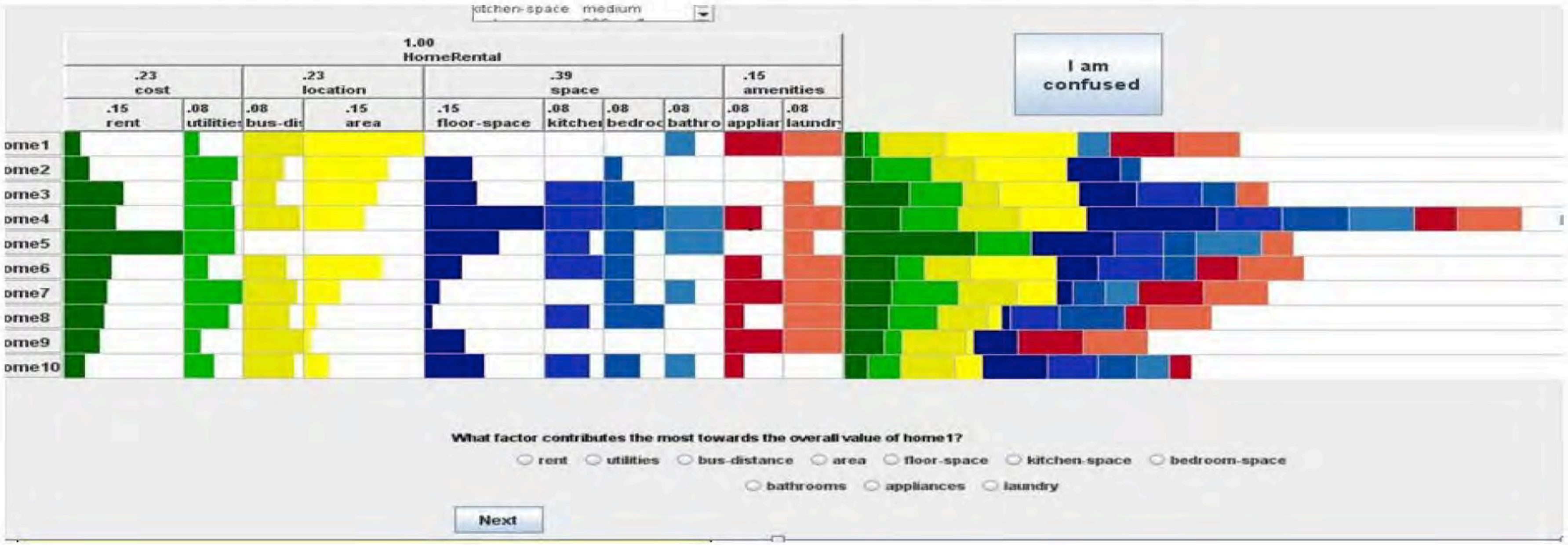
Another trigger for when to adapt could be detecting user confusion during visualization processing

In This Talk

- Example applications of ET in user modeling for
  - User-Adaptive Visualizations
    - relevant cognitive abilities
    - confusion
  - Intelligent Tutoring Systems (ITS)

When to provide adaptive interventions?

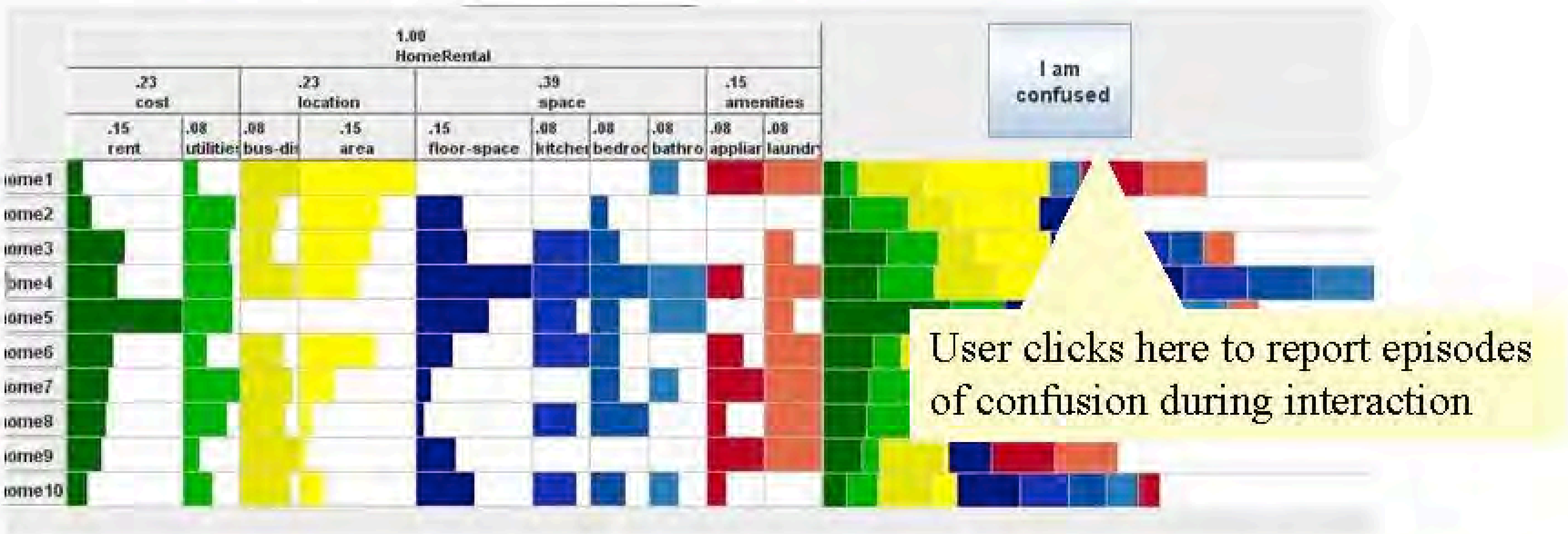
- Investigated predictors of user's confusion while using Value Chart: interactive visualization for decision making (Lalle, Conati et al, IJCAI 2016)



ValueCharts

Collecting episodes of confusion

- To train confusion classifiers, we need a dataset with episodes of user confusion
- But collecting labels for transient user states is challenging
  - We collected user self-reports for confusion via a non-intrusive button



- Approach previously used to collect labels for other user affective states
  - E.g., motivation, emotions (deVincente & Pain 1999, Conati & MacLaren 2009)

User Study

10’ practice with Value Chart

Instructions on reporting confusion

→

Battery of 5 types of tasks, of varying complexity

8 tasks per type, 40 tasks per user in total

- Participant gaze tracked with a Toby T-120 eye-tracker

User Study

10' practice with Value Chart

Instructions on reporting confusion

Battery of 5 types of tasks, of varying complexity

8 tasks per type, 40 tasks per user in total

"[you should click the confusion button] if you feel that you want to ask the experimenter a question about something; if you are confused about the interface; if you are confused about the wording of a question. ...These are just a few examples, to show that confusion can occur in many unforeseeable ways, [which are all] OK reasons to click the confusion button."

User Study

10' practice with Value Chart

Instructions on reporting confusion

→

Battery of 5 types of tasks, of varying complexity

8 tasks per type, 40 tasks per user in total

→

Confusion episodes replayed to participants and discussed with experimenter

- Participant gaze tracked with a Toby T-120 eye-tracker

Data

- 136 users, 5440 tasks
  - 80 users (59%) reported confusion at least once (avg = 1.4, stdev=1.9)
  - never more than one click per task
- 112 clicks on the confusion button
  - Confusion in ~2% of the tasks

Classification labels

"Confusion" task	**"No-confusion" task**
Confusion reported during task	Confusion not reported during task

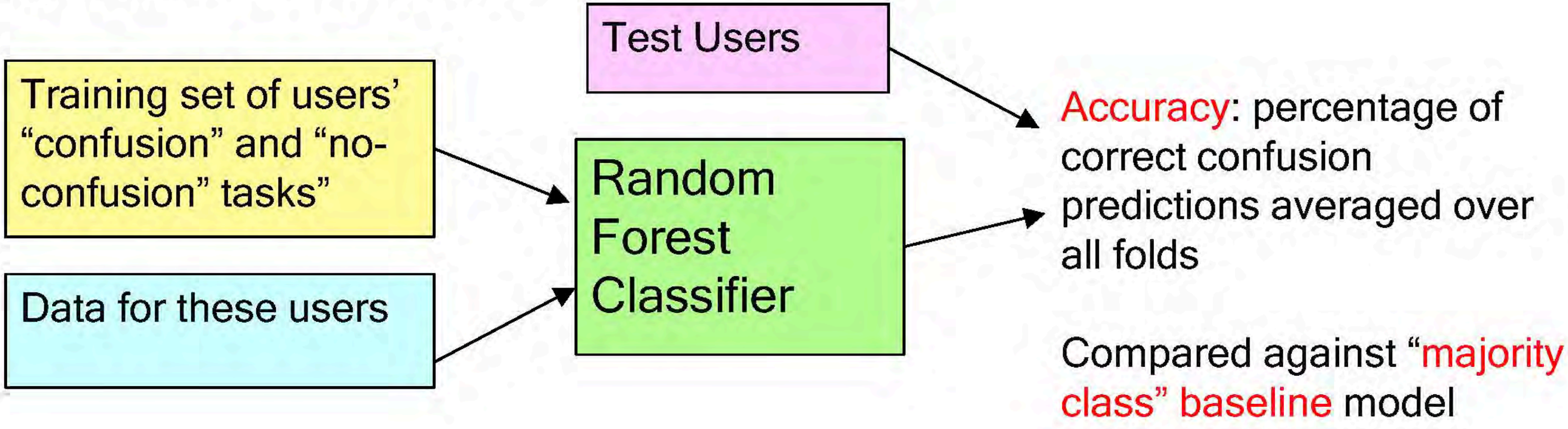
Features sets for classification

- **4 individual feature sets:**
  - Gaze 149 features
  - Pupil 6 features
  - Head Distance: 6 features
  - Mouse Events: 32 features

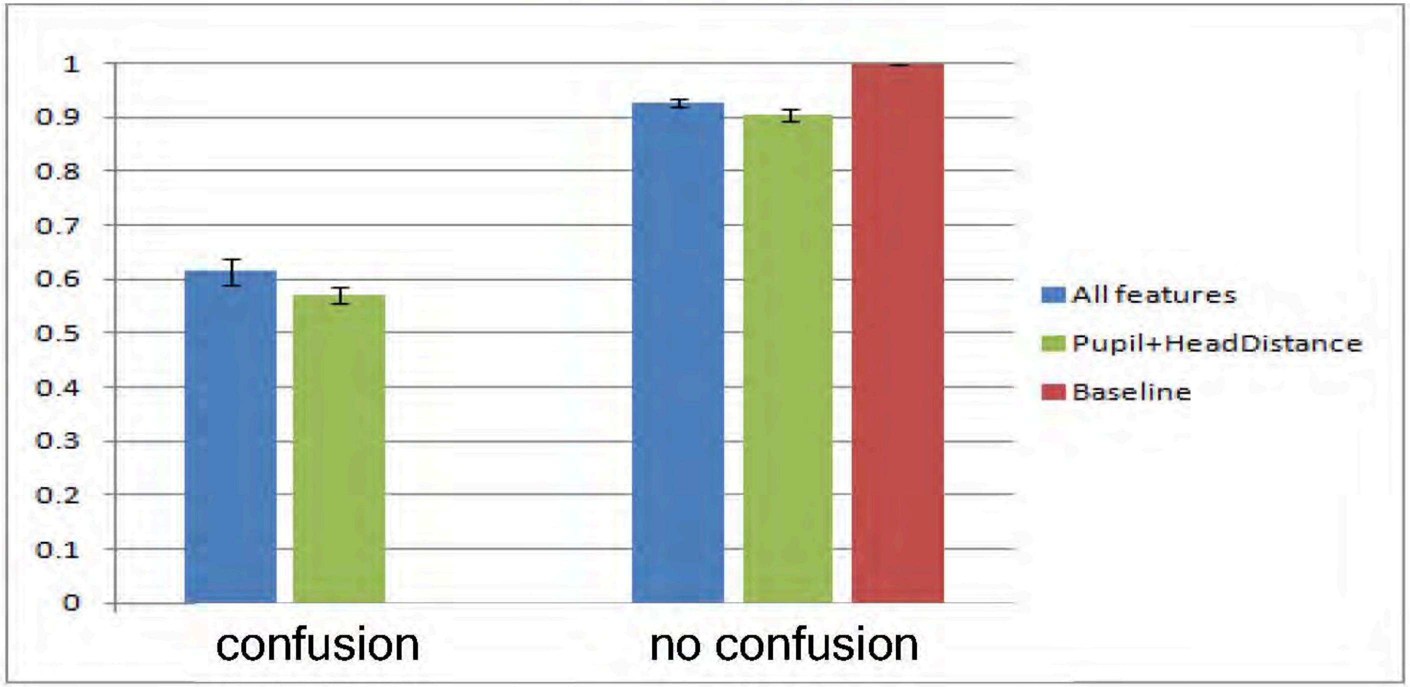
- **Four combined sets:**
  - Gaze + Pupil + Head (eye-tracking only)
  - Head Distance + Mouse (no eye-tracking needed)
  - Head Distance + Pupil (vis independent)
  - All features

Classification Experiments

- Classify tasks as “confusion” or “no-confusion”
  - Random Forest trained with 10-folds cross-validation over users
  - For all 4 feature sets
  - For incremental percentages of data
- Baseline: majority class
  - Always predicts *“no-confusion”*



Results



- Best classifier uses **all features**
  - 61% instances of confusion correctly predicted
  - 92% accuracy for no confusion
    - » Very few false positives - limited intrusiveness
- Pupil + Head Distance second best (layout independent)

Lots of Exciting Research Questions

- How to intervene to reduce confusion?
- Can we capture the reasons for confusion?
- What other cues/states should be consider to decide when to adapt?

In This Talk

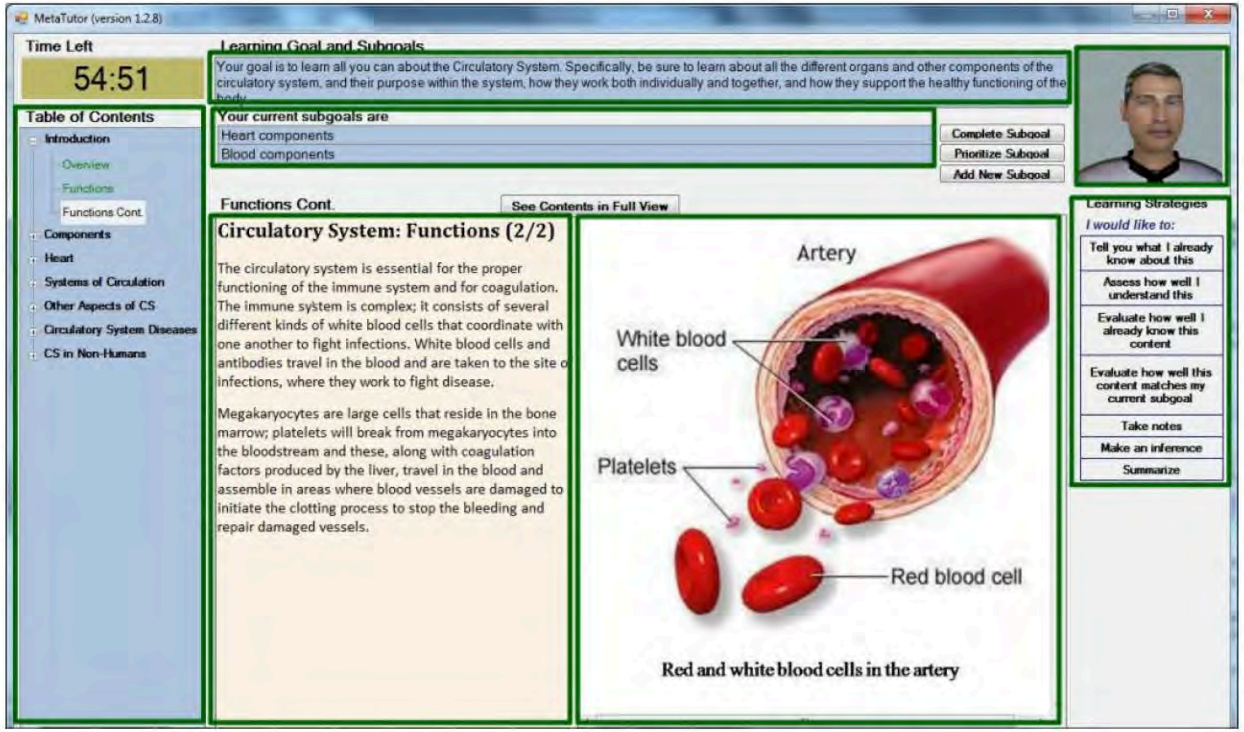
- Example applications of ET in user modeling for
  - User-Adaptive Visualizations
    - relevant cognitive abilities
    - confusion
  - Intelligent Tutoring Systems (ITS)

Eye-Tracking for ITS

- Evidence that eye-tracking can help predict
  - Learning (e.g. Kardan and Conati UMAP 2013)
  - Mind Wandering (e.g. Bixter and D' Mello 2015)
  - Boredom (e.g. Jaques and Conati, 2013)

Here

- Predict affect and achievement goals with Meta-Tutor



- ITS with pedagogical agents (PA) that prompt and provide feedback on good learning strategies in science *[Azevedo et. al, 2012, 2018]*

Achievement Goals

- **Achievement goals** *[Elliot & Murayama 2008]*
  - Mastery-oriented ⟶ Focus on developing competence and skills
  - Performance-oriented ⟶ Focus on performing better than their peers
- A facet of motivation: affect several aspects of academic performance
- Measured via a standard questionnaire, AG-R, *[Elliot & Murayama 2008]*
- Students achievement goals collected during a study with Meta-Tutor *[Azevedo et al 2013]*

User Study

62 university students

- Study tasks with MetaTutor (~145 min):
  - PF group with full agents' prompts and feedback
  - Control group without
- Gaze tracked (SMI Red 250)
- Periodic self-reports of emotions (~14')

Achievement Goals and MetaTutor

- Mastery-oriented students reported more negative affect with PAs interventions than performance-oriented students *[Lalle et al. IVA 2016]*

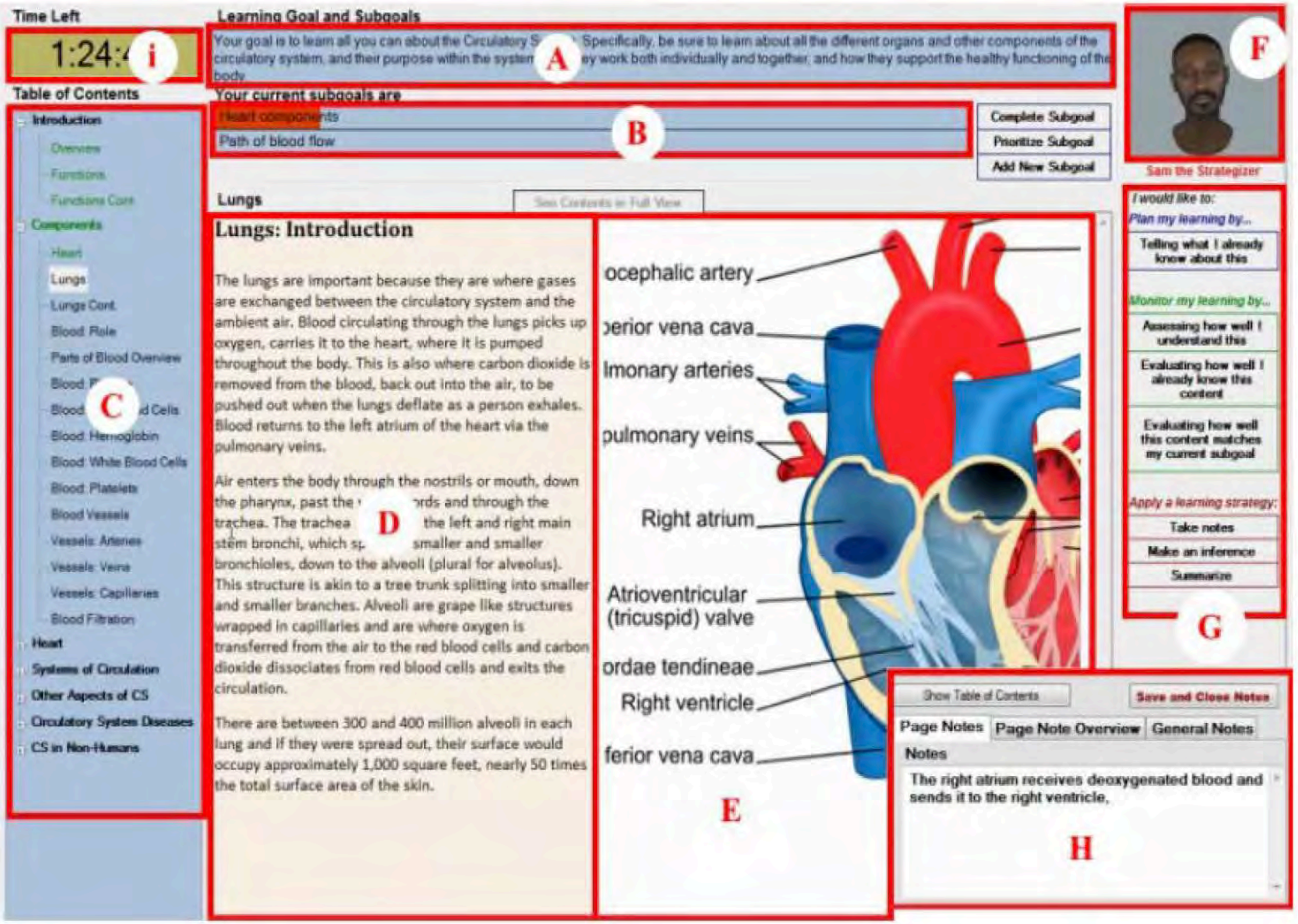
- They also learned less *[Duffy and Azevedo 2015]*

Can we detect in real time if a student is
- Mastery-oriented, and
- experiencing negative affect

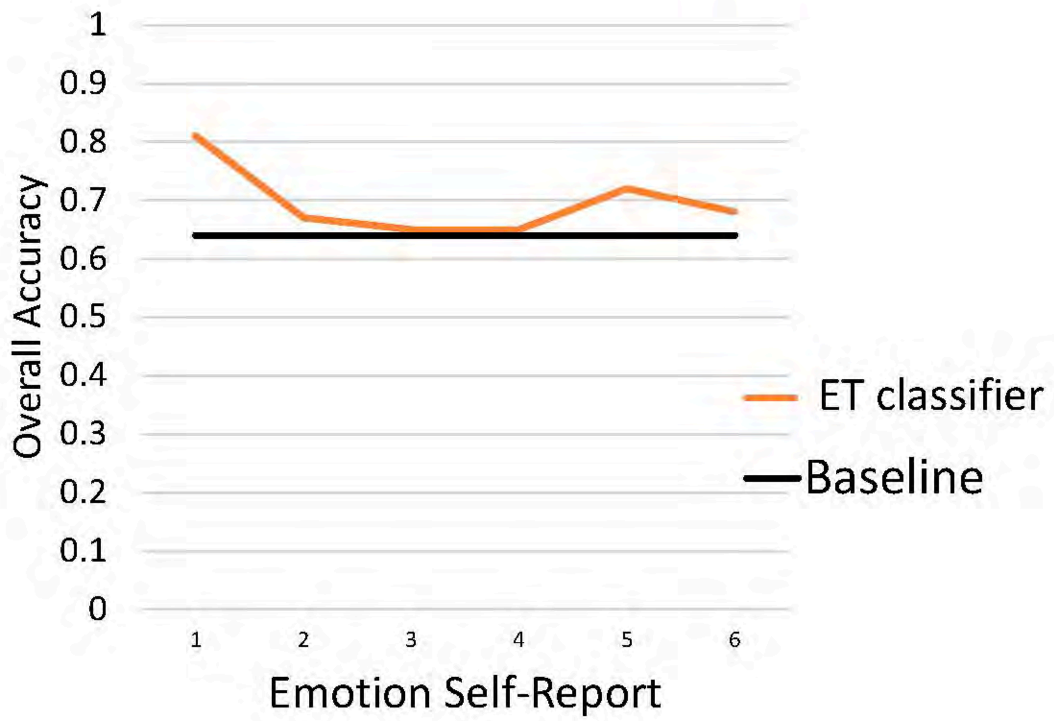
so that PAs can adjust their prompt's and feedback accordingly?

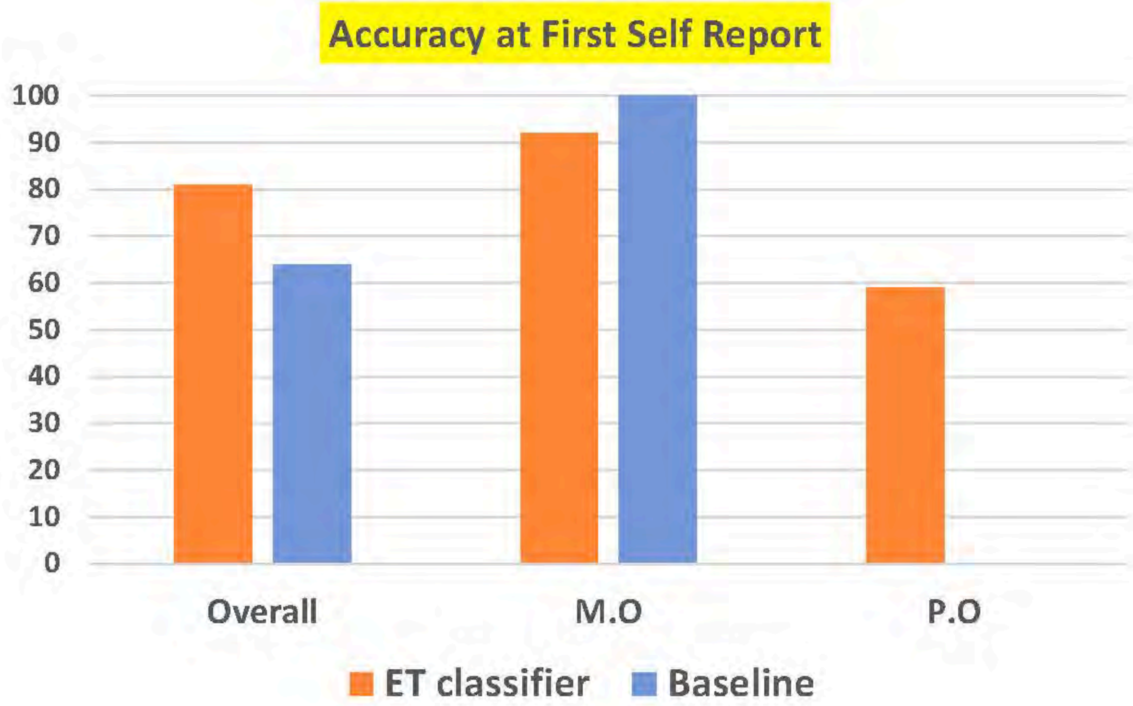
Classification

- Gaze and Head features similar to previous studies
  - Gaze tracked over 9 AOIs



Results: Achievement Goals

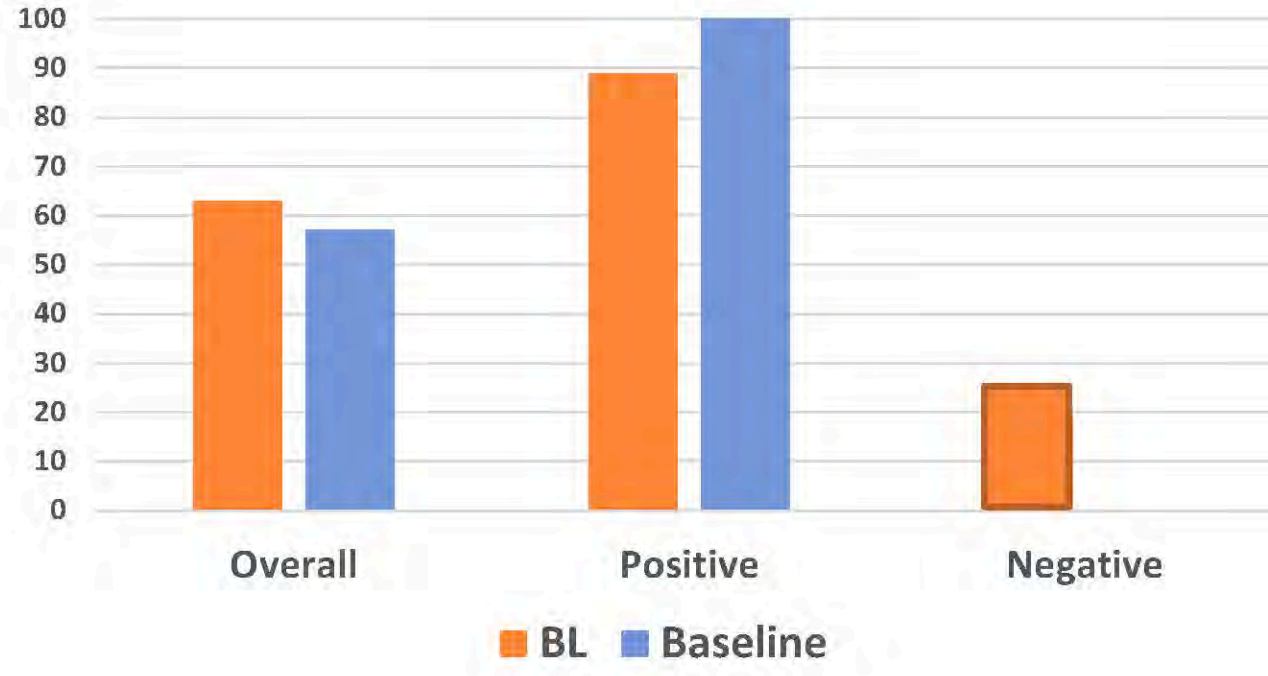




- ET classifier beats majority class baseline (predicts M.O.)
- Best accuracy at first self-report (about 14’ of interaction)
  - 92% accuracy on mastery-oriented
  - 59% accuracy on performance-oriented

Results: Valence

- ET classifier beats baseline (predicts positive)
- 88% accuracy on positive reports
  - Only 12% of positive self-reports misclassified: limited intrusiveness
- Still able to capture 26% of negative self-report
  - The rest are missed opportunities to rectify negative affect
  - but would not otherwise interfere with the interaction.

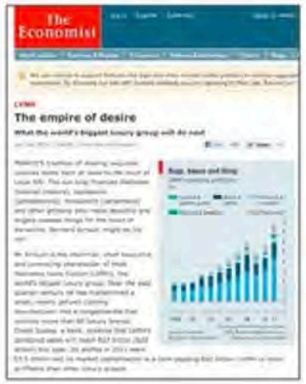


Current Work

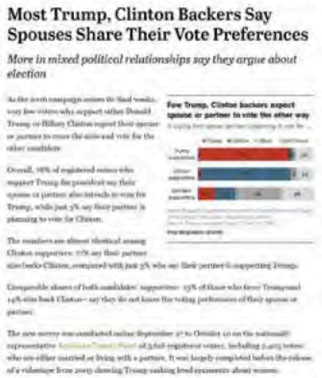
- Improve classification accuracy
  - ensemble models, include other sensor data (e.g., EDA)
- Adaptive interventions for Mastery students
  - Make them more relevant
  - Reduce frequency
- Predict personality traits also found to influence affect *[Lalle' et al IVA 2016]*
  - E.g. Neuroticism

Conclusions

- Eye Tracking: great potential for user modeling and user-adaptive interaction
  - Helps predict several user properties/states in very different contexts
- Many more to be explored
  - e.g. support for processing multimodal information (text , graphics, videos)


The Economist

The empire of desire


Most Trump, Clinton Backers Say Spouses Share Their Vote Preferences

More in mixed political relationships say they argue about election

- Exciting new possibilities with mobile and more affordable high end ET!

Thanks To

- Sebastien Lalle
- Abed Rahman
- Dereck Toker
- Ben Steichen
- Matthew Gingerich
- Enamul Hoque
- Giuseppe Carenini
- Roger Azevedo

And to all of you for your kind attention !