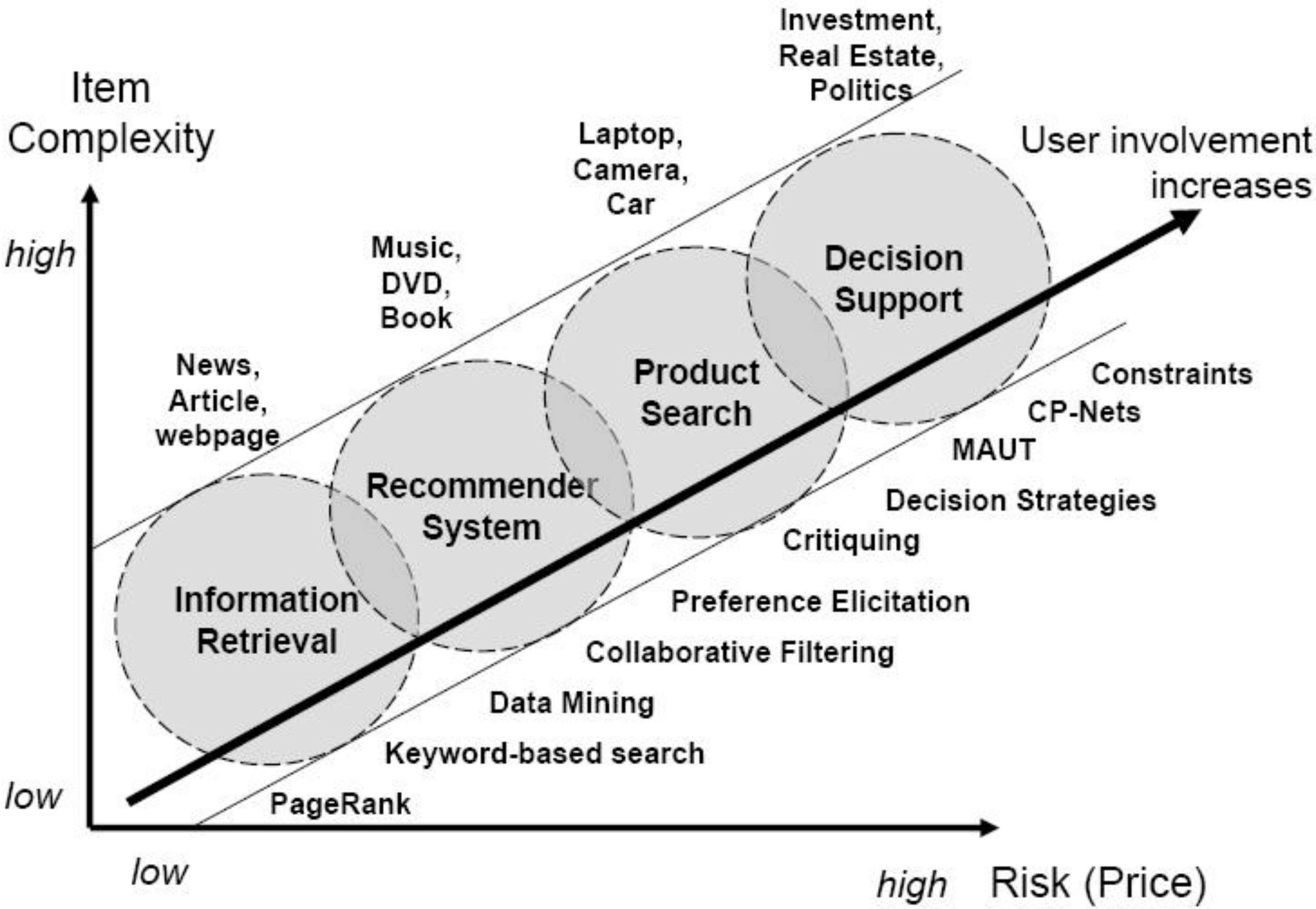
Critiquing-based Recommender Systems and User Experiences

Dr. Li CHEN
Department of Computer Science
Hong Kong Baptist University

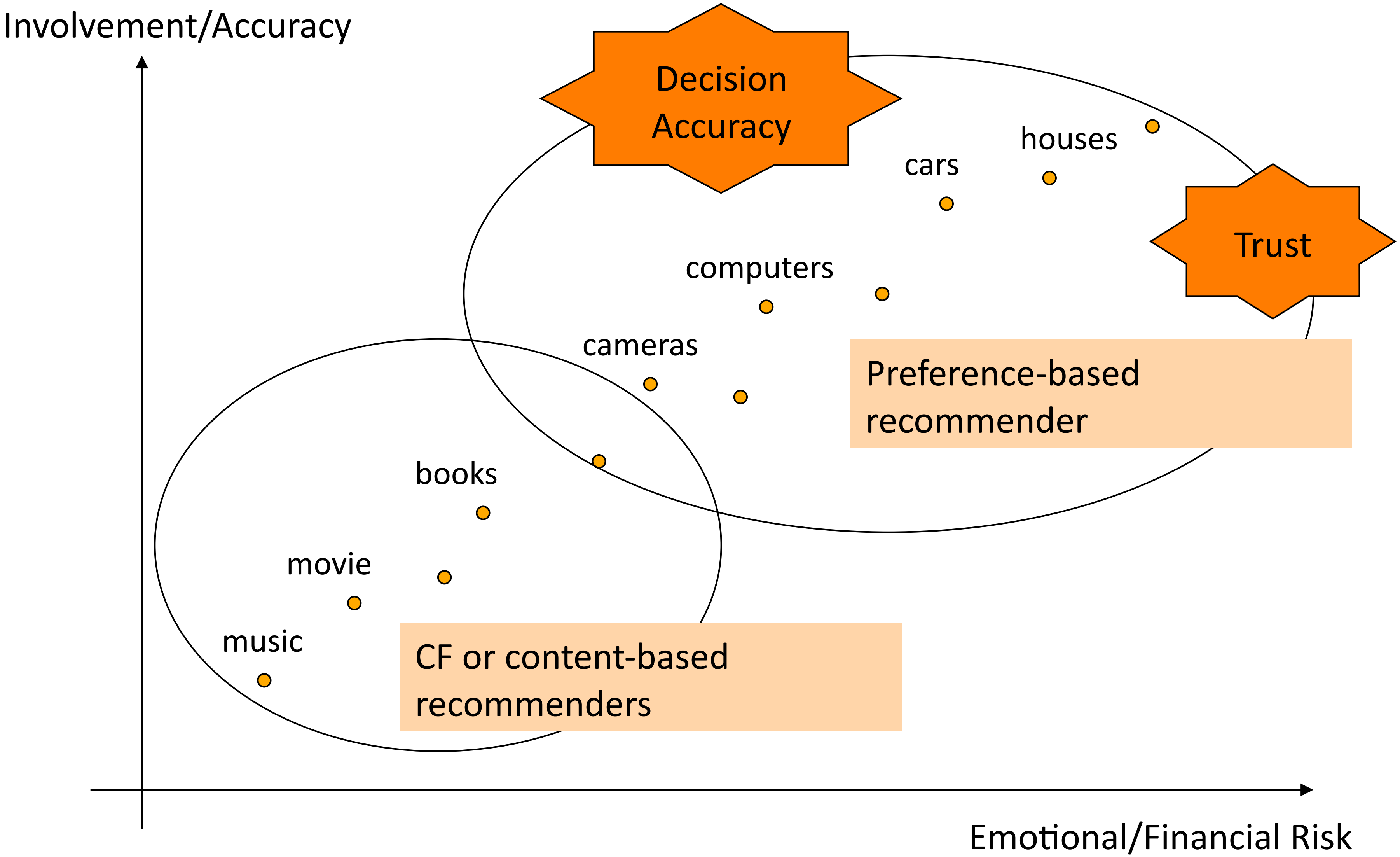
Homepage: http://www.comp.hkbu.edu.hk/~lichen/
Email: lichen@comp.hkbu.edu.hk

Outline

- What is critiquing-based recommender system and Why?
- Development history
- User experiences
- Conclusion


Item
Complexity
high
low
low
high
Risk (Price)
User involvement
increases
Investment,
Real Estate,
Politics
Laptop,
Camera,
Car
Music,
DVD,
Book
News,
Article,
webpage
Information
Retrieval
Recommender
System
Product
Search
Decision
Support
Constraints
CP-Nets
MAUT
Decision Strategies
Critiquing
Preference Elicitation
Collaborative Filtering
Data Mining
Keyword-based search
PageRank

For high-risk products



Challenge 1 – Adaptive decision maker

Unfamiliar product domain

Complex decision environment with overwhelming information

How do people make decisions in unfamiliar and complex environment?

- Adaptive and constructive nature of user preferences (Payne *et al.*, 1993)
- Tend to use non-compensatory decision strategies (e.g., elimination-by-aspects)
- Tradeoff avoidance due to emotional and cognitive reasons (Hogarth, 1987; Payne et al., 1999)
- Decision meta-goals: maximize the accuracy and minimize the effort (Bettman *et al.*, 1998)

http://www.youtube.com/watch?v=CKM9u65kZHg

Challenge 2 – Trust building in online environment

Replaced by

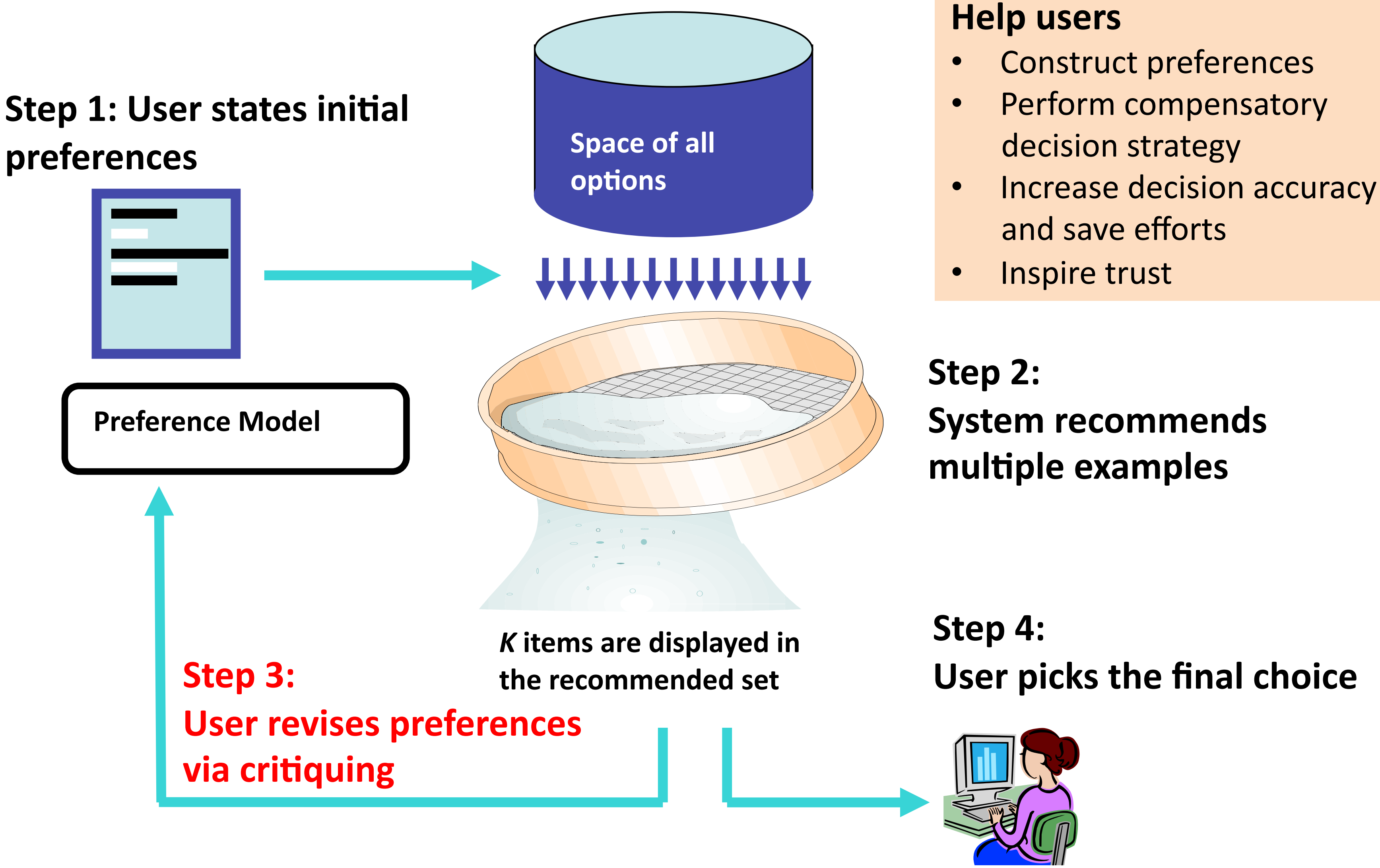
Trust is difficult to build and easy to lose

- Lack of face-to-face interaction in the online environment
- Impede customers from performing particular actions (like continue to transact, purchase, return, etc.) (Jarvenpaa *et al.*, 2000)
- Key factor to the success of e-commerce (Gefen, 2000)

Critiquing-based interaction: A **user-feedback** mechanism

Help users

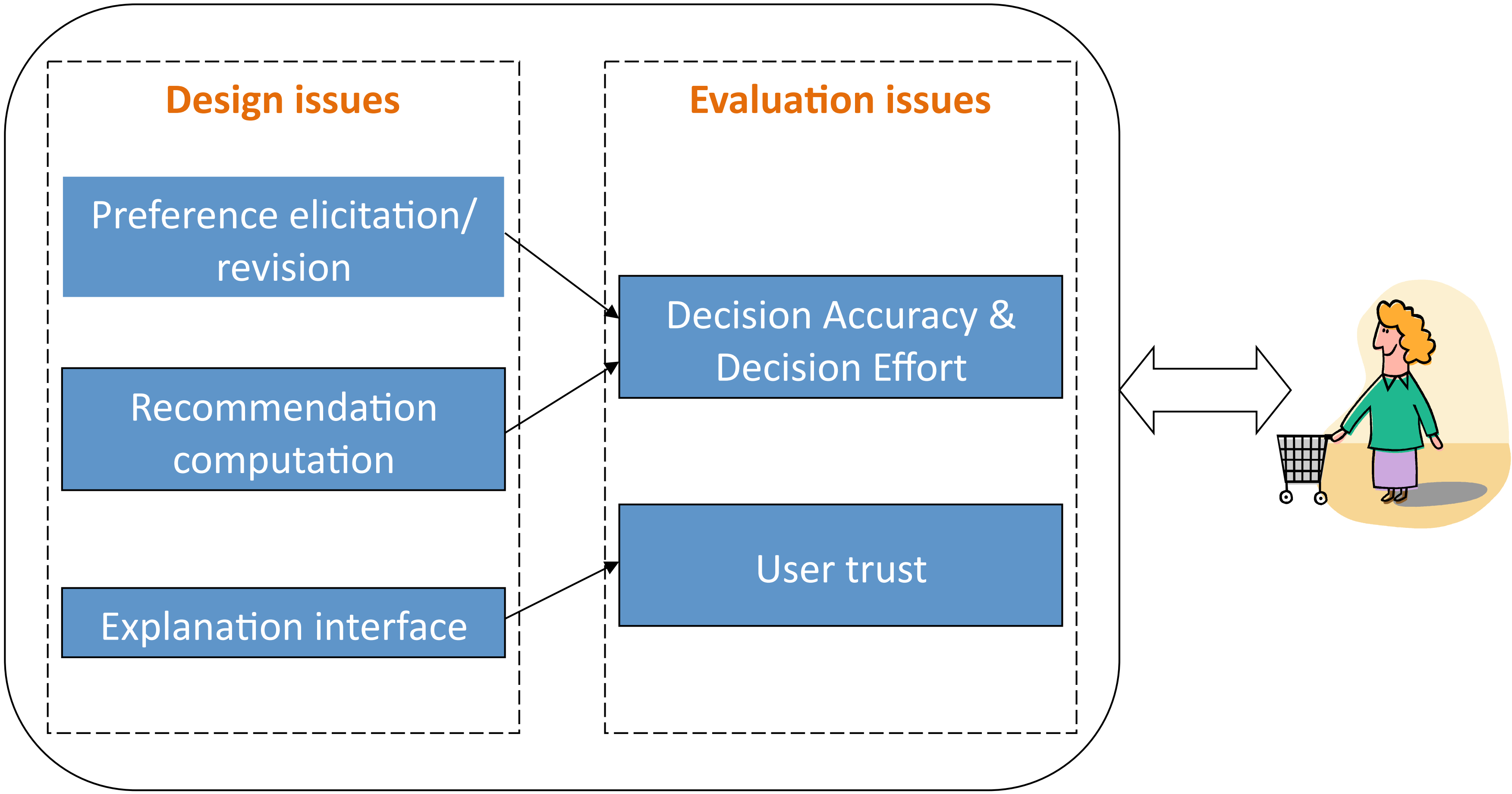
- Construct preferences
- Perform compensatory decision strategy
- Increase decision accuracy and save efforts
- Inspire trust



Why critiquing support?

- Particularly assist users in handling with two preference conditions
  - Preference incompleteness
    - Users are usually unable to accurately state their preferences up front (“*Adaptive decision maker*”, *Payne et al. 1993*)
    - **Solution**: to elicit users’ preferences on attributes via stimulating them to provide feedback (in form of critiques)
  - Preference conflict
    - No product satisfies all of the user’s stated preferences, e.g., “cheaper, faster and longer battery life”
    - **Solution:** to support users to make tradeoffs, i.e., obtaining the gains on important attributes while accepting the losses on less important ones
    - Making tradeoffs is a crucial aspect of high-quality, rational decision making (Frisch and Clemen, 1994) - **compensatory decision strategy**
    - Tradeoff making can **increase** users’ decision accuracy up to **57%** (Pu and Chen, EC’05)

Research Issues



Outline

- What is critiquing-based recommender system and Why?
- Development history (1997-2013)
- User experiences
- Conclusion

Representative Works

- Natural language dialog
- Graphical user interfaces
  - System-suggested critiques
  - User-initiated critiquing
  - Hybrid critiquing

1st type: Natural Language Dialog

- **ExpertClerk** (Shimazu, IJCAI'01)
- **Adaptive Place Advisor** (Thompson *et al.*, 2004)
- **Speech-based Critiquing** (Grasch et al., ACM RecSys'13)

ExpertClerk (Shimazu, IJCAI’01)

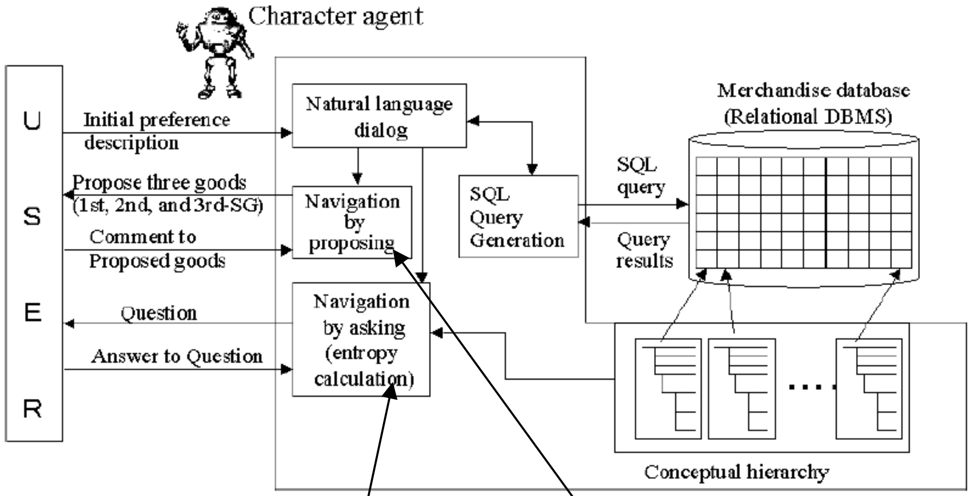


Figure 3: The ExpertClerk System

Switch between navigation by proposing and navigation by asking

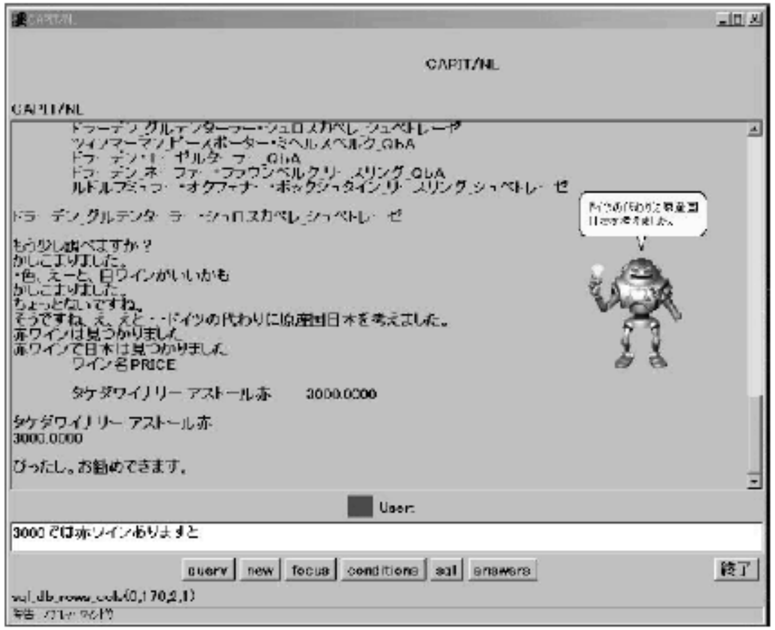


Figure 4: The ExpertClerk screen image

Cont.

- Method
  - First ask a few questions to identify shopper's initial preference (buying points)
  - ***<u>Navigate by proposing</u>***
    - Show three contrasting samples with explanations of their selling point
    - e.g., "this is twice as expensive as those because it is made of silk and the other two made of polyester"
  - Observe shopper's reactions on likes/dislikes and why
    - e.g., "this one is too dark for me compared with the other two"
  - ***<u>Navigate by asking</u>***
    - If too many matching goods, narrow down them by asking appropriate questions using entropy
  - Pick new samples fit shopper's responses
  - Repeat until shopper finds an appropriate good

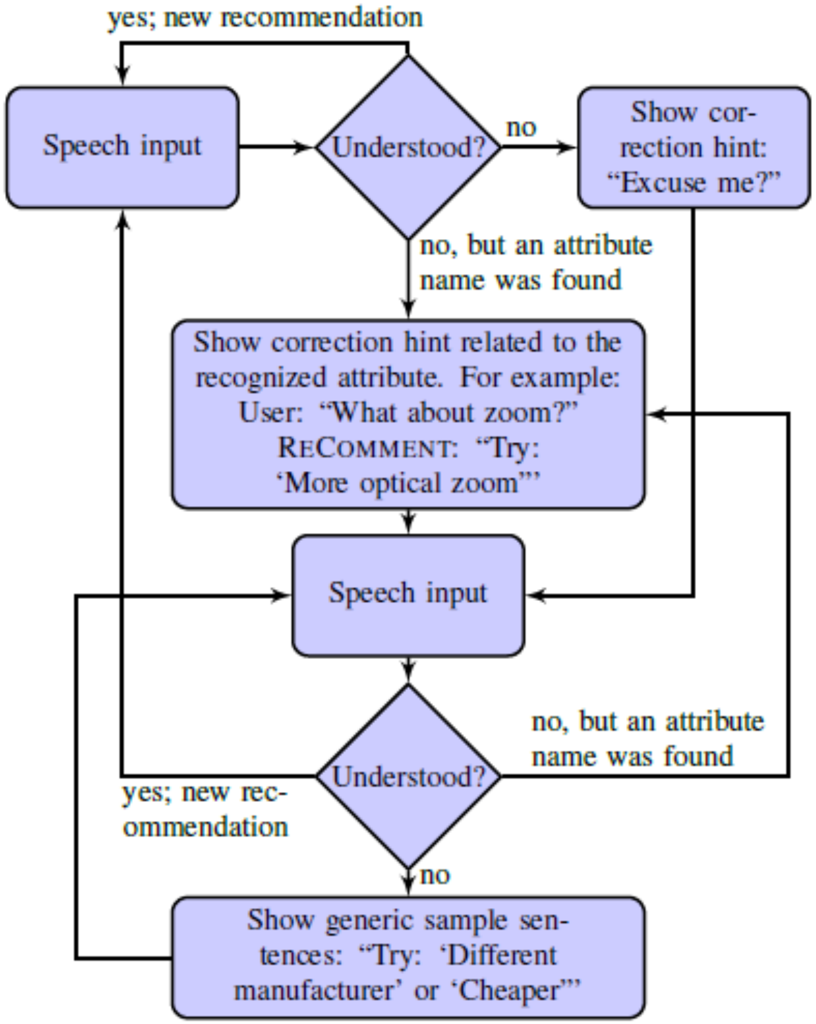
Adaptive Place Advisor (Thompson *et al.*, 2004)

1. Inquirer: Where do you think I should eat tonight?
2. Advisor: What type of food would you like?
3. Inquirer: What types are there?
4. Advisor: You can say things like Chinese, Indian, and Mediterranean.
5. Inquirer: Oh, maybe a cheap Indian place.
6. Advisor: What kind of parking would you like?
7. Inquirer: I don't care, as long as it's in Palo Alto.
8. Advisor: I'm sorry, I don't know of any restaurants like that, would you like to search for any price?
9. Inquirer: No, I think I'd like Chinese instead.
10. Advisor: I know three inexpensive Chinese restaurants in Palo Alto. One is Mandarin Gourmet on 420 Ramona. How does that sound?[7]
11. Inquirer: No, what else do you have?
12. Advisor: How about Jing-Jing Szechwan Hunan Gourmet on 443 Emerson Street?
13. Inquirer: Sure, that sounds fine.

Method

- A probabilistic representation of the user's preferences, i.e., the query, is expanded
- The system incrementally refines this query, based on the user's critiques to the attributes and items offered during a conversation

Speech-based Critiquing (Grasch et al., RecSys'13)



2nd type: System-Suggested Critiques

- **FindMe** (Burke *et al.*, 1997)
- **Dynamic Critiquing** (McCarthy *et al.,* IUI'05)
- **MAUT-based Compound Critiques** (Zhang *et al.,* AH'06)
- **Preference-based Organization** (Chen and Pu, UM'07)

FindMe (Burke *et al.*, 1997)

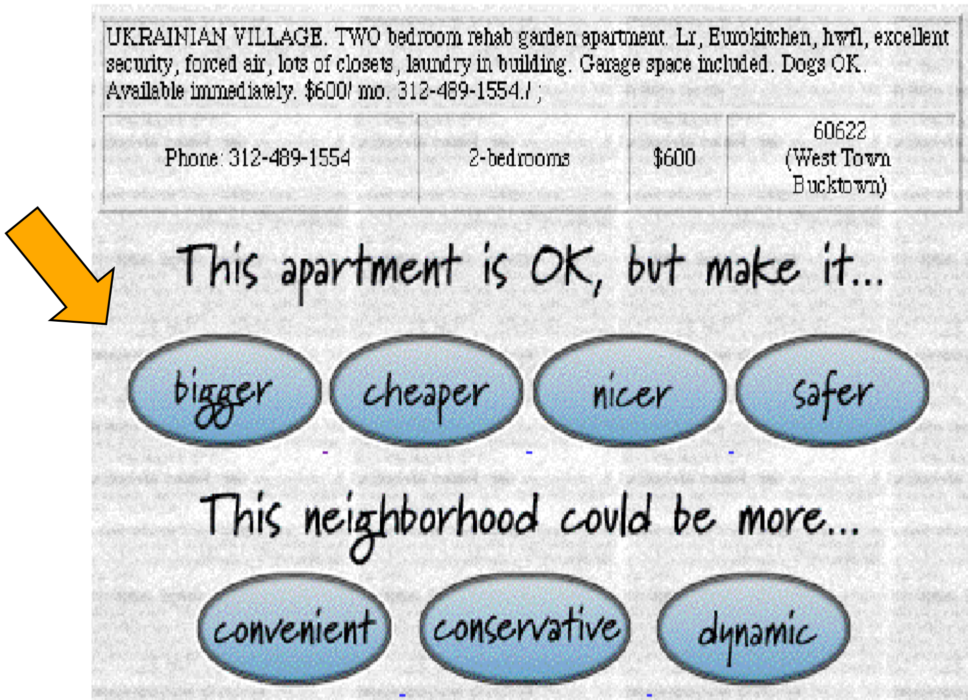


Figure 1: Tweaking an apartment in RentMe

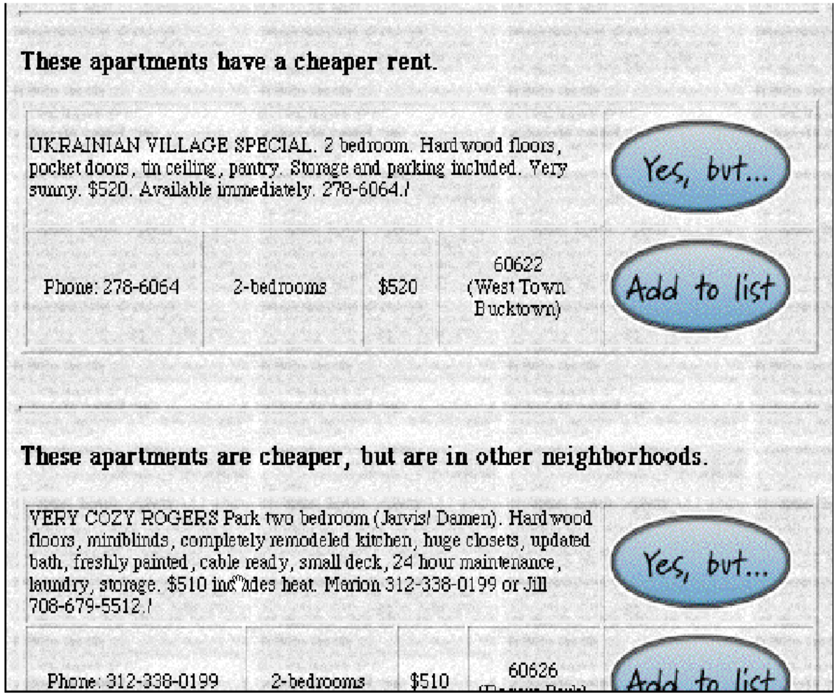
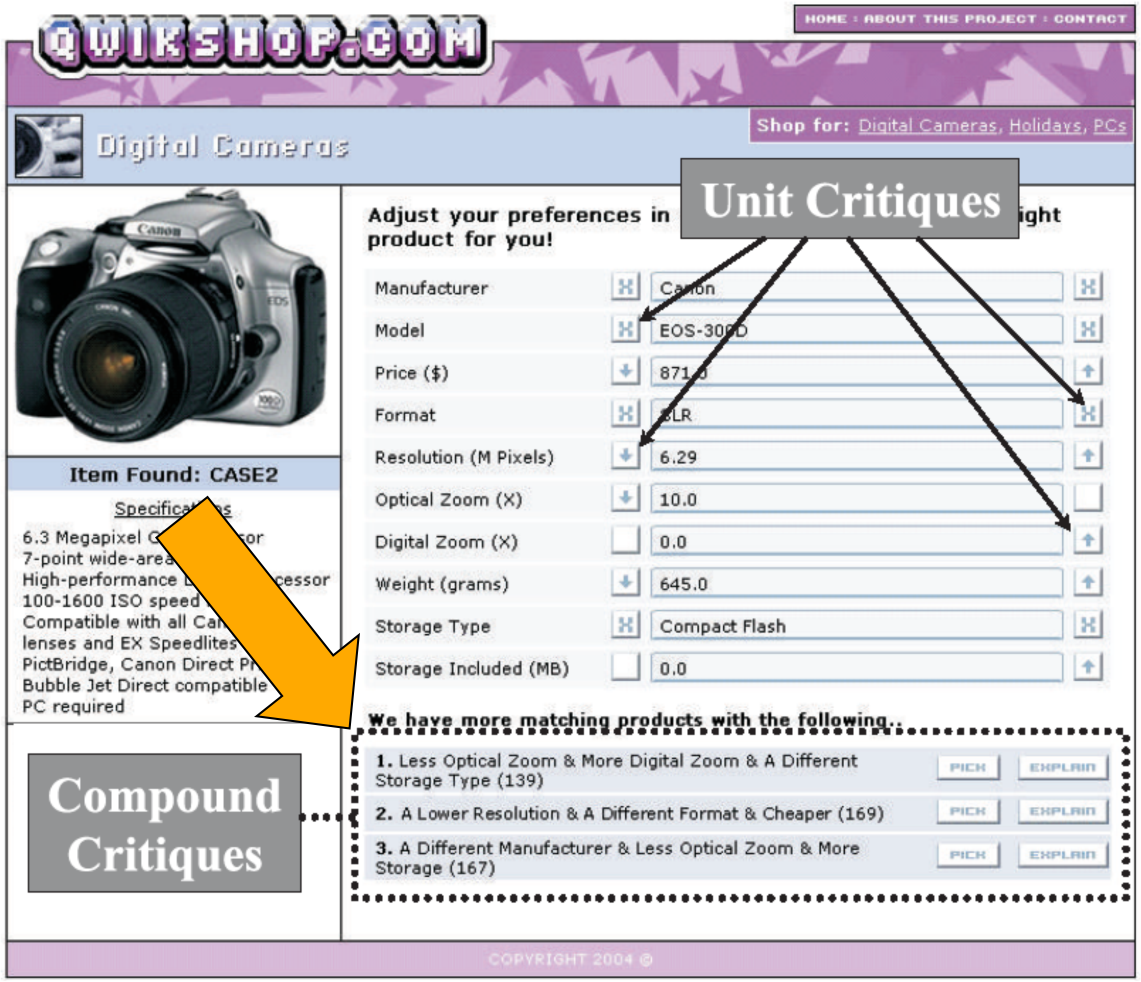


Figure 2: The result of applying the "cheaper" tweak

Method
- The entry point is the user's initial query, e.g., [600<price<650,neighborhood = 'Bucktown', size=2]
- Each tweak is treated as a constraint with the highest priority
- A SQL query is passed to database
- If no enough entities return, drop lower priority constraint

Dynamic Critiquing (McCarthy *et al.,* IUI'05)

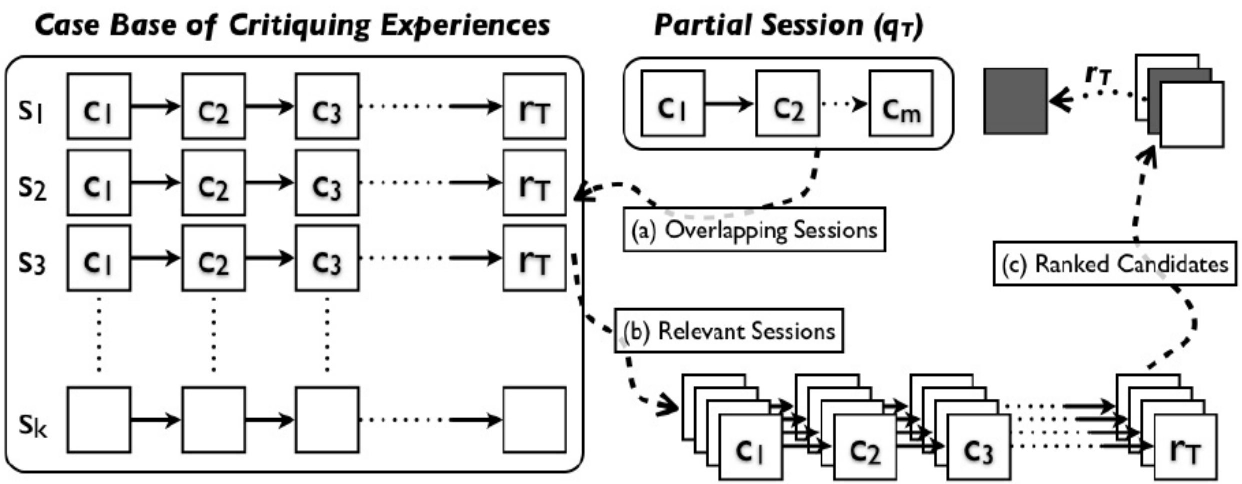


Method

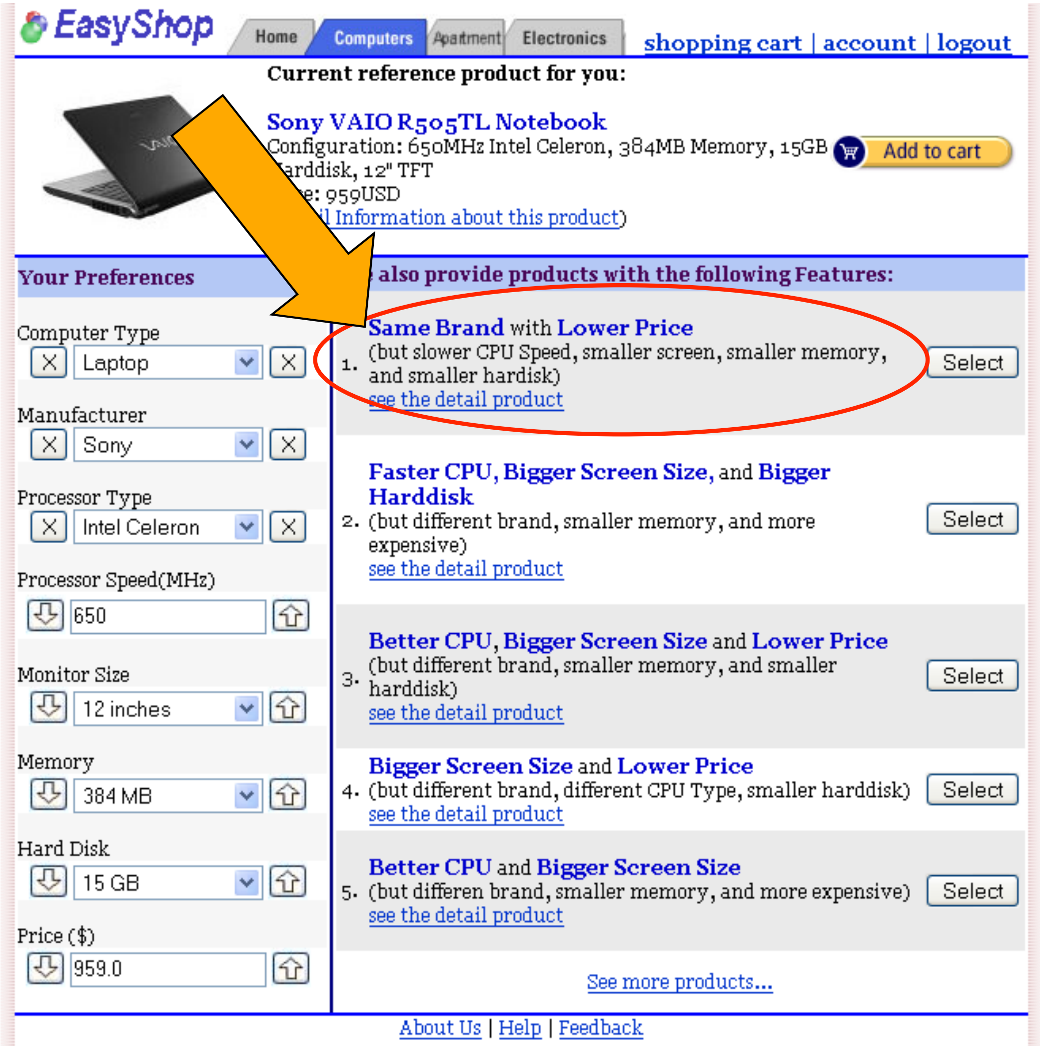
- Dynamically generate a set of compound critiques (each over multiple attributes) through frequent pattern mining
- The recommended item should satisfy the user picked critique as well as being most similar to the previous recommendation

Cont.: two extensions

- Incremental Critiquing (Reilly *et al.*, 2005)
  - The recommended item must be <u>*additionally*</u> compatible with the user's previously selected critiques
- Experience-based Critiquing (McCarthy *et al.*, ICCBR'10)



MAUT-based Compound Critiques (Zhang *et al.,* AH'06)



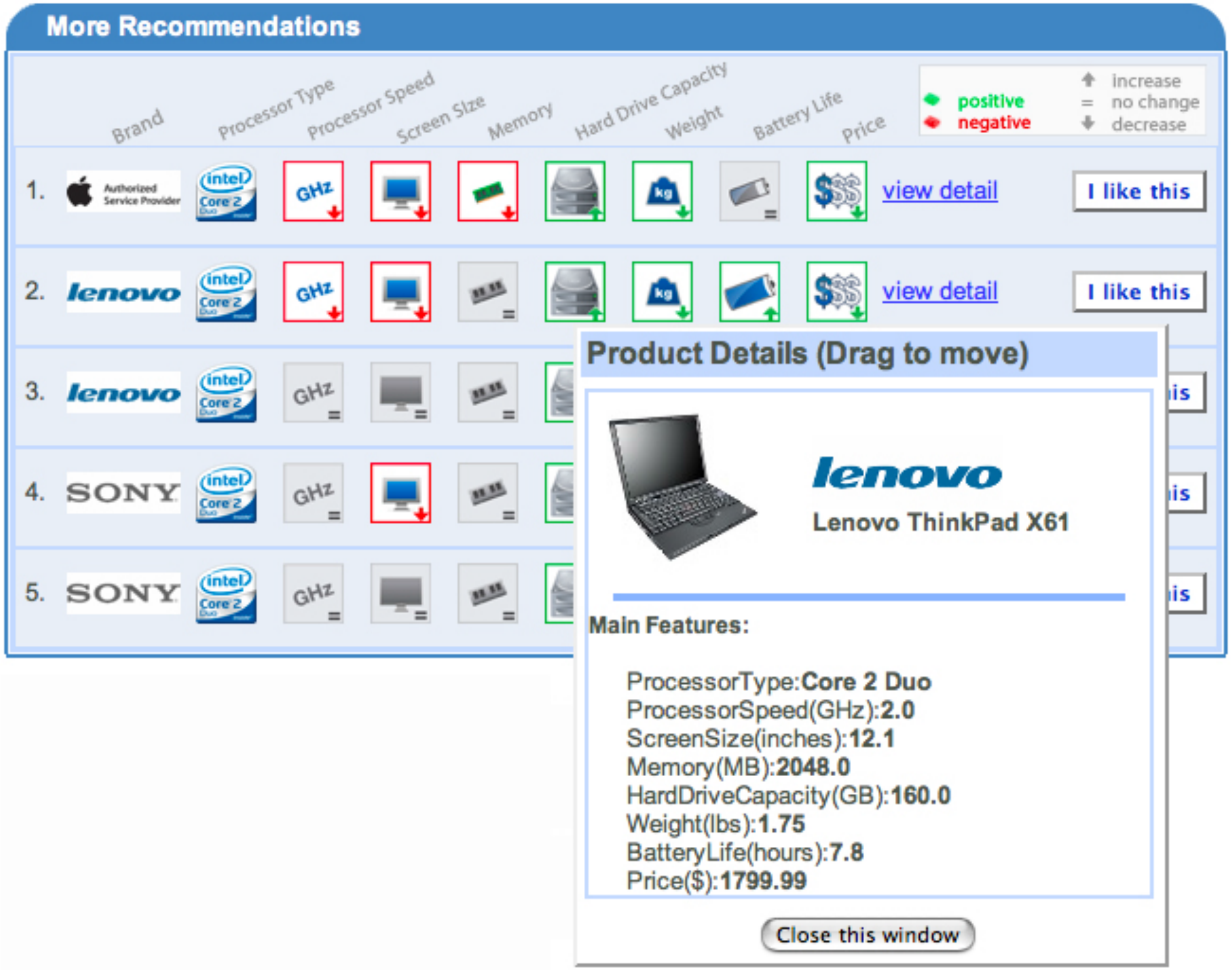
Method

- Model the user's preferences based on the Multi-Attribute Utility Theory (MAUT)

$$U(\langle x_1, \cdots, x_n \rangle) = \sum_{i=1}^{n} w_i V_i(x_i)$$

- Rank products according to their utilities and the top-k ones are presented as compound critiques (except the ranked first one)

Extension: visual critiquing (Zhang *et al.*, EC'08)



Text-only critiques are replaced with meaningful icons

Preference-based Organization (Chen and Pu, UM'07)

Manufacturer	Price	MegaPixels	Optical zoom	Memory type	Flash memory	LCD screen size	Depth	Weight	
The top candidate according to your preferences									
Canon	$242.00	5.0 MP	3x	CompactFlash Card	32 MB	1.8 in	1.37 in	8.3 oz	choose
We have more products with the following									
they are cheaper and lighter, but have fewer megapixels									
Nikon	$167.95	4 MP	3x	SD Memory Card	14 MB	1.8 in	1.4 in	4.6 oz	choose
Canon	$230.00	4.1 MP	3x	CompactFlash Card	32 MB	1.5 in	1.09 in	6.53 oz	choose
Canon	$180.00	[illegible]	2x	SD Memory Card	16 MB	2 in	0.83 in	4.06 oz	choose
Canon	$219.18	4.2 MP	4x	MultiMedia Card	16 MB	1.8 in	1.51 in	6.35 oz	choose
Canon	$163.50	3.2 MP	4x	MultiMedia Card	16 MB	1.8 in	1.5 in	6.3 oz	choose
Canon	$199.40	3.2 MP	2.2x	SD Memory Card	16 MB	1.5 in	1.4 in	5.8 oz	choose
they have more megapixels and bigger screens, but are more expensive									
Sony	$365.00	7.2 MP	3x	Internal Memory	32 MB	2.5 in	1.5 in	6.9 oz	choose
Canon	$439.99	7.1 MP	3x	SD Memory Card	32 MB	2 in	1.04 in	6 oz	choose
Fuji	$253.00	6.3 MP	4x	XD-Picture Card	16 MB	2 in	1.4 in	7.1 oz	choose
Sony	$336.00	7.2 MP	3x	Internal Memory	32 MB	2 in	1 in	5 oz	choose
Nikon	$304.18	7.1 MP	3x	Internal Memory	13.5 MB	2 in	1.4 in	5.3 oz	choose
Olympus	$334.00	7.4 MP	5x	XD-Picture Card	32 MB	2.0 in	1.7 in	7.1 oz	choose
they are lighter and thinner, but have less flash memory									
Pentax	$238.99	5.3 MP	3x	Internal Memory	10 MB	1.8 in	0.8 in	3.7 oz	choose
Canon	$273.18	4.0 MP	3x	SD Memory Card	16 MB	2 in	0.82 in	4.59 oz	choose
Nikon	$329.95	5.1 MP	3x	Internal Memory	12 MB	2.5 in	0.8 in	4.2 oz	choose
Canon	$316.18	5.3 MP	3x	SD Memory Card	16 MB	2 in	0.81 in	4.59 oz	choose
Casio	$386.00	7.2 MP	3x	Internal Memory	8.3 MB	2.5 in	0.88 in	4.48 oz	choose
Fuji	$309.18	6.3 MP	3x	XD-Picture Card	16 MB	2.5 in	1.1 in	5.5 oz	choose
they have more optical zoom with different memory type, but are thicker and heavier									
Panasonic	$386.00	5.0 MP	12x	SD Memory Card	16 MB	1.8 in	3.34 in	11.52 oz	choose
Konica Minolta	$349.99	5.0 MP	12x	SD Memory Card	16 MB	2 in	3.3 in	12 oz	choose
Fuji	$259.18	4.23 MP	10x	XD-Picture Card	16 MB	1.5 in	3.1 in	11.9 oz	choose
Olympus	$253.00	4.0 MP	10x	XD-Picture Card	16 MB	1.8 in	2.7 in	9.9 oz	choose
Olympus	$284.99	4.0 MP	10x	XD-Picture Card	16 MB	1.8 in	2.7 in	10.6 oz	choose
Nikon	$259.18	4.2 MP	8.3x	Internal Memory	13.5 MB	1.8 in	2.2 in	9 oz	choose

Critique suggestion

Satisfying items

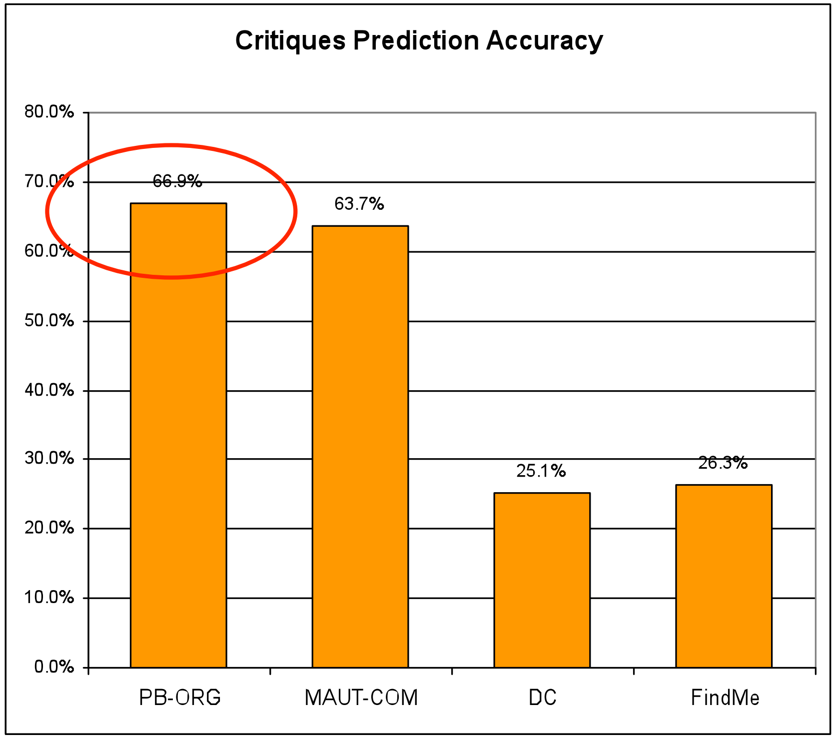
Interface design guidelines

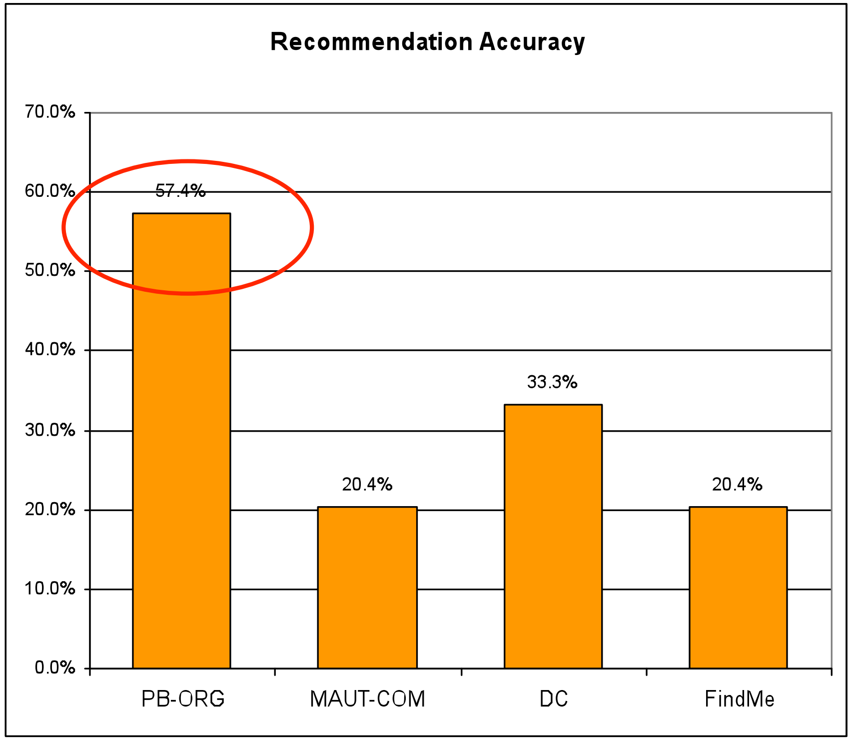
1. Propose improvements and compromises (i.e., attributes tradeoff) in the critique
2. keep the number of attributes in the critique under five to avoid information overload
3. Include actual products (up to six) under each critique
4. Diversify the proposed critiques and their contained products

Comparison with others

	Dynamic critiques	Critiques typical of the remaining products	Critiques adaptive to user preferences	Diversity within critiques and their contained products
Preference-based organization	√	√	√	√
MAUT-based compound critiques	√	×	√	×
Dynamic critiquing	√	√	×	Partially (only critiques)
FindMe	×	×	×	Partially (only critiques)

Experiment



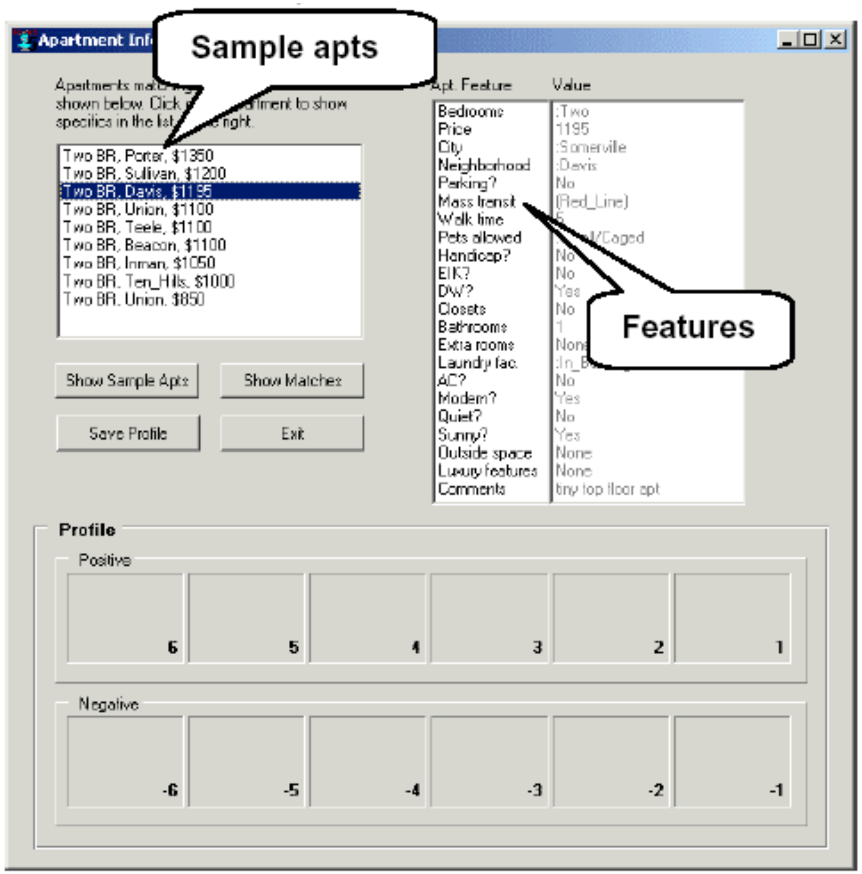


➔ The preference-based organization algorithm achieves **the highest accuracy** (significantly) in terms of both critique predication and recommendation computation

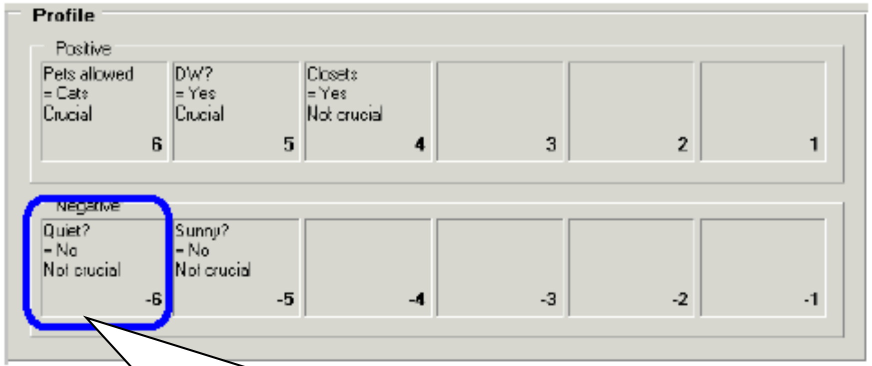
3rd type: User-Initiated Critiquing

- **Apt Decision** (Shearin and Lieberman, IUI'01)
- **Example Critiquing** (Pu and Chen, EC'05; Chen and Pu, AAAI'06)
- **Flat Finder** (Viappiani *et al.*, 2007)

Apt Decision (Shearin and Lieberman, IUI'01)



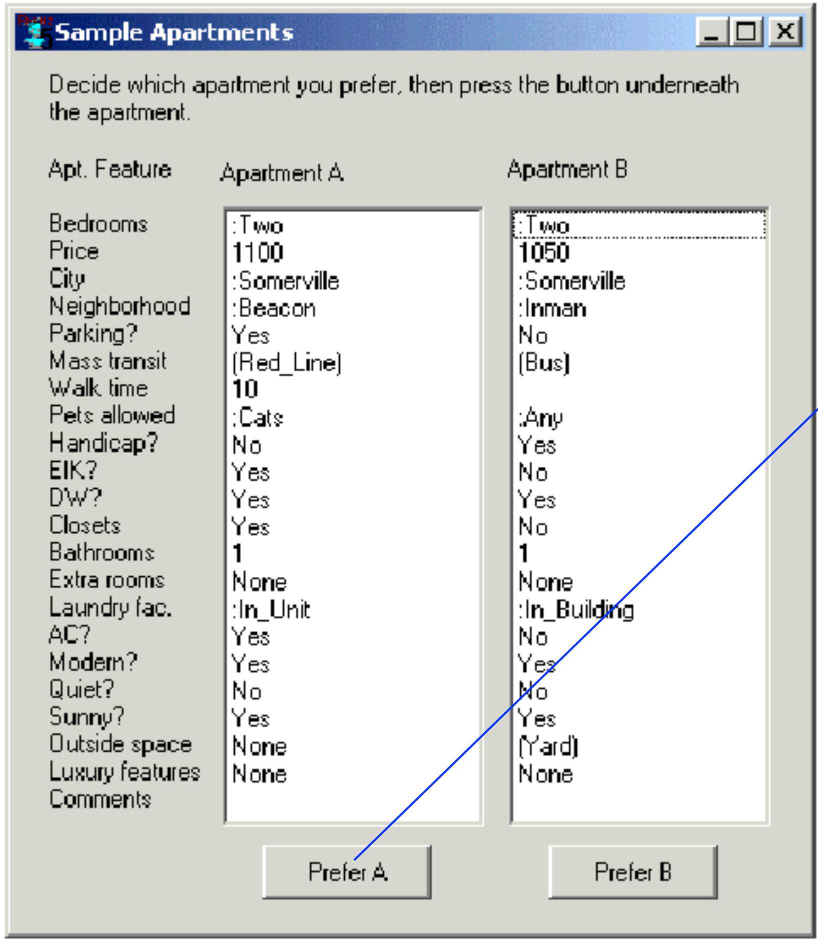
Drag features onto slots in the profile, which contains twelve weight slots: six positive weights (1 to 6) and six negative weights (-1 to -6)



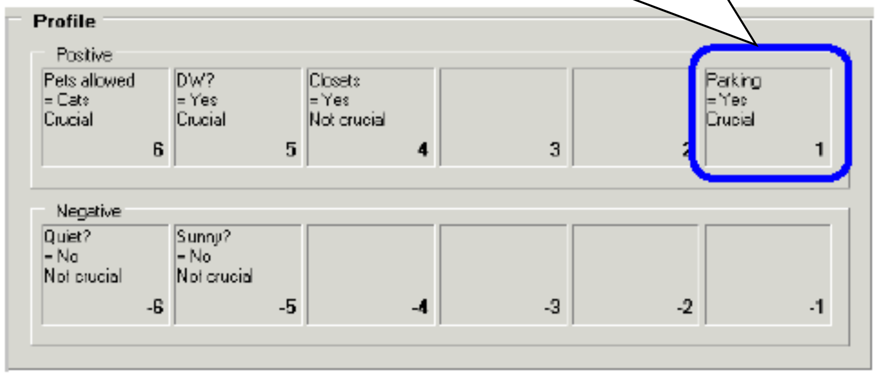
For example, the leftmost negative slot indicates that the user feels very strongly about the fact that this selected apartment is not quiet

User profile: weighted feature vector, e.g., {(pets allowed, cats, 6(must have)), (closets, yes, 4(neutral)),…,}

Cont.

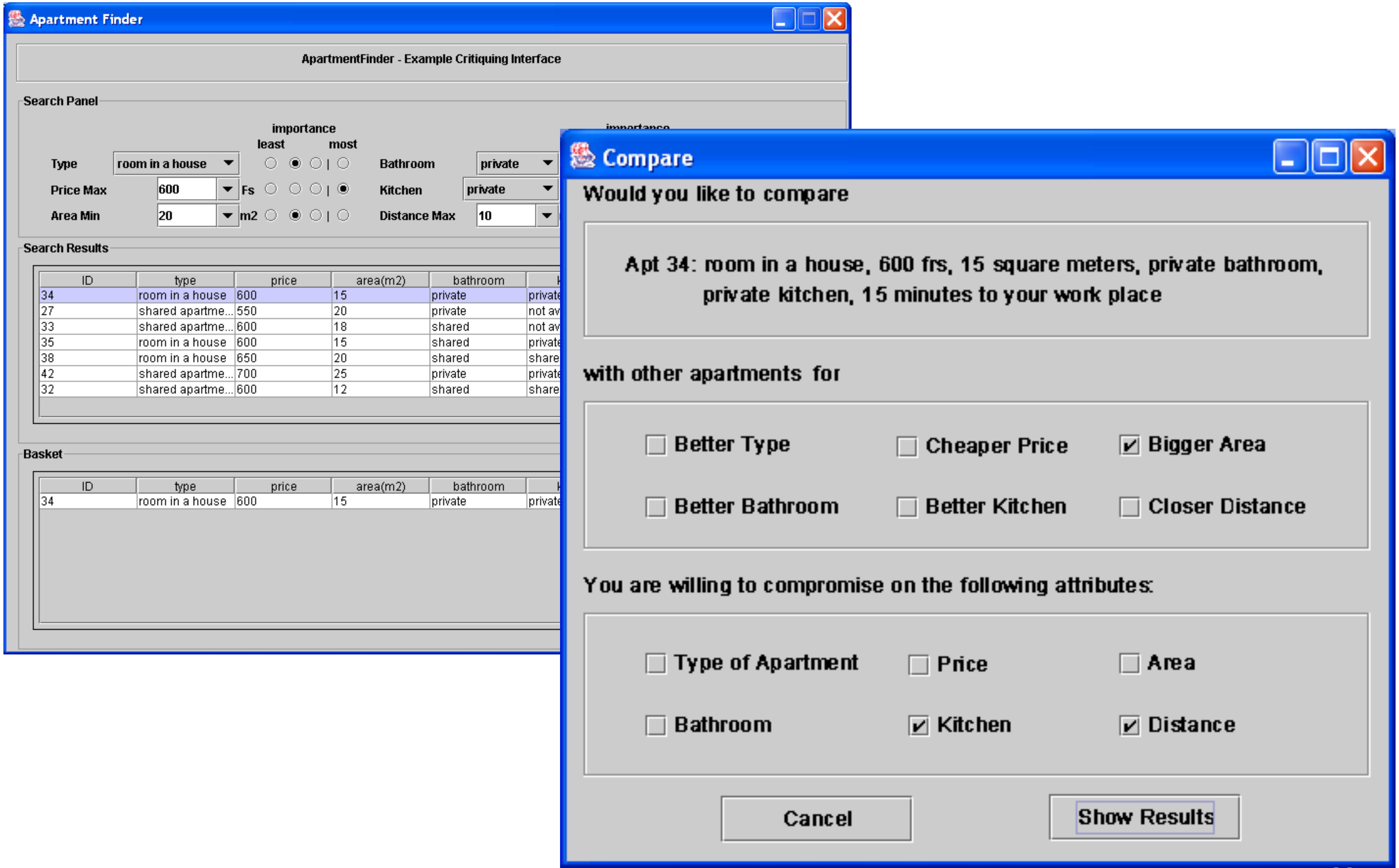


User chooses an apartment, and the features unique to the chosen apartment but not present in the profile will be added to the right side of the profile



Profile expansion: pairwise preference among pairs of sample apartments

Example Critiquing (Pu and Chen, EC'05)



Cont.

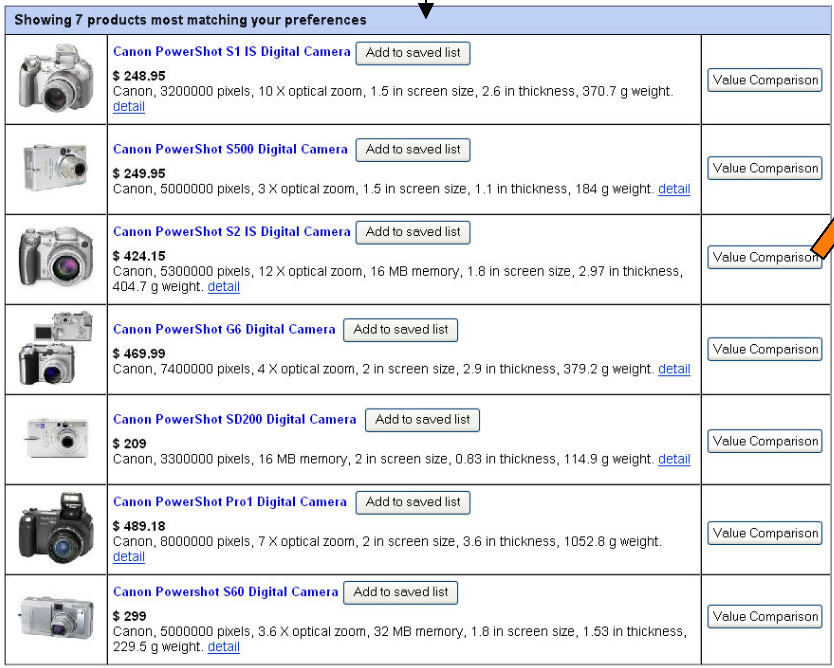
- Initial preference elicitation
  - Any preferences
  - Default preferences
- *How many examples to show?*
  - Multiple items for users to select the final choice or the one to be critiqued
  - For a moderate number of preferences, the amount falls between 5 and 20 (Faltings *et al.,* 2004)
- *What example to show?*
  - Combined strategy: Elimination-by-Aspect (EBA) (for hard constraints) plus Multi-Attribute Utility Theory (MAUT)
  - Show partial satisfaction set to help resolve preference conflicts

Cont.

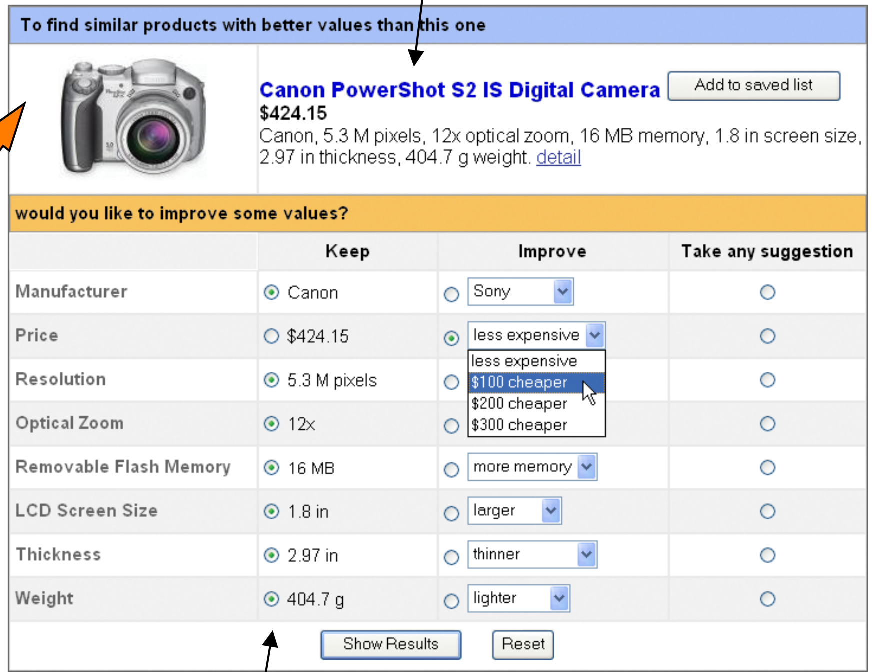
- *How to support tradeoff making?*
  - Three types
    - Value tradeoff: change a particular attribute's preference value
    - Utility tradeoff: change the weight of a preference
    - Outcome tradeoff: add new preferences
  - Complexity of tradeoff task: *(optimize, compromise)*
    - Simple tradeoff: (1,1)
      - e.g., ({price}, {size of room})
    - Complex tradeoff: (m,n) (m or n > 1)
      - e.g., ({price}, {size of room, distance to work}) (three different ways to compromise two attributes in order to gain on one attribute)

Extension: Chen and Pu, AAAI'06

Multiple recommended items

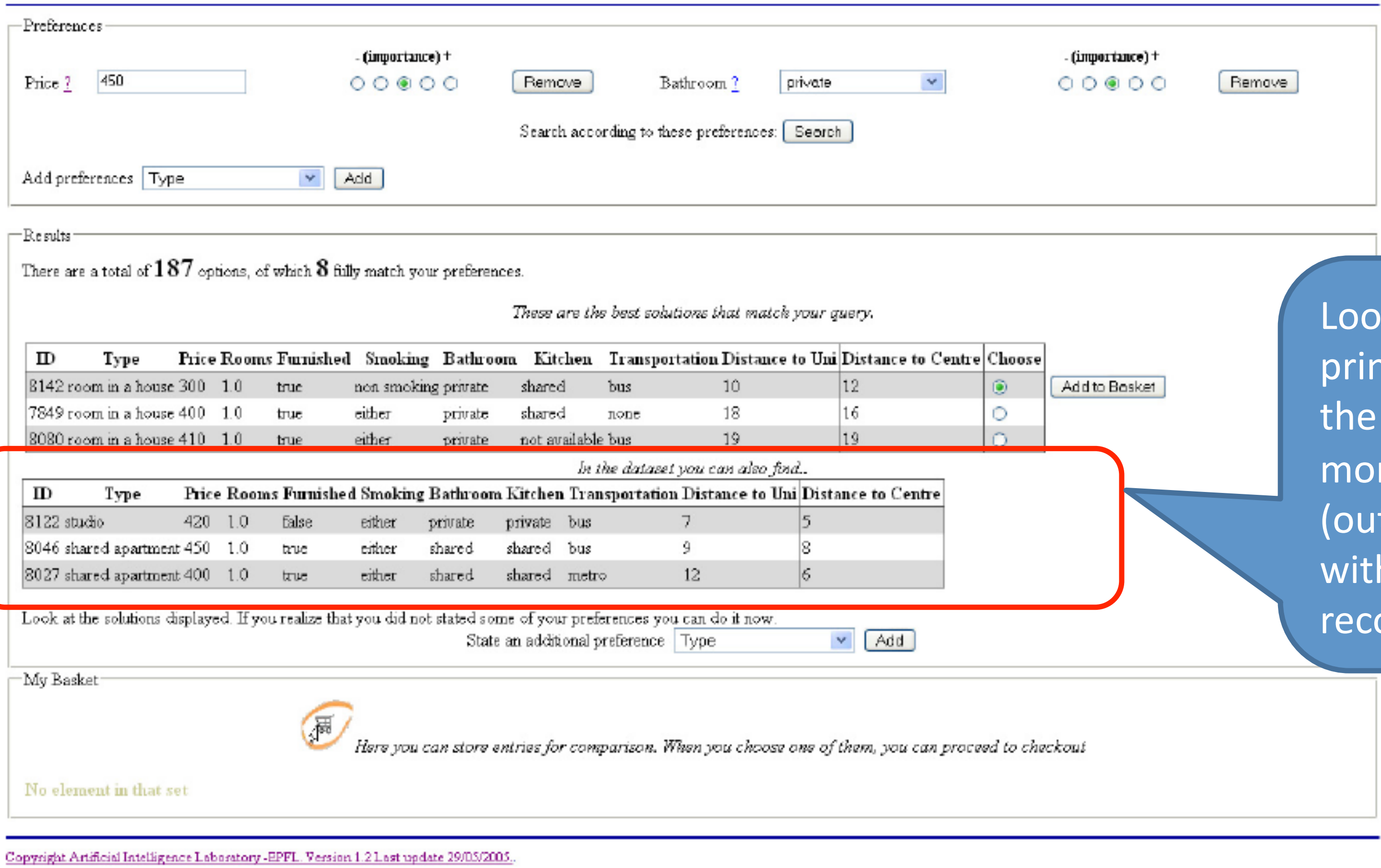


The product chosen to be critiqued



User-initiated critiquing facility

Flat Finder (Viappiani *et al.*, 2007)



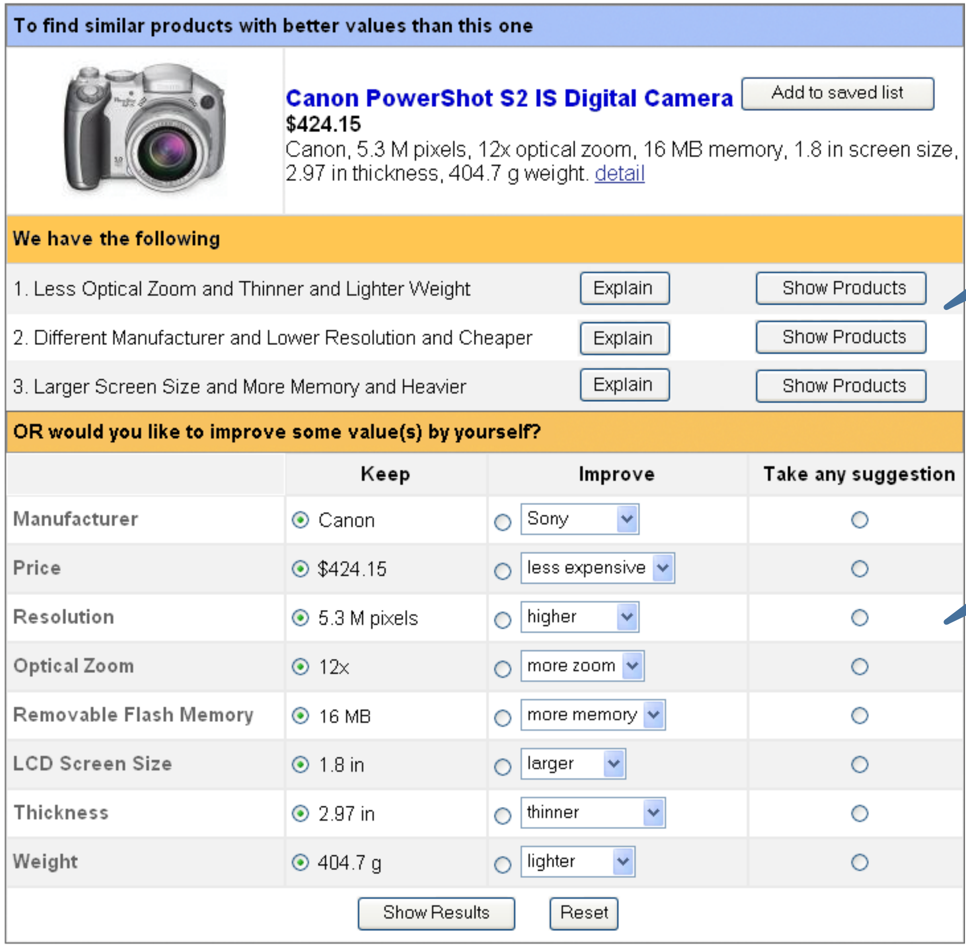
Look-ahead principle: stimulate the user to express more preferences (outcome tradeoff) with the predictive recommendations

Representative Works

- Natural language dialog
- Graphical user interfaces
  - System-suggested critiques
  - User-initiated critiquing
  - **<u>Hybrid critiquing</u>**

4th type: Hybrid Critiquing

Hybrid of *system-suggested critiques* and *user-initiated critiquing* (Chen and Pu, IUI'07)



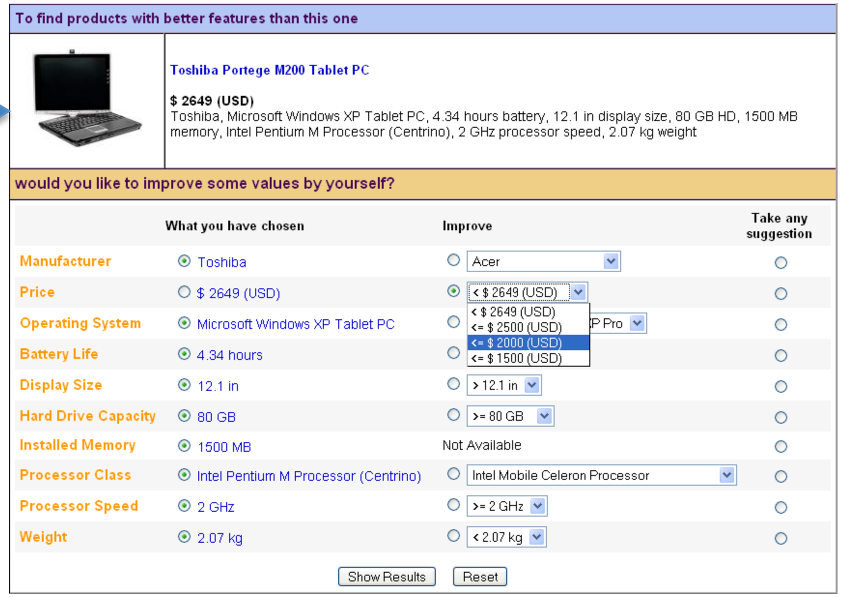
System-suggested critiques

User-initiated critiquing

Extension: Chen and Pu, RecSys'07

Preference-based organization

User-initiated critiquing support

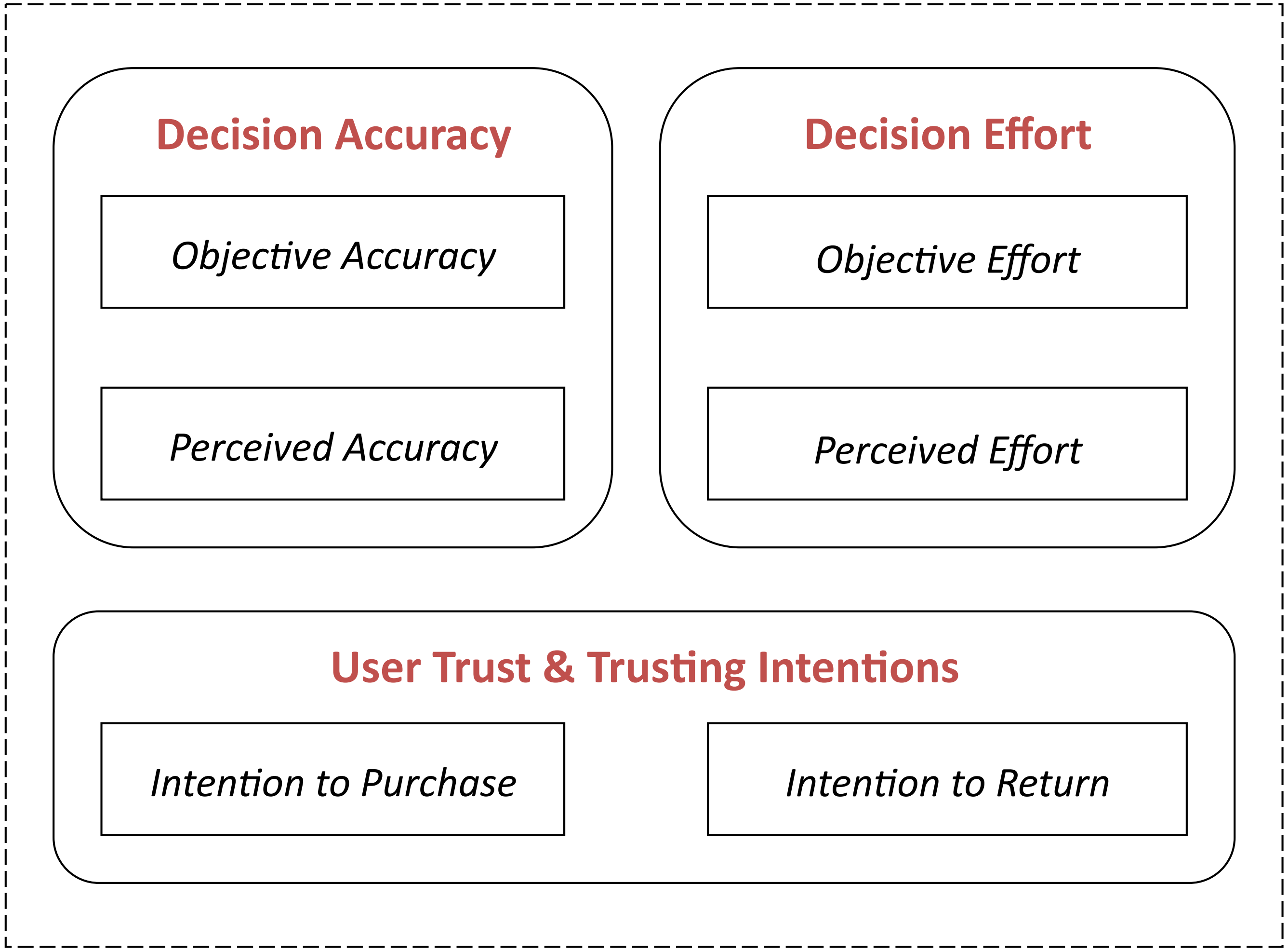


If suggested critiques and products do not interest the user in the organization interface, she could switch to make the self-initiated critiquing

Outline

- What is critiquing-based recommender system and Why?
- Development history
- User experiences
- Conclusion

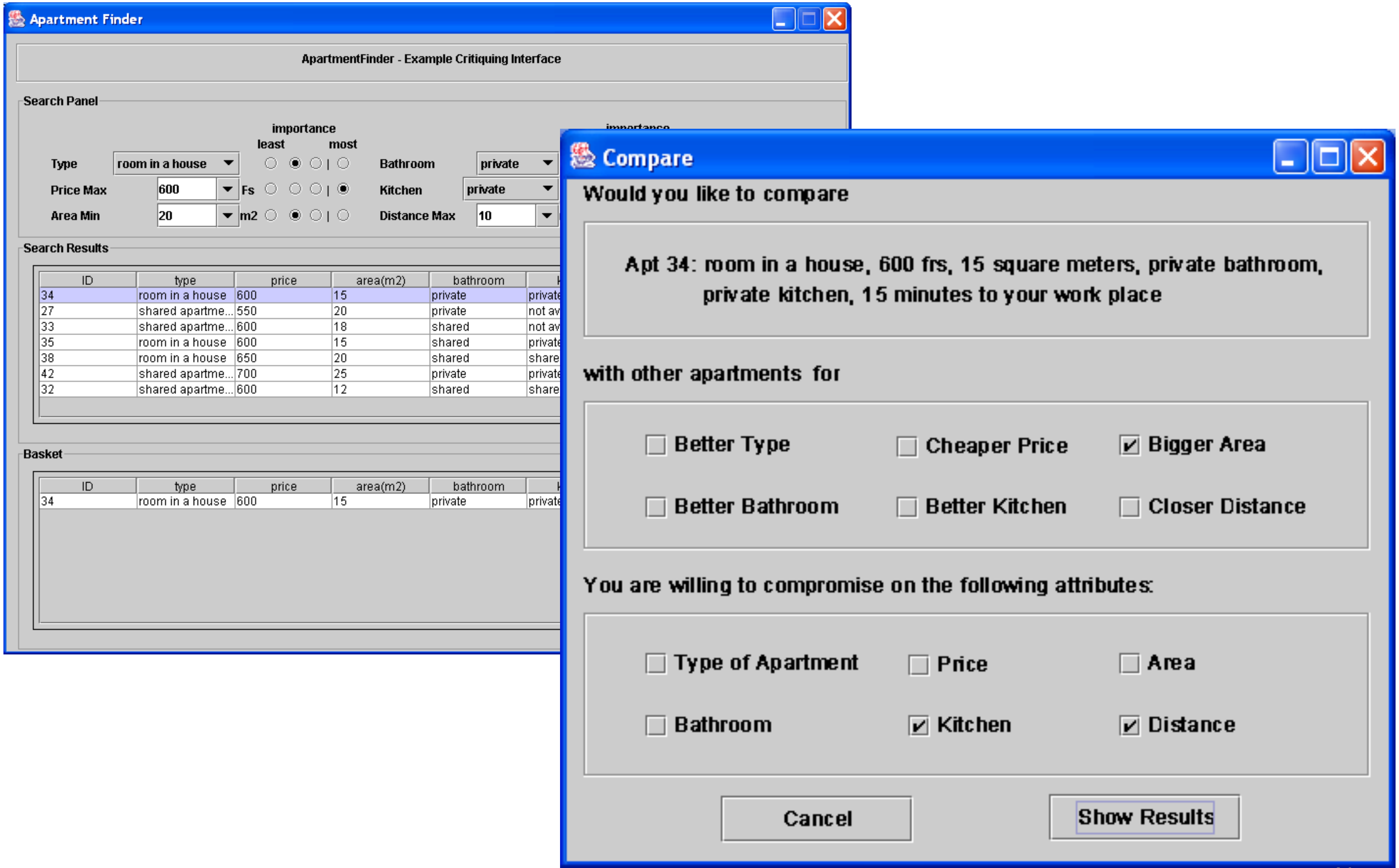
User Evaluation Framework



Experiments

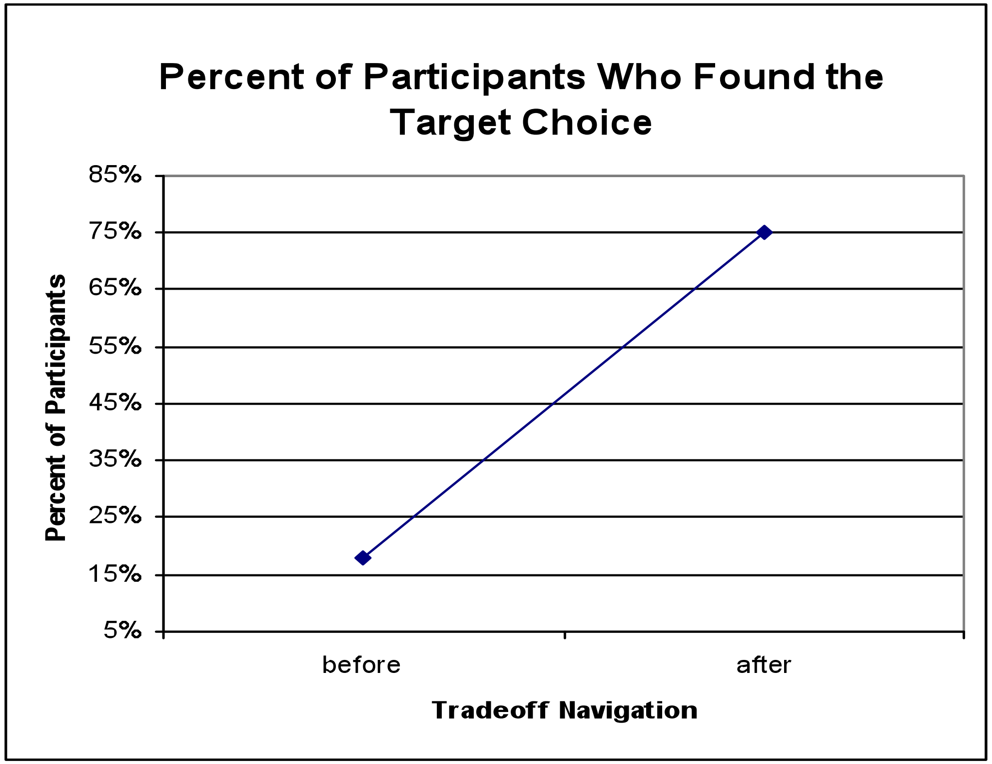
1. Example critiquing vs. non-critiquing based system (EC'05)
2. Example critiquing vs. dynamic critiquing (AAAI'06)
3. Organization interface vs. ranked list (IUI'06, UMAP'10)
4. Evaluation of hybrid critiquing (IUI'07, RecSys'08)

Example Critiquing (Pu and Chen, EC'05)



Experiment 1: Example critiquing vs. non-critiquing based system (EC'05)

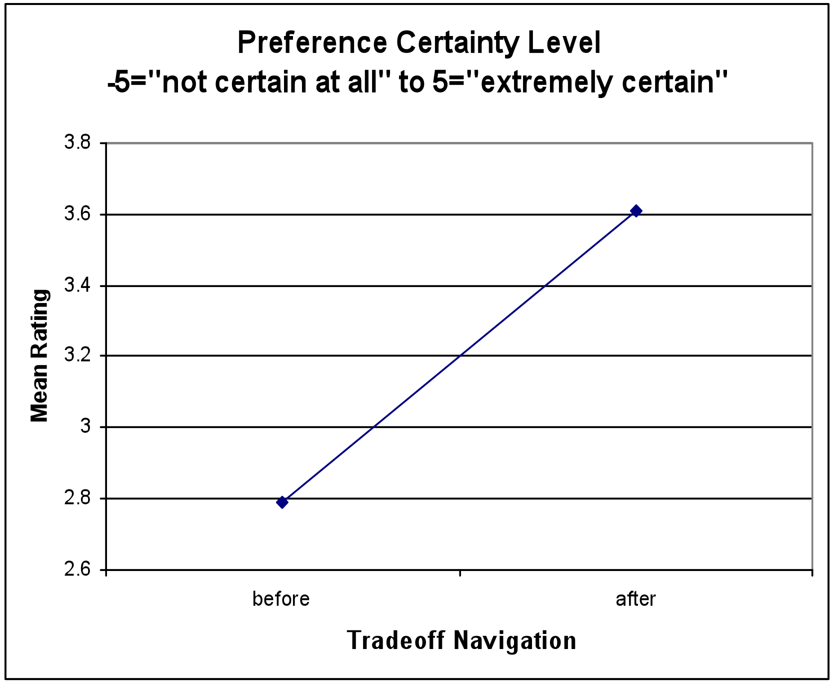
Participants: 28; Material: Apartment Finder



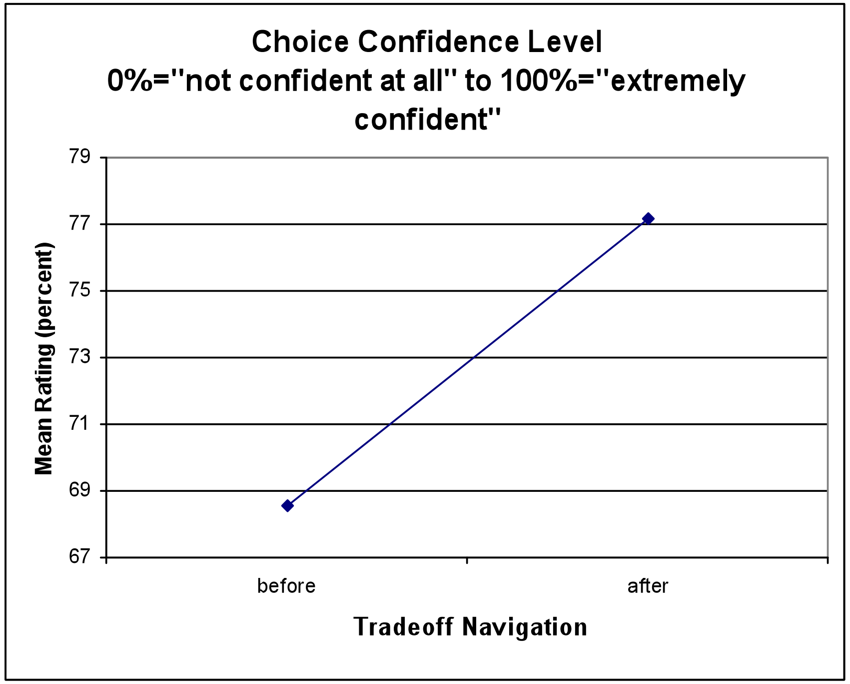
Effect of example critiquing on improving decision accuracy

➔ Tradeoff navigation process with the support of example-critiquing can significantly improve users' decision accuracy by **up to 57%**

Preference certainty



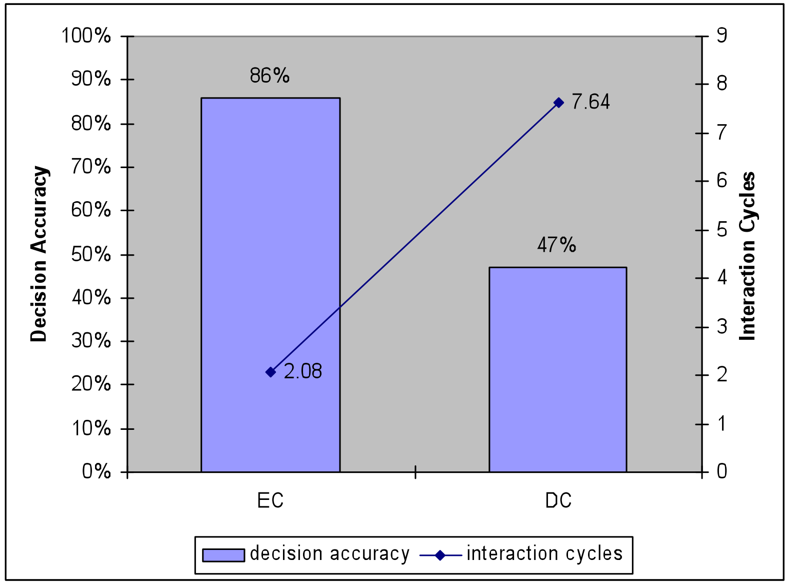
Decision confidence

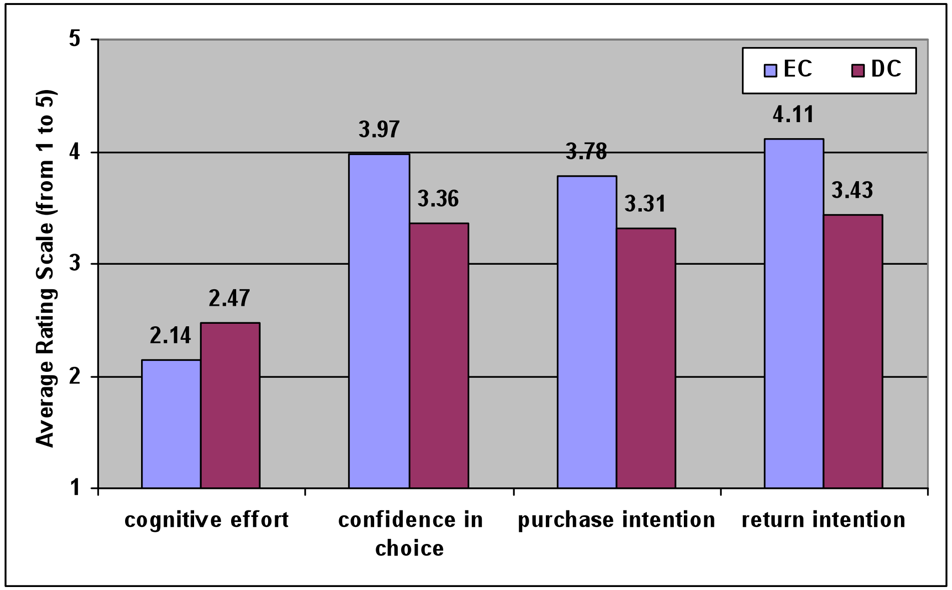


➔ Tradeoff navigation process with the support of example-critiquing can significantly improve users' preference certainty and decision confidence

Experiment 2: Example critiquing vs. dynamic critiquing (AAAI'06)

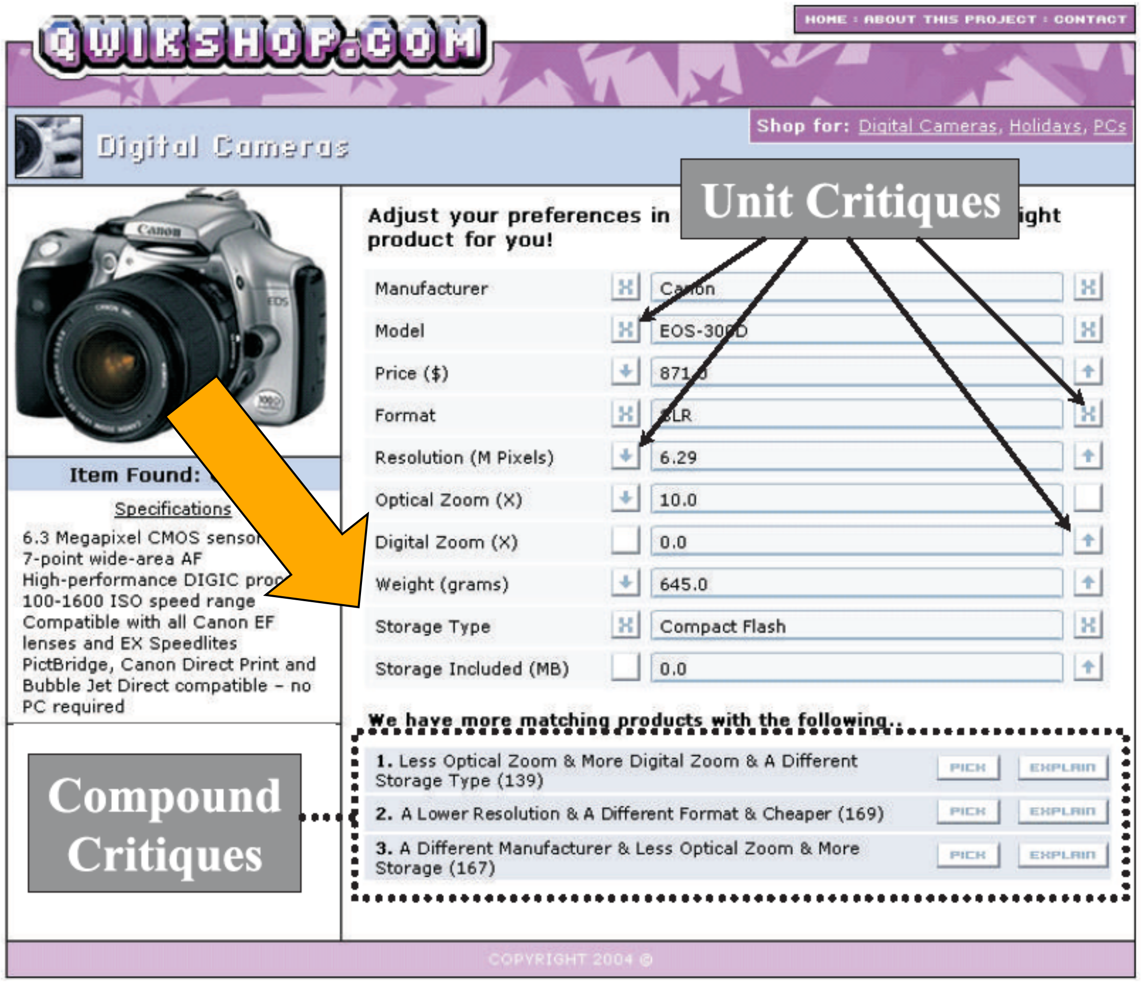
Participants: 36; Material: Online product finder; Procedure: within-subjects





➔ Example critiquing significantly outperforms dynamic critiquing regarding **objective/ subjective accuracy**, **objective interaction effort** and **perceived effort**, **purchase and return intentions**

Dynamic Critiquing (McCarthy *et al.*, IUI'05)



Method

- Dynamically generate a set of compound critiques (each over multiple attributes) through association rule mining
- The recommended item should satisfy the user picked critique as well as being most similar to the previous recommendation

Experiment 3: Organization interface vs. ranked list (IUI'06)

The most popular product

	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
◉	—	$2'095.00	1.67 GHz	4.5 hour(s)	512 MB	80 GB	38.6 cm	2.54 kg

We also recommend the following products because

they are cheaper and lighter, but have lower processor speed

	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
○	—	$1'499.00	1.5 GHz	5 hour(s)	512 MB	80 GB	33.8 cm	1.91 kg
○	—	$1'739.99	1.5 GHz	4.5 hour(s)	512 MB	80 GB	38.6 cm	2.49 kg
○	—	$1'625.99	1.5 GHz	5 hour(s)	512 MB	80 GB	30.7 cm	2.09 kg
○	—	$1'426.99	1.5 GHz	5 hour(s)	512 MB	60 GB	30.7 cm	2.09 kg
○	—	$1'929.00	1.2 GHz	4 hour(s)	512 MB	60 GB	26.9 cm	1.41 kg
○	—	$1'595.00	1 GHz	5.5 hour(s)	512 MB	40 GB	26.9 cm	1.41 kg

they have higher processor speed and bigger hard drive capacity, but are heavier

	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
○	—	$1'220.49	1.8 GHz	5 hour(s)	1 GB	100 GB	38.1 cm	2.95 kg
○	—	$2'148.99	2 GHz	4 hour(s)	1 GB	100 GB	39.1 cm	2.9 kg
○	—	$1'379.00	3.3 GHz	2 hour(s)	512 MB	100 GB	43.2 cm	4.31 kg
○	—	$2'235.00	1.8 GHz	2.5 hour(s)	1 GB	100 GB	43.2 cm	3.99 kg
○	—	$2'319.00	1.7 GHz	4.5 hour(s)	512 MB	100 GB	43.2 cm	3.13 kg
○	—	$2'075.00	1.8 GHz	1.67 hour(s)	512 MB	100 GB	43.2 cm	4.4 kg

they have longer battery life and lighter weight, but smaller display size

	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
○	—	$1'529.00	1.7 GHz	6.5 hour(s)	512 MB	80 GB	33.8 cm	1.77 kg
○	—	$1'599.00	1.7 GHz	6.5 hour(s)	512 MB	80 GB	33.8 cm	1.91 kg
○	—	$1'125.00	1.5 GHz	6 hour(s)	512 MB	80 GB	30.7 cm	2 kg
○	—	$2'099.99	1.2 GHz	9 hour(s)	512 MB	60 GB	26.9 cm	1.41 kg
○	—	$1'649.00	1.1 GHz	8.5 hour(s)	512 MB	40 GB	26.9 cm	1.36 kg
○	—	$969.00	1.2 GHz	6 hour(s)	256 MB	39 GB	30.7 cm	2.22 kg

they are cheaper, but heavier

	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
○	—	$1'179.00	3.2 GHz	2 hour(s)	512 MB	80 GB	39.1 cm	3.62 kg
○	—	$1'425.00	1.6 GHz	5.5 hour(s)	512 MB	80 GB	39.1 cm	2.86 kg
○	—	$1'190.00	3.2 GHz	1 hour(s)	512 MB	80 GB	39.1 cm	3.72 kg
○	—	$1'629.00	1.8 GHz	5.8 hour(s)	512 MB	60 GB	38.1 cm	2.81 kg
○	—	$627.10	1.6 GHz	1.5 hour(s)	256 MB	40 GB	38.1 cm	2.81 kg
○	—	$520.00	1.13 GHz	3.5 hour(s)	128 MB	30 GB	35.8 cm	2.59 kg

Organization interface

The most popular product

		Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
◉		—	$2'095.00	1.67 GHz	4.5 hours	512 MB	80 GB	38.6 cm	2.54 kg

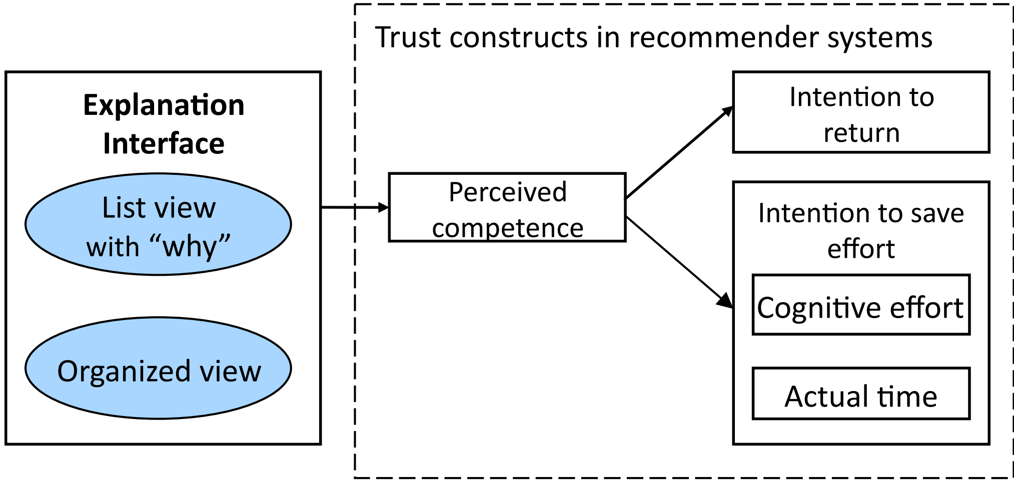
We also recommend the following products

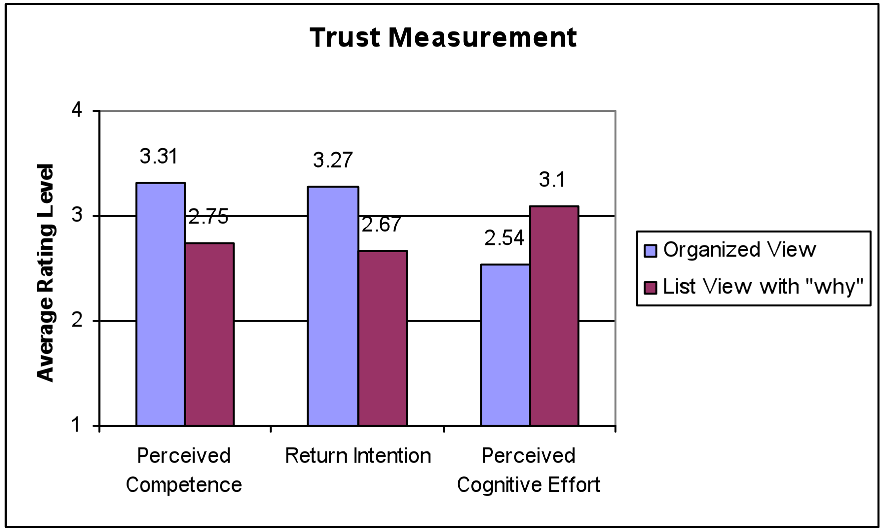
		Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
○	Why?	—	$1'220.49	1.8 GHz	5 hours	1 GB	100 GB	38.1 cm	2.95 kg
○	Why?	—	$2'148.99	2.0 GHz	4 hours	1 GB	100 GB	39.1 cm	2.90 kg
○	Why?	—	$1'379.00	3.3 GHz	2 hours	512 MB	100 GB	43.2 cm	4.31 kg
○	Why?	—	$1'179.00	3.2 GHz	2 hours	512 MB	80 GB	39.1 cm	3.62 kg
○	Why?	—	$1'529.00	1.7 GHz	6.5 hours	512 MB	80 GB	33.8 cm	1.77 kg
○	Why?	—	$1'599.00	1.7 GHz	6.5 hours	512 MB	80 GB	33.8 cm	1.91 kg
○	Why?	—	$1'425.00	1.6 GHz	5.5 hours	512 MB	80 GB	39.1 cm	2.86 kg
○	Why?	—	$2'235.00	1.8 GHz	2.5 hours	1 GB	100 GB	43.2 cm	3.99 kg
			$1'190.00	3.2 GHz	1 hours	512 MB	80 GB	39.1 cm	3.72 kg
			$1'125.00	1.5 GHz	6 hours	512 MB	80 GB	30.7 cm	2 kg
○	Why?	—	$2'319.00	1.67 GHz	4.5 hours	512 MB	100 GB	43.2 cm	3.13 kg
○	Why?	—	$1'499.00	1.5 GHz	5 hours	512 MB	80 GB	33.8 cm	1.91 kg
○	Why?	—	$1'739.99	1.5 GHz	4.5 hours	512 MB	80 GB	38.6 cm	2.49 kg
○	Why?	—	$1'629.00	1.8 GHz	5.8 hours	512 MB	60 GB	38.1 cm	2.81 kg
○	Why?	—	$1'625.99	1.5 GHz	5 hours	512 MB	80 GB	30.7 cm	2.09 kg
○	Why?	—	$1'426.99	1.5 GHz	5 hours	512 MB	60 GB	30.7 cm	2.09 kg
○	Why?	—	$2'099.99	1.2 GHz	9 hours	512 MB	60 GB	26.9 cm	1.41 kg
○	Why?	—	$2'075.00	1.8 GHz	1.67 hours	512 MB	100 GB	43.2 cm	4.4 kg
○	Why?	—	$1'649.00	1.1 GHz	8.5 hours	512 MB	40 GB	26.9 cm	1.36 kg
○	Why?	—	$627.10	1.6 GHz	1.5 hours	256 MB	40 GB	38.1 cm	2.81 kg
○	Why?	—	$969.00	1.2 GHz	6 hours	256 MB	39 GB	30.7 cm	2.22 kg
○	Why?	—	$520.00	1.13 GHz	3.5 hours	128 MB	30 GB	35.8 cm	2.59 kg
○	Why?	—	$1'929.00	1.2 GHz	4 hours	512 MB	60 GB	26.9 cm	1.41 kg
○	Why?	—	$1'595.00	1.0 GHz	5.5 hours	512 MB	40 GB	26.9 cm	1.41 kg

This product has higher processor speed and bigger hard drive capacity but is heavier

Ranked list

Results





Participants: 72; Material: online product finder; Procedure: within-subjects

Items in the **Perceived Competence** construct	Mean	
	Organized view	List view with "why"
I felt comfortable using the interface;	3.24	2.78
This interface enabled me to compare different products very efficiently.	3.38	2.72
Cronbach's alpha = 0.84		

Items in the **Intention to Return** construct	Mean	
	Organized view	List view with "why"
If I had to buy a product online in the future and an interface such as this was available, I would be very likely to use it;	3.11	2.56
I don't like this interface, so I would not use it again (*reverse scale*).	3.40	2.79
Cronbach's alpha = 0.91		

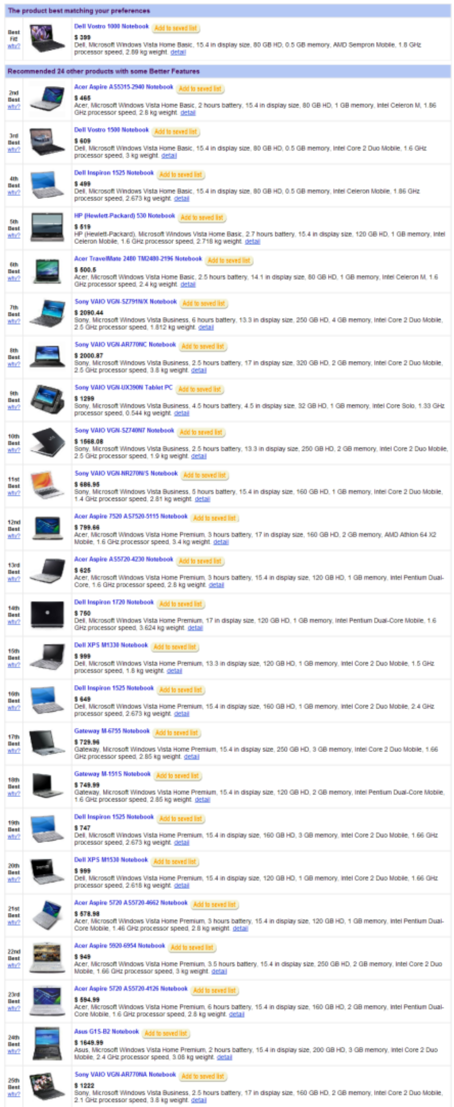
Items in the **Cognitive Effort** construct	Mean	
	Organized view	List view with "why"
I easily found the information I was looking for (*reverse scale*);	2.47	3.07
Selecting a product using this interface required too much effort.	2.61	3.14
Cronbach's alpha = 0.73		

Eye-Tracking Study (Chen and Pu, UMAP’10, IUI’11)

Three different layouts:

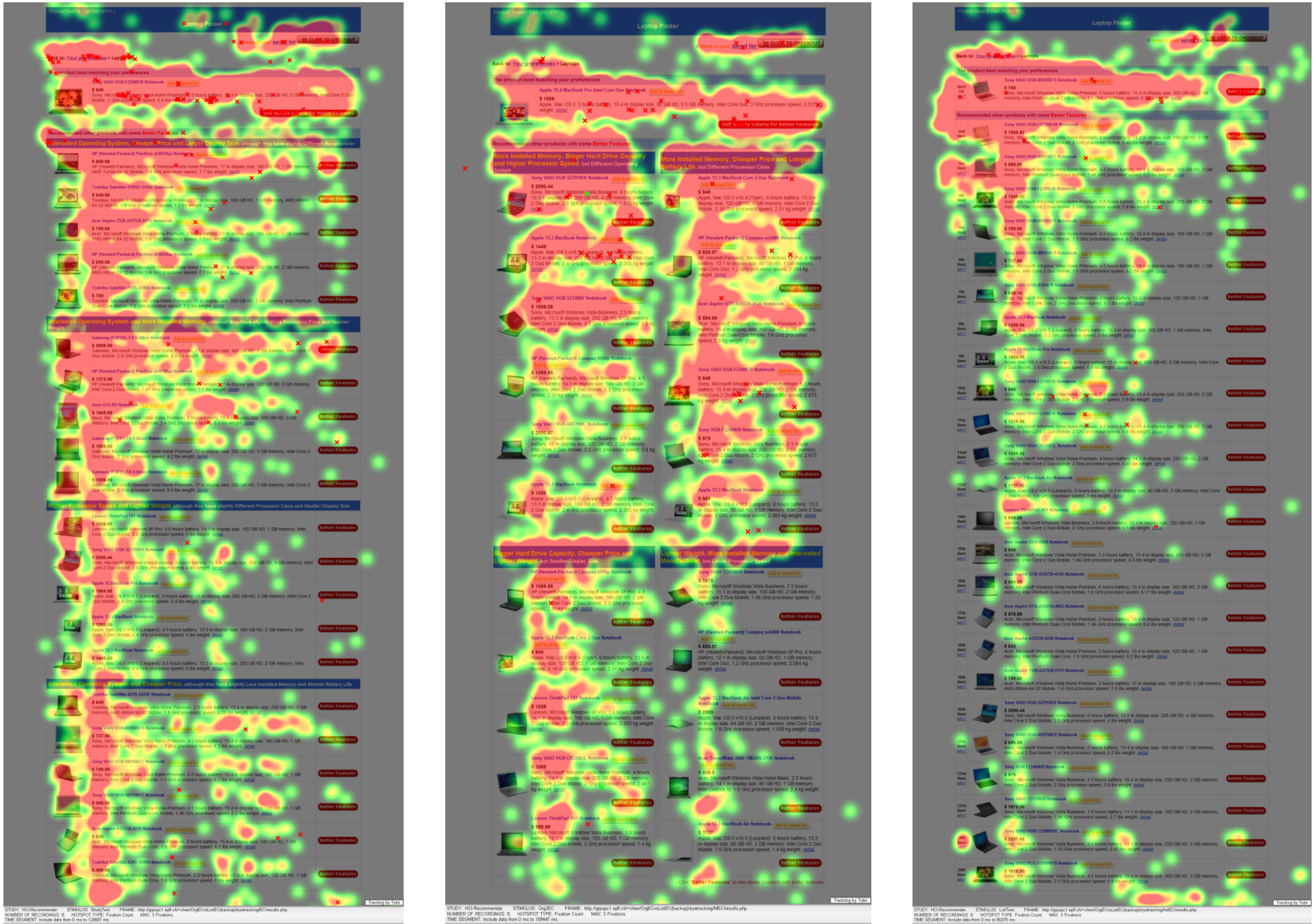
ORG1

ORG2



LIST

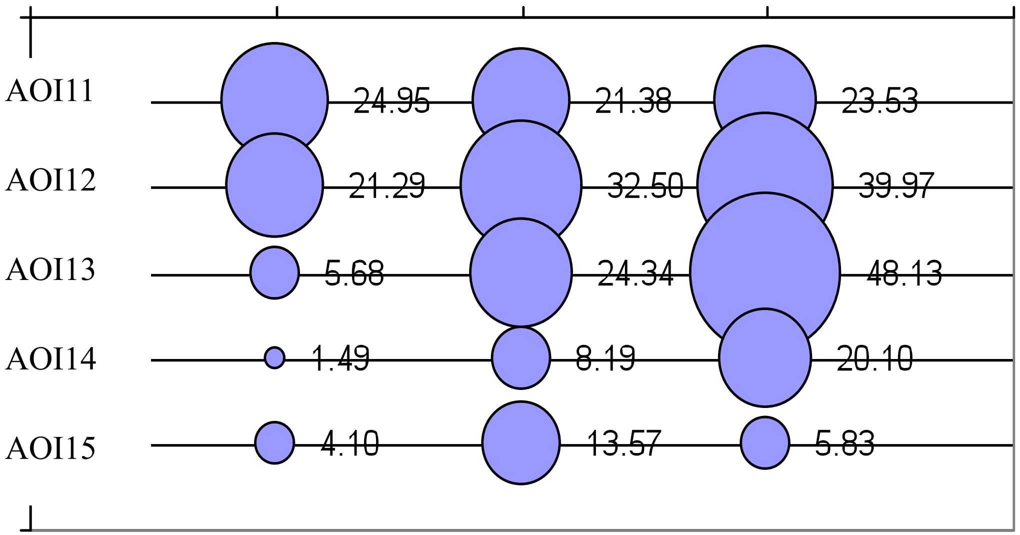
Hotspot plot



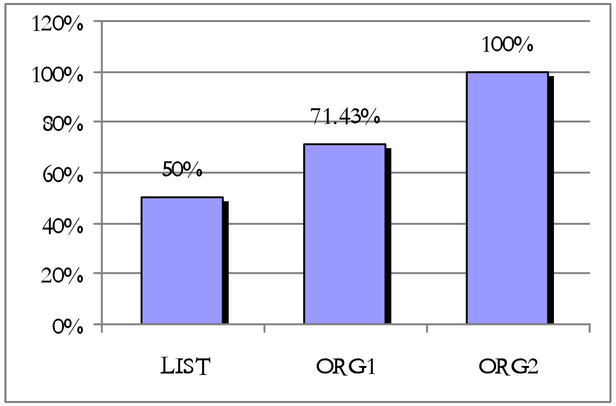
ORG1 ORG2 LIST

AOI analysis & decision quality

Fixation duration on each AOI



Percent of users who have finally made product choice



Distribution of users' choices among AOIs

	Average selections	Top item (AOI1)	AOI2	AOI3	AOI4	AOI5
LIST	1.33	25%	75%			
ORG1	1.86	23%	31%	15%	8%	23%
ORG2	3.2	12.5%	37.5%	37.5%	12.5%	

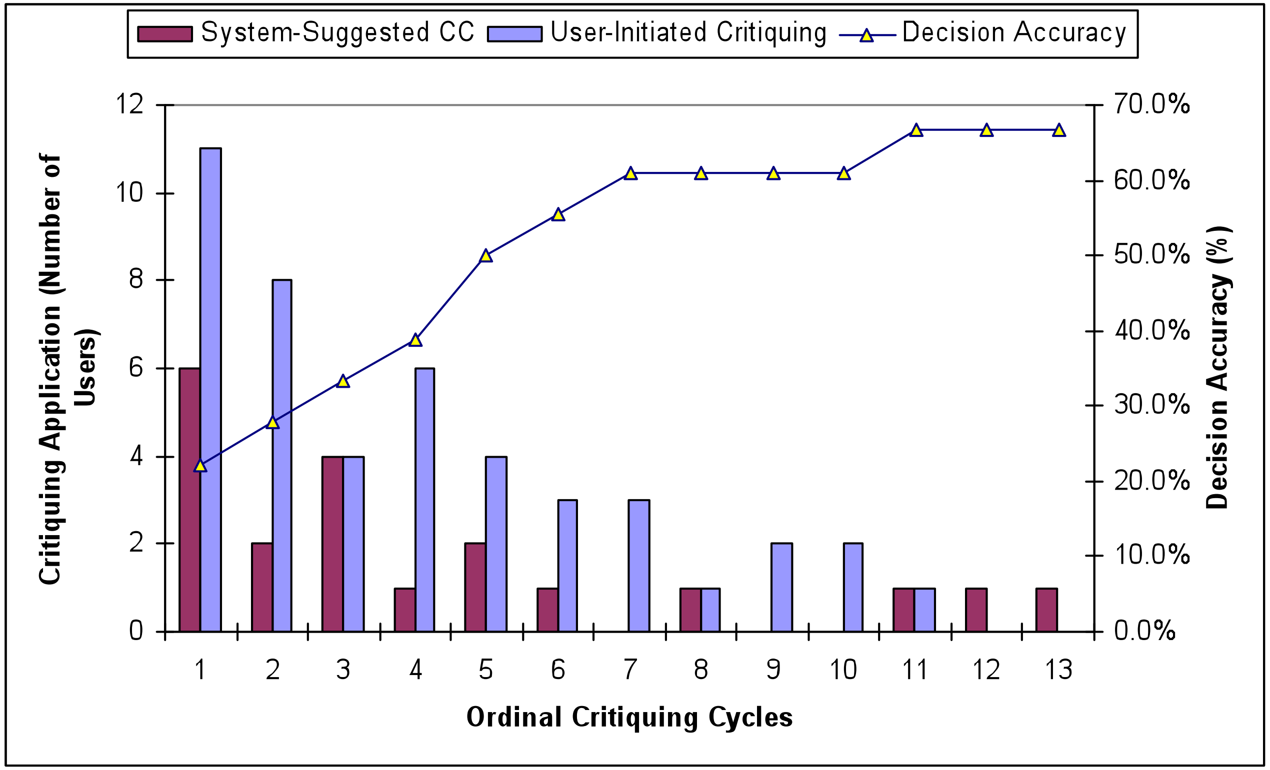
Cross-Cultural Study (Chen and Pu, RecSys'08)

	Oriental Culture (60)	*Western Culture (60)*
Nation	China (60)	Switzerland (41); Other European countries (19)
Gender	Female (23); Male (37)	Female (15); Male (45)
Average age	21~30 (57); >30 (3)	<21 (14); 21~30 (44); >30 (2)
Major/ job domain	Computer, mathematics, environment, electronics, architecture, etc.	Computer, education, mechanics, electronics,, architecture, etc.
Computer knowledge	4.34 (advanced)	4.08 (advanced)
Internet usage	4.83 (almost daily)	4.98 (almost daily)
e-commerce site visits	3.69 (1-3 times a month)	3.36 (a few times every 3 months)
e-shopping experiences	3.25 (a few times every 3 months)	2.92 (a few times every 3 months)

- People from different cultural backgrounds basically performed similar regarding both objective performance and subjective perceptions
- Significant favor of ORG against LIST
- Stronger among Chinese participants

Experiment 4: Evaluation of hybrid critiquing (IUI'07)

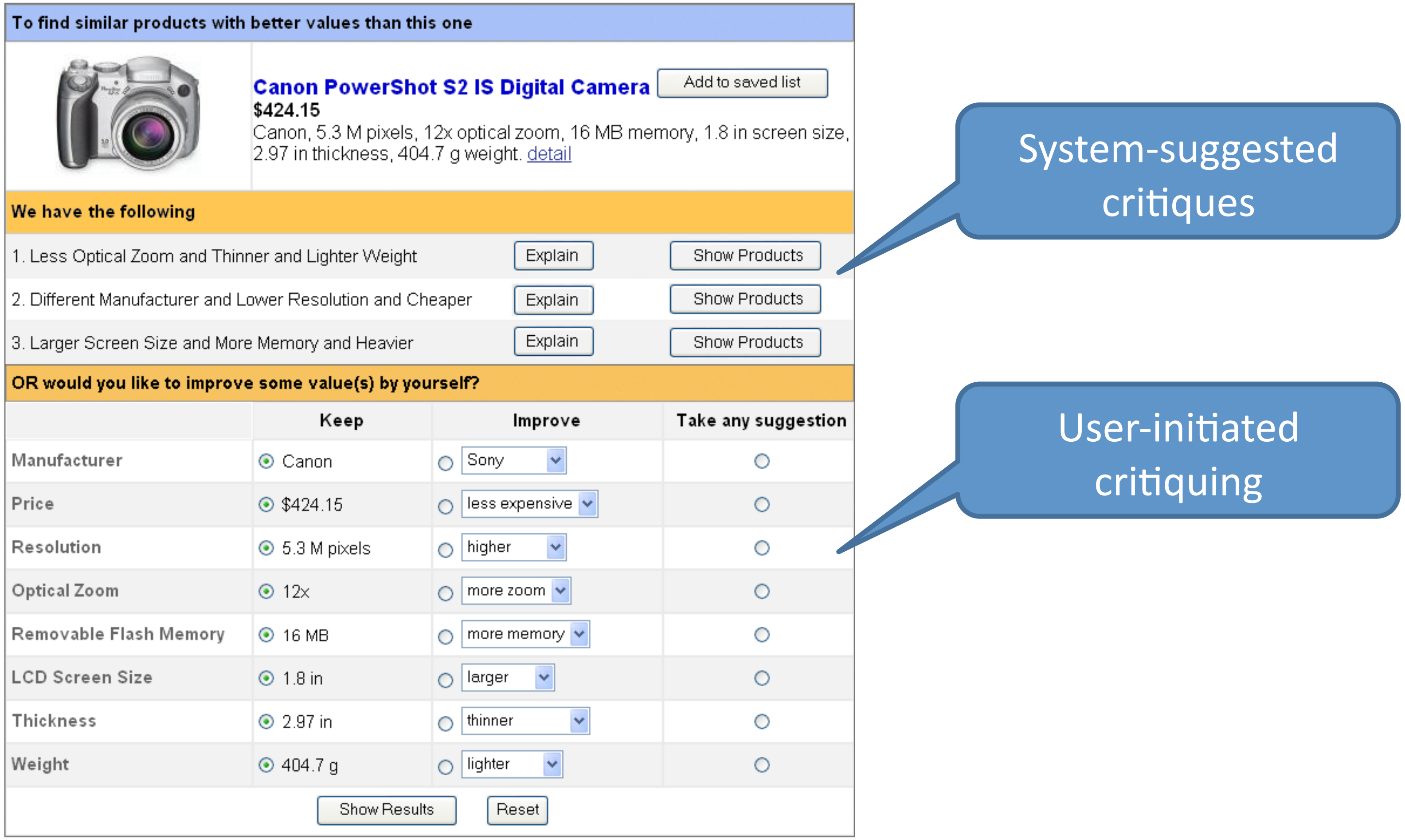
Participants: 18



➔ Users behaved more active in creating their own criteria with the self-initiated critiquing aid, relative to their application of the system-suggested critiques

4^{th} type: Hybrid Critiquing

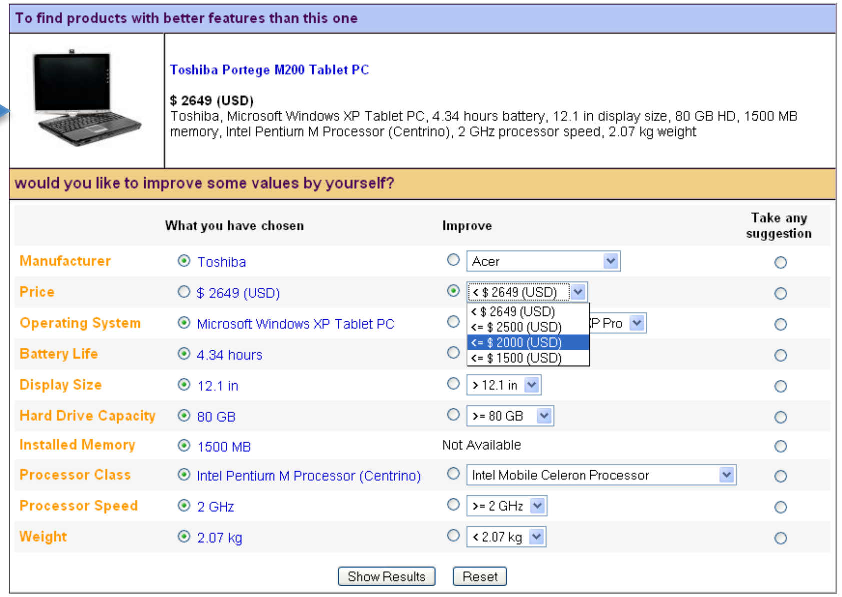
Hybrid of ***system-suggested critiques*** and ***user-initiated critiquing*** (Chen and Pu, IUI'07)



Extension: Chen and Pu, RecSys'07

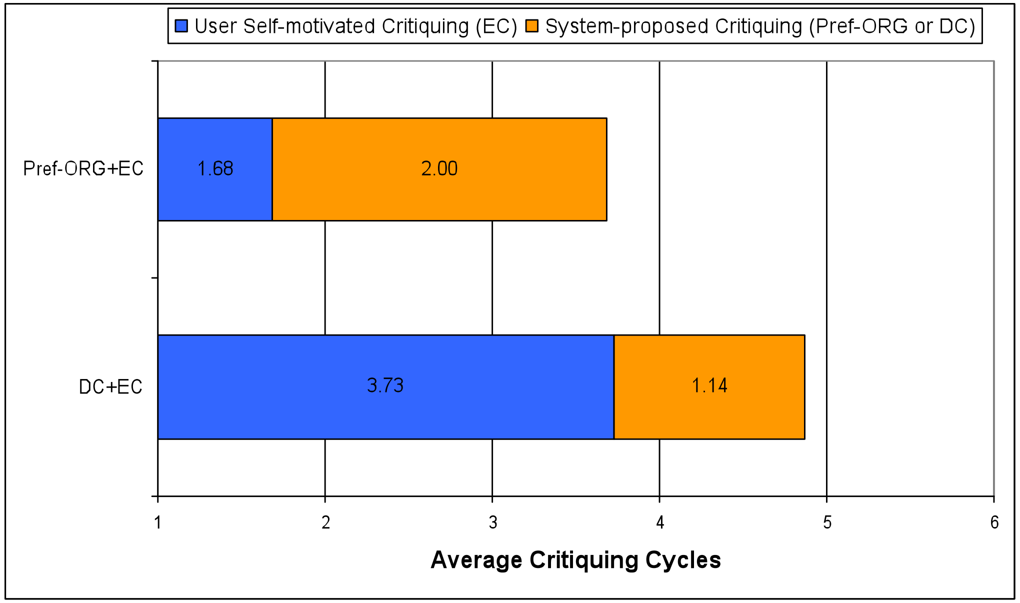
Preference-based organization

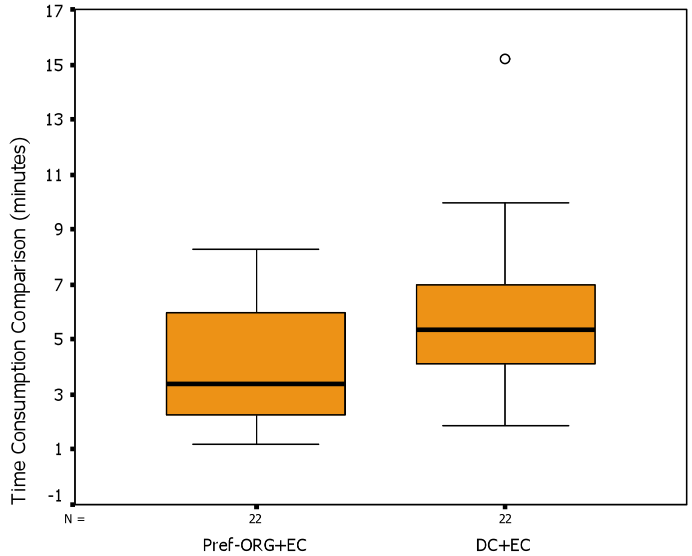
User-initiated critiquing support



If suggested critiques and products do not interest the user in the organization interface, she could switch to make the self-initiated critiquing by clicking the button

Participants: 44; between-group experiment procedure





→ The integration of the preference-based recommendation organization in hybrid critiquing can effectively help **increase the suggested critiques' application frequency** and significantly contribute **to saving users' task time and interaction effort**

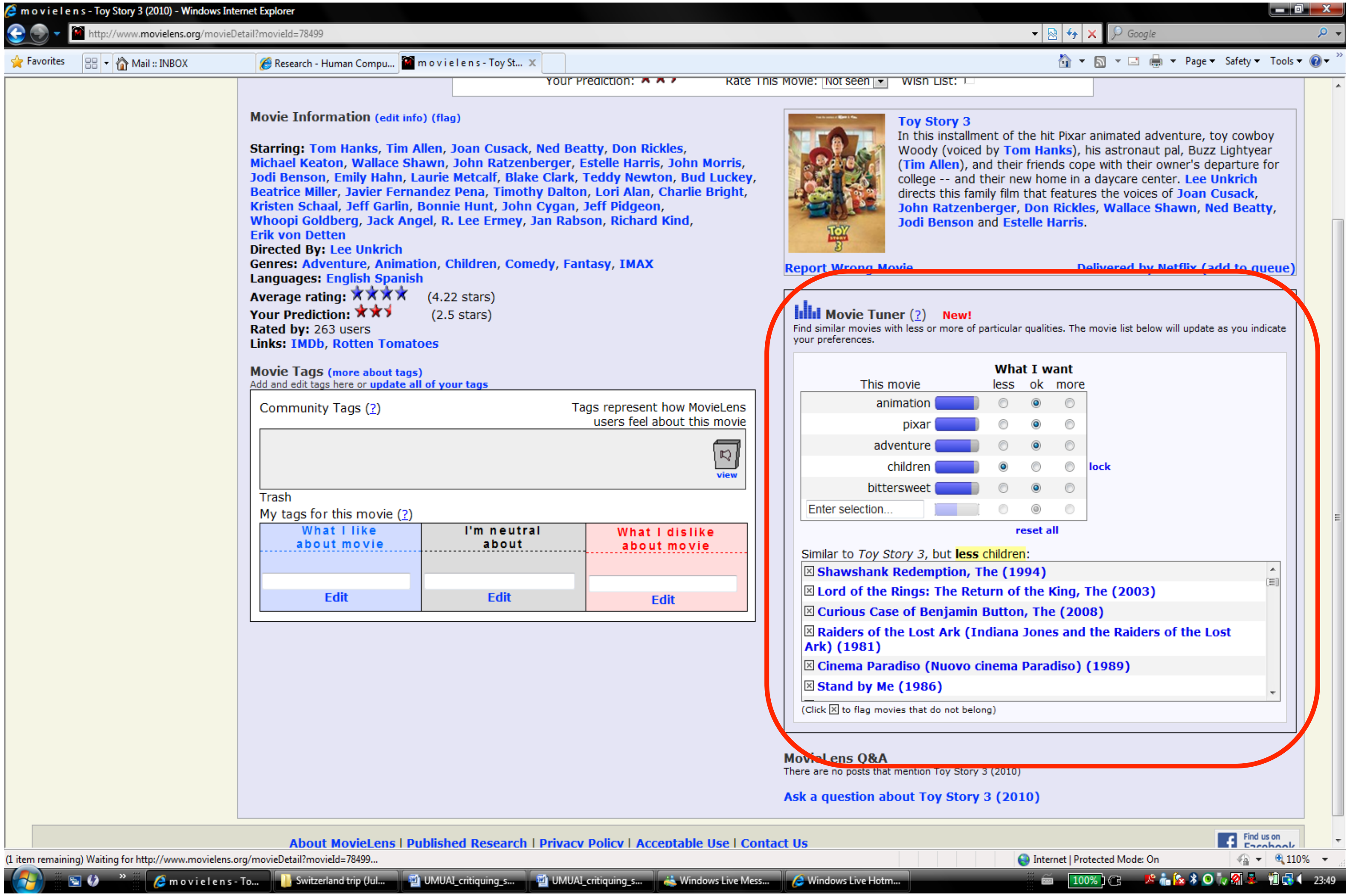
Outline

- What is critiquing-based recommender system and Why?
- Development history
- User experiences
- Conclusion

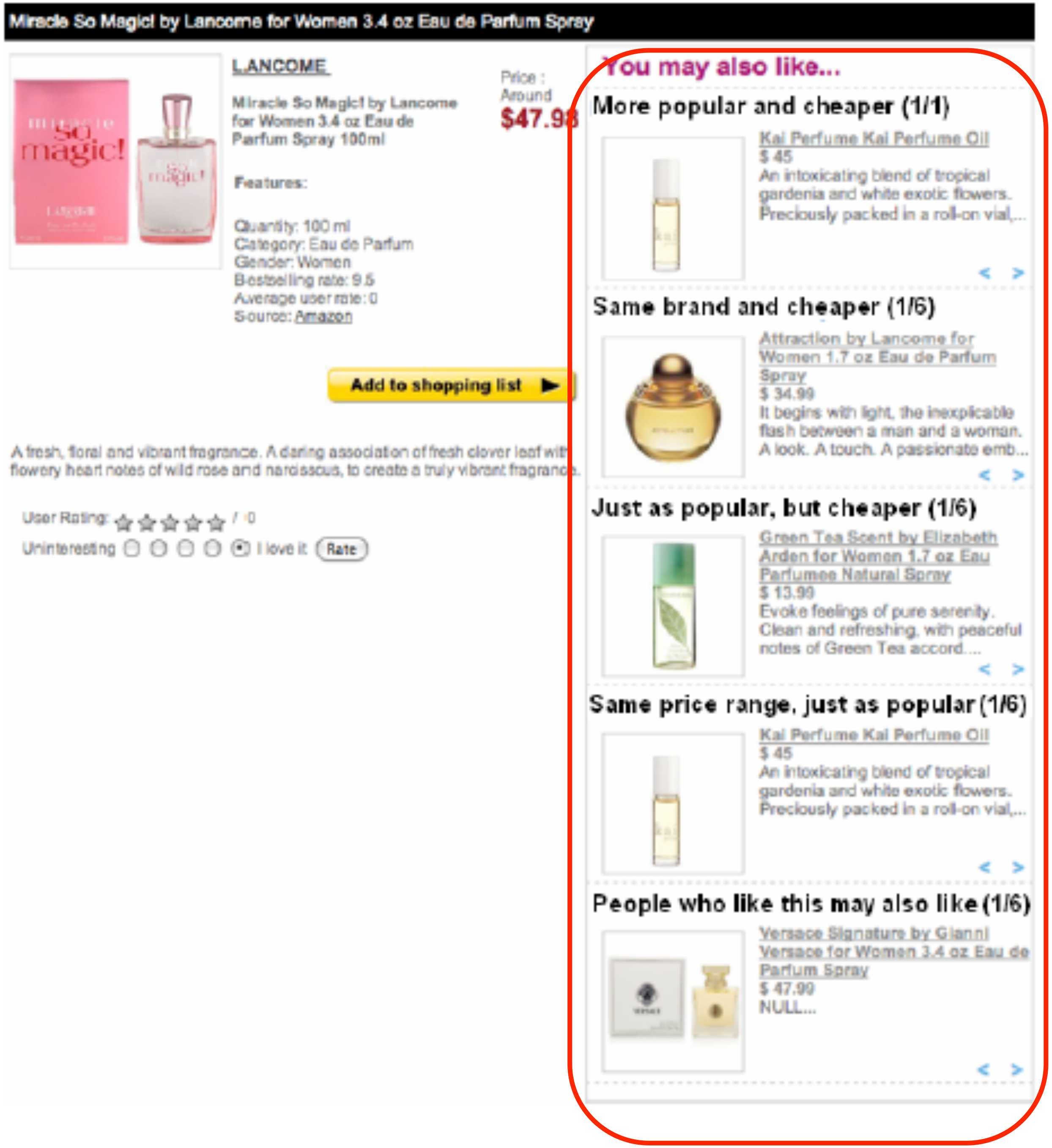
What can be concluded?

- Design guidelines
  - Li Chen and Pearl Pu. Interaction Design Guidelines on Critiquing-based Recommender Systems. *User Modeling and User-Adapted Interaction Journal (**UMUAI**)*, vol. 19 (3), pages 167-206, 2009.
  - Li Chen and Pearl Pu. Critiquing-based Recommenders: Survey and Emerging Trends. *User Modeling and User-Adapted Interaction Journal (**UMUAI**)*, vol. 22(1), pages 125-150, 2012.
- User-centric evaluation framework for RS
  - Pearl Pu, Li Chen and Rong Hu. A User-Centric Evaluation Framework for Recommender Systems. In *Proceedings of the 5th ACM Conference on Recommender Systems (**RecSys'11**)*, pages 157-164, Chicago, IL, USA, October 23-27, 2011.

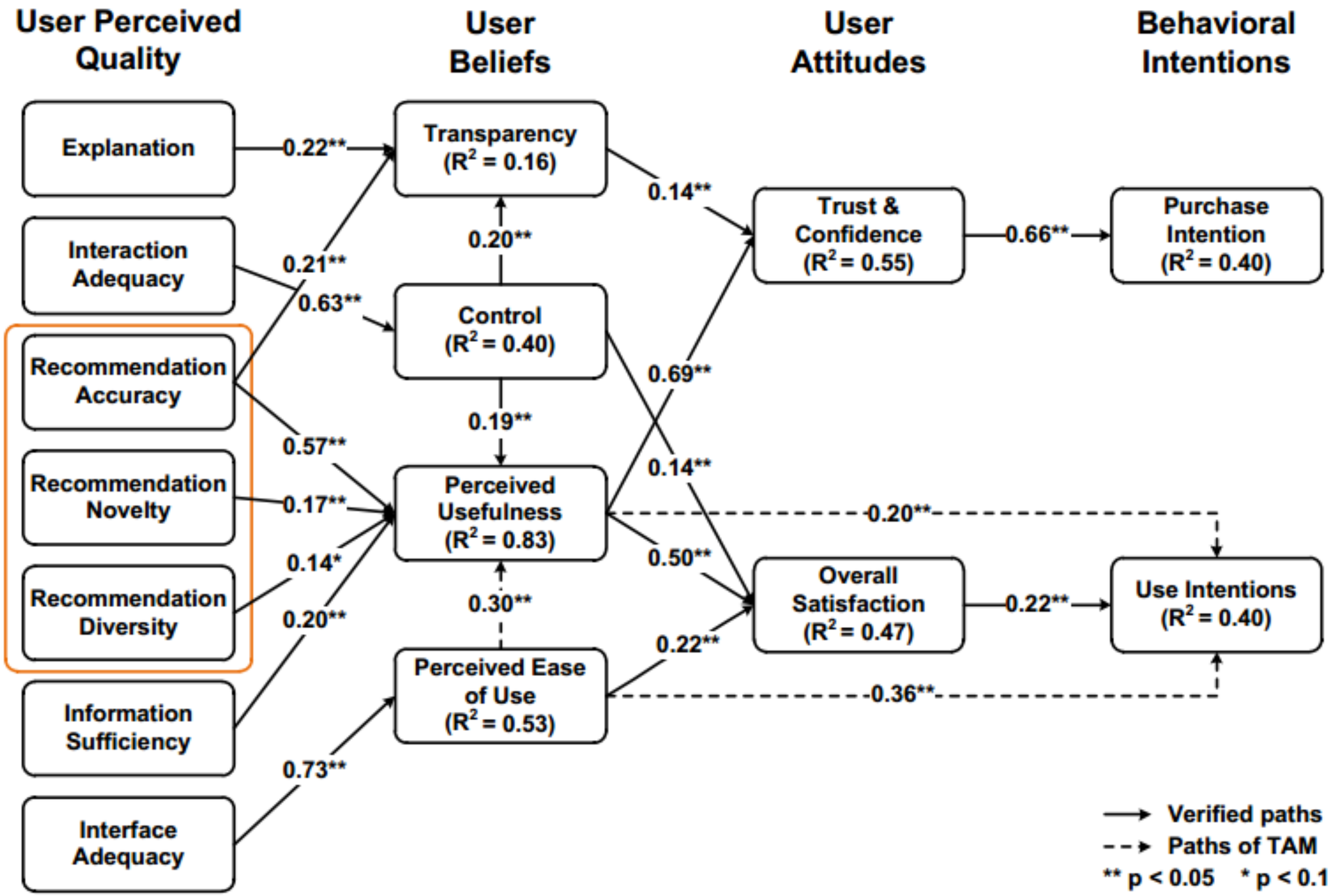
Critiquing in MovieLens (Vig *et al.*, IUI'11)



Editorial Picked Critiques (Pu *et al.*, RecSys'09)



ResQue: **Re**commender **s**ystems' **Qu**ality of user **e**xperience



32 questions and 15 constructs

1. Recommendation Accuracy
The items recommended to me matched my interests.

2. Recommendation Novelty
The items recommended to me are novel.
The recommender system helped me discover new products.

3. Recommendation Diversity
The items recommended to me are diverse.

4. Interface Adequacy
The labels of the recommender interface are clear.
The labels of the recommender interface are adequate.
The layout of the recommender interface is attractive.
The layout of the recommender interface is adequate.

5. Explanation
The recommender explains why the products are recommended to me.

6. Information Sufficiency
The information provided for the recommended items is sufficient for me to make a purchase/download decision.

7. Interaction Adequacy
The recommender allows me to tell what I like/dislike.
I found it easy to tell the system what I like/dislike.
I found it easy to inform the system if I dislike/like the recommended item.

8. Perceived Ease of Use
I became familiar with the recommender system very quickly.
I easily found the recommended items.

9. Control
I feel in control of modifying my taste profile.
The recommender allows me to modify my taste profile.
I found it easy to modify my taste profile in the recommender.

10. Transparency
I understood why the items were recommended to me.

11. Perceived Usefulness
The recommender helped me find the ideal item.
Using the recommender to find what I like is easy.
The recommender gave me good suggestions.

12. Overall Satisfaction
Overall, I am satisfied with the recommender.

13. Confidence & Trust
I am convinced of the items recommended to me.
I am confident I will like the items recommended to me.
The recommender made me more confident about my selection/decision.
The recommender can be trusted.

14. Use Intentions
I will use this recommender again.
I will use this recommender frequently.
I will tell my friends about this recommender.

15. Purchase Intention
I would buy the items recommended, given the opportunity.

End of my talk

- Thanks!
- Q&A

- Call For Papers
  - 2014 5th International Workshop on Social Recommender System (in WWW'14) (http://users.soe.ucsc.edu/~jwang30/srs2014/)
  - 2014 ACM Conference on Recommender System (http://recsys.acm.org/recsys14/)