From

“*Combining Unsupervised and Supervised Classification to Build User Models for Exploratory Learning Environments*”*

To FUMA:

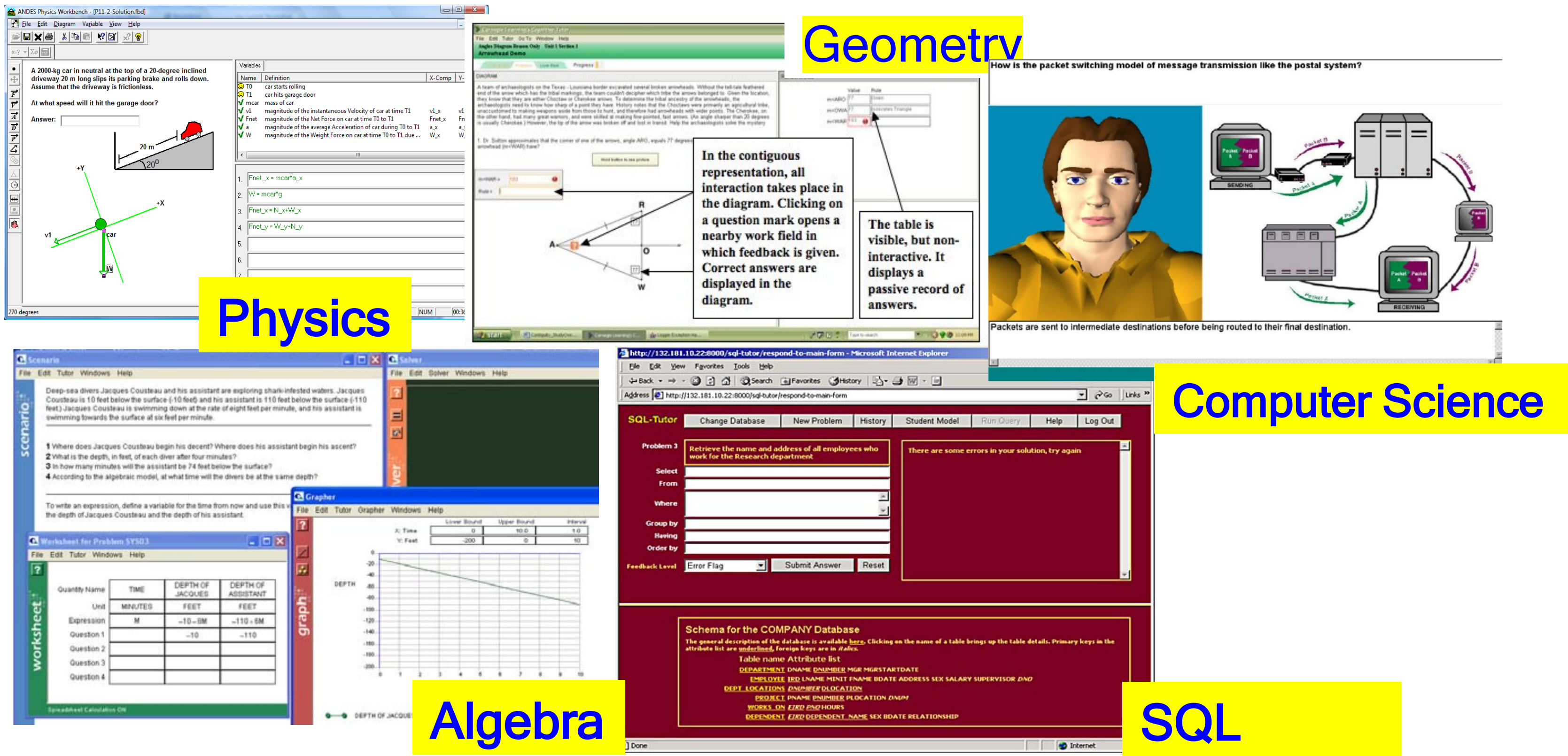
Framework for User Modeling and Adaptation

Cristina Conati
Department of Computer Science
University of British Columbia

*** With Saleema Amershi, now Microsoft Research**

Great Progress on AI-driven Support to Problem Solving

[DuBulay, Mitrovic, Yacef; Handbook of AI in Education 2023]

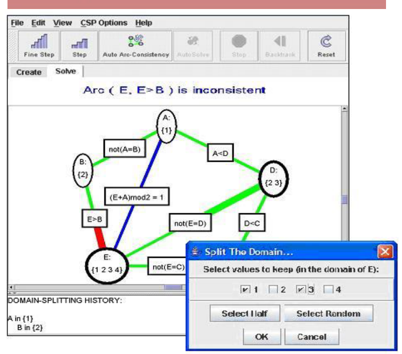


- Well defined problem solutions => guidance on problem solving steps
- Clear definition of correctness => basis for feedback

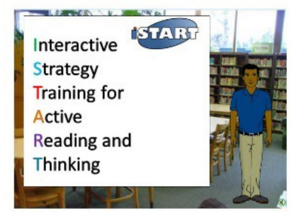
Beyond Problem Solving

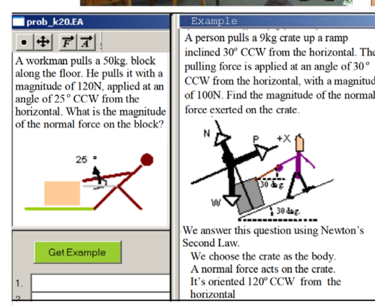
- Exploratory Learning Environments (ELEs) that support active learning via student-driven exploration

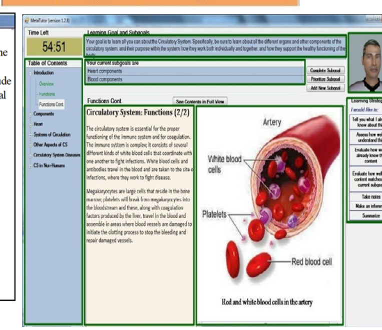
Educational Games and Simulations



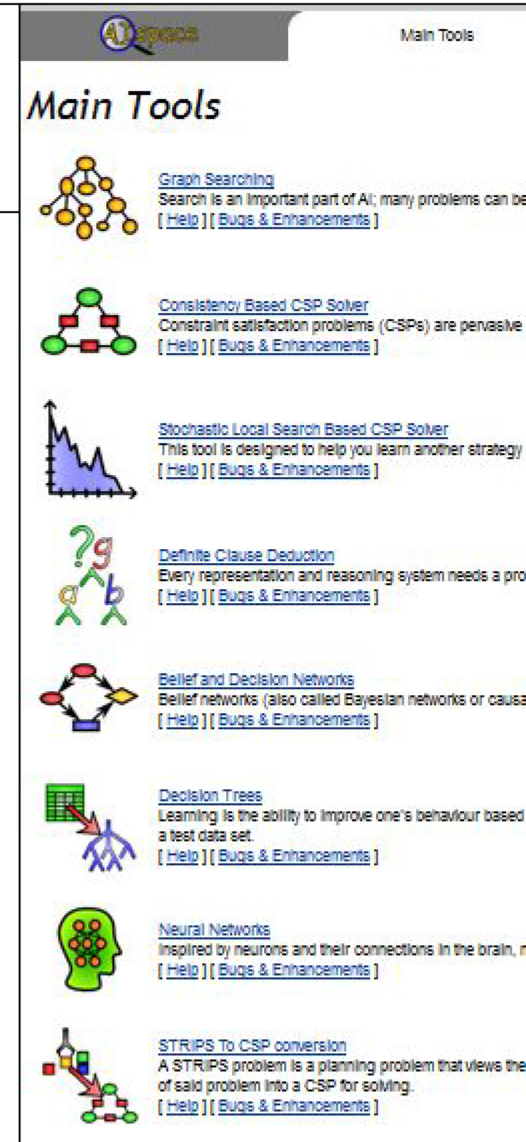
Meta-Cognitive Tutors







An Example

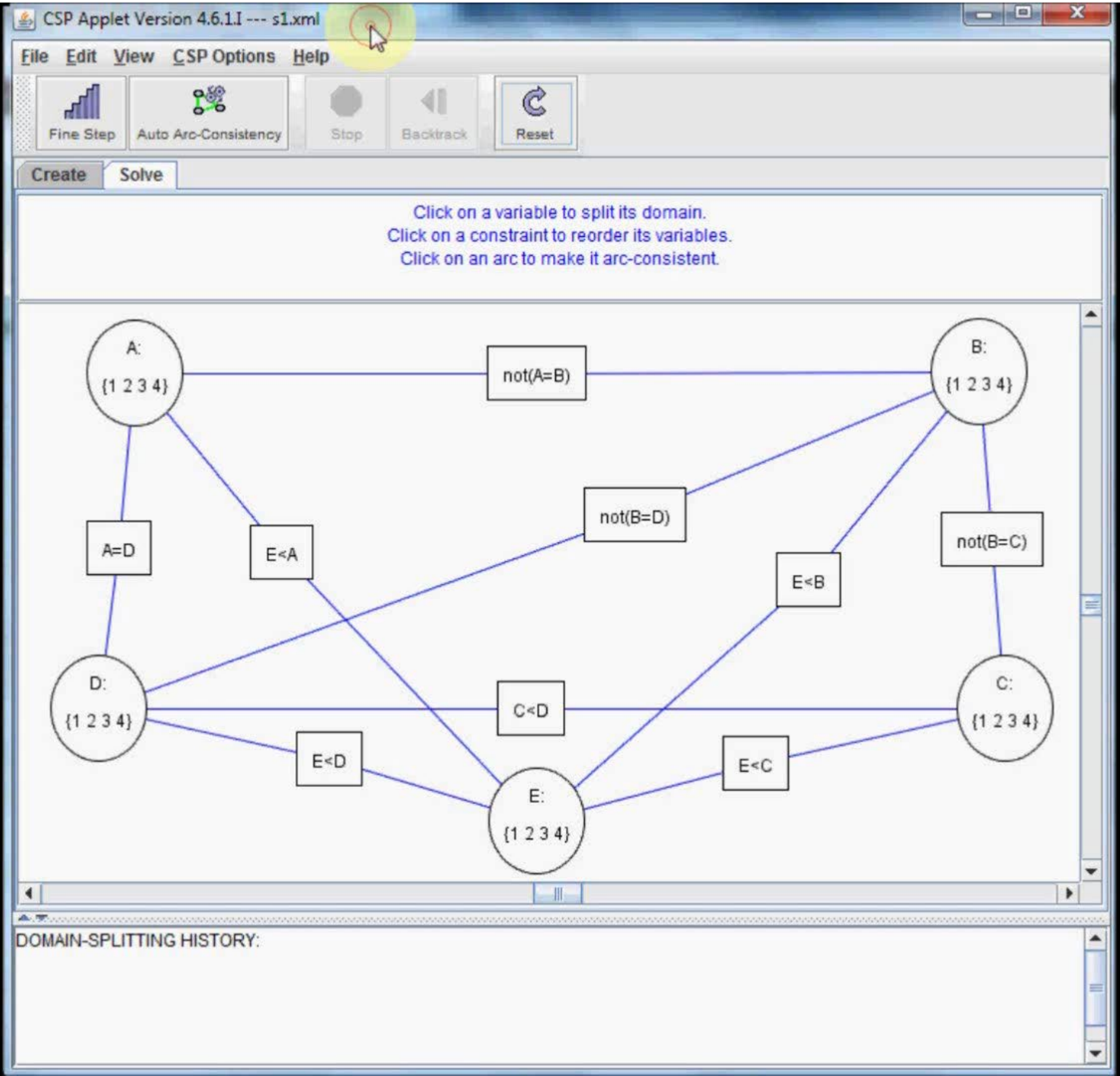


Tools for Learning Artificial Intelligence

AIspace

- AISpace (Amershi et al., 2007)
  - Suite of interactive simulations of common Artificial Intelligence algorithms
  - Used regularly in our AI courses
  - Google “AISpace” if you want to try it out
- CSP (Constraint Satisfaction Problems) Applet
  - visualizes the working of the AC3 algorithm

The ACSP applet



AI-driven Support in ELE

- Not all students learn well from exploratory activities [e.g.,Van Joolingen et al., 2007]
  - Important to provide support for those students who need help.
  - While maintaining student initiative and engagement

- Challenge: No clear definition of correct/effective behaviors

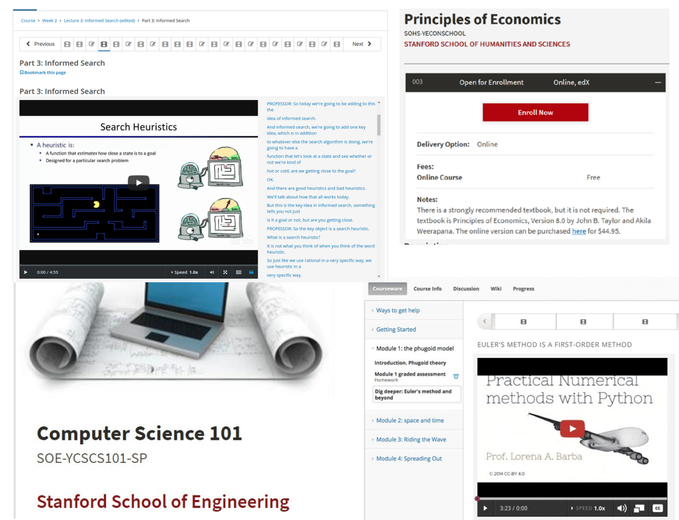
what behaviors should drive personalized support?

how to provide such support effectively and unobtrusively?

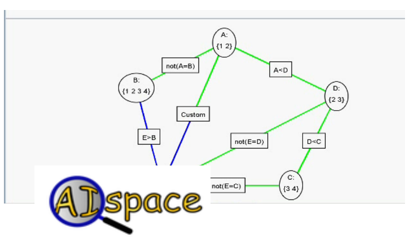
FUMA for Data-Driven Personalization

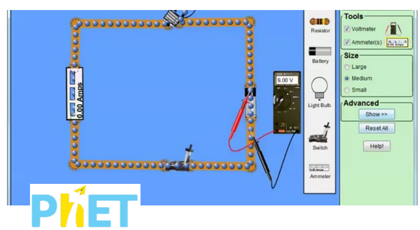
- FUMA (Framework for User Modeling and Adaptation)
  - First version proposed by Amershi and Conati 2009 (ToT Award 2022)
  - Learn from data what user behaviors should trigger personalized help
  - Recognize and react to these behaviors in real-time during interaction
- Evaluated in several ELEs

Four MOOCs [Lallé et al., AIED 2020]



Two Interactive Simulations Kardan Conati UMAP 2015 Fratamico et el JAIED 2017]





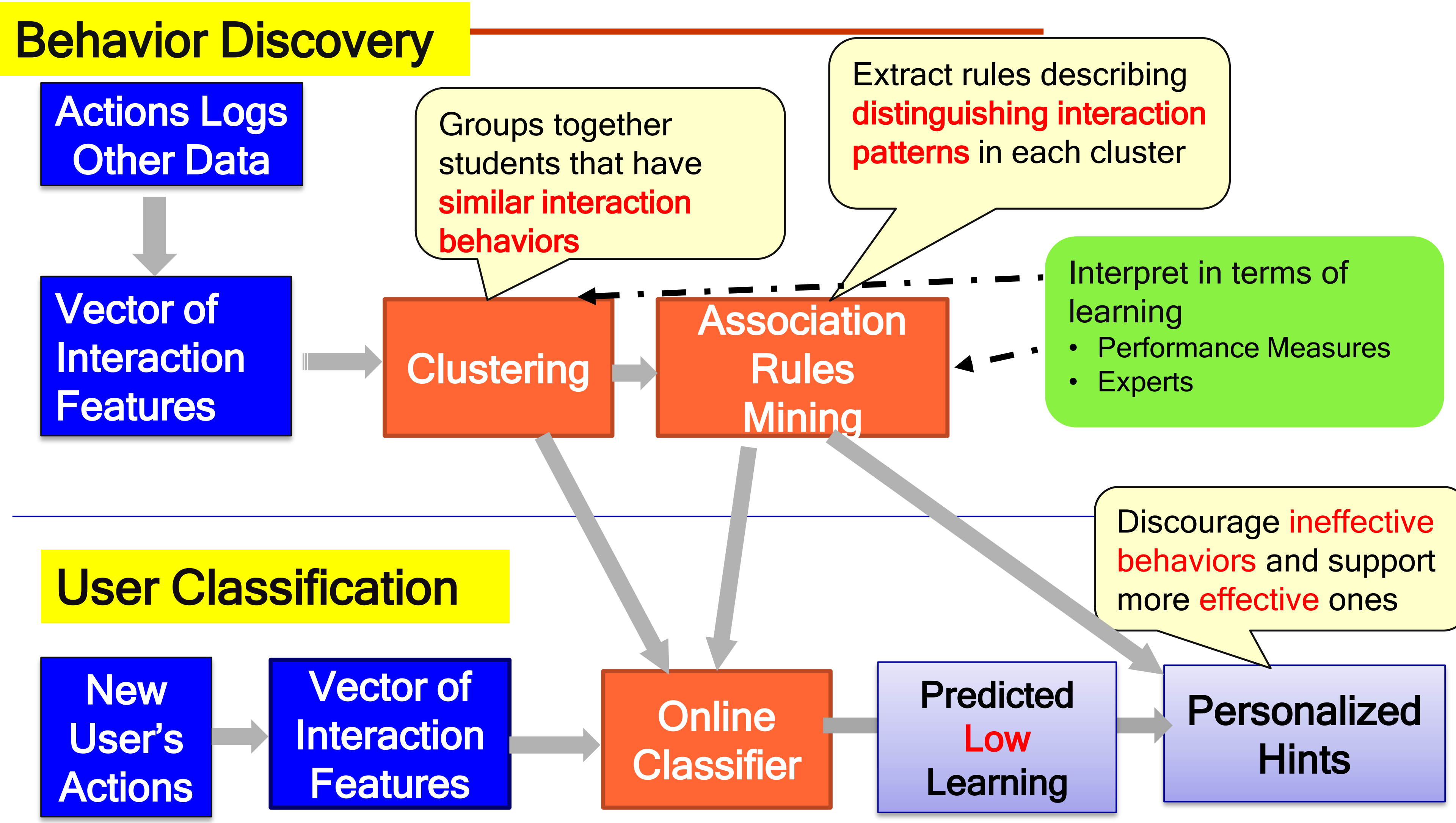
Environment for Game Design
[Lallé et al., LAK 2021, AIED 2023, Yalcin et al TiiS 2022]

Overview

- Overview of FUMA and initial results with the CSP applet
- Extension to other data and environments
  - Challenges and lessons learned
- What's next?

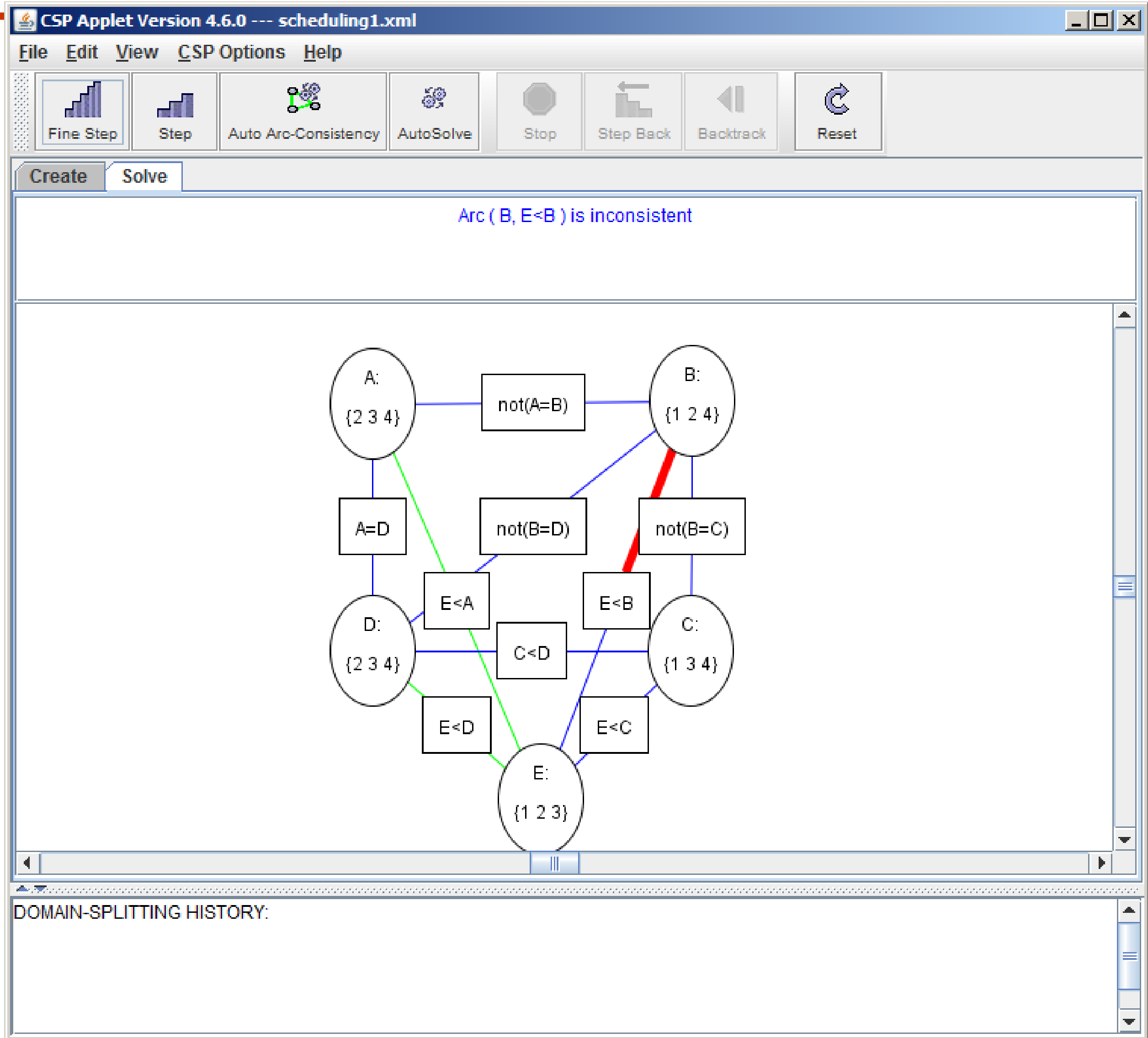
FUMA

[Amershi and Conati 2009, Kardan and Conati 2011, 2015]



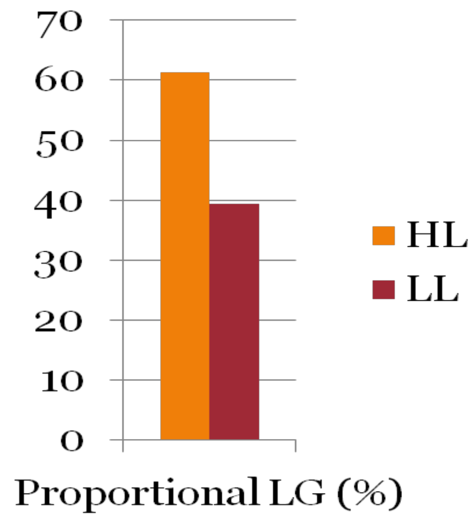
Test Bed - CSP Applet

[Amershi and Conati 2009, Kardan and Conati 2011, 2015]



Behavior Discovery

- Dataset:
  - 64 subjects, 13,000+ actions, 17+ hour
- 7 types of actions → 21 features
  - Action frequency
  - Time between actions (Mean and SD)
- Found two clusters with different learning
  - lower learning (LL) and higher learning (HL)



- Sample Rules

HL members:

- Use *Direct Arc Click* action **frequently** (R1).

LL members:

- Use *Direct Arc Click* **sparsely** (R3)
- Leave **little time** between a *Direct Arc Click* and the next action (R2)

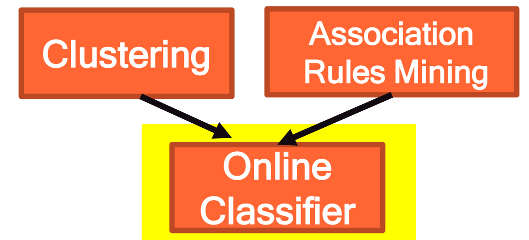
From Behavior Patterns to Hints

Intervention Code	Intervention Description
DAC_fr	Using Direct Arc Click more often
DAC_PA	Spending more time after performing Direct Arc Clicks
Reset_fr	Using Reset less frequently
AAC_fr	Using Auto Arc-consistency less frequently
DS_fr	Using Domain Splitting less frequently (only when appropriate)
FS_PA	Spending more time after performing Fine Steps
BT_fr	Using Back Track less frequently (only when appropriate)
FS_fr	Using Fine Step less frequently
Reset_PA	Spending more time after performing after resetting for planning

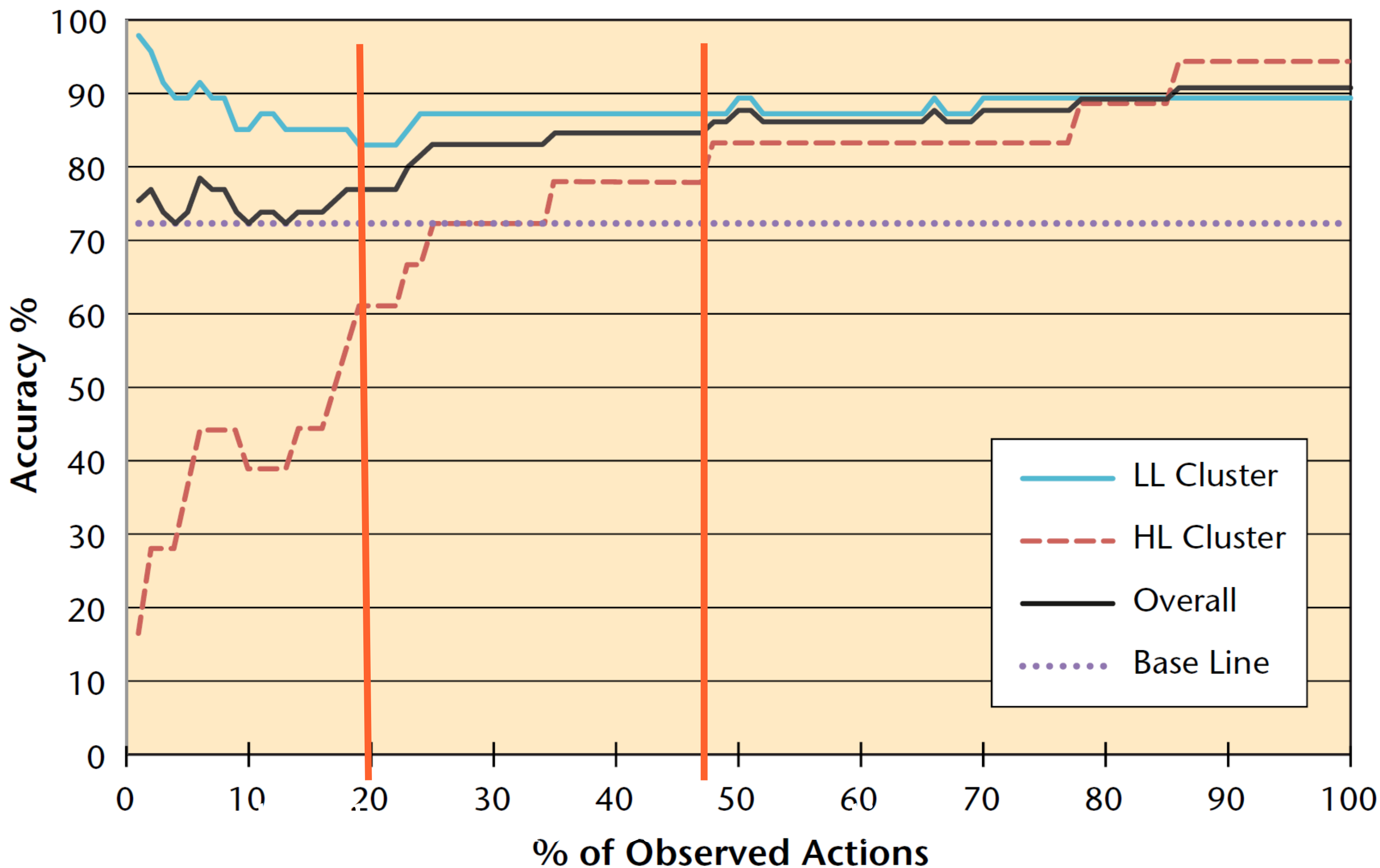
Table 2. Description of hints

Classifier Evaluation on CSP Applet

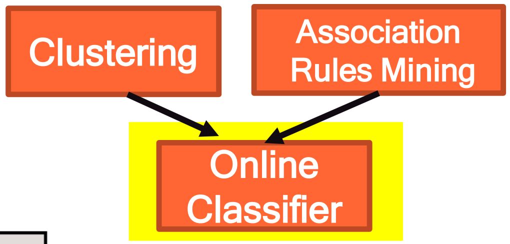
[Kardan and Conati 2012]

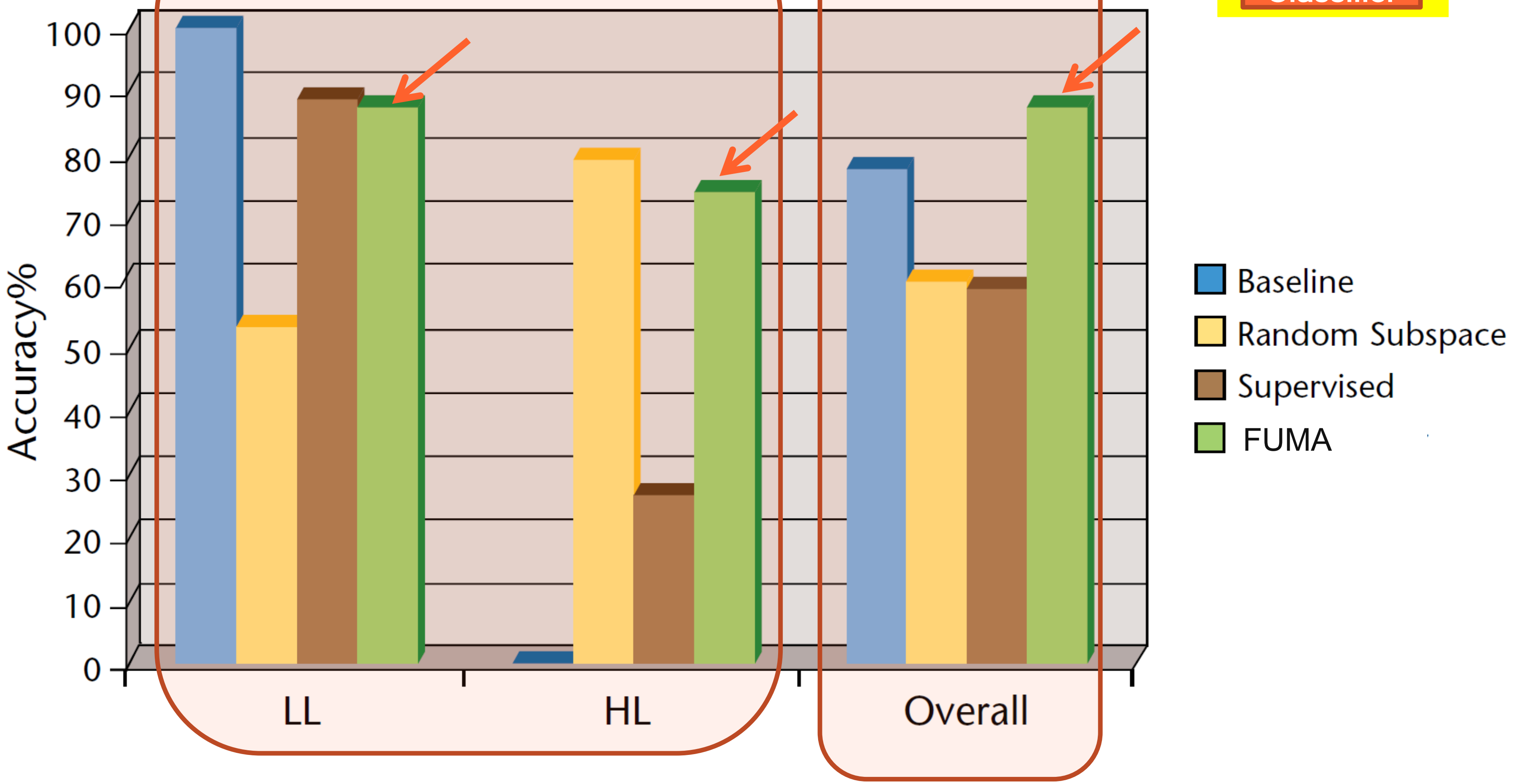


Accuracy as a function of observed actions



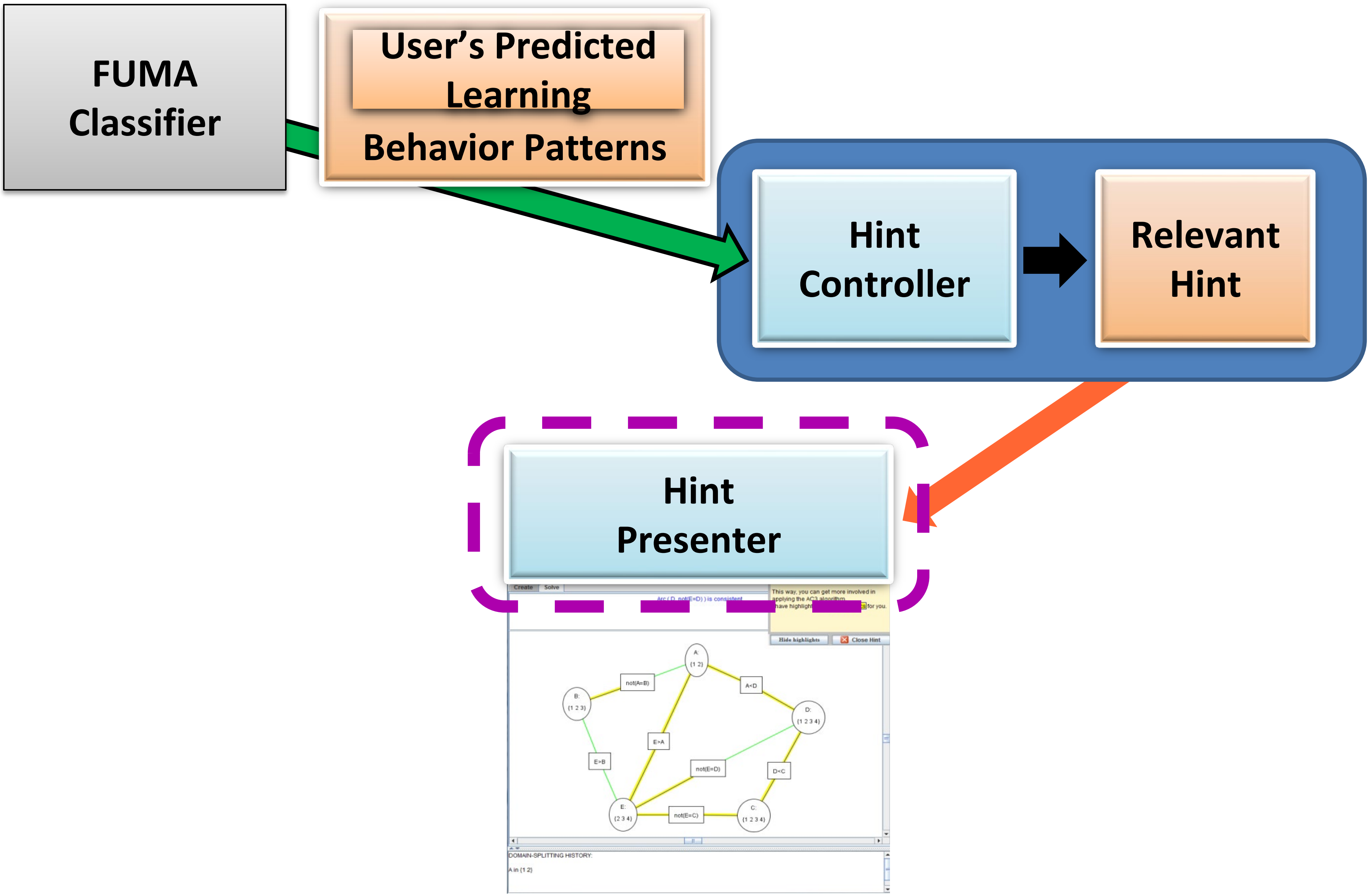
Average Overtime Accuracy





Providing Personalized Support

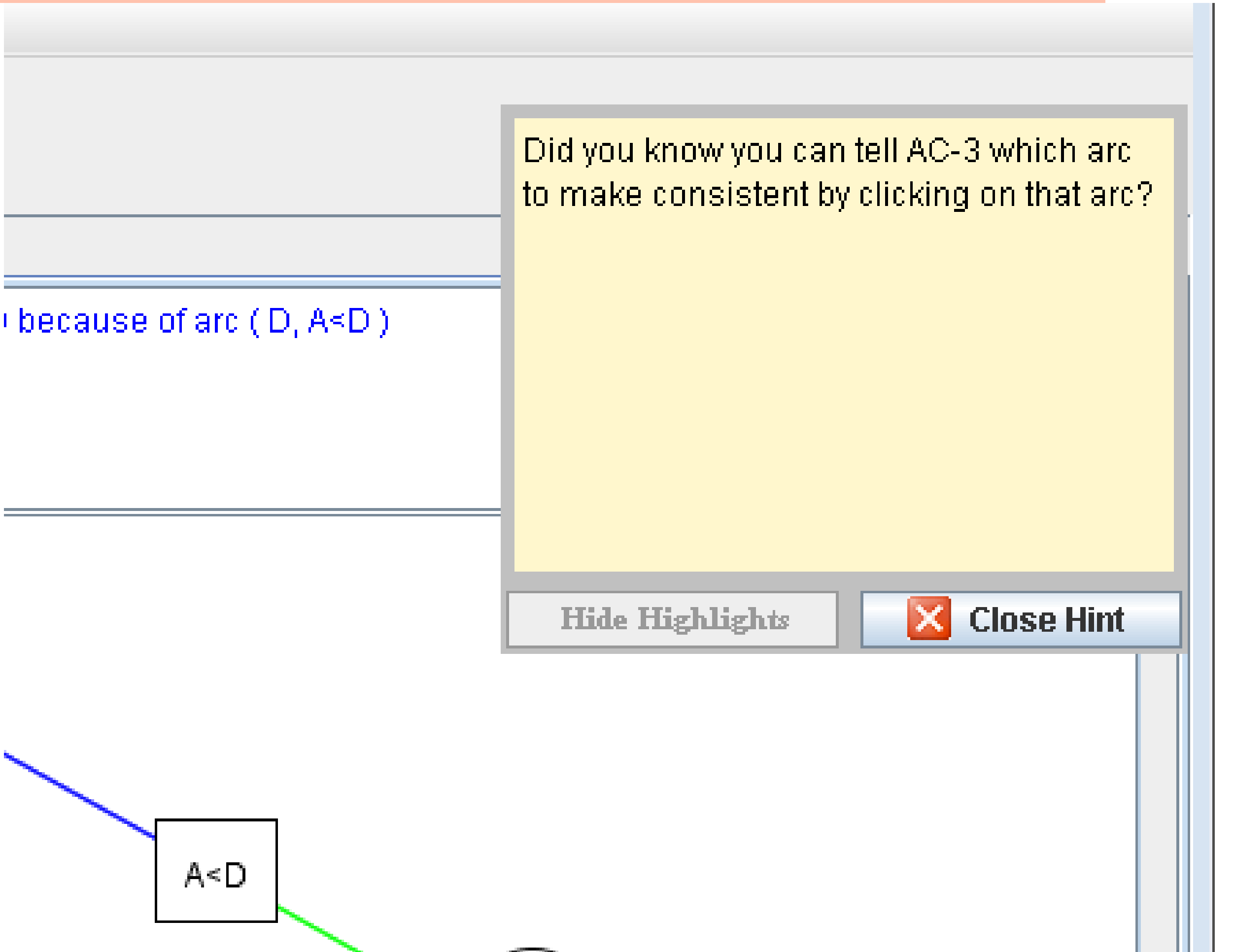
(Kardan and Conati CHI 2015)



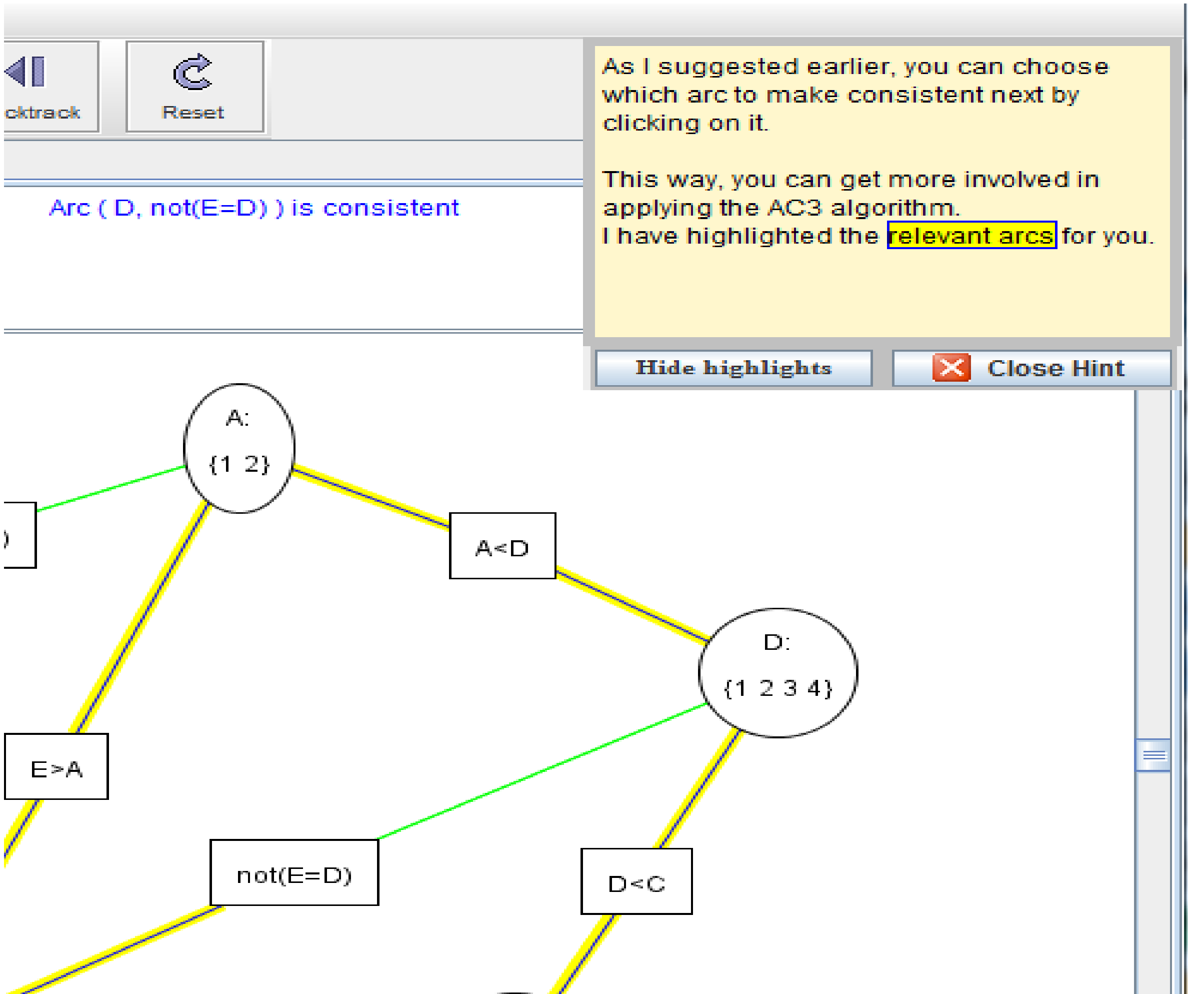
Incremental Hints: Level 1

Classifier User Model detects a Low Learner that

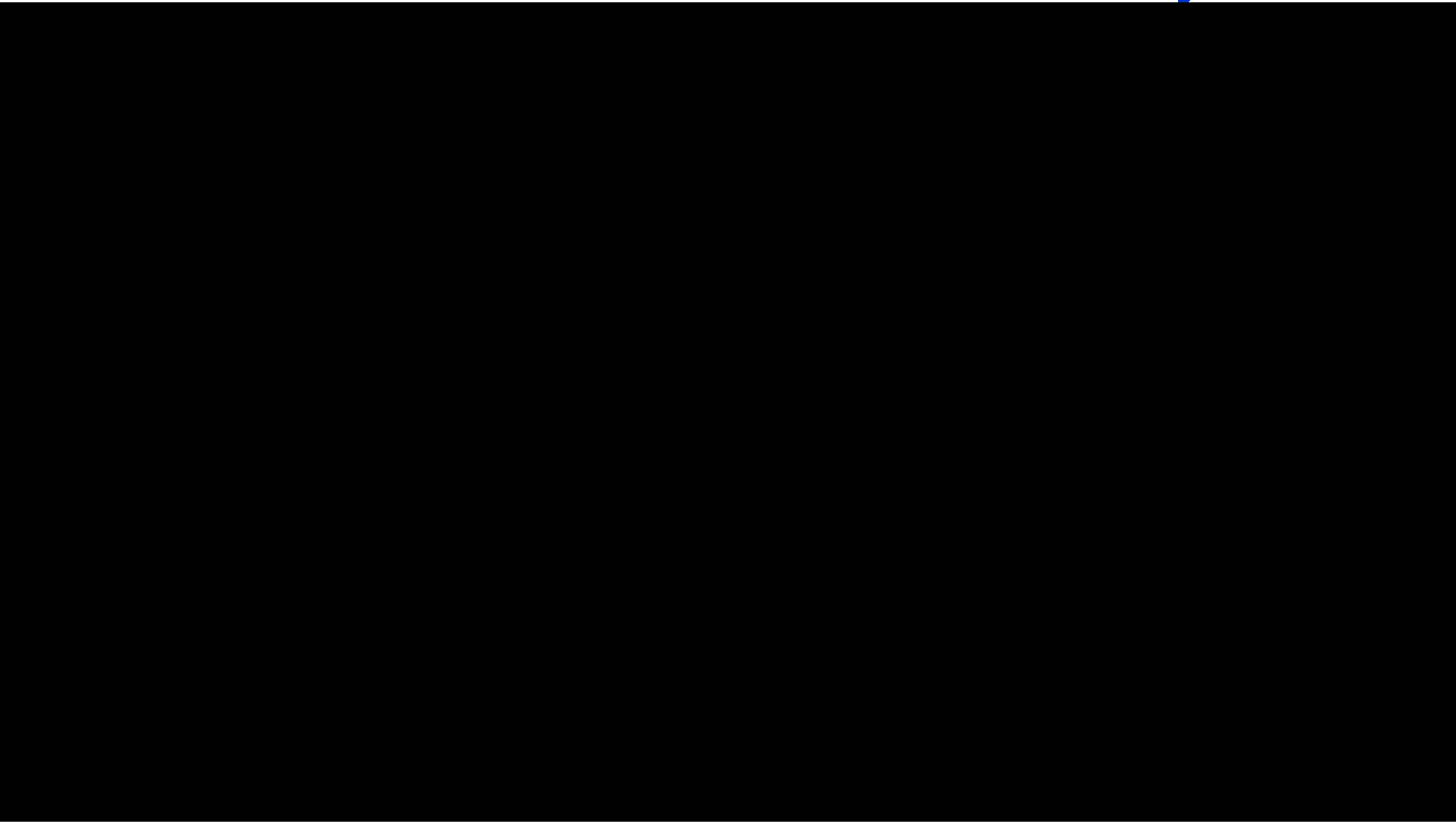
- Uses *Direct Arc Click* sparsely (R3)



Incremental Hints: level 2



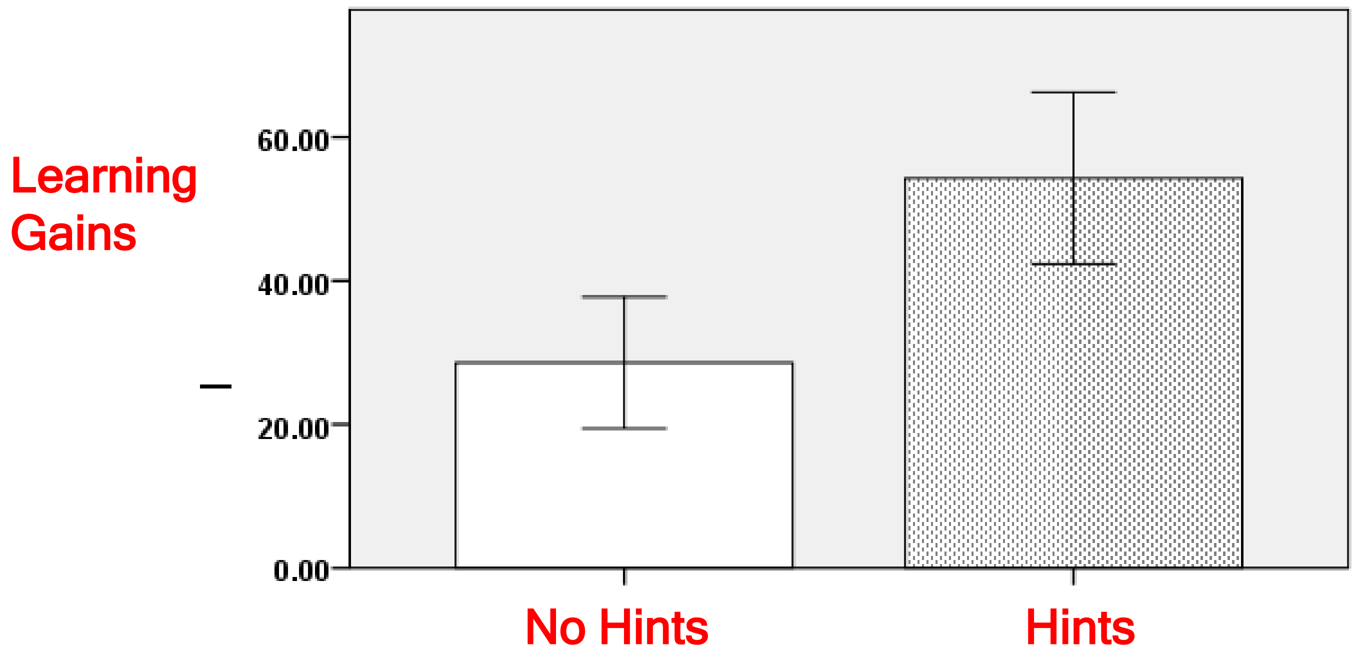
How to Delivet the Hints Effectively?



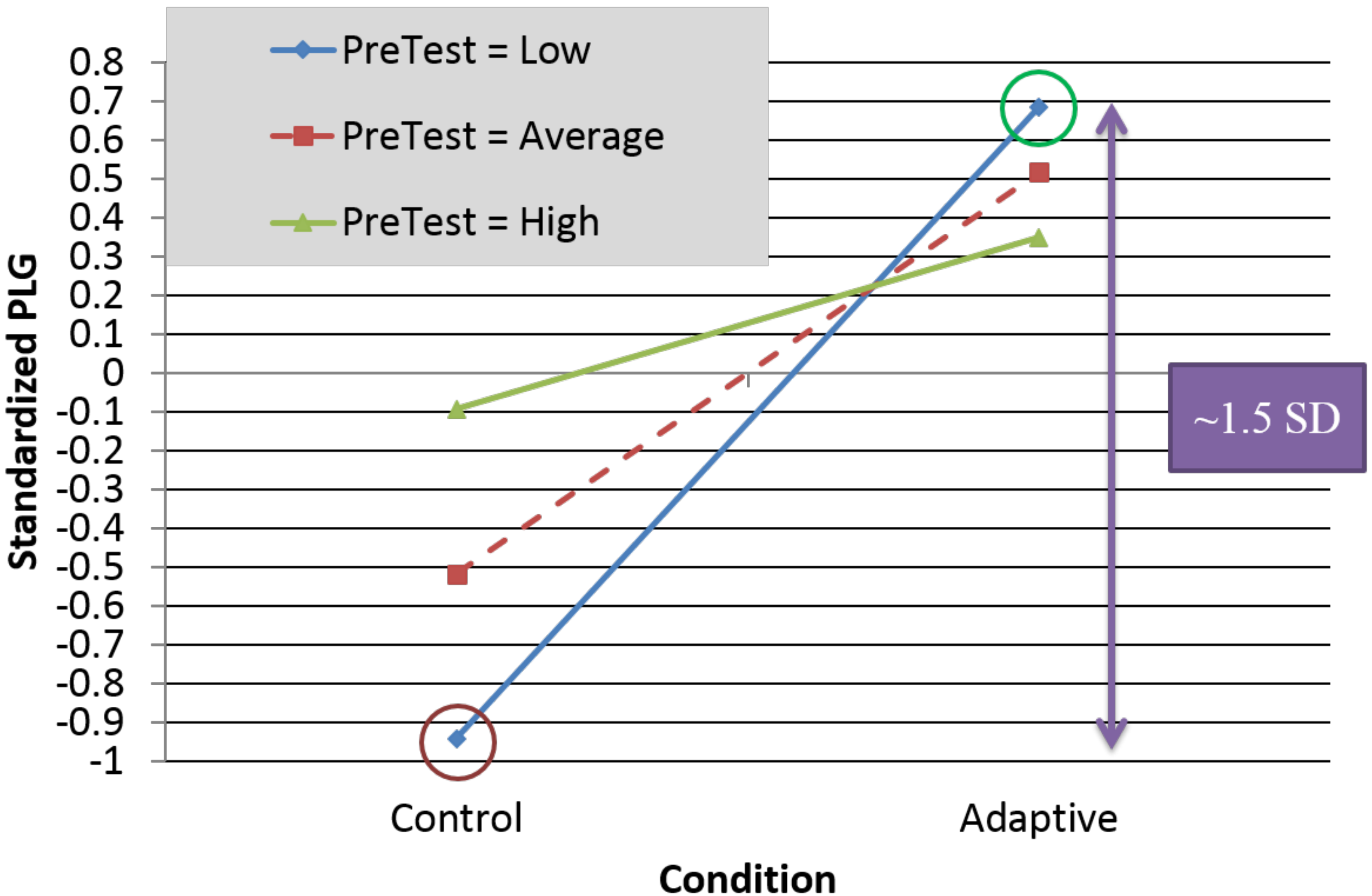
Evaluation

(Kardan and Conati CHI 2015)

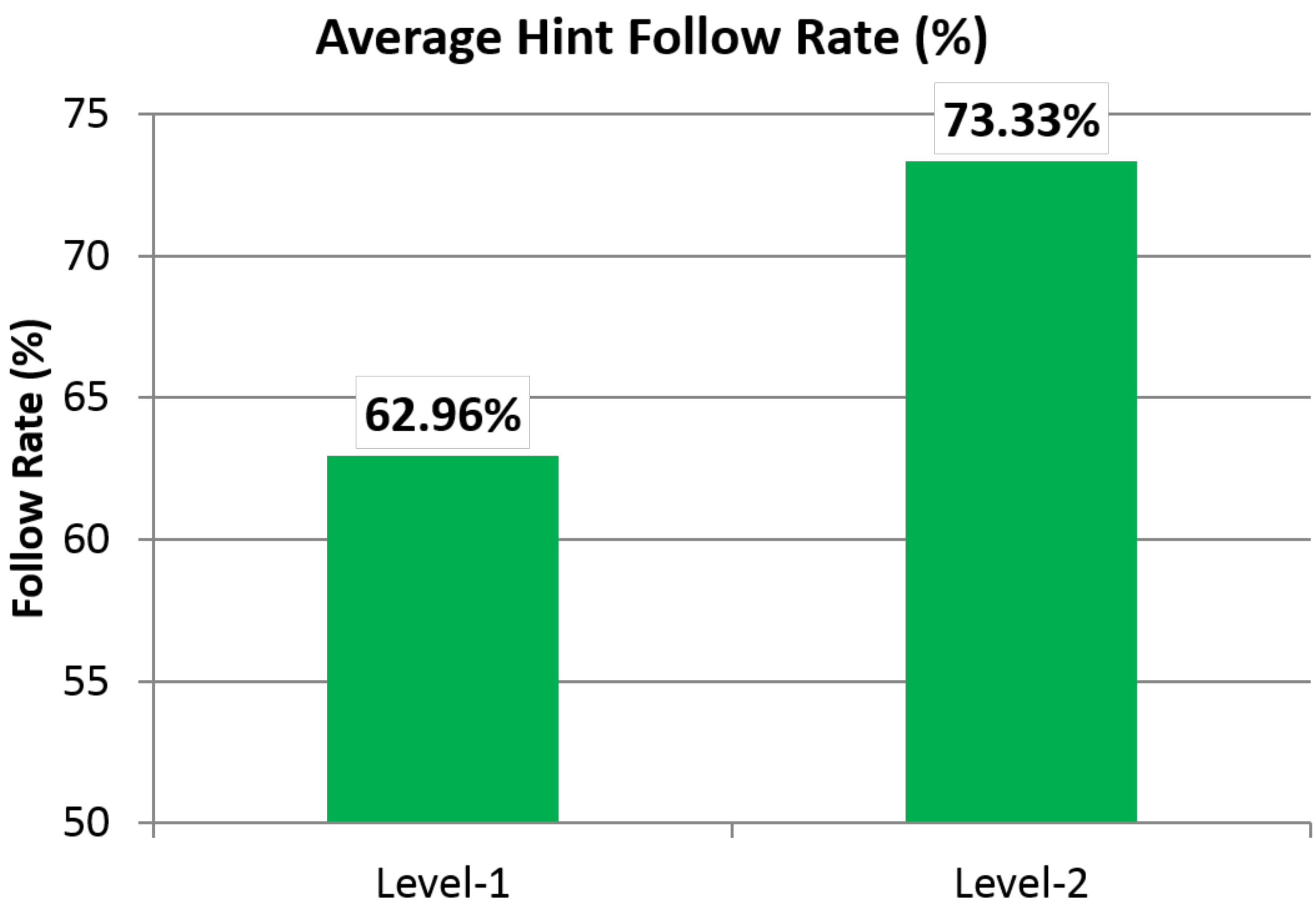
- User study :
  - Two groups of 18 students worked with the CSP applet
  - One group with personalized hints, and one without
- Students in the ACSP group learned more



Learning Gain: PreTest×Condition



Results: Acceptance of Interventions

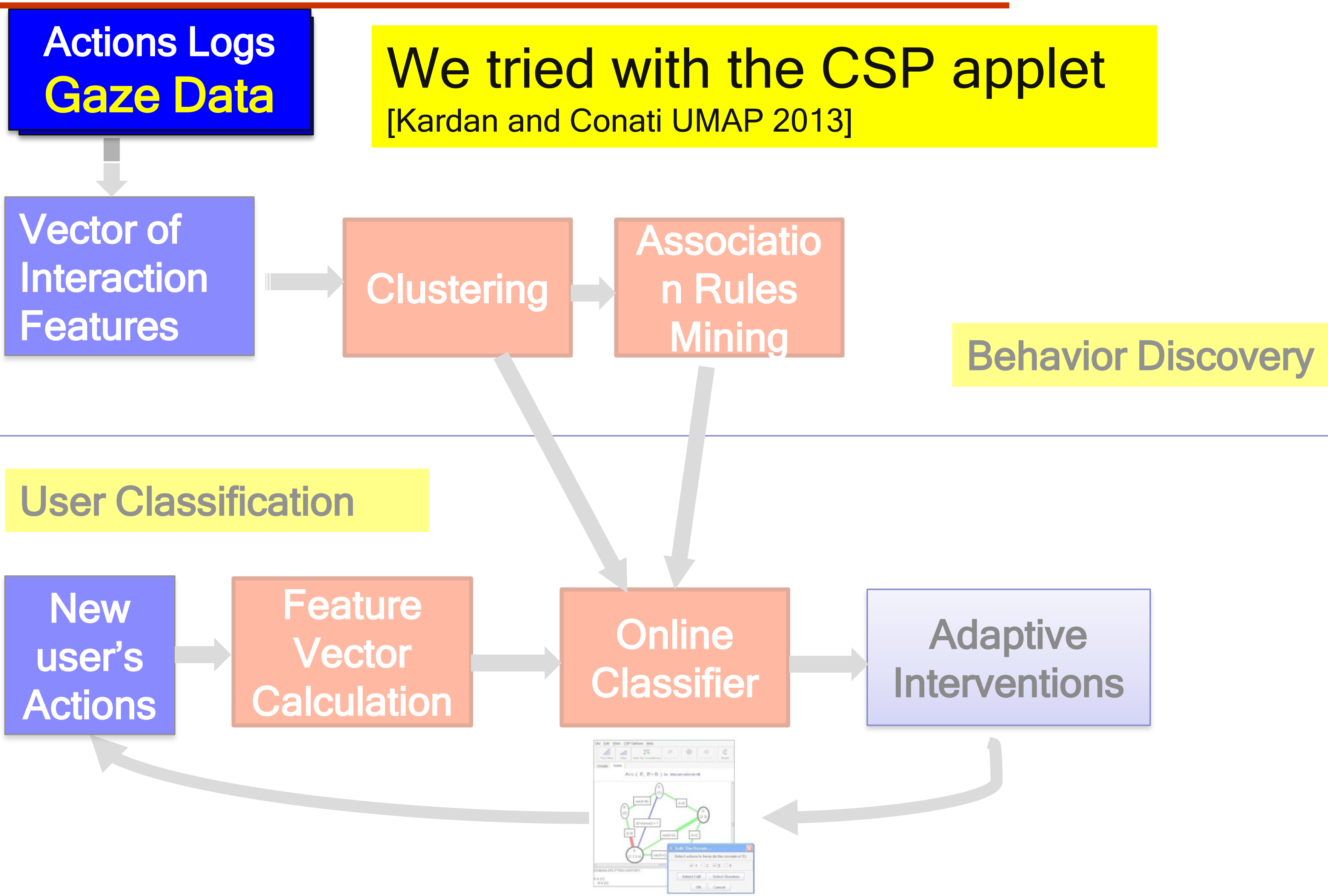


Overview

- Overview of FUMA and initial results with ACSP applet [Amershi and Conati 2009, Kardan and Conati 2012, 2015]
- Extensions to
  - multimodal data
  - More complex OELEs
- What's next?

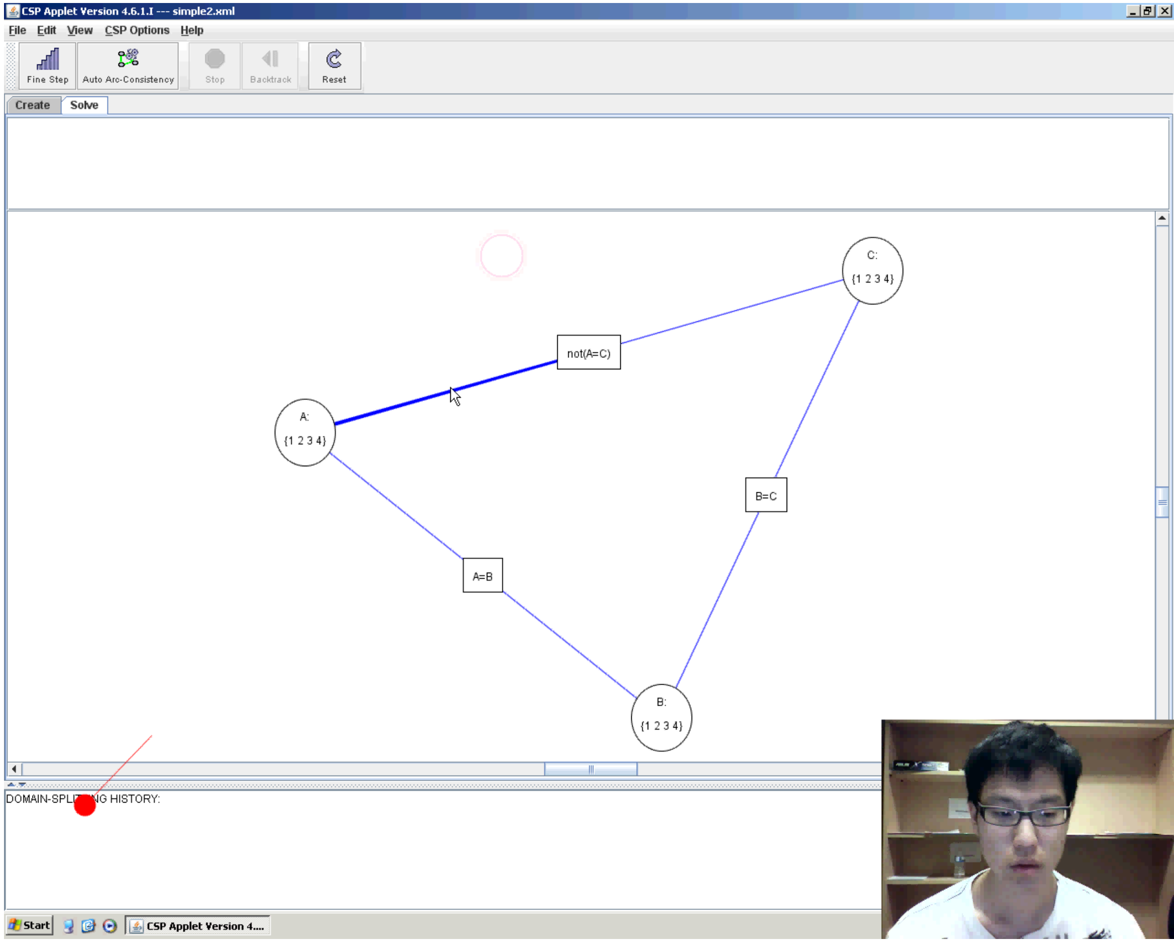
Experimenting With Multimodal Data

We tried with the CSP applet
[Kardan and Conati UMAP 2013]

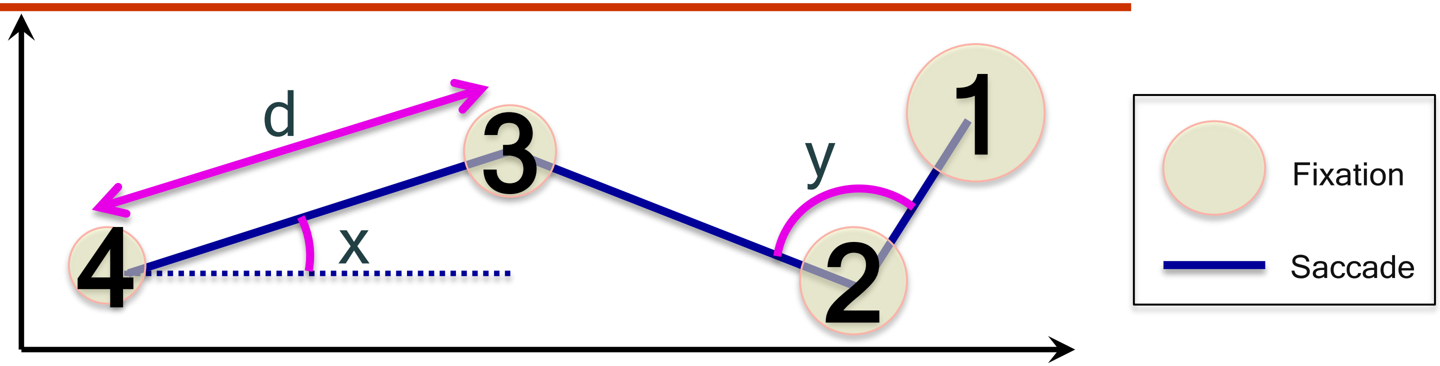


User Study to Collect Gaze Data

- 45 participants
- Tobii T120 eye tracker to capture user gaze



Eye-tracking measures

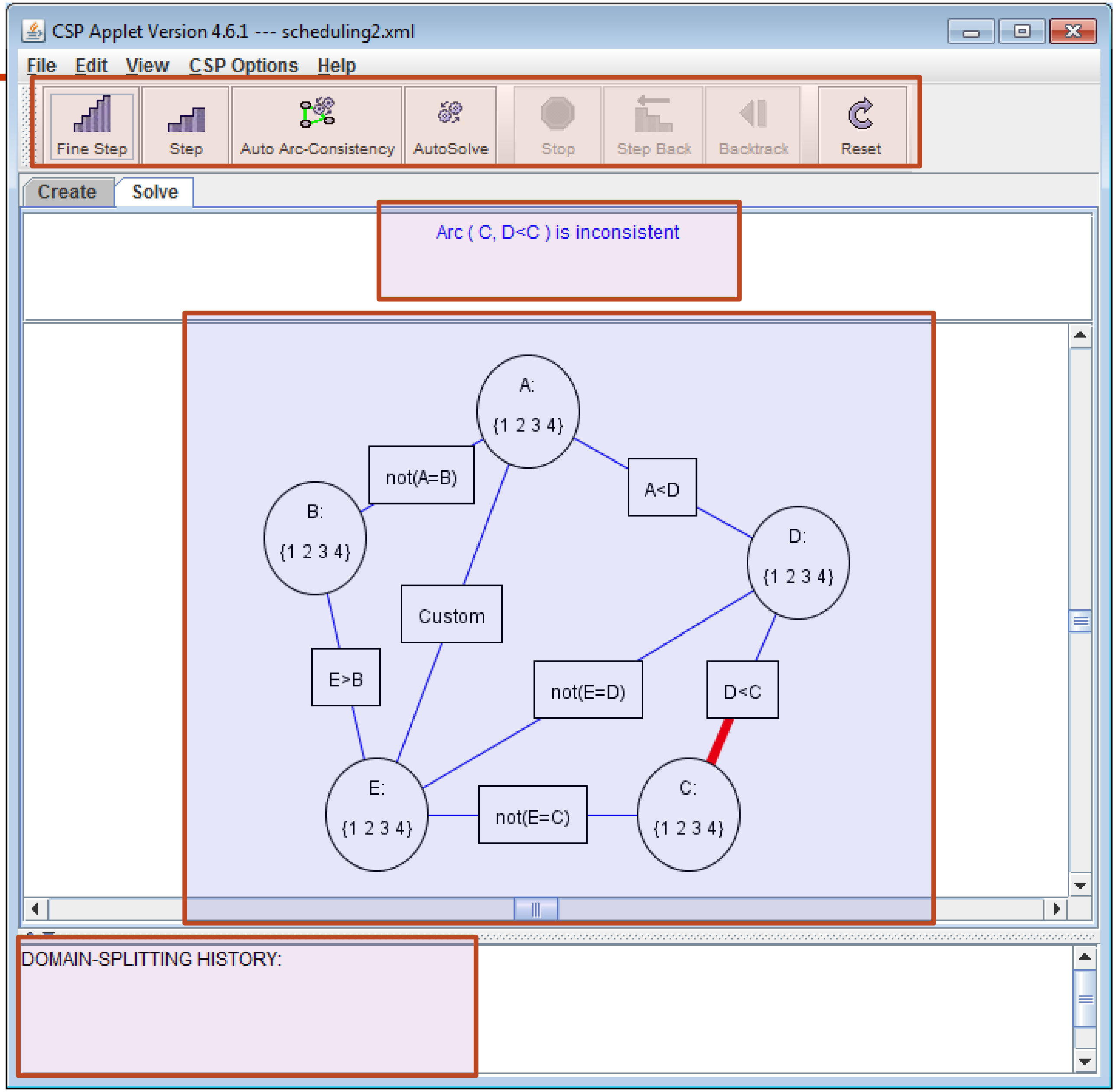


General Measures

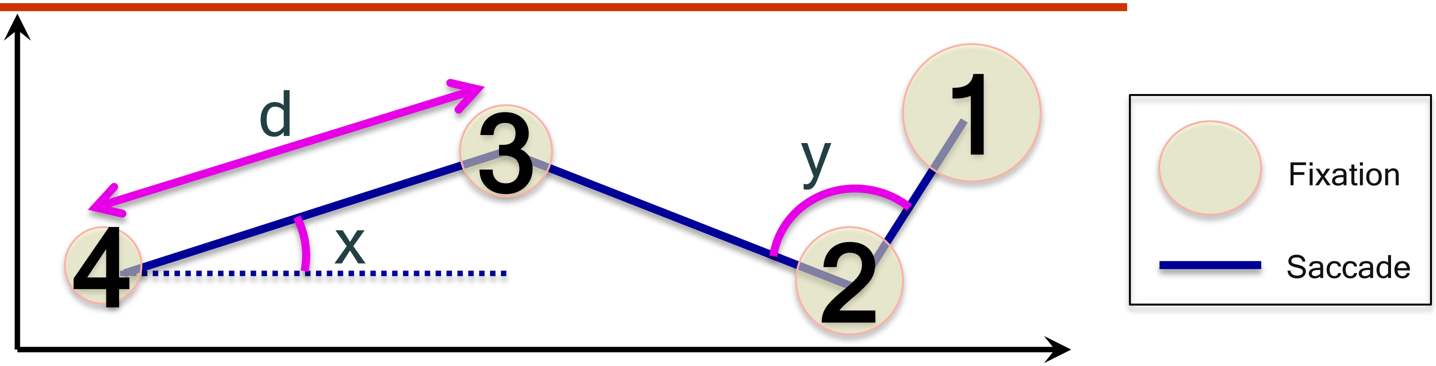
- Number of Fixations
- Fixation rate
- Fixation Duration
- Saccade Length (d)
- Relative Saccade Angles (y)
- Absolute Saccade Angles (x)
- ……………….

Measures specific to Areas of Interest (AOI)

Areas Of Interest



Eye-tracking measures



General Measures

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

Measures specific to Areas of Interest (AOI)

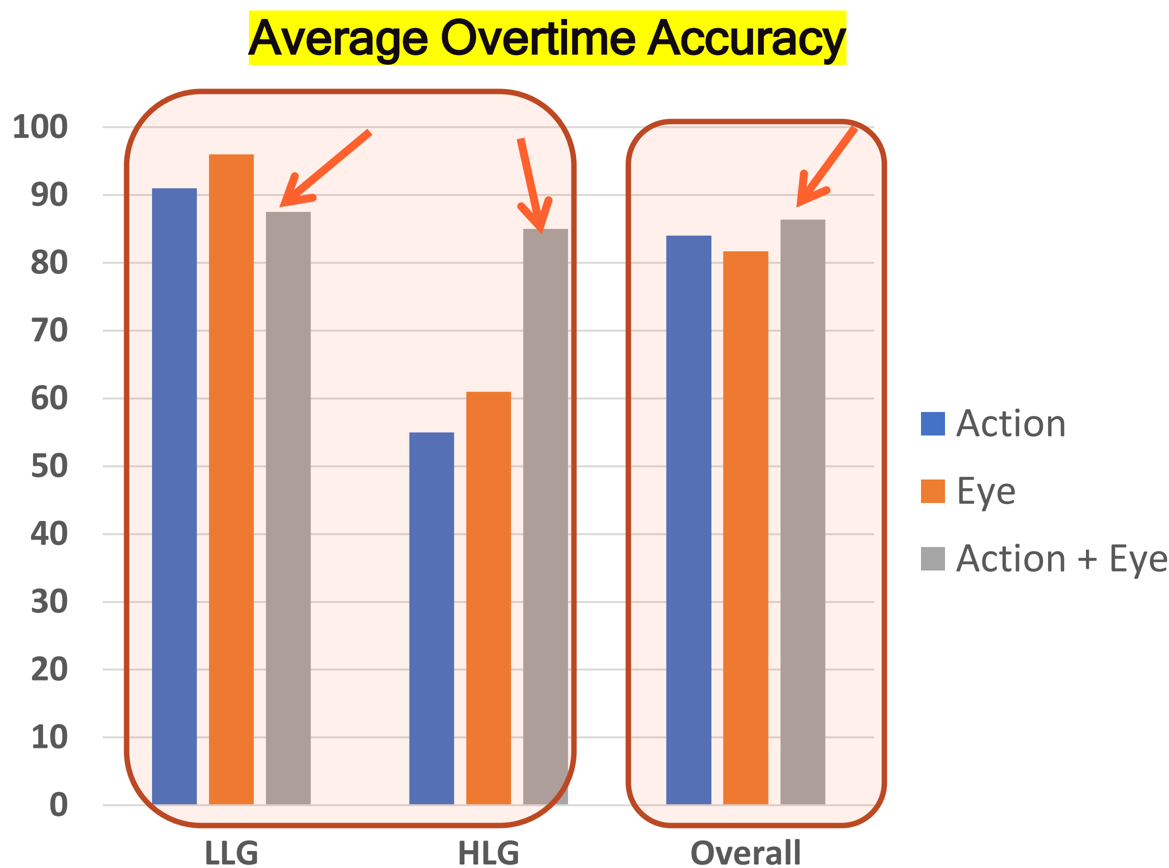
- Proportional number of fixations
- Proportional time spent
- Time to first fixation
- Transitions between two AOIs
- …………………..

51 features based on summary statistics (e.g. mean, st.dev.) of these measures

Apply FUMA to Action and Gaze Data

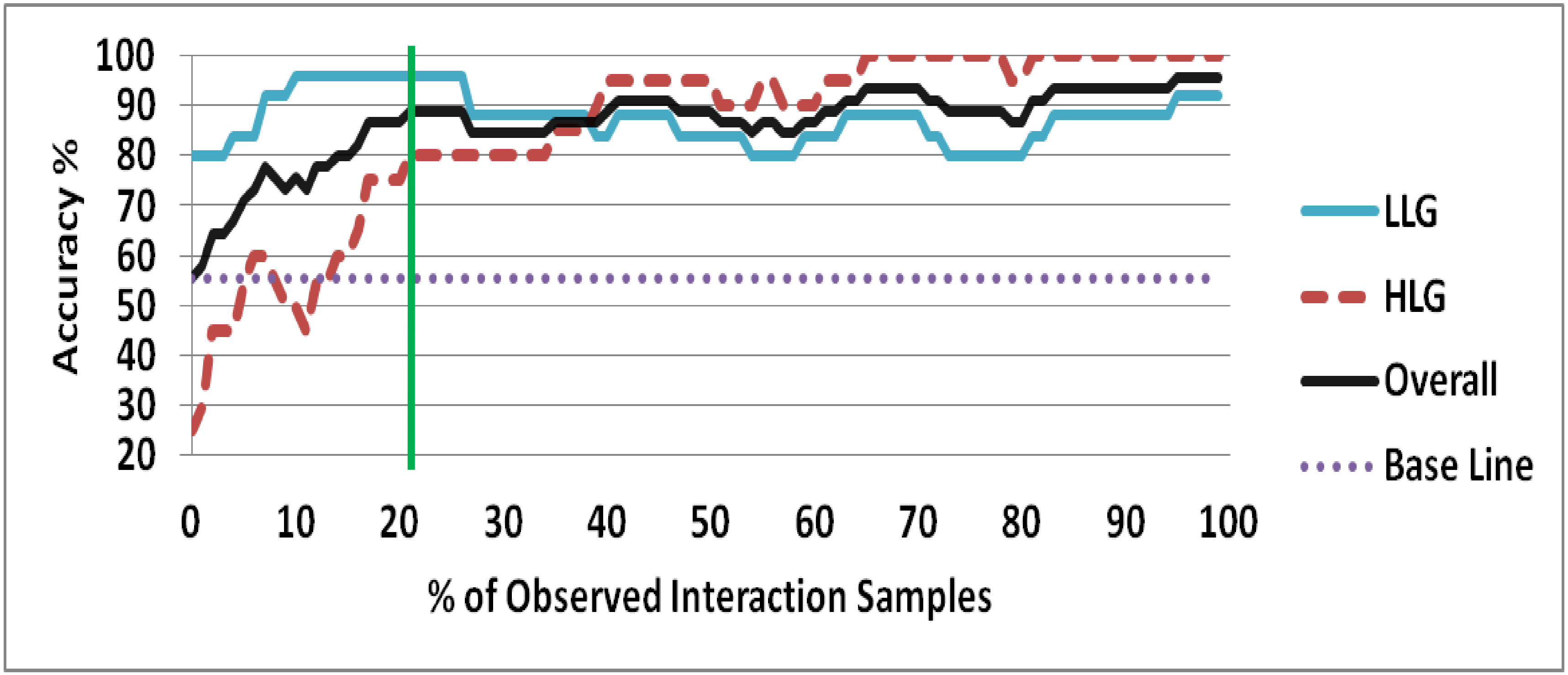
- Still found 2 clusters
  - Higher and Lower Learners

- Compared classifiers based on
  - Action only
  - Gaze only
  - Gaze + action



Merging Action and Gaze Data

[Kardan and Conati 2013)]



- Action + Gaze classifier achieves 80% classification accuracy over both classes after seeing 22% of the data

Multimodal Data: Lessons Learned

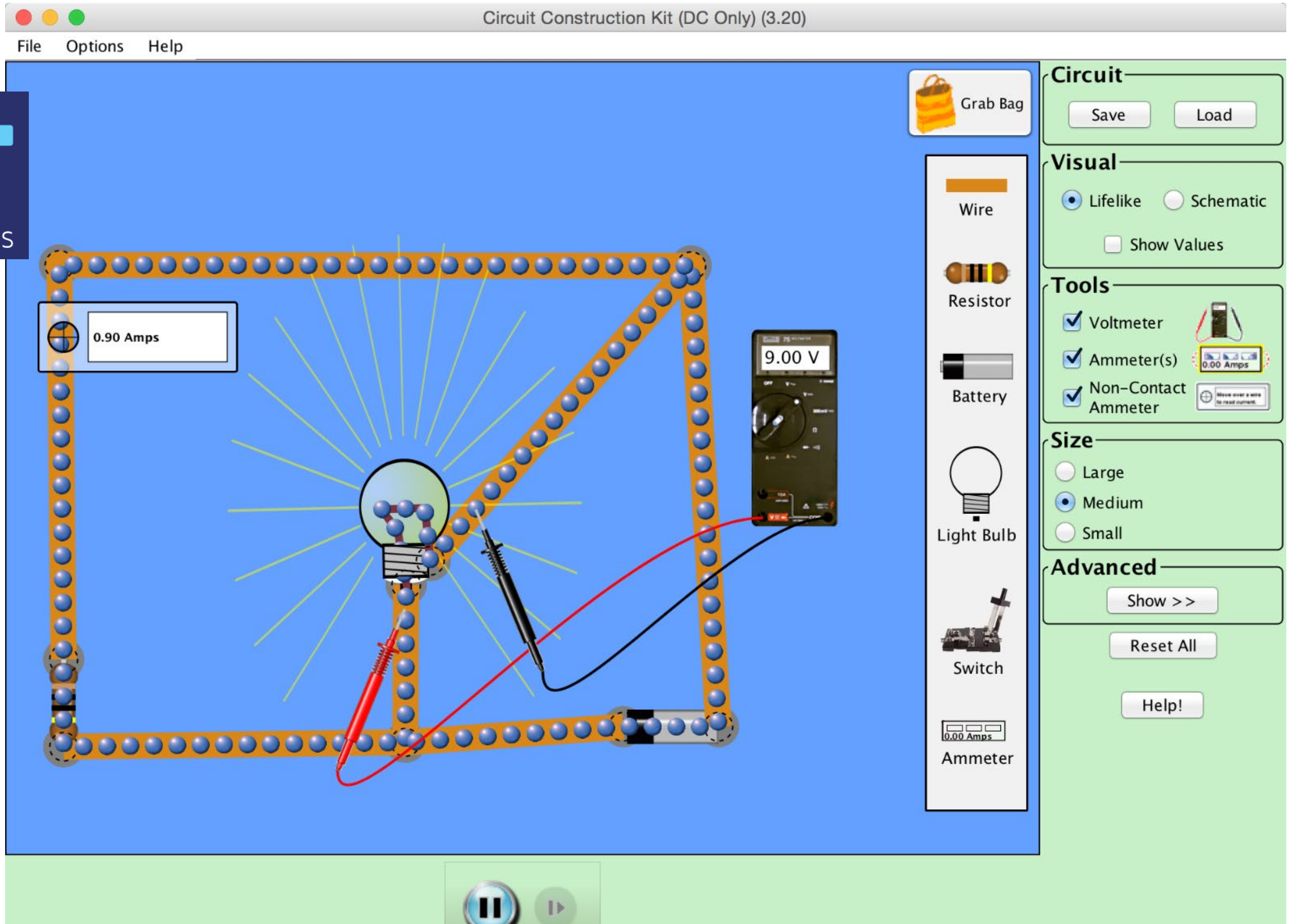
- Combining action and gaze data increases classification accuracy
- But the associations rules from these multimodal clusters are harder to turn into actionable hints
  - They may include features such as average saccade angles or fixation rate
- Solutions to investigate
  - Use multimodal data for classification/user modeling, but only action features to build hints
  - Use only higher-level gaze features (E.g. transitions between AOI)
  - Other?
- More future work:
  - investigate the tradeoff between classification accuracy and rule interpretability with other multimodal data

Overview

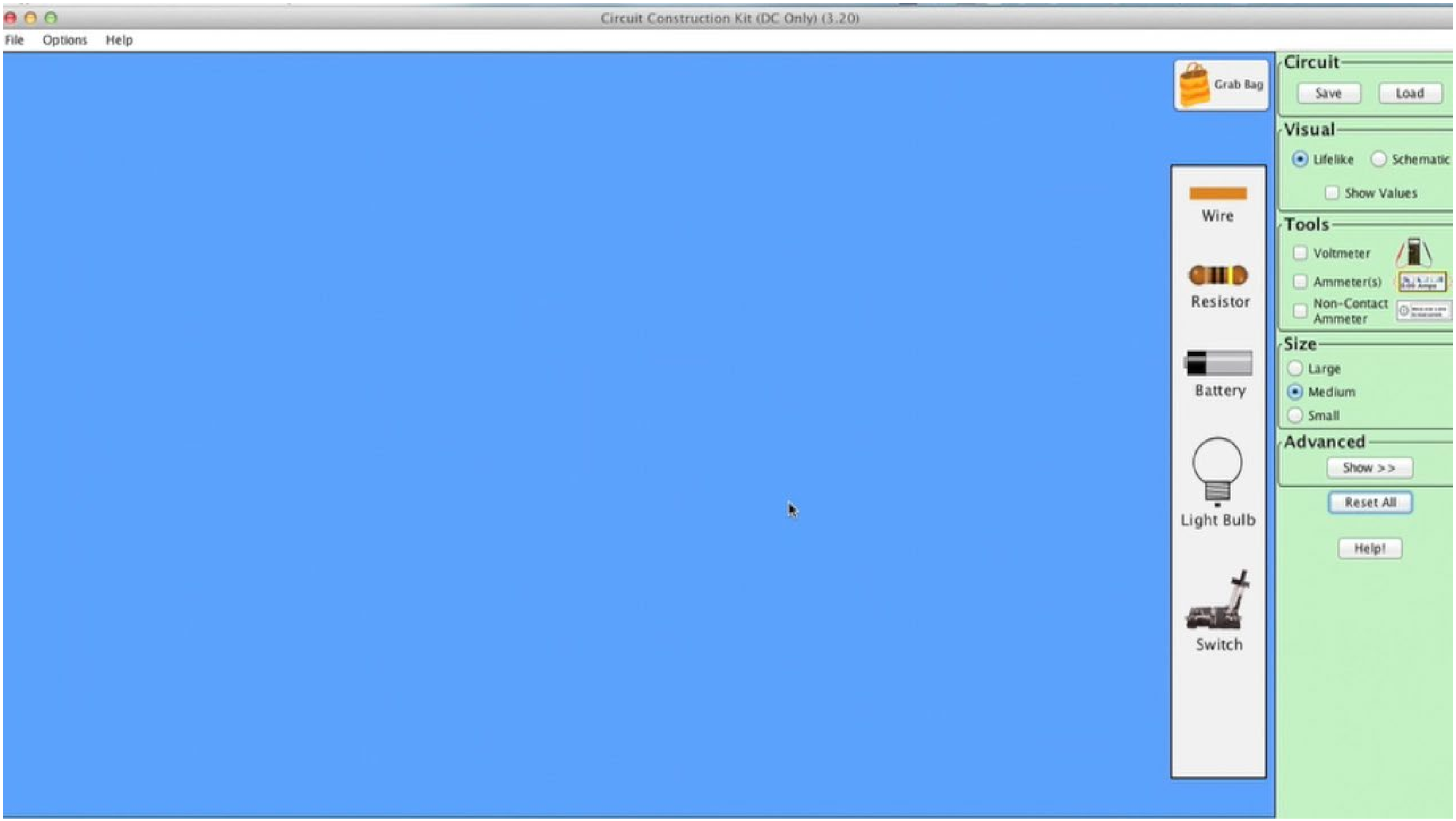
- Overview of FUMA and initial results with ACSP applet [Amershi and Conati 2009, Kardan and Conati 2012, 2015]
- Extensions to
  - Multimodal data
  - More complex OELEs
- What's next?

PhET DC Circuit Construction Kit (CCK)

- Part of large suite of simulations developed of U. of Colorado
- Allows students to explore building electrical circuits



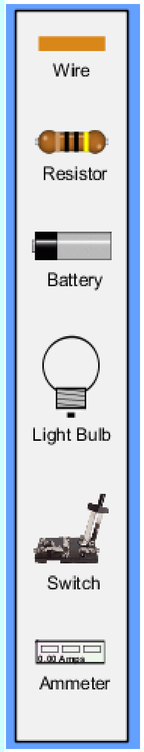
Interaction Demo



Complex Interaction

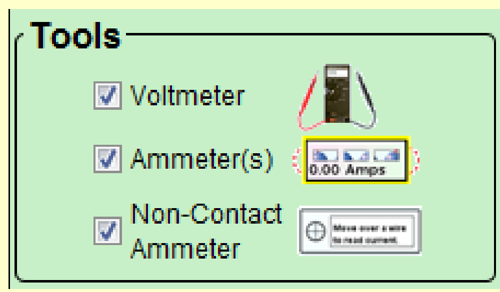
25 **actions**, eg:

- **On circuit elements**
  - Add
  - Move
  - Remove
  - Join
- **Measurement**
  - Voltage
  - Current
- **On Interface**
  - Simulation settings
  - Window



22 **components**, eg:

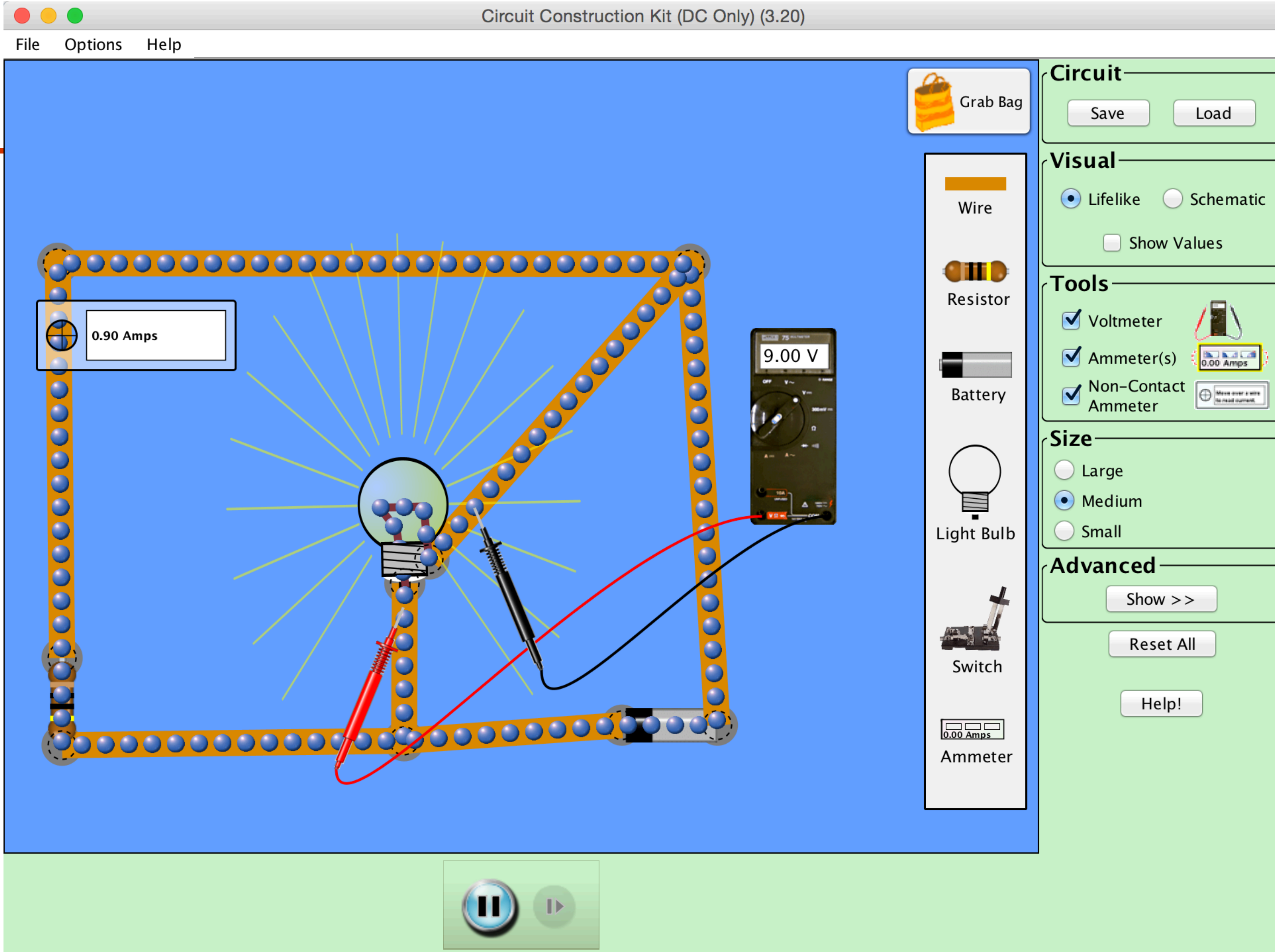
- Basic circuit elements
- Measurement tools



Context-dependent **outcomes**

- Light intensity change
- Current change
- Fire
- Measurement Reading change
- None

- Many ways to interact
- Context plays an important role
  - » Different outcomes depending on the state of the circuit



Circuit Construction Kit (DC Only) (3.20)
File
Options
Help
Grab Bag
Wire
Resistor
Battery
Light Bulb
Switch
0.00 Amps
Ammeter
0.90 Amps
9.00 V
Circuit
Save
Load
Visual
Lifelike
Schematic
Show Values
Tools
Voltmeter
Ammeter(s)
Non-Contact Ammeter
Size
Large
Medium
Small
Advanced
Show >>
Reset All
Help!

Layered Representation to Capture Complex Interaction with FUMA

[Fratamico at al. AIED 2015, JAIED 2017]

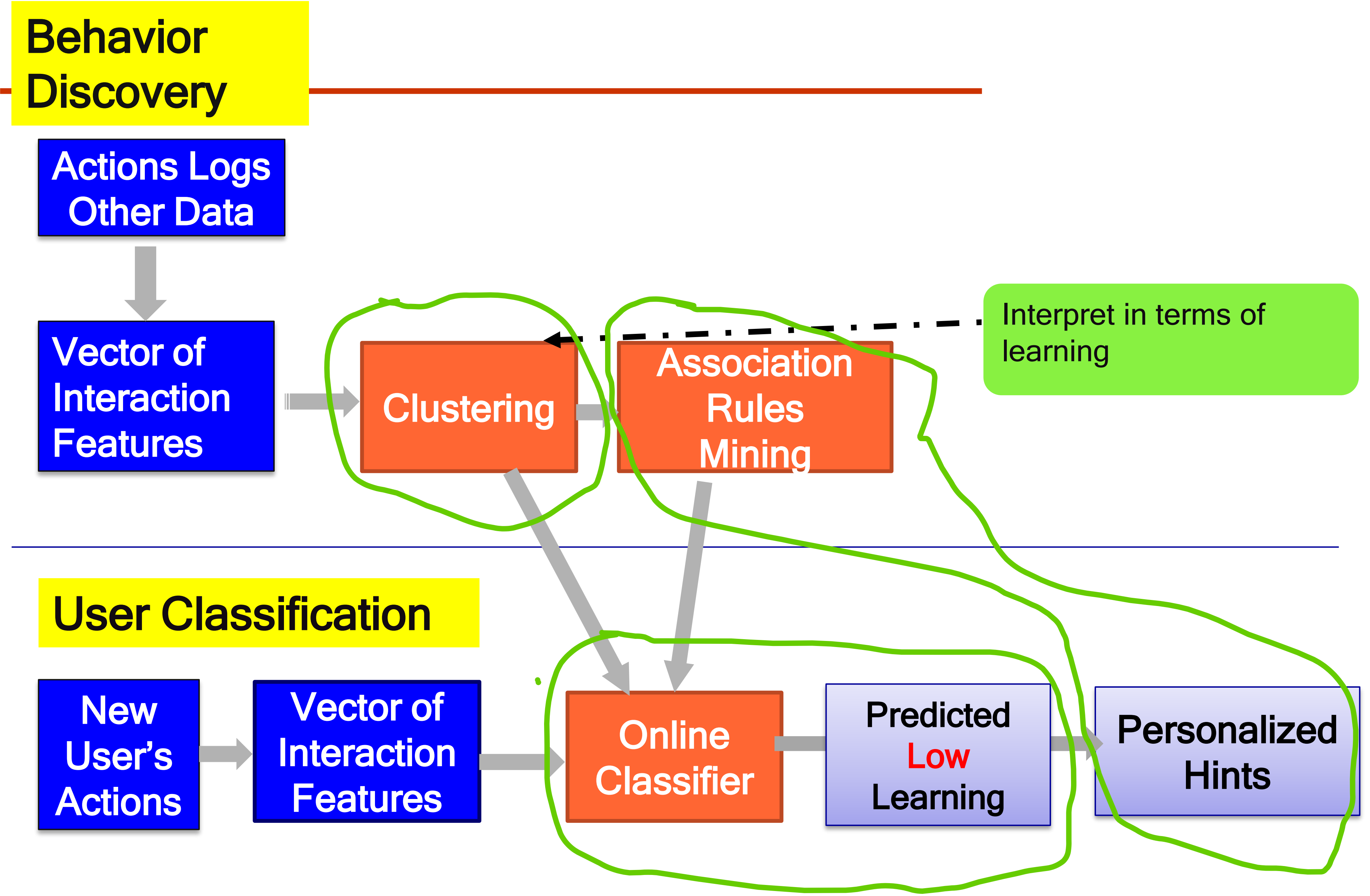
4 layers:

- Actions (A), Components (C), Outcomes (O)
  - from logs
- Families (F): engineered
  - Abstract actions into 8 more general activities that students can perform in CCK, e.g.
    - » Build (add, changeResistance, join)
    - » Revise (changeResistance, join, split, remove)
    - » Test (startMeasure, endMeasure, traceMeasure)

Representing the User Interaction

- Different combinations of the 4 layers represent interaction-events at different granularities, e.g.:
  - All 4 layers (OFAC)
    - » **current_change.revise.join.wire**
    - » Student generated a current change while **revising** the circuit by **joining** two **wires**
  - Outcome, Action, Component: (OAC)
    - » **current_change.join.wire**
    - » Does not include high level information on family
- Tested FUMA on 11 of these combinations, based on
  - Quality of the derived clusters
  - Classification accuracy
  - Usefulness of the generated association rules for adaptive interventions

FUMA



Behavior Discovery
Actions Logs
Other Data
Vector of Interaction Features
Clustering
Association Rules Mining
Interpret in terms of learning
User Classification
New User's Actions
Vector of Interaction Features
Online Classifier
Predicted Low Learning
Personalized Hints

User Study

Data collected from a lab study with CCK

- 96 UBC students taking a first year physics course
- Were given a general learning goal:
  - *Explore how resistors affect the behavior of circuits by exploring different combinations of resistors and resistances*
- Collected pre and post test data

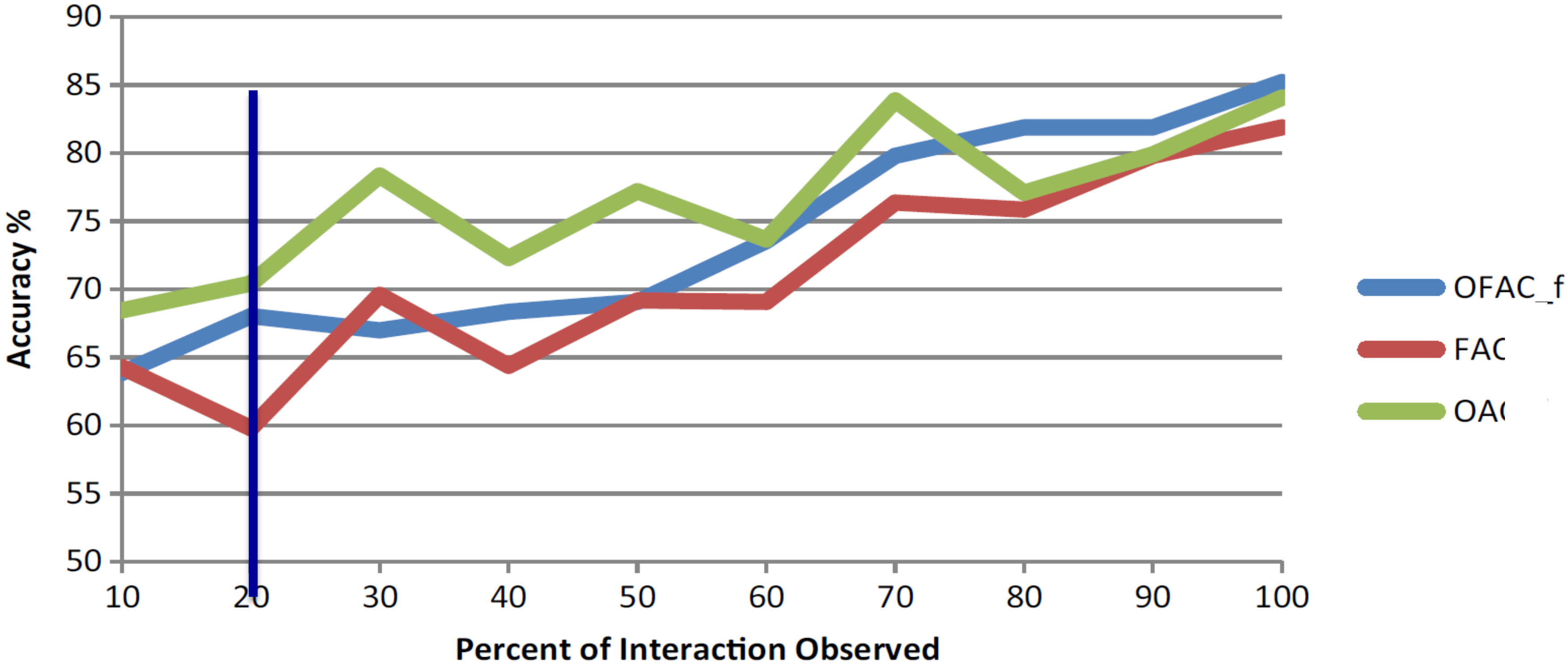
Quality of Clusters

- 3 of the 11 feature sets generated clusters (2) with significant difference in learning gains

Feature Set	Effect Size (partial η^2)
Family.Action.Component (FAC)	.041
Outcome.Action.Componenet (OAC)	**.076**
Outcome.Family.Action.Component (OFAC)	.065

OAC achieves the best cluster quality in term of highest difference in learning gains

Classification Accuracy



- OAC is the best classifier
  - Achieves 70% accuracy after seeing 20% of interaction data (~ 5 min)
  - OFAC gets there after seing 50% of the data

Generated Association Rules

- All features sets identified 4 general behavior patterns that instructors confirmed to impact learning with CCK
  - test frequently
  - frequently change resistance of resistors
  - pause to reflect in between actions
  - limit the usage of light bulbs and changes to their light intensity
- OFAC generated more specific rules (22)
  - Against the 15 generated by OAC
- Better suited to provide incremental feedback, e.g.
  - Start at the “Family” level; (e.g. “Test more”)
  - Incrementally go into more detail on how to do it

Summary of Results

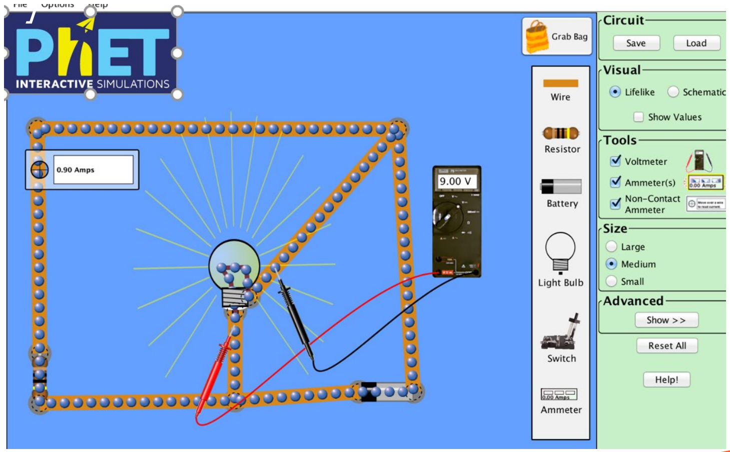
- OAC best for classification accuracy, specifically for providing timely hints
- OFAC best for usefulness of the generated association rules
  - can provide richer hints

- Need to empirically explore tradeoff between there two factors
- Investigate if this tradeoff exist with other complex ELEs

FUMA: Evaluated in several ELEs

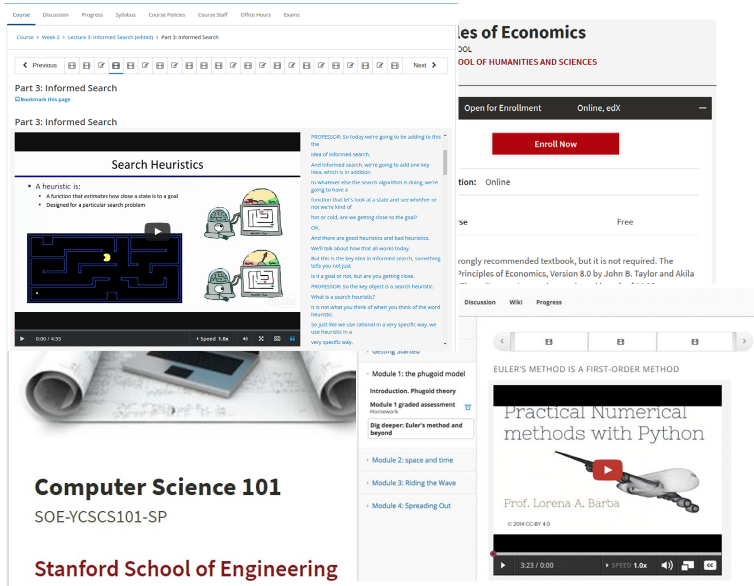
PHET Circuit Construction Kit (CCK)

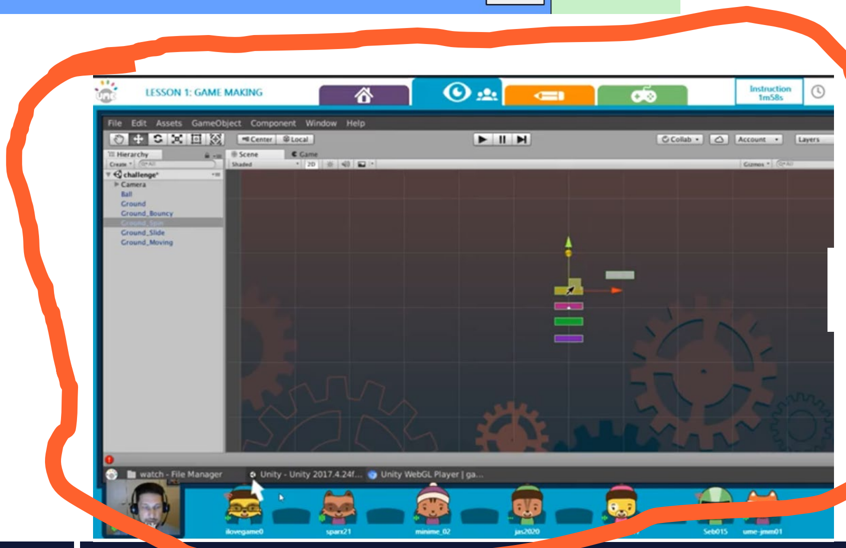
Kardan et al., 2014; Fratamico et al., 2017;



Four MOOCs

Lallé et al., 2020,





ELE for Game Design

Lallé et al., 2021,

Unity-CT

Collaboration with UME Academy:

- Use the popular Unity game engine to teach Computational Thinking (CT) skills to K-12 kids

- Free-form interaction to create small games

https://ume.academy

Unity-CT

Collaboration with UME Academy:

- Use the popular Unity game engine to teach Computational Thinking (CT) skills to K-12 kids

- Free-form interaction to create small games

https://ume.academy

Unity-CT

Collaboration with UME Academy:

- Use the popular Unity game engine to teach Computational Thinking (CT) skills to K-12 kids

- Classes facilitated by a UME instructor

- Can we have AI agents that helps with this facilitation?

https://ume.academy

FUMA for Unity CT [Lalle et al LAK 2021]

Actions Logs
Other Data

Vector of Interaction Features

Clustering

Association Rules Mining

Behavior Discovery

User Classification

New user's Actions

Vector of Interaction Features

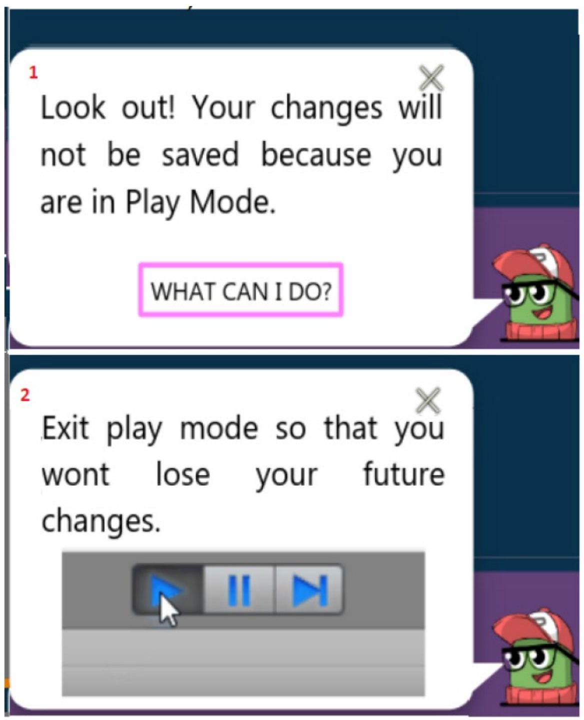
Online Classifier

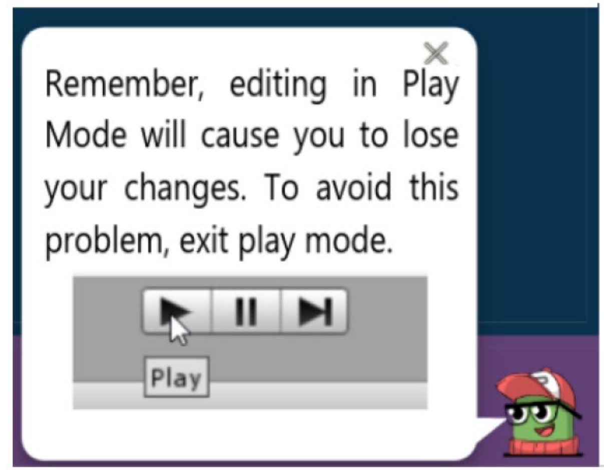
Predicted Low Learning

Personalized Hints

Current Work

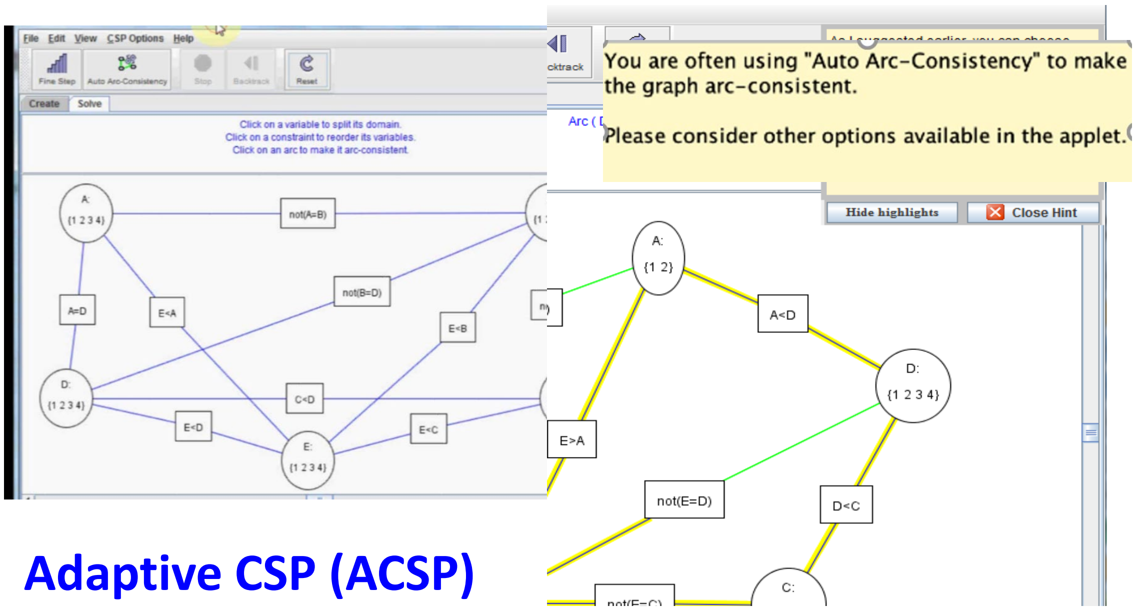
- Designing the delivery of adaptive support, with UME UX expert and instructors
- For instance, what to do about repeated hints (Yalcin et al. IUI 2022, AIED 2023)





Explaining FUMA Hints

[Conati et al., AI Journal 2021]



Adaptive CSP (ACSP)

FUMA-driven hints shown to improve student learning
[Kardan and Conati, CHI 2015]

Evidence that these hints are more effective if the system can explain why and how they were generated

And that hint explanations may be even more effective if they are personalized to specific student characteristic

Conclusions

- FUMA: data-driven framework for user modeling and personalization to support learning with ELEs
- Evaluation with several ELEs show that FUMA can
  - Identify clusters with behaviors representative of student performance
  - Classify student performance with good accuracy, early enough to generate help when needed
  - Drive the design of personalized help from the detected behaviors
- Initial evidence that FUMA-driven interventions can help learning
  - And that their effectiveness can be improved with explanations

Future Work

- Apply FUMA to other OLEs
- Experiment with multimodal data
- More evidence that FUMA-driven hints foster learning
- These hints are shallow.
  - How do they compare against richer, knowledge-based hints?
- Consider student affect for hint provision
- Look at collaborative activities
- Continue investigating the value of personalized explanations of FUMA-driven hints

Thanks To

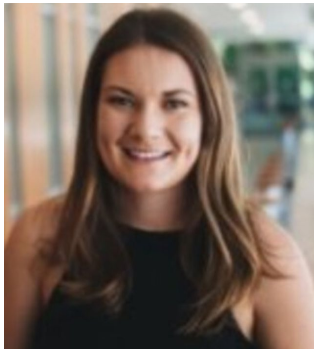
Saleema Amershi

Oswald Barral

Lauren Fratamico

Sebastien Lalle

Samad Kardan



Vanessa Putnam

Ido Roll

Nilay Yalcin

And to all of you for your kind attention !