A Methodological Framework for Decision-Theoretic Software Customization Assistance

PhD Thesis

Bowen Hui
Department of Computer Science, University of Toronto
June 15th 2011

Need for Software Customization

- One-size-fits-all:
  - Cluttered interfaces, *bloat*-ware
  - *Dis*satisfied users
  - Most affected users
    - People with cognitive, sensory, motor impairments
    - Elderly/Children
    - Novices

- Recognize varying user needs and preferences

Interface Customization

- Objectives
  - Minimize user effort
  - Maximize interaction experience

Commercial Apps	Research Apps
Windows “Start” menu	Auto meeting scheduler
Mini toolbar in Office 2007	Folder prediction
Auto text completions	Adaptive menus

- Different benefits and costs involved

Where do I start?

– Which existing research results can I draw upon?
– What are components in such a system?
– What steps are involved?

Thesis Contributions

- Decision-theoretic framework and guidelines
  - Techniques from interdisciplinary fields
  - Data collection, simulation testing, user evaluation
- Formal model of user types, characteristics, goals
  - Probabilistic models
  - Fast, online inference
  - Explains individual preferences
- Formal model of interaction cost
  - Models of interaction factors
  - Account for objective and subjective utility
  - Characteristics parameters to capture evolving preferences
  - New method for eliciting experienced utility

What aspects of the system should I make adaptive?

- Which tasks do users need help with?
- How do I model individual differences?
- What does the system need to know?

Thesis Contributions

- Decision-theoretic framework and guidelines
  - Techniques from interdisciplinary fields
  - Data collection, simulation testing, user evaluation
- Formal model of user types, characteristics, goals
  - Probabilistic models
  - Fast, online inference
  - Explains different interaction preferences
- Formal model of interaction cost
  - Models of interaction factors
  - Account for objective and subjective utility
  - Characteristics parameters to capture evolving preferences
  - New method for eliciting experienced utility

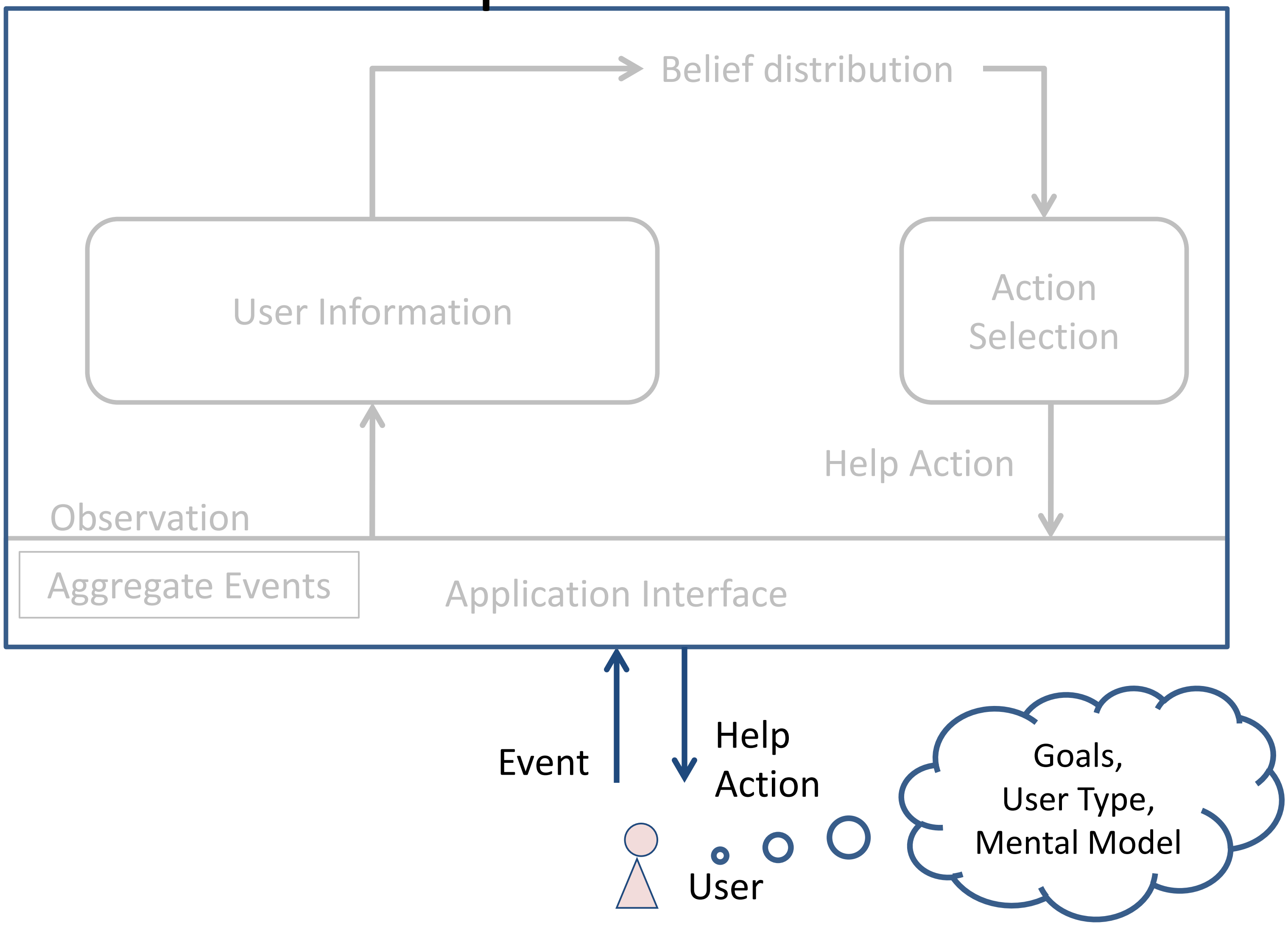
How do I know if it works?

– For different user groups?
– Can I anticipate “impact” before adapting the interface?
– What if user preferences change over time?

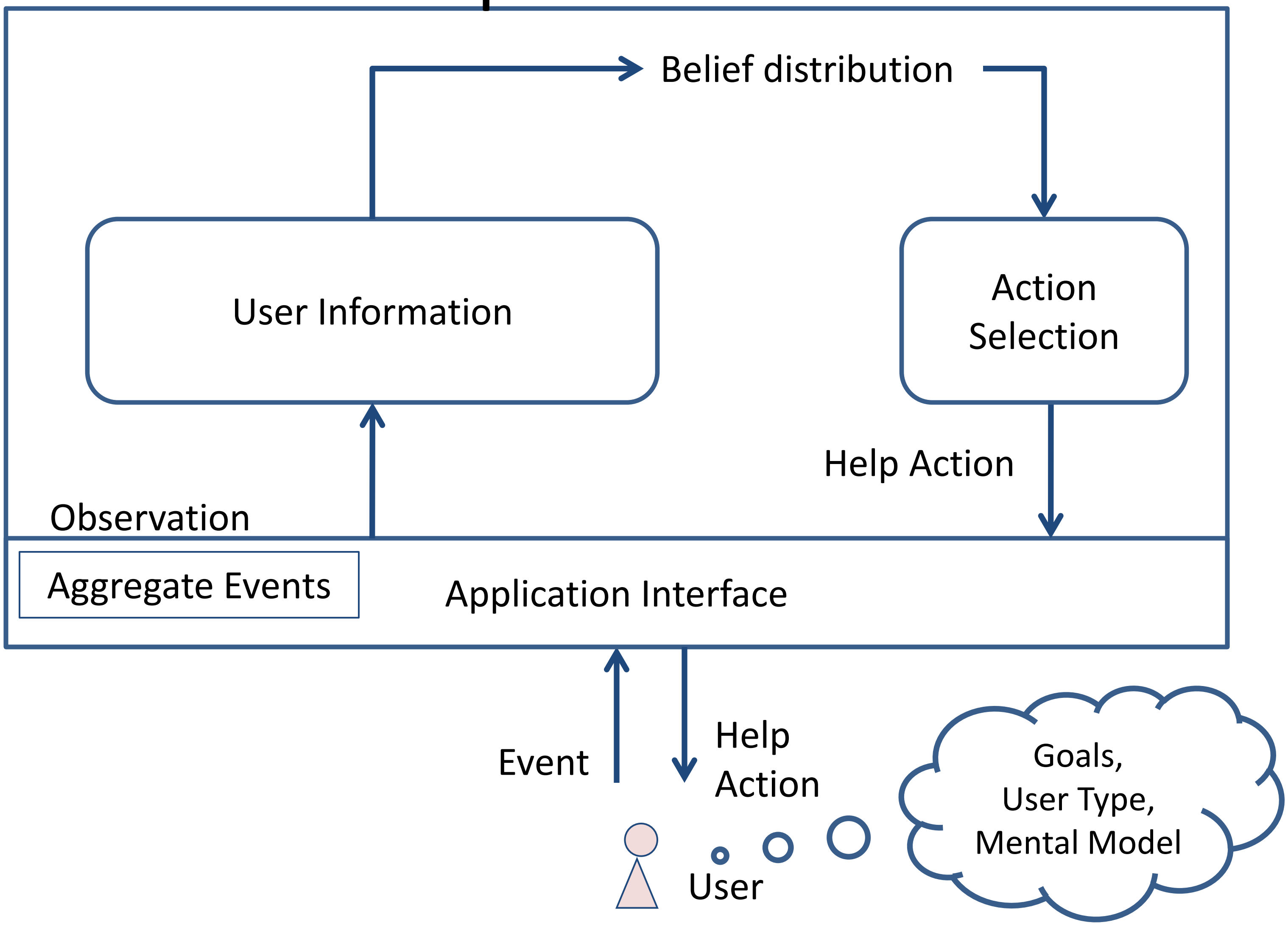
Thesis Contributions

- Decision-theoretic framework and guidelines
  - Techniques from interdisciplinary fields
  - Data collection, simulation testing, user evaluation
- Formal model of user types, characteristics, goals
  - Probabilistic models
  - Fast, online inference
  - Explains individual preferences
- Formal model of interaction cost
  - Models of interaction factors
  - Account for *objective* and *subjective* utility
  - New method for eliciting *experienced* utility
  - Characteristics parameters to capture evolving preferences

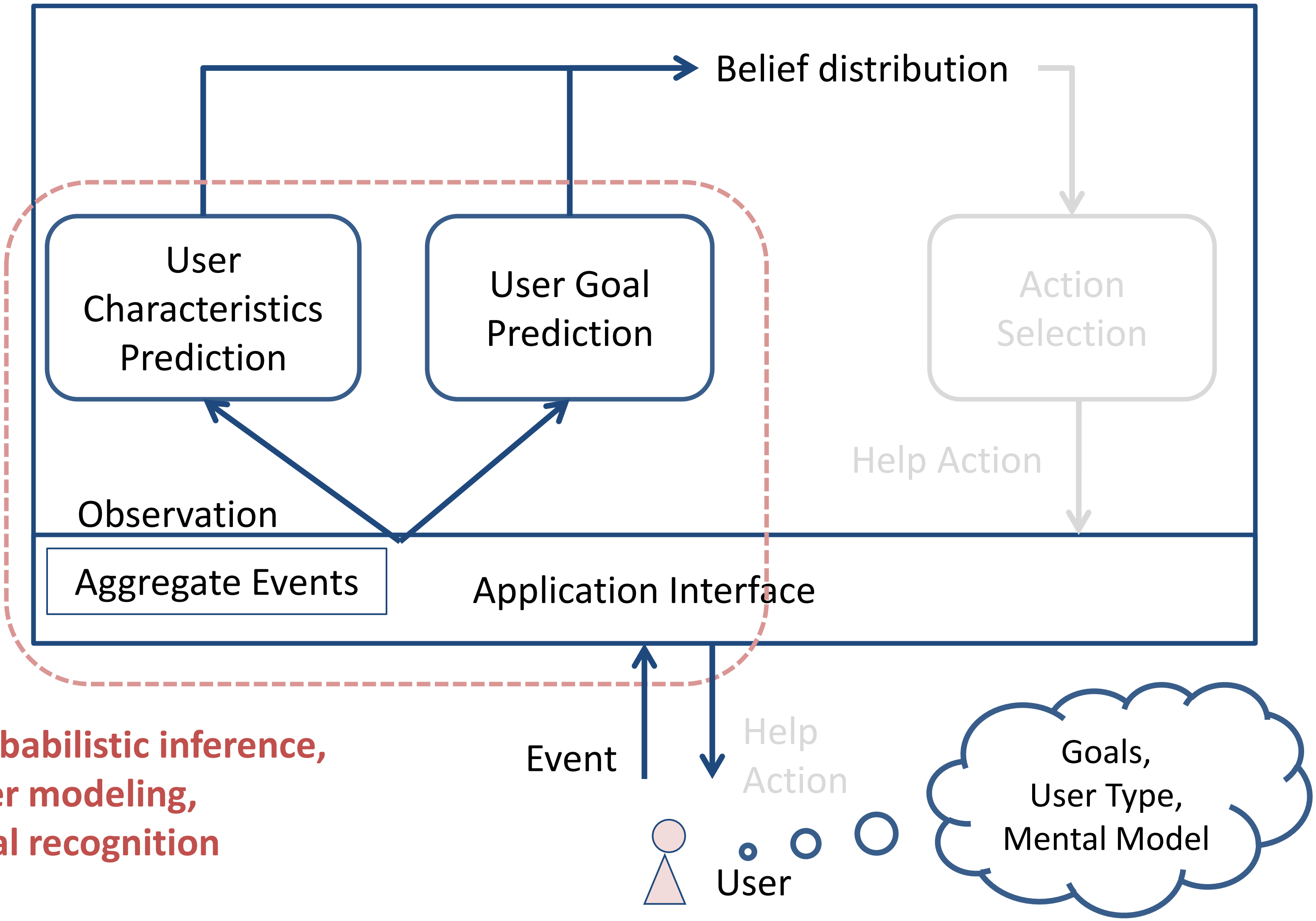
Software Customization Assistance (SCA) Development Architecture



Software Customization Assistance (SCA) Development Architecture

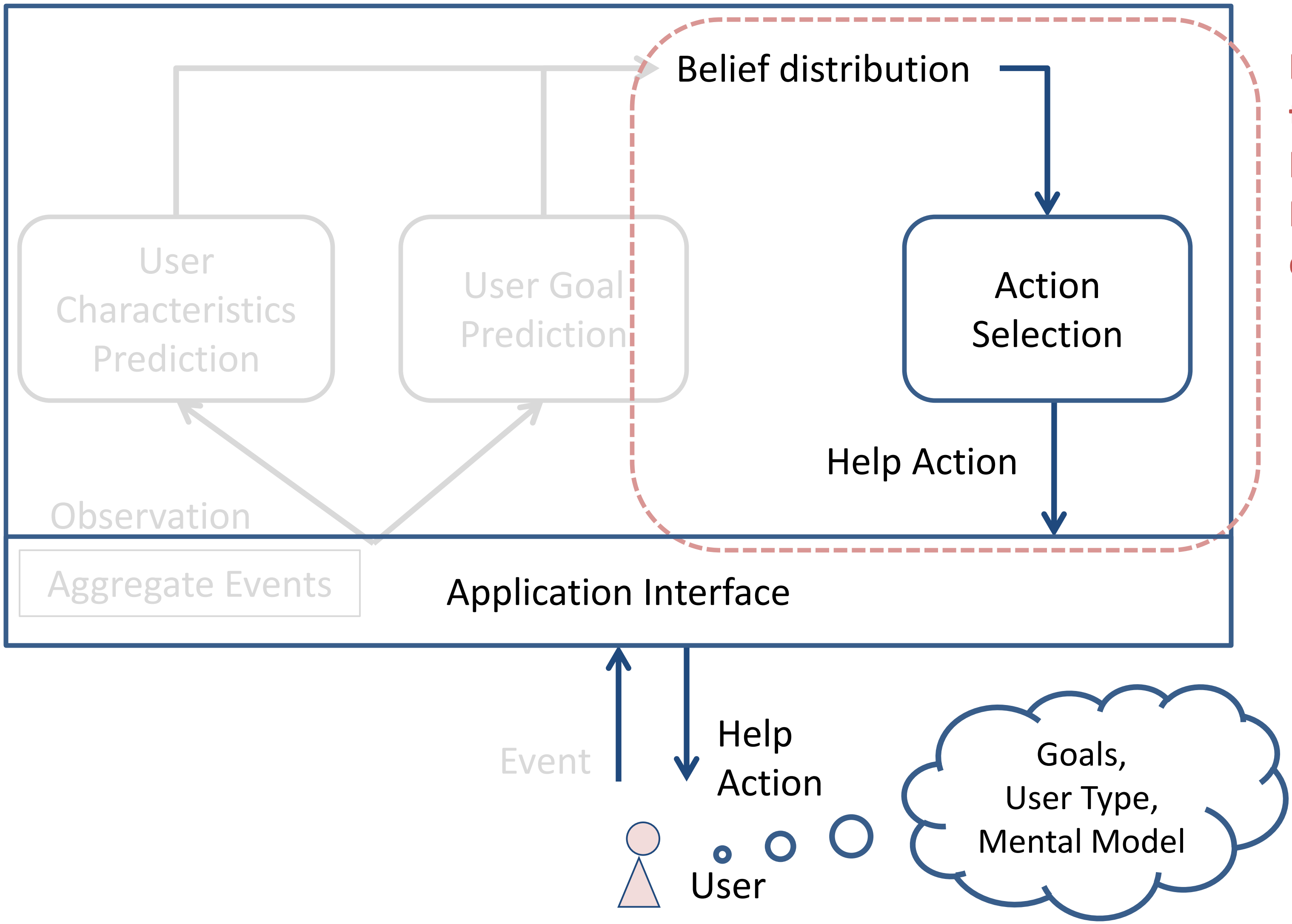


SCA Development Architecture



Probabilistic inference,
User modeling,
Goal recognition

SCA Development Architecture

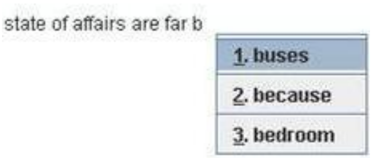


Decision theory, HCI, Preference elicitation

Case Studies

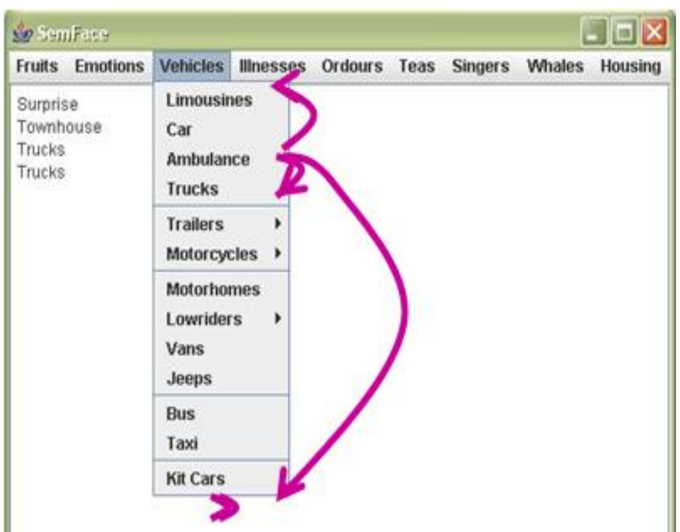
Word completion [Hui & Boutilier 06; ACM Finals 07]

Bayesian user characteristics model



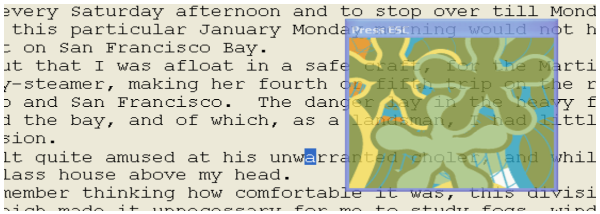
Adaptive menu [Hui et al. 09]

Probabilistic mental model



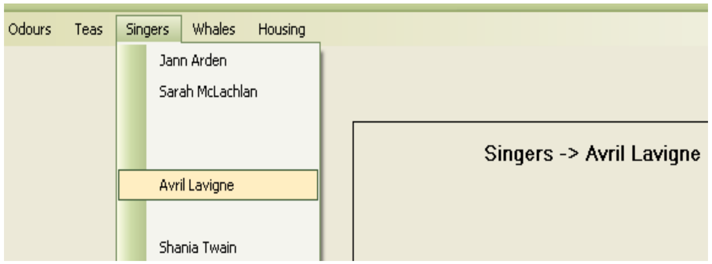
Typing [Hui et al. 08]

Occlusion model



Menu selection [Hui et al. 08]

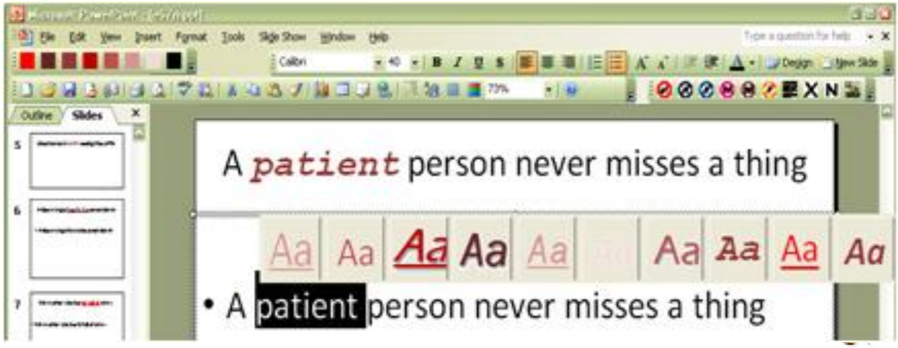
Bloat model



Highlighting toolbar [Hui & Boutilier 08]

Goal model

Experiential elicitation



Case Studies

Word completion [Hui & Boutilier 06; ACM Finals 07]

Bayesian user characteristics model

- Data collection: 45 users
- Simulation: Always, Never, MEU, Thresh
- Usability: 4 users
- Analysis: standard error, EM, KL-divergence, factor analysis

Adaptive menu [Hui et al. 09]

Probabilistic mental model

- Data collection: 48 users
- Simulation: Best-Static, Split-4, JES, WER(.1), WER(.5), WER(.9)
- Usability: 8 users
- Analysis: correlation, factorial ANOVA, regression (linear, log, Gaussian)

Typing [Hui et al. 08]

Occlusion model

- Data collection: 12 users
- Analysis: factor analysis, regression (linear, quadratic, cubic)

Menu selection [Hui et al. 08]

Bloat model

- Data collection: 12 users
- Simulation: Static, Random, MDP
- Analysis: standard error, factor analysis, regression (linear, quadratic)

Highlighting toolbar [Hui & Boutilier 08]

Goal model

- Data collection: *online-adaptation*
- Simulation: Static, Freq-Char, Freq-only, Goal-Char, Goal-only
- Usability: 12 users

Experienced utility elicitation

- Data collection: 38 users
- Analysis: t-test, Hotelling's T2 test

Case Studies

Word completion [Hui & Boutilier 06; ACM Finals 07]

Bayesian user characteristics model

- Data collection: 45 users
- Simulation: Always, Never, MEU, Thresh
- Usability: 4 users
- Analysis: standard error, EM, KL-divergence, factor analysis

Adaptive menu [Hui et al. 09]

Probabilistic mental model

- Data collection: 48 users
- Simulation: Best-Static, Split-4, JES, WER(.1), WER(.5), WER(.9)
- Usability: 8 users
- Analysis: correlation, factorial ANOVA, regression (linear, log, Gaussian)

Typing [Hui et al. 08]

Occlusion model

- Data collection: 12 users
- Analysis: factor analysis, regression (linear, quadratic, cubic)

Menu selection [Hui et al. 08]

Bloat model

- Data collection: 12 users
- Simulation: Static, Random, MDP
- Analysis: standard error, factor analysis, regression (linear, quadratic)

Highlighting toolbar [Hui & Boutilier 08]

Goal model

- Data collection: *online-adaptation*
- Simulation: Static, Freq-Char, Freq-only, Goal-Char, Goal-only
- Usability: 12 users

Experienced utility elicitation

- Data collection: 38 users
- Analysis: t-test, Hotelling's T2 test

Case Studies

Word completion [Hui & Boutilier 06; ACM Finals 07]

Bayesian user characteristics model

- Data collection: 45 users
- Simulation: Always, Never, MEU, Thresh
- Usability: 4 users
- Analysis: standard error, EM, KL-divergence, factor analysis

Adaptive menu [Hui et al. 09]

Probabilistic mental model

- Data collection: 48 users
- Simulation: Best-Static, Split-4, JES, WER(.1), WER(.5), WER(.9)
- Usability: 8 users
- Analysis: correlation, factorial ANOVA, regression (linear, log, Gaussian)

Typing [Hui et al. 08]

Occlusion model

- Data collection: 12 users
- Analysis: factor analysis, regression (linear, quadratic, cubic)

Menu selection [Hui et al. 08]

Bloat model

- Data collection: 12 users
- Simulation: Static, Random, MDP
- Analysis: standard error, factor analysis, regression (linear, quadratic)

Highlighting toolbar [Hui & Boutilier 08]

Goal model

- Data collection: *online-adaptation*
- Simulation: Static, Freq-Char, Freq-only, Goal-Char, Goal-only
- Usability: 12 users

Experienced utility elicitation

- Data collection: 38 users
- Analysis: t-test, Hotelling's T2 test

Case Studies

Word completion [Hui & Boutilier 06; ACM Finals 07]

Bayesian user characteristics model

- Data collection: 45 users
- Simulation: Always, Never, MEU, Thresh
- Usability: 4 users
- Analysis: standard error, EM, KL-divergence, factor analysis

Adaptive menu [Hui et al. 09]

Probabilistic mental model

- Data collection: 48 users
- Simulation: Best-Static, Split-4, JES, WER(.1), WER(.5), WER(.9)
- Usability: 8 users
- Analysis: correlation, factorial ANOVA, regression (linear, log, Gaussian)

Typing [Hui et al. 08]

Occlusion model

- Data collection: 12 users
- Analysis: factor analysis, regression (linear, quadratic, cubic)

Menu selection [Hui et al. 08]

Bloat model

- Data collection: 12 users
- Simulation: Static, Random, MDP
- Analysis: standard error, factor analysis, regression (linear, quadratic)

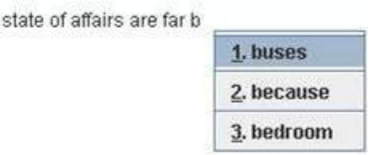
Highlighting toolbar [Hui & Boutilier 08]

Goal model

- Data collection: *online-adaptation*
- Simulation: Static, Freq-Char, Freq-only, Goal-Char, Goal-only
- Usability: 12 users

Experienced utility elicitation

- Data collection: 38 users
- Analysis: t-test, Hotelling's T2 test

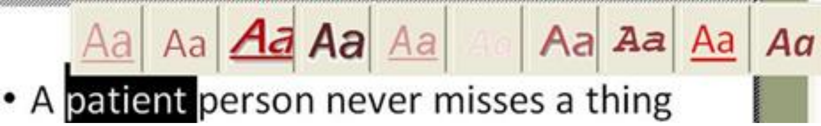


Findings – Char. Model

- **Inference is feasible**
  - 36 user types
  - Convergence between 20-150 words
  - System behaviour adapts to user characteristics

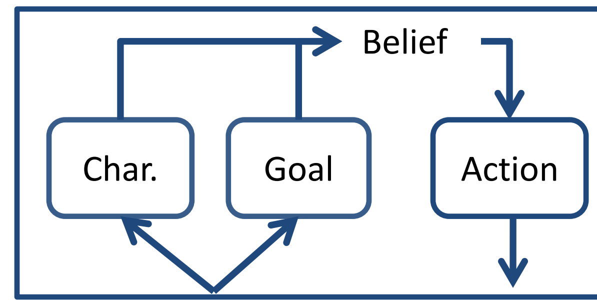
- **Users perceive help utility**
  - Joint expected savings vs. bigrams (ML)
    - ~2% more exact word matches
    - ~11% more character savings
    - *Greedy*: ~6.4% more character savings
  - Users accept 20% *inexact* matches

E.g., “the nu”:
- number
- nuclear
- nurses



Findings – Goal Model

- **Adaptive help saves user effort**
  - Average event reduction
    Freq: 13%
    Goal: 22%
    Goal-Char: 21% (w.r.t. NeverHelp)
  - % Suggestions accepted
    Goal: 3.5x
    Goal-Char: 3.3x (w.r.t. Freq)



- **Users like Goal/Goal-Char better than baseline**
  - Personalization and Helpfulness:
    Goal >> Freq ($p < 0.05$)
    Goal-Char >> Freq ($p < 0.05$)
  - Incremental inference more suited for sequential task

Summary

- **Intelligent assistance as decision-theoretic planning problem**
  - Propose POMDP-SCA framework
  - Model both user characteristics and user goals
- **Formal user models and parameter acquisition experiments**
  - Incorporate user behaviour and feedback
  - Explain interaction preferences
  - Monitor changes in characteristics over time
  - Recognize personalized goals
- **Learning interaction preferences**
  - Learn interaction cost models
  - Trade off adaptation benefits with interaction costs
  - Elicit experiential preferences
  - Use joint action selection for wider coverage of true goal