

# Improving User Confidence in Decision Support Systems for Electronic Catalogs

*PhD Candidate:* David Portabella Clotet  
*Supervisor:* Martin Rajman

EFPL – IC – IIF – LIA  
December 20, 2007

# We make decisions...

- Where to go for holidays
- Your vote in elections
- Shopping  
e.g. which photo-camera to buy
- ...

# Objective of this thesis

Improve *decision confidence*  
in decision support systems.

*Decision confidence:*

the certainty of the decision maker that  
she has made the best decision.

# Motivation

- Confidence -> commit to decision
- Confidence -> justify to others
- E-commerce is a big and growing market

Fill research gap:

- Previous research has focused **mainly** on improving decision accuracy

# Contents

- Previous work
- Improving confidence
- User study
- Conclusions

# How we should make accurate decisions

Prescriptive Decision Theory

e.g. Weighted additive strategy (WADD)

[Neumann47, Fishburn70, Kenney76]

Main difficulty: We have conflicting objectives

Based on *utilities* (a decision analyst is required)



Want to know more?

I want to become rich. How to decide which lottery to buy?

Decision theory under uncertainty...

# How we actually make decisions

Descriptive Decision Theory

e.g. Lexicographic (LEX), Elimination by aspects (EBA)

[Tversky72, Payne93]

We use *constraints*

Tradeoff: Accuracy  $\Leftrightarrow$  Effort

Tradeoff: Accuracy  $\Leftrightarrow$  Confidence

# Revolution... Computers!

Now we can make more accurate decisions  
with less effort...




Want to know more?

Reduce effort by using *parametric* utilities...

# Decision support systems

- Matching list [Rabbit84]
  - Conflict [Kaplan82]
  - Corrective [Motro90]
  - Ranking list
  - Diversity [Linden97, Shimazu01, Smyth01]
  - Suggestions [Linden97, Faltings04, Viappiani06]
  - System-proposed critiquing [Burke96, Smyth04, Pu06]
- ➡ Want to know more?
- Visualization feedback
  - Collaborative filtering: People that bought this also bought X

# User studies on accuracy

	There exists user studies
Matching list (+EBA, +LEX)	Yes
Conflict and corrective	No 
Ranking list	Yes
Diversity	Yes
Suggestions	Yes
System-proposed critiquing	Yes

# Improving confidence

# Our hypothesis

Even if the machine makes the computation for us,  
we will not feel *confident*  
with *utility*-based decision strategies.

# Accuracy $\Leftrightarrow$ Confidence





Eliciting the *utility* for color







Balancing the preferences

# Properties for building confidence 1/2

- Constraint-based (ease to internalize)
- Universally valid (reusable)
- Negatory (support for falsifying hypotheses)
- Unique in its class (ease of verification)

# Properties for building confidence 2/2

		Constraint-based	Universally valid	Negatory	Unique in its class
 Conflict 		<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>
 Corrective 		<b>Yes</b>	No	<b>Yes</b>	<b>Yes</b>
Ranking List		Depends on the alg.	No	No	No
Diversity		No	No	No	No
System-proposed critiquing		Depends on the alg.	No	No	No

# New type of explanation (CBTE)

# Confidence Building Tradeoff Explanations (CBTE)

Explain the query tradeoffs

in terms of conflicts and relaxation

that potentially improve confidence.

Application

There are 0 items (out of 186) matching your criteria.

Selection criteria		Reset
rent	<= 400 CHF	clear
Nbr of rooms	>=	clear
Location	one of	clear
time to unil	<=	clear
time to center	<=	clear
kitchen	<input type="radio"/> yes <input type="radio"/> no	clear
bathroom	<input type="radio"/> yes <input type="radio"/> no	clear
furnished	<input type="radio"/> yes <input checked="" type="radio"/> no	clear
bus	<input type="radio"/> yes <input type="radio"/> no	clear
metro m1	<input checked="" type="radio"/> yes <input type="radio"/> no	clear
parking	<input type="radio"/> yes <input type="radio"/> no	clear

Matching Items						
rent	Nbr of roo...	Location	time to unil	time to ce...	kitchen	ba

Considered items		Reset

### Trade-off analysis

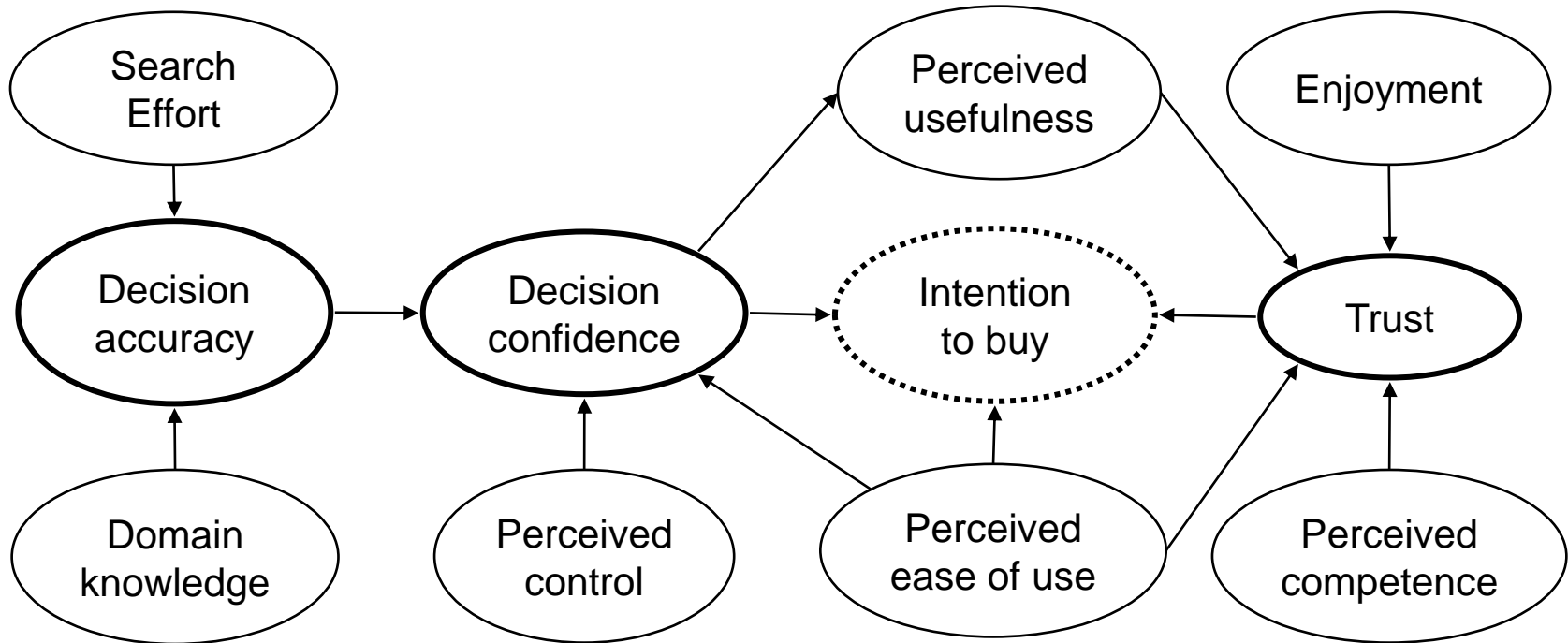
There are items if you [change rent <= 475 CHF \(instead of <= 400 CHF\)](#) or [clear furnished](#)

---

There are no items with: [rent <= 400 CHF](#) and [furnished = no](#)  
 To solve: [change rent <= 415 CHF \(instead of <= 400 CHF\)](#) or [clear furnished](#)

# User study: Methodology

# Research model



# User study

## Compare:

- Baseline (Matching list)
- CBTE
- Ranking list (PWADD)
- CBTE + Ranking list



Want to know more?

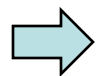
The baseline system is quite simple, yet most current websites have an extremely poor design

# Methodology

Users: 13 participants / group

Stimuli: Catalog of used cars from Comparis<sup>©</sup>:  
– 200 cars, defined with 35 attributes

Task: Find your most preferred car  
(open question)



Want to know more?

We turned the 2-factor experiment into a 1-factor experiment.  
3 groups instead of 4 groups of participants.

# Results: Subjective measures

7-point Likert scale: 1-Strongly disagree, 4-Neutral, 7-Strongly agree

**< 0.05**: statistically significant difference

Construct	Mean			Statistical significance (wilcoxon p-value)		
	G1 CBTE	G2 PWADD	G3 both	G1-G2	G1-G3	G2-G3
+ Perceived ease of use and usefulness	5.40	4.71	5.66	0.003	0.045	0.003
+ Confidence after validation	5.94	5.06	5.67	0.004	0.028	0.006
+ Confidence before validation	5.73	4.95	5.50	0.015	0.116	0.126
+ Trust	5.00	4.33	5.39	0.008	0.047	0.003
+ Domain knowledge	4.73	4.90	5.23	0.554	0.043	0.050
+ Perceived control	4.09	3.45	4.55	0.018	0.091	0.008

# Results: decision accuracy

Mean			Chi-square p-value		
G1	G2	G3	G1-G2	G1-G3	G2-G3
69.2%	61.5%	69.2%	0.68	0.68	1

## Conclusion:

Same decision accuracy for the three systems ~60-70%  
(no statistically significant different)



Want to know more?

Decision accuracy was objectively measured

# Results: Perceived effort

## Nasa TLX test

Mean			Wilcoxon p-value		
G1	G2	G3	G1-G2	G1-G3	G2-G3
45.7	43	38	0.42	0.50	0.23

### Conclusion:

Same perceived effort

(no statistically significant different)



Want to know more?

The Nasa TLX test is not very useful for measuring effort on user interfaces

# User study: Interpretation of the results

# Decision accuracy

Baseline: 23% - 38%

CBTE, PWADD: 61%, 69%

-> Equivalent amount of information

CBTE+PWADD: 69% 🤖

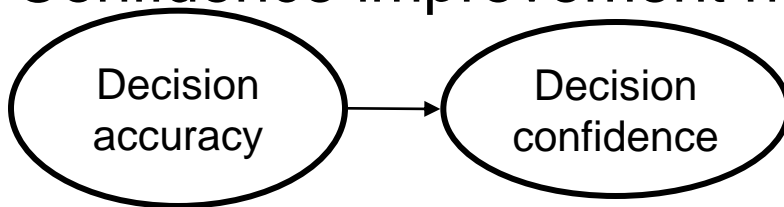
-> Equivalent type of information

# Main result: Decision confidence

CBTE compared to PWADD:

equal decision accuracy, better decision confidence

Confidence improvement not only because:



Validity evidence for DCEBS framework 😊

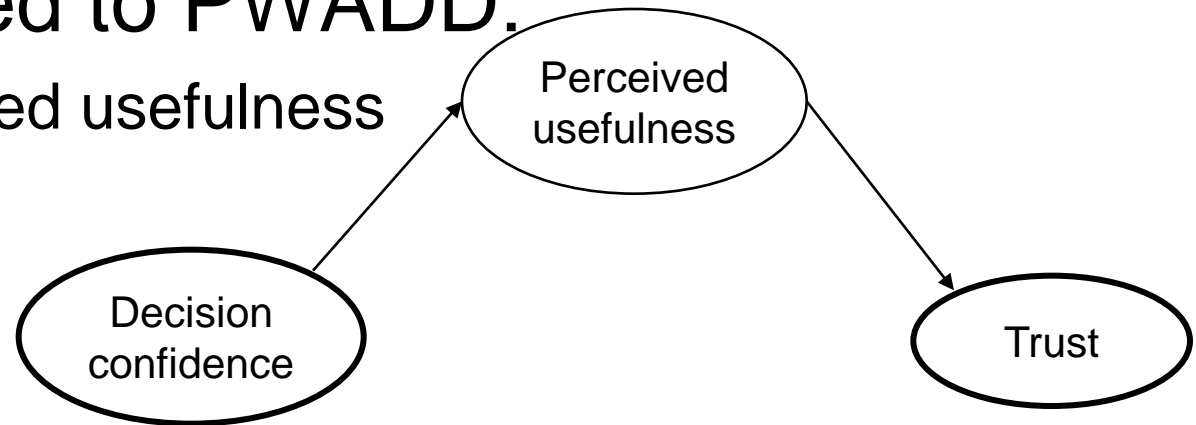
Decision justification:

CBTE helps to build a qualitative argumentation for one's decision.

# Perceived usefulness & Trust

## CBTE compared to PWADD:

- Better Perceived usefulness
- Better Trust



Want to know more?

Combining two systems can lead to strange results.

- Individual heterogeneity [Ariely00]
- Dynamic heterogeneity [Ariely00]

# Perceived control

CBTE = baseline

PWADD < baseline

Ranking list removes the possibility to sort items by attribute.

Previous studies: Users like to feel control of the system.

+ Perceived control -> +Perceived usefulness?

# Accurate objectives

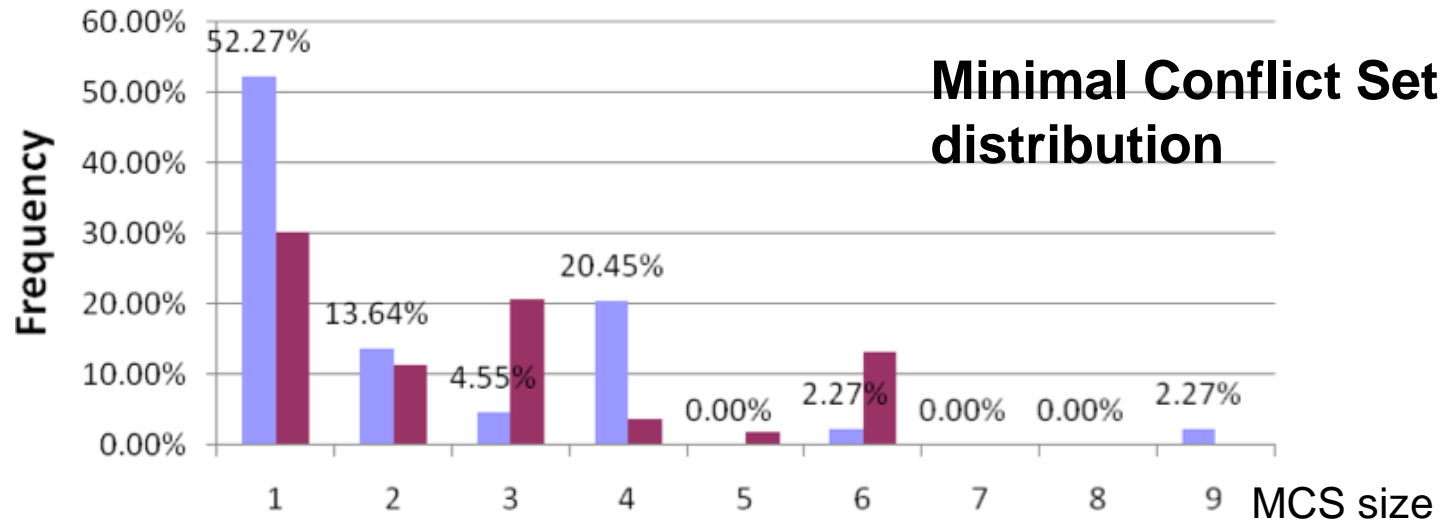
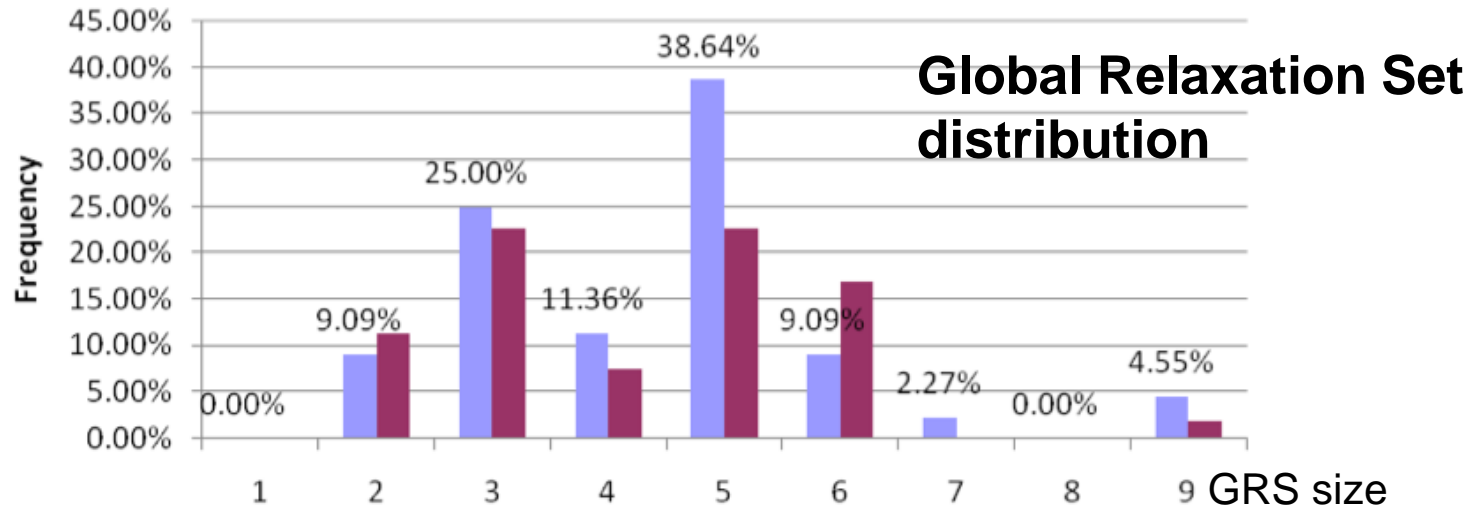
Participants with at least one failing query:

- 100% in CBTE 😊
- 69.2% PWADD
- 46.2% baseline

CBTE: triggers more accurate objectives

CBTE: not intimidated by empty answers

# CBTE size



CBTE: size not problematic

# Limitations of the user study

- Few participants (13 per group)
- Generalization not guaranteed:  
participants, stimuli, task
- The use of tutorials



Want to know more?

We also got some results through simulations of human behavior. They are less convincing, but less costly.

# User study: conclusions

Accuracy and confidence: two different goals

- 😊 CBTE: better decision confidence  
(based on DCEBS framework)
- 😊 CBTE: explanation size not problematic
- 😊 CBTE: triggers more accurate objectives
- 😨 CBTE and PWADD: equivalent type of information  
for decision accuracy

# Conclusions

# Contributions

- A unified framework: Query-feedback search, including decision confidence (DCEBS)
- A new type of feedback improving decision confidence (CBTE)
- An experimental evaluation through a user study



Want to know more?

Open-source software, catalog and collected data available

# Future work

- Remove assumption “homogeneous set of attributes”
- Remove assumption “decision theory under certainty”
- Computationally efficient
- New types of feedback,  
e.g. tradeoff contrasts [Simonson92]
- Integrating CF to content-based search
- Produce constraint-based feedback from utility functions

# Take away message

“Conflict and corrective feedback  
improve decision confidence”

Questions?