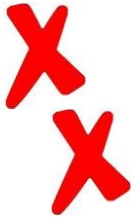
Who's Asking for Help? A Bayesian Approach to Intelligent Assistance

Bowen Hui and Craig Boutilier
{bowen,cebly}@cs.utoronto.ca

IUI 2006, January 29th – February 1st, Sydney, Australia

Need for software customization

- increasing complexity
  - lost in interface/functionality
  - repeated customization effort



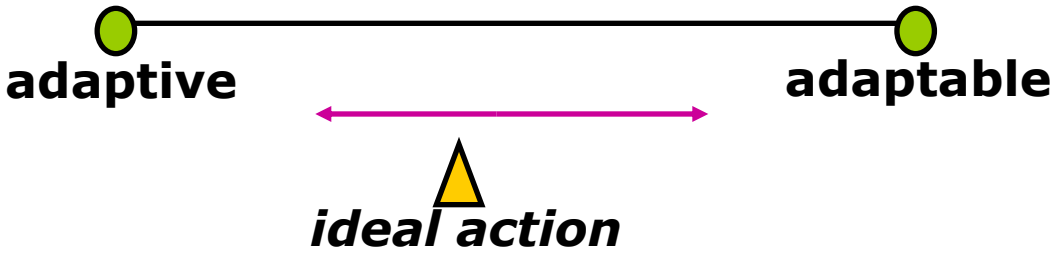
- most affected users
  - cognitive, sensory, motor impairments
  - elderly
  - children
  - novices

Approaches to customization

- adaptable – more user effort
- adaptive – when, what, how?
- hybrid – fixed, doesn't learn about user

Approaches to customization

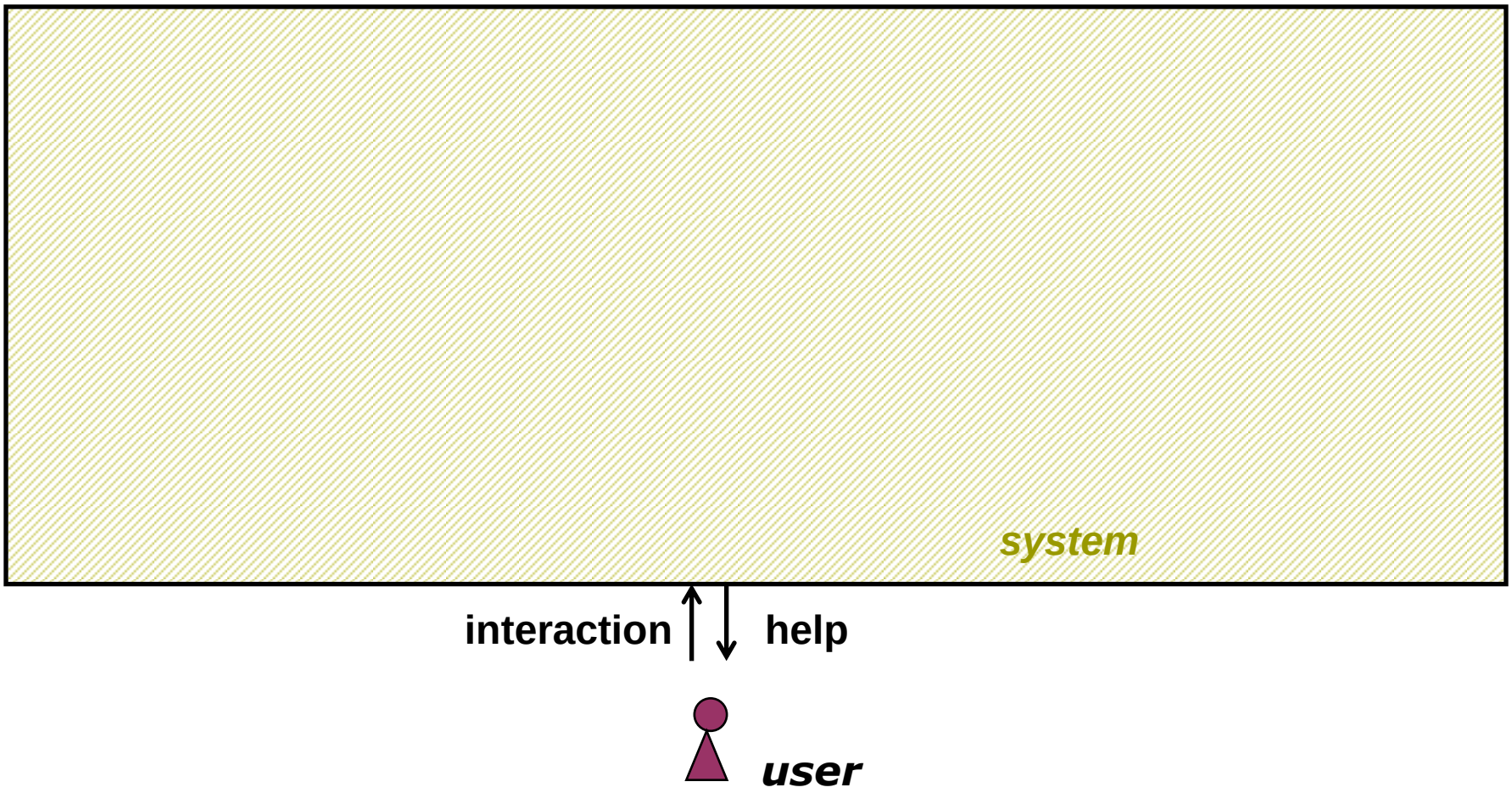
- adaptable – more user effort
- adaptive – when, what, how?
- hybrid – fixed, doesn't learn about user
- continuum:



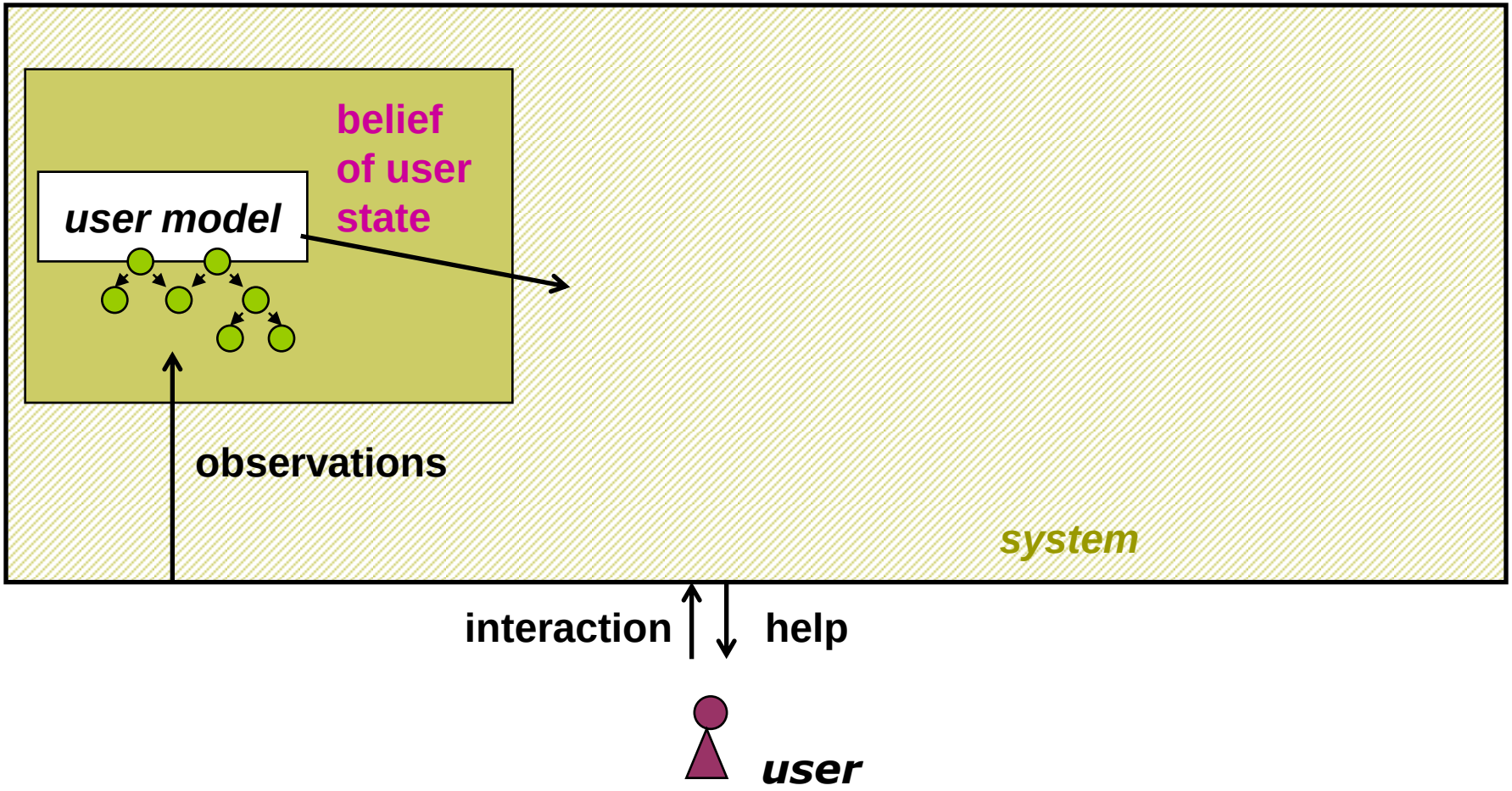
Decision-theoretic assistance

- identify relevant user features
- individual cost models
  - interruptions
  - quality of actions
  - partial suggestions
  - “off-topic” help
- utility of predictions
- act w.r.t. uncertainty

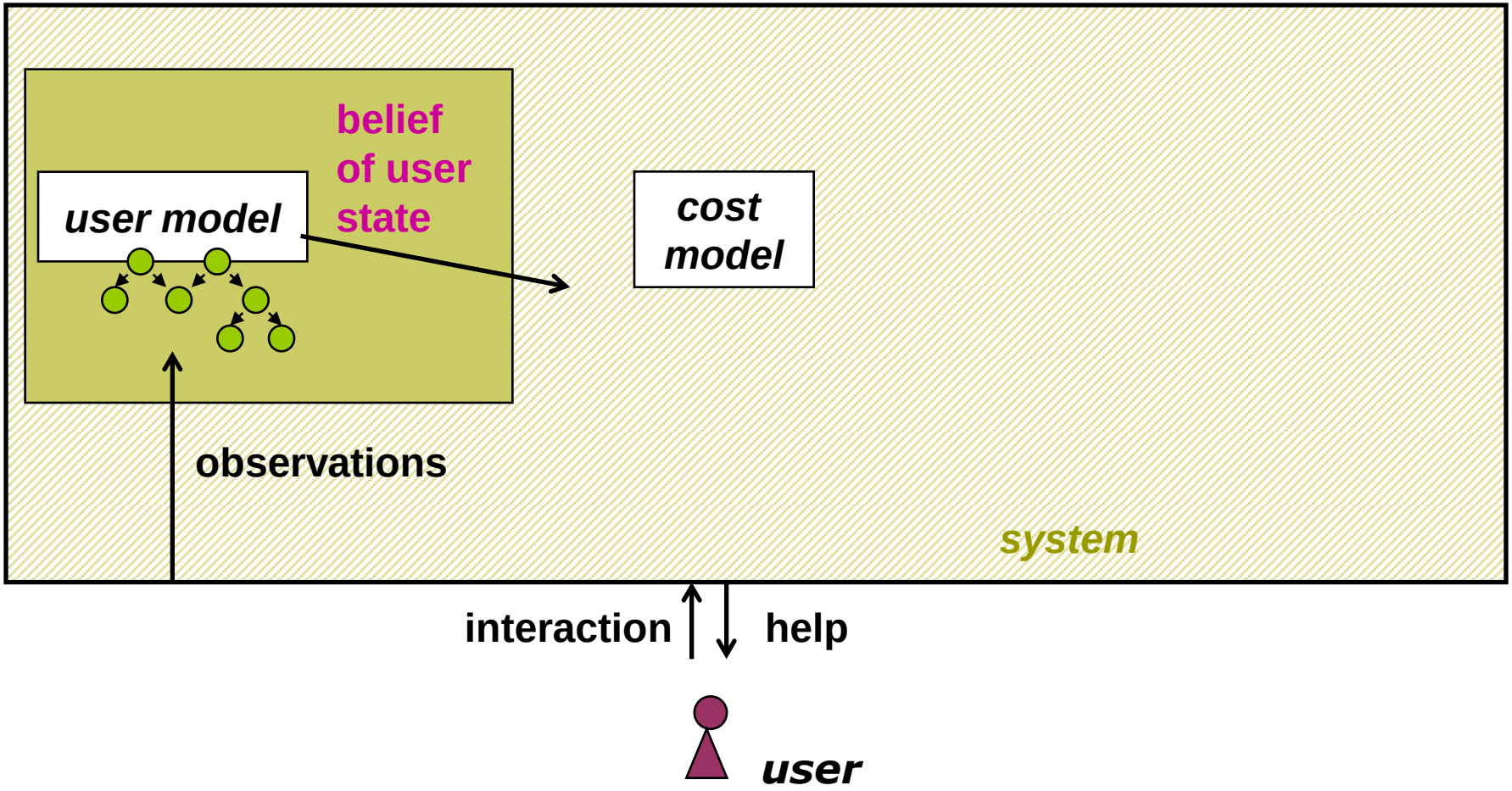
Decision-theoretic assistance



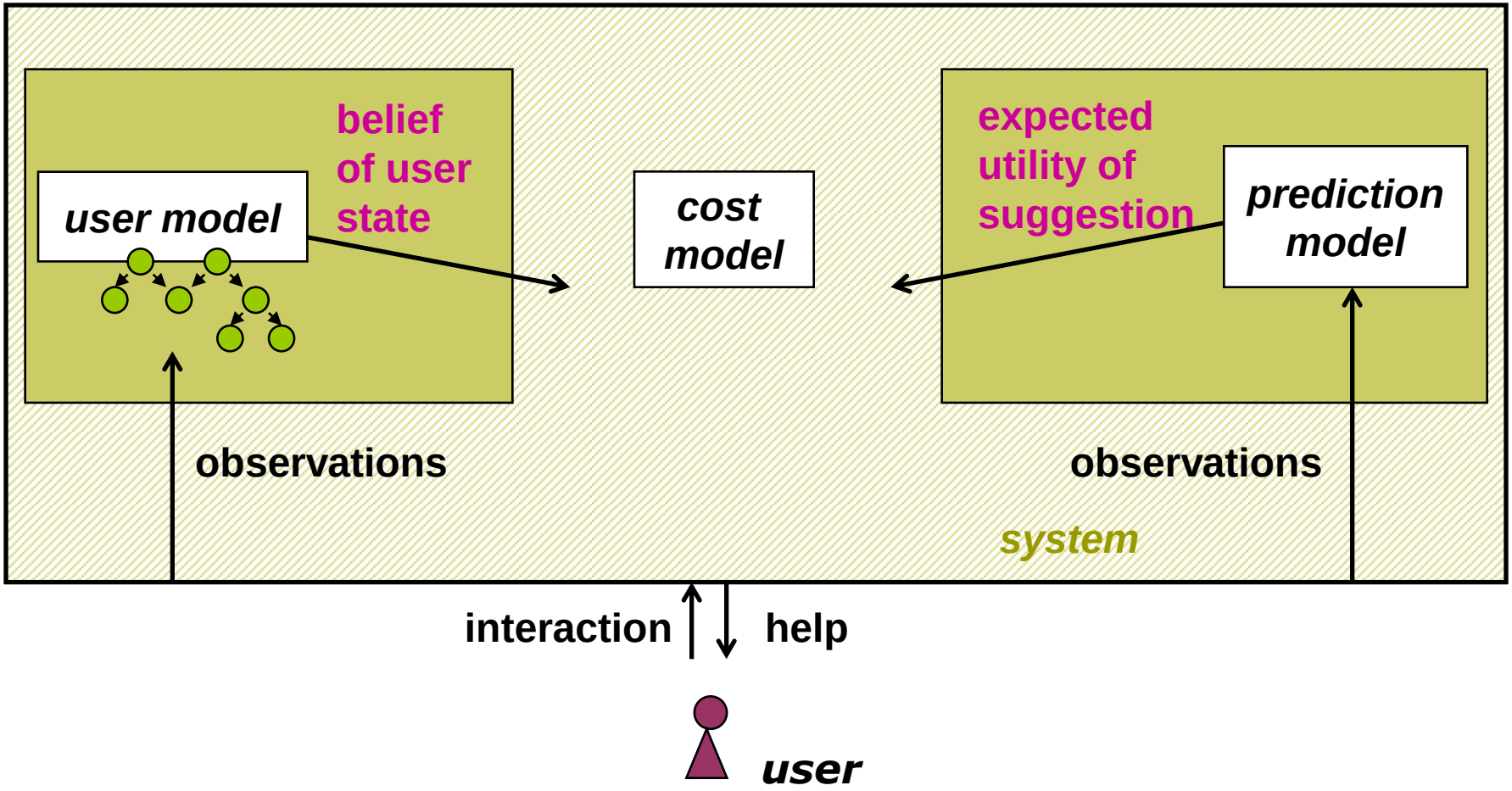
Decision-theoretic assistance



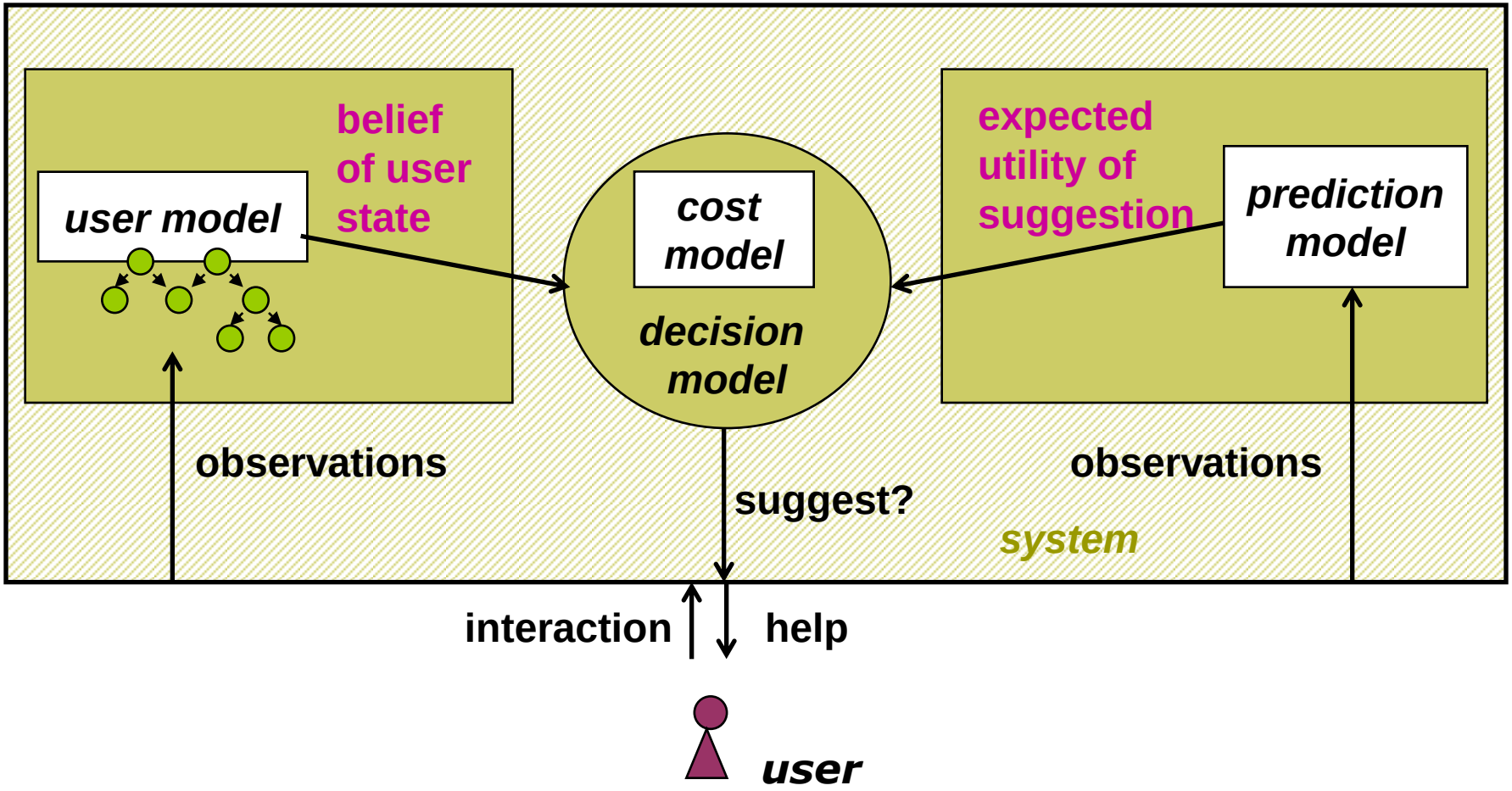
Decision-theoretic assistance



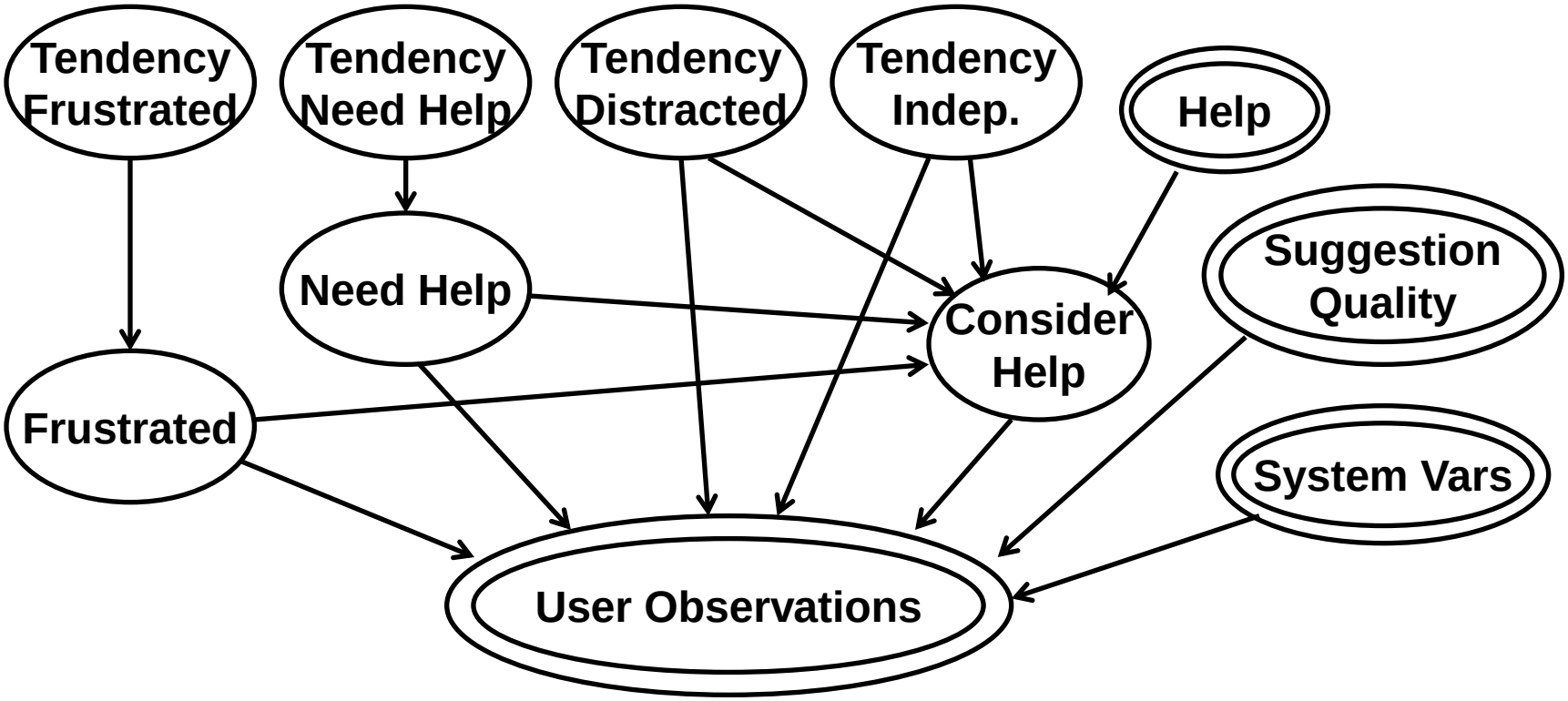
Decision-theoretic assistance



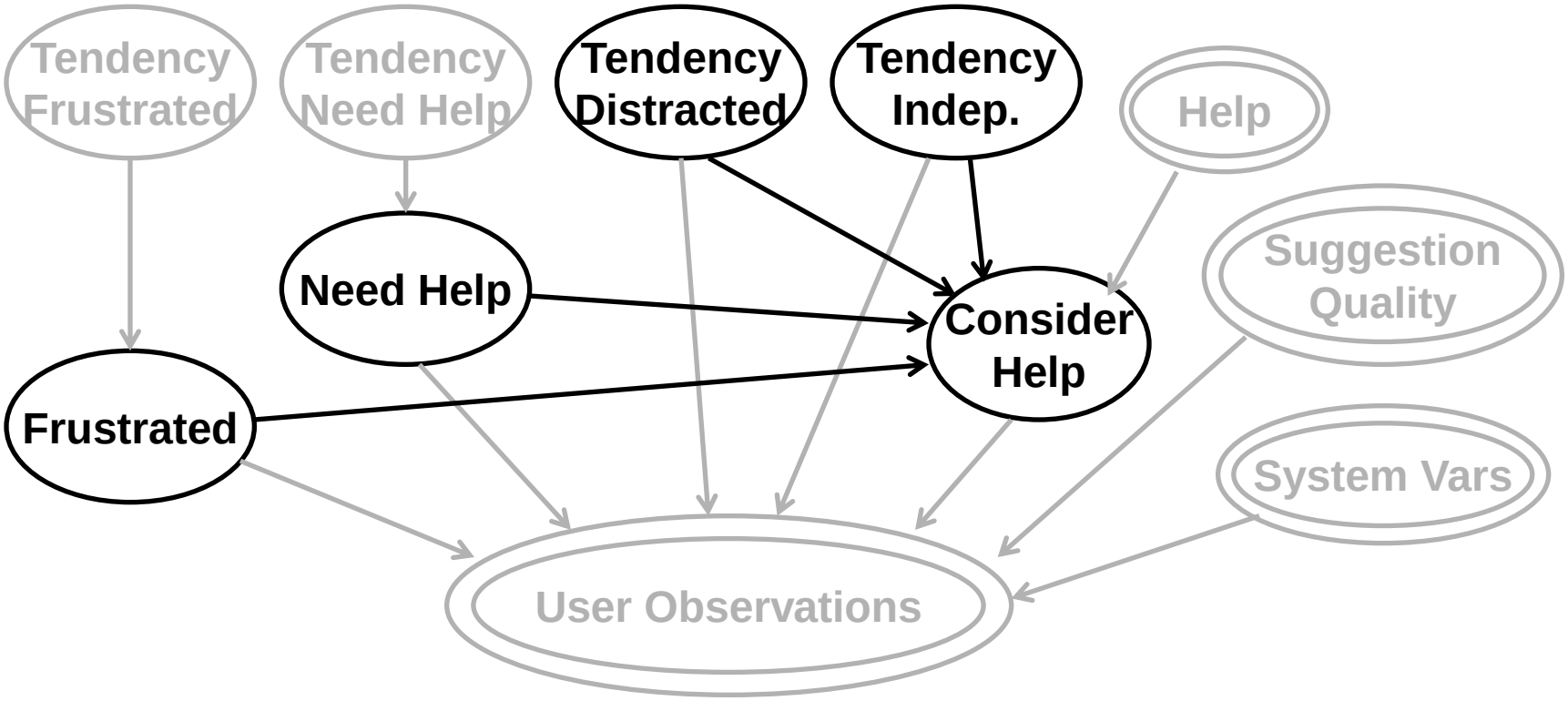
Decision-theoretic assistance



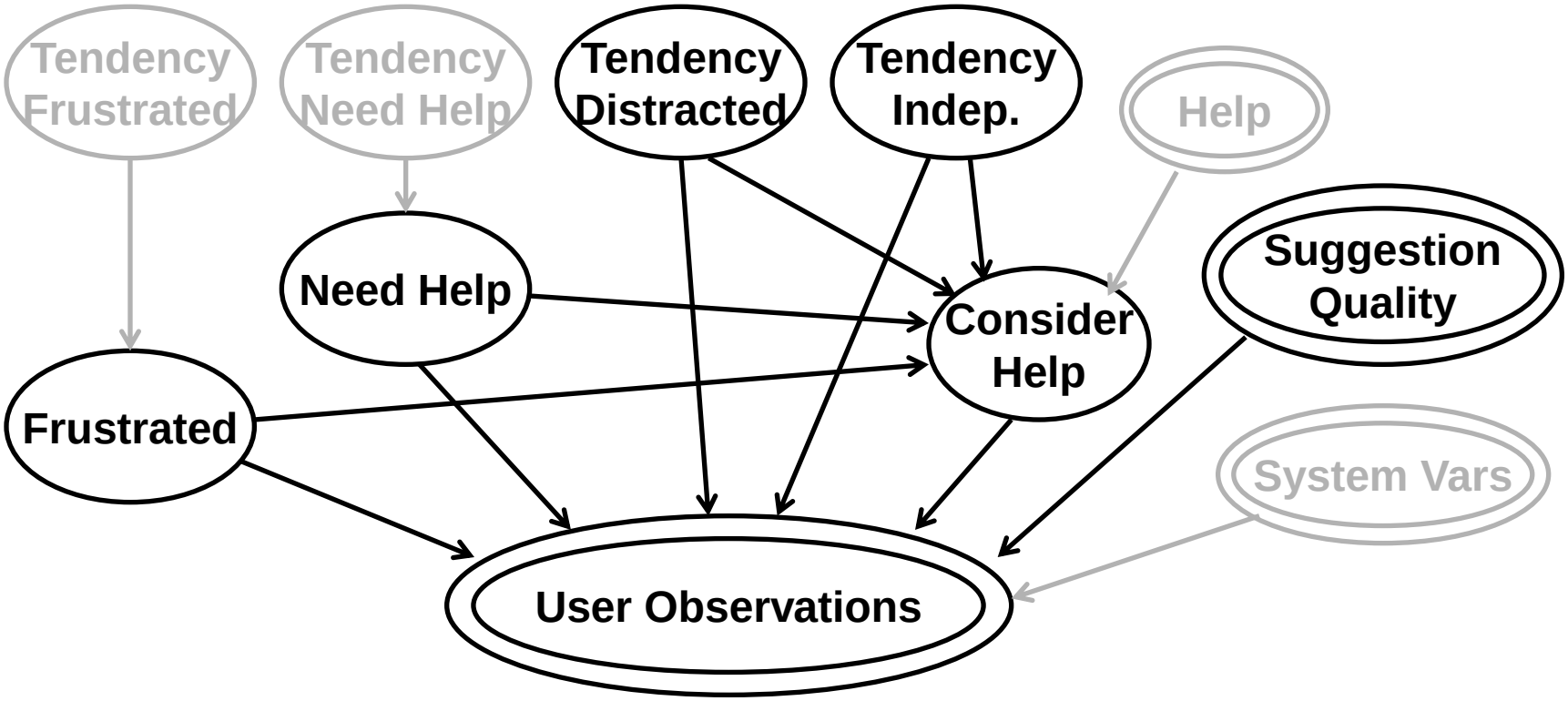
Generic user model



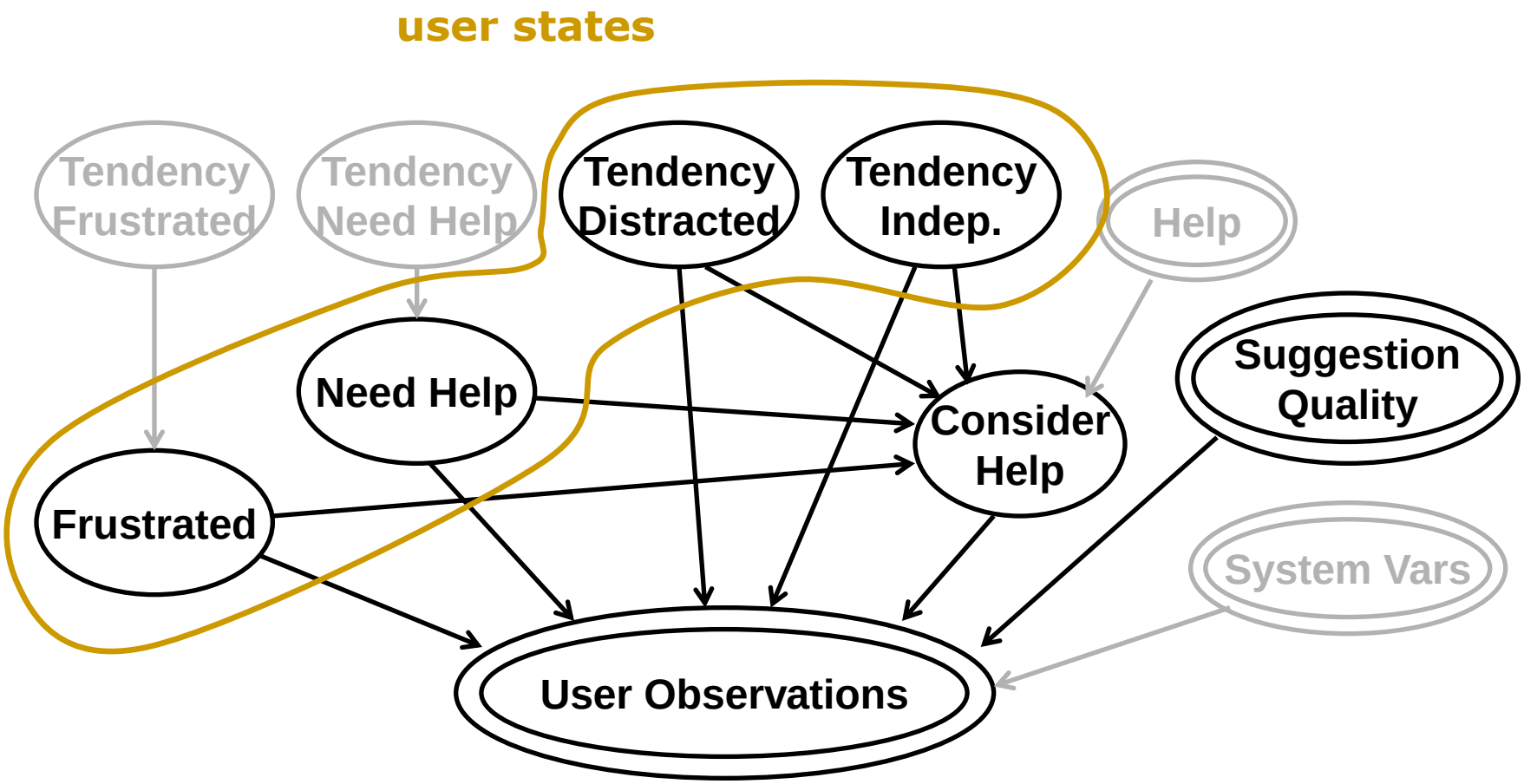
Generic user model



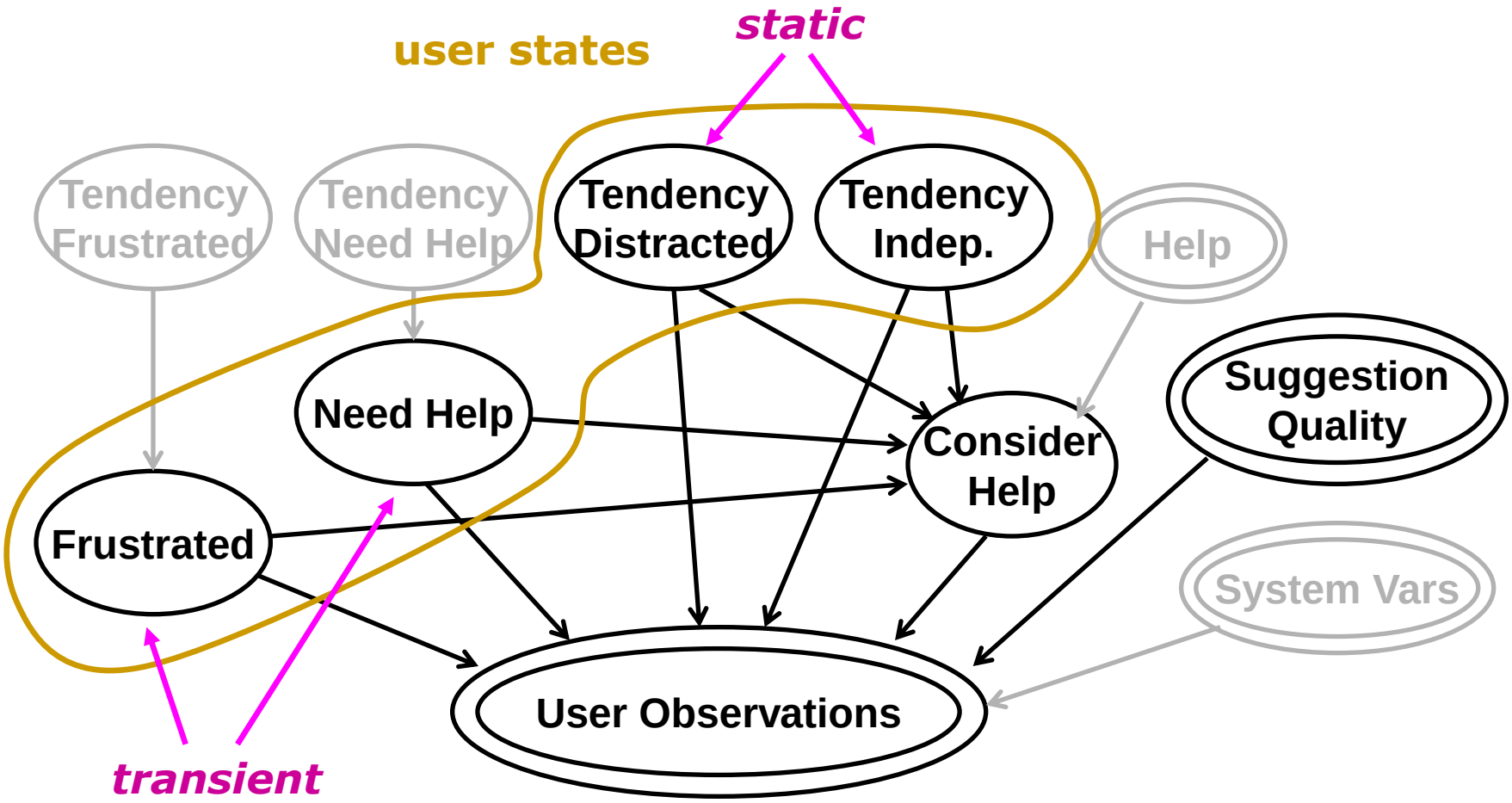
Generic user model



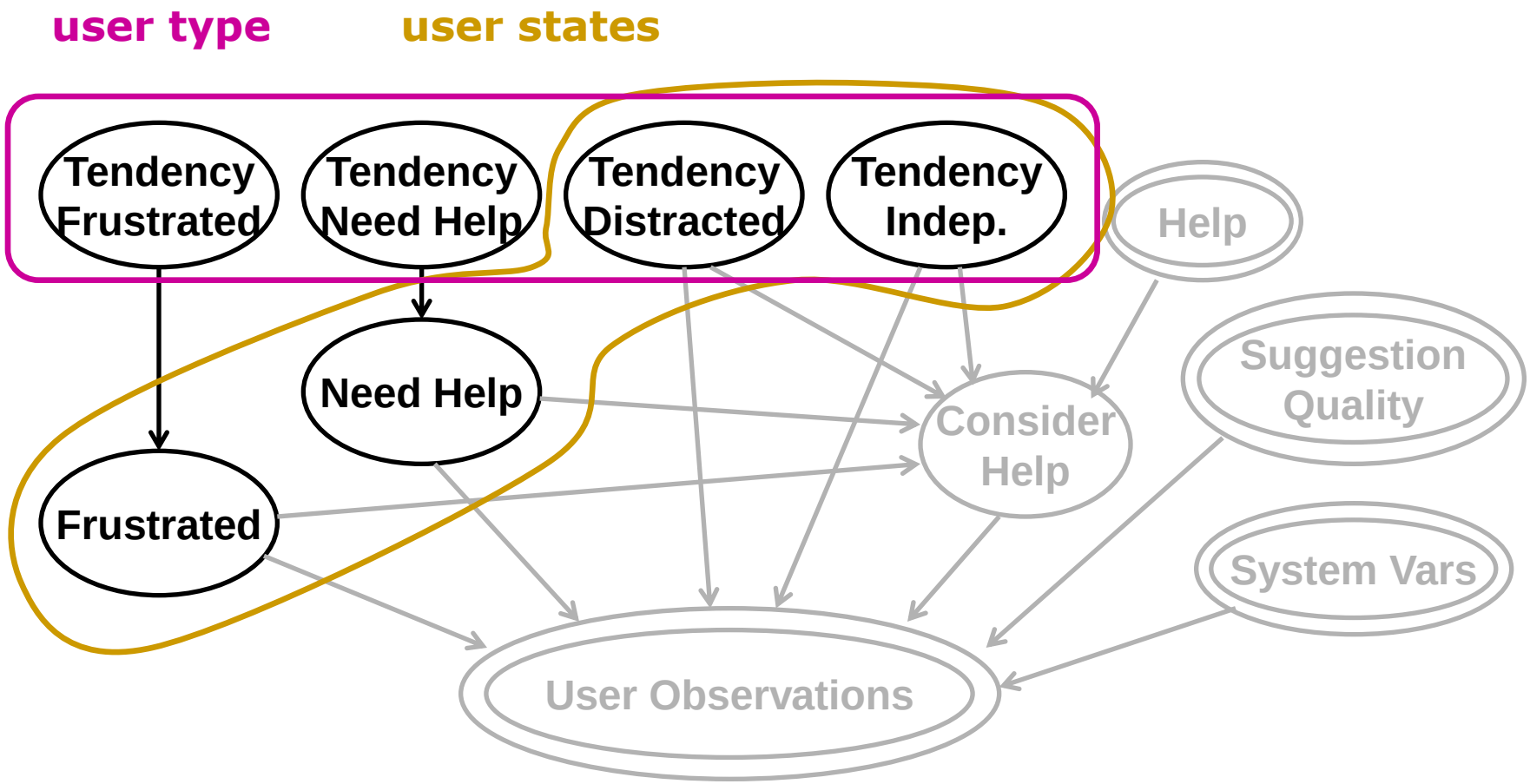
Generic user model



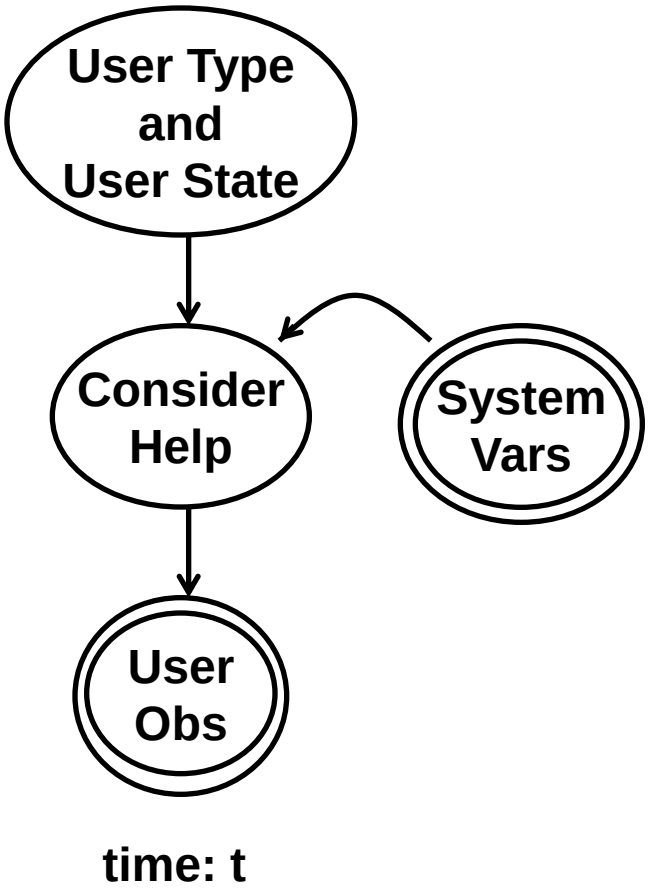
Generic user model



Generic user model

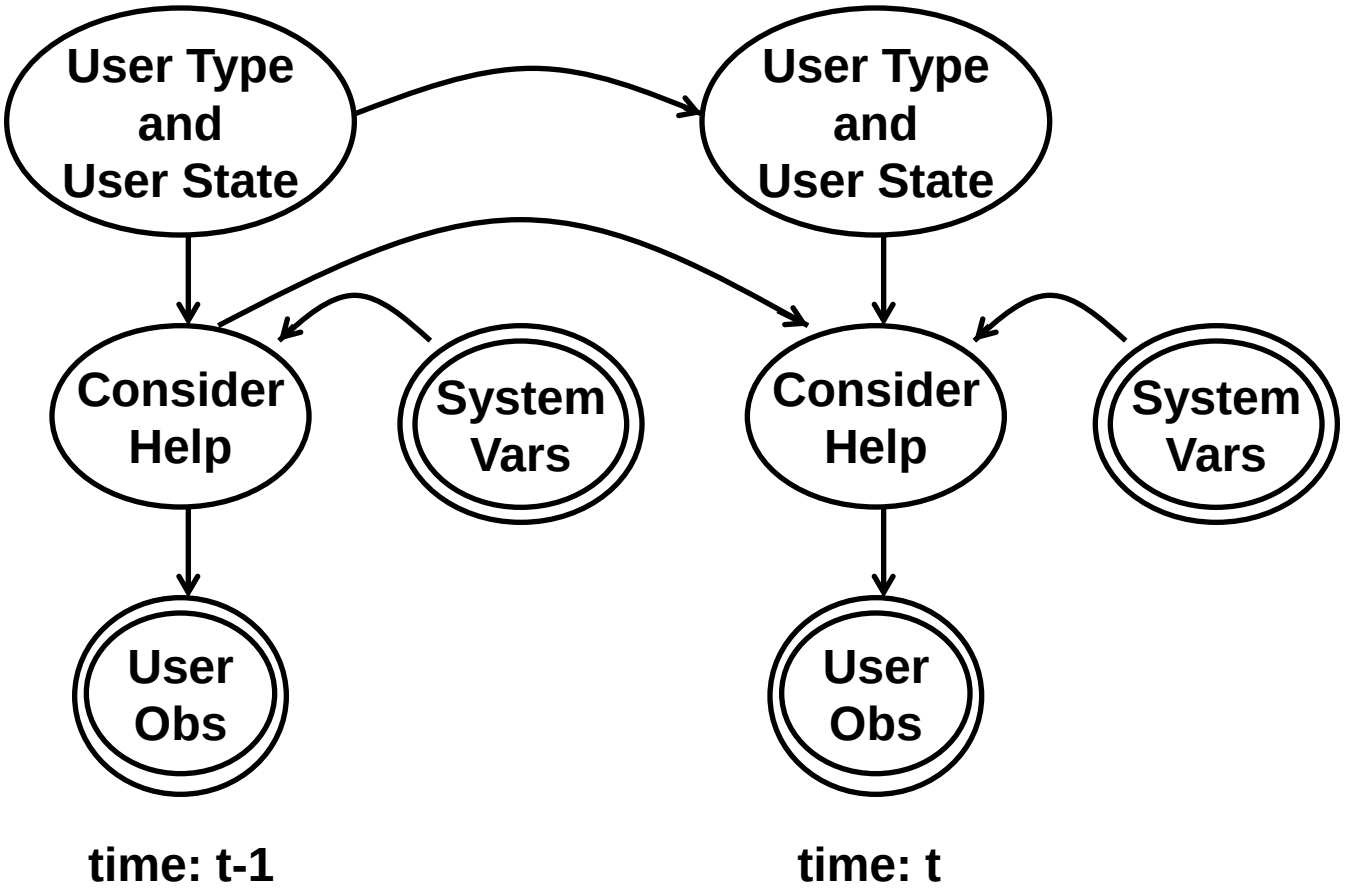


Dynamic Bayesian network (DBN)

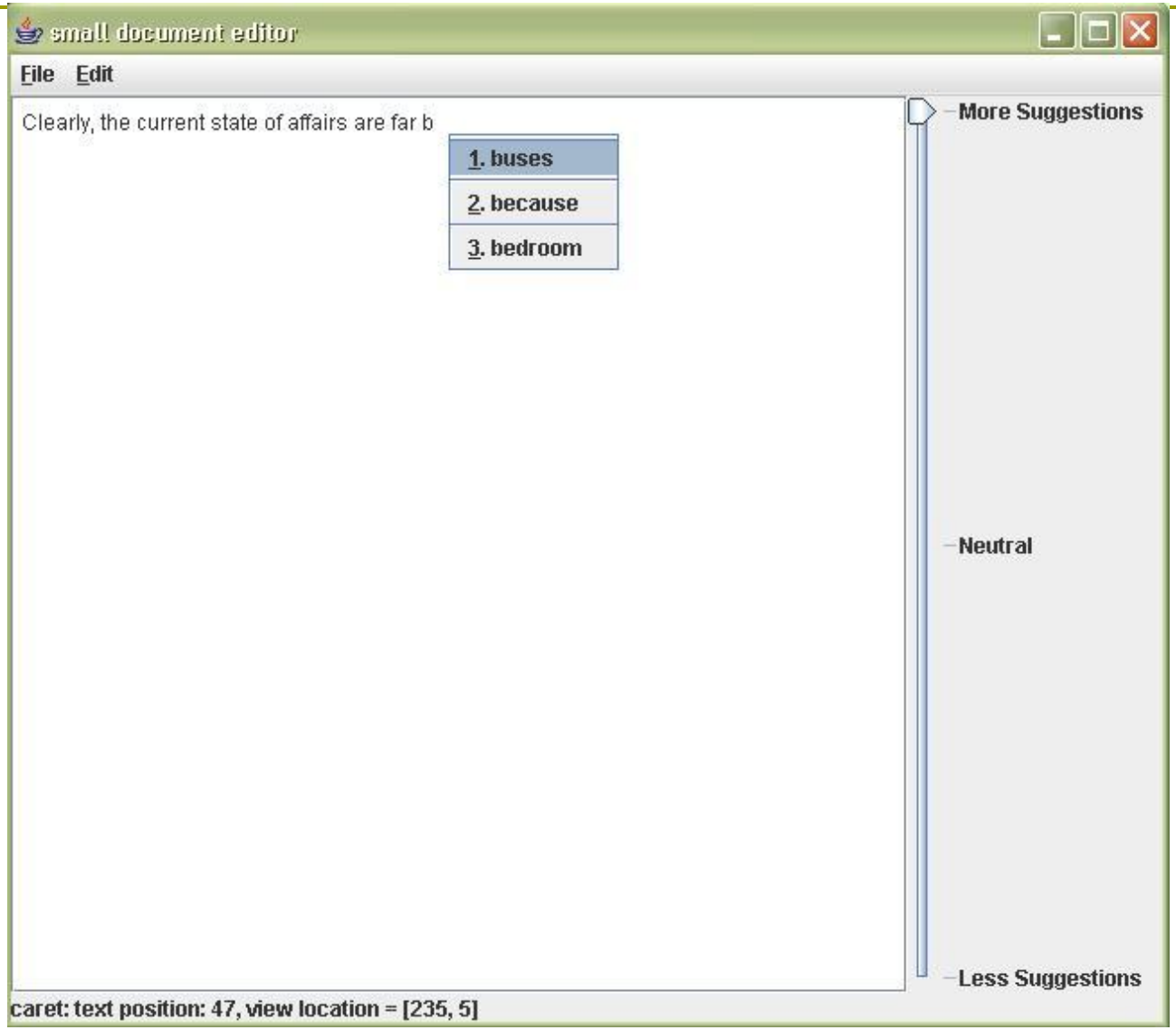


Dynamic Bayesian network (DBN)

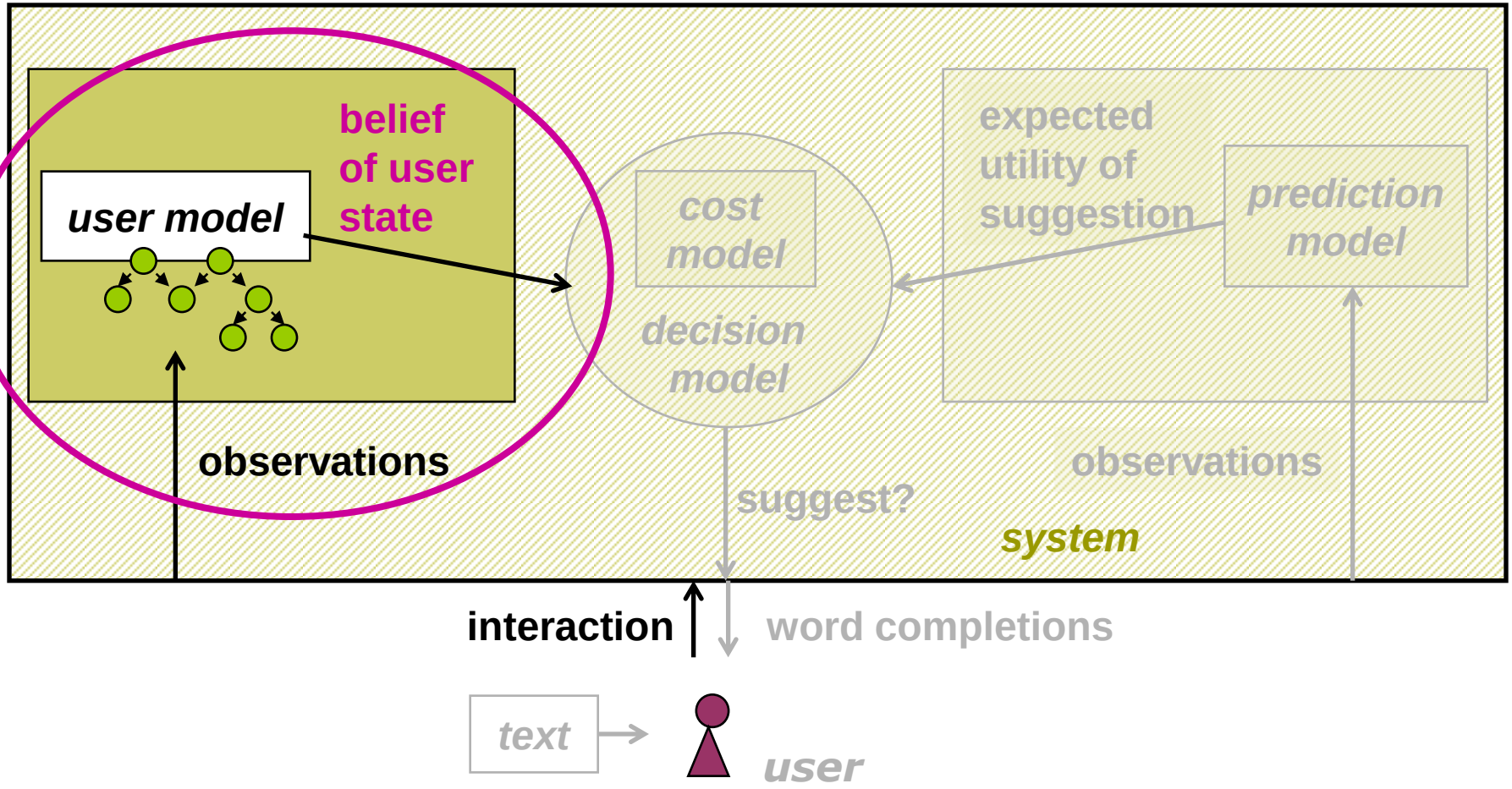
- capture user dynamics



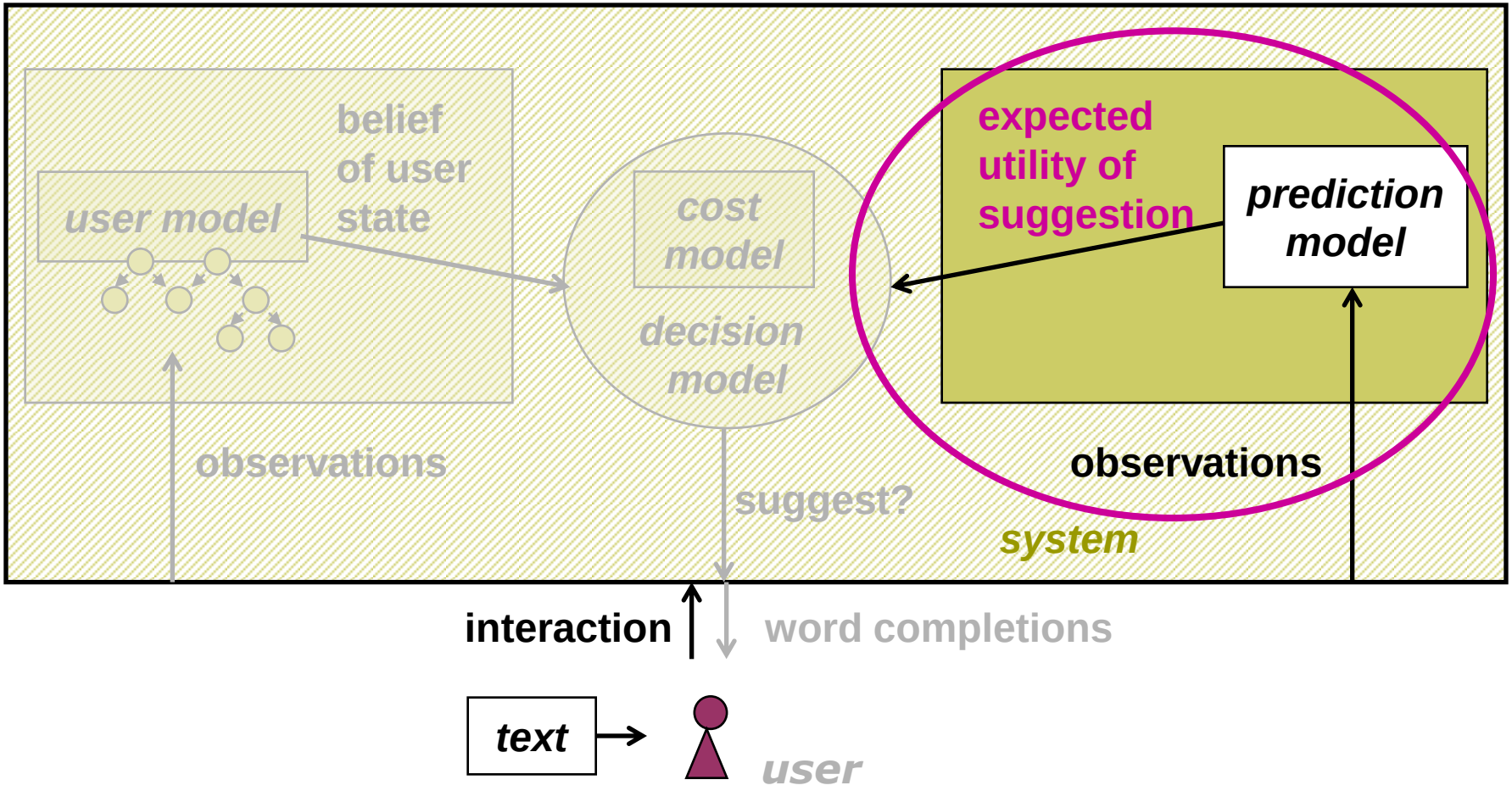
Word prediction prototype



Word prediction prototype



Word prediction prototype



Joint expected savings

- plausible set: $\{c_1,\ldots,c_K\}$
- each p_k corresponds to c_k

} **bigram**

Joint expected savings

- plausible set: $\{c_1,\ldots,c_K\}$
- each p_k corresponds to c_k

} **bigram**

- utility $U(c_{offer}|s)$ = #(same) – #(erase) – #(add)
  - U("apple"|"application") = 4 – 1 – 7 = –4

Joint expected savings

- plausible set: $\{c_1,\ldots,c_K\}$
- each p_k corresponds to c_k

} **bigram**

- utility $U(c_{offer}|s)$ = #(same) – #(erase) – #(add)
  - U("apple"|"application") = 4 – 1 – 7 = –4

- expected savings $ES(c_{offer}) = \sum_{i=1}^{K} U(c_{offer}|c_i)p_i$

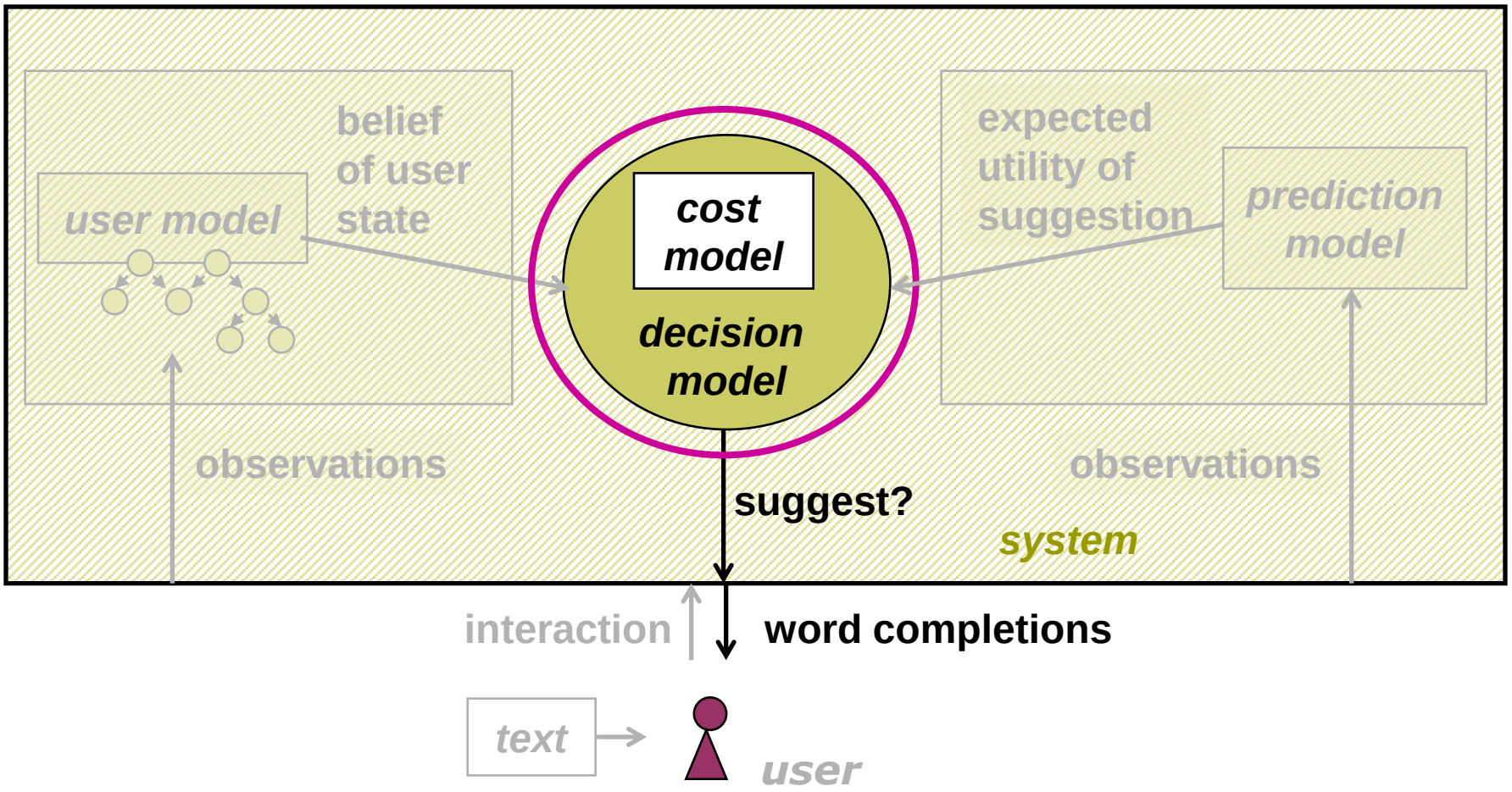
Joint expected savings

- plausible set: $\{c_1,\ldots,c_K\}$
- each p_k corresponds to c_k

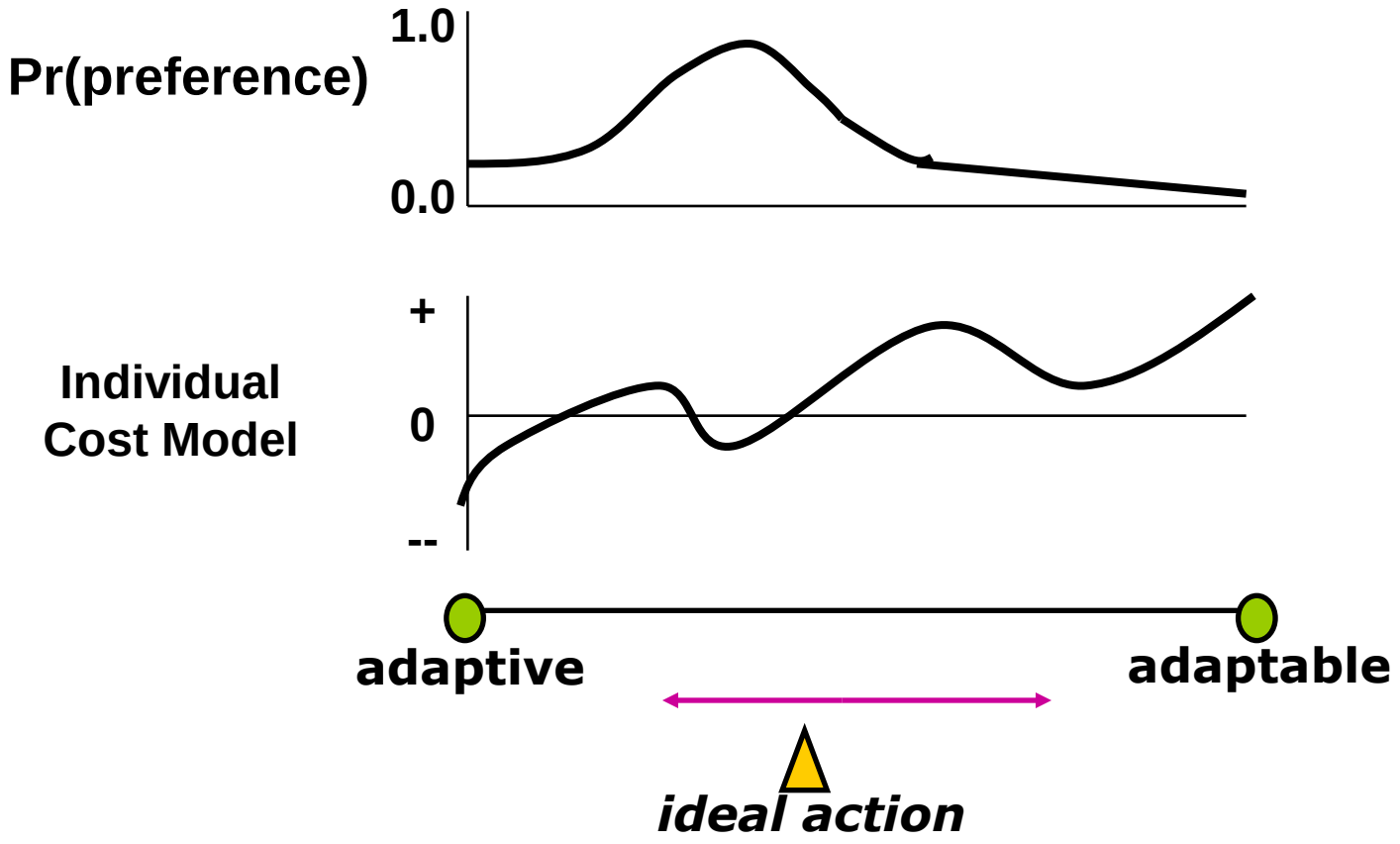
} **bigram**

- utility $U(c_{offer}|s)$ = #(same) – #(erase) – #(add)
  - U("apple"|"application") = 4 – 1 – 7 = –4
- expected savings $ES(c_{offer}) = \sum_{i=1}^{K} U(c_{offer}|c_i)p_i$
- joint $JES(c_1,\ldots,c_J) = \sum_{i=1}^{K} \underset{c_j}{\text{argmax}}\, U(c_j|c_i)p_i$

Word prediction prototype



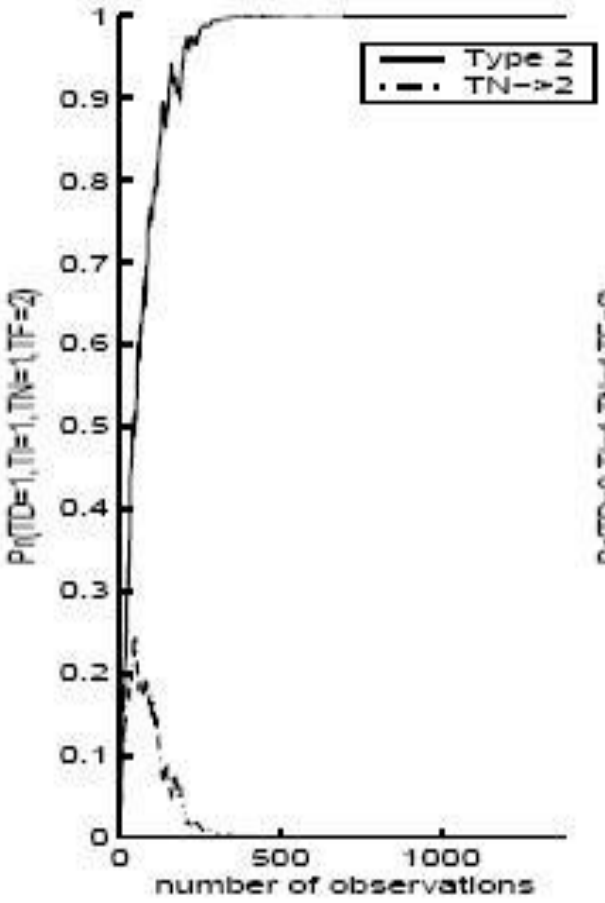
Bayesian action selection

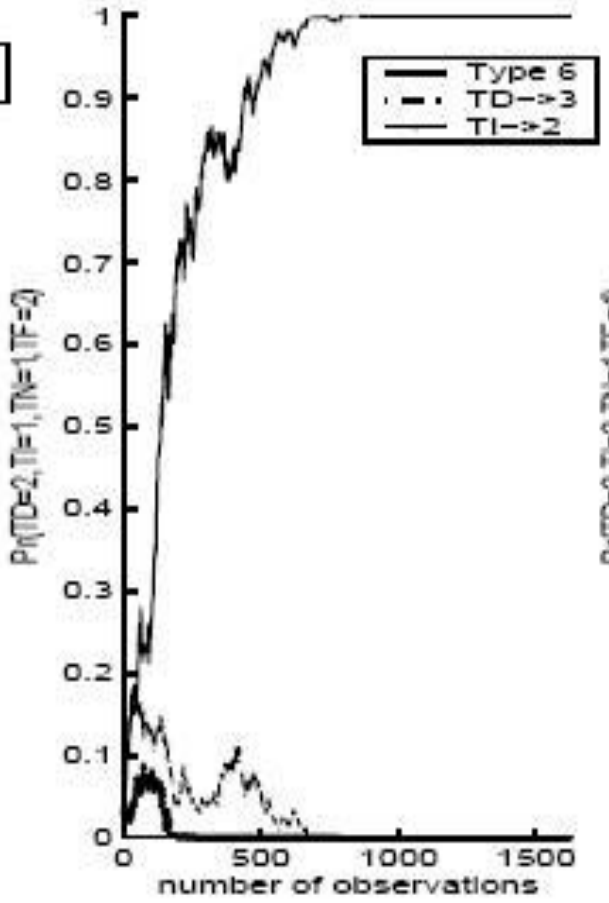


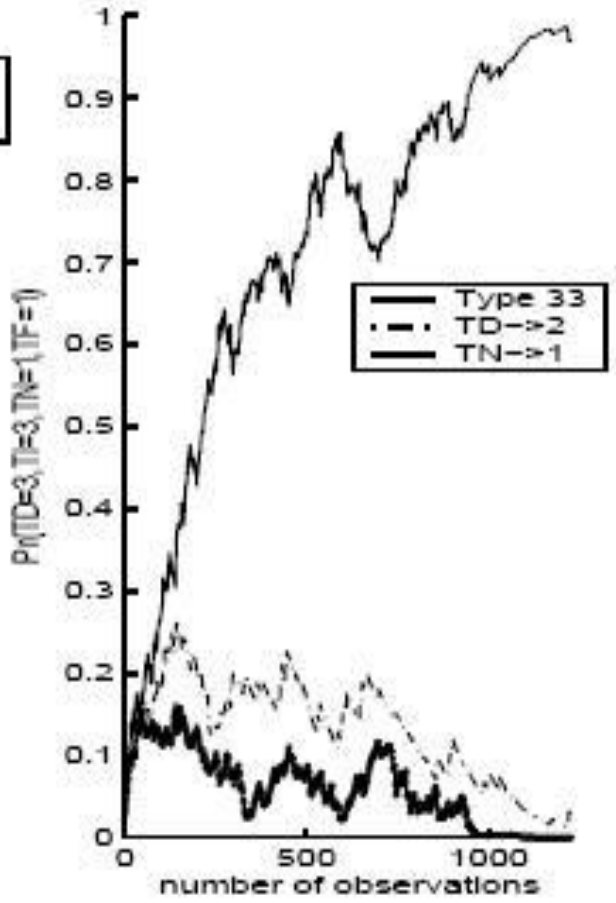
Simulations

- user model
  - converge to user type?
- prediction model
  - value of utility definition?
- cost model
  - effect of personalization?
- short text ~ 300 words

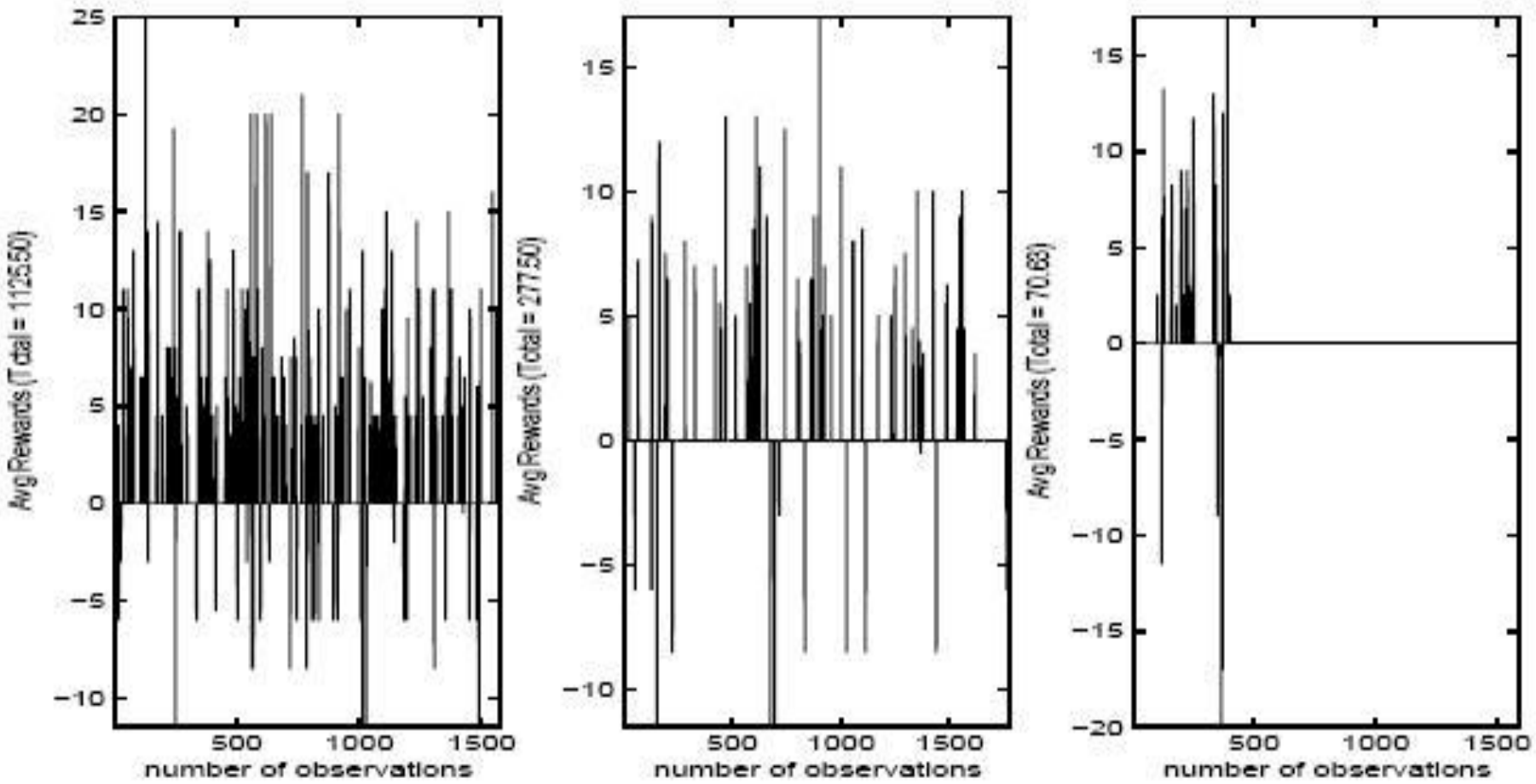
User convergence







Impact of personalization



Policy comparison

- average reward per time step for

{Frustrable, Needy, Distractible, Independent}

User Type	ALWAYS	MEU	THRESHOLD	NEVER
{_F,¯N,_D,_I}	1.64	0.93	0.91	0
{¯F,_N,−D,_I}	0.46	0.62	0.65	0
{_F,_N,¯D,−I}	-0.64	0.39	0.31	0
{¯F,_N,_D,¯I}	-10.89	-1.89	-2.93	0
{¯F,_N,¯D,¯I}	-6.04	-0.07	-1.43	0

Policy comparison

- average reward per time step for

{Frustrable, Needy, Distractible, Independent}

User Type	ALWAYS	MEU	THRESHOLD	NEVER
{_F,¯N,_D,_I}	1.64	0.93	0.91	0
{¯F,_N,−D,_I}	0.46	0.62	0.65	0
{_F,_N,¯D,−I}	-0.64	0.39	0.31	0
{¯F,_N,_D,¯I}	-10.89	-1.89	-2.93	0
{¯F,_N,¯D,¯I}	-6.04	-0.07	-1.43	0

Policy comparison

- average reward per time step for

{Frustrable, Needy, Distractible, Independent}

User Type	ALWAYS	MEU	THRESHOLD	NEVER
{_F,¯N,_D,_I}	1.64	0.93	0.91	0
{¯F,_N,−D,_I}	0.46	0.62	0.65	0
{_F,_N,¯D,−I}	-0.64	0.39	0.31	0
{¯F,_N,_D,¯I}	-10.89	-1.89	-2.93	0
{¯F,_N,¯D,¯I}	-6.04	-0.07	-1.43	0

Learning user model parameters

- what kind of behavioural observations?
- 3 conditions
- elicitation
  - online questions (F,N)
  - post-questionnaire (TF,TD,TI)
  - vocab and typing speed (TN)
- 45 participants

Pilot usability experiment

- what do people prefer systems do?
- 4 conditions
  - ALWAYS
  - NEVER
  - THRESHOLD
  - MEU
- 4 participants
- short text ~ 70 words

Preliminary usability results

- typing speed 4-8 words/min (¯Needy,_Indep)

Preliminary usability results

- typing speed 4-8 words/min (¯Needy,_Indep)

- subjective: ALWAYS >> adaptive >> NEVER
- typing effort: NEVER >> adaptive >> ALWAYS

Preliminary usability results

- typing speed 4-8 words/min (¯Needy,_Indep)

- subjective: ALWAYS >> adaptive >> NEVER
- typing effort: NEVER >> adaptive >> ALWAYS

- subjective: MEU >> THRESH, quality
- num accept, per correct help: THRESH >> MEU
- 20% partial accepts

Preliminary usability results

- typing speed 4-8 words/min (¯Needy,_Indep)

- subjective: ALWAYS >> adaptive >> NEVER
- typing effort: NEVER >> adaptive >> ALWAYS

- subjective: MEU >> THRESH, quality
- num accept, per correct help: THRESH >> MEU
- 20% partial accepts

users perceive *utility* of suggestions

Summary of contributions

- HCI as decision making problem
- generic model of user preferences
- learn user features online
- individual costs/rewards
- Bayesian action selection