Understanding the Effect of Adaptive Preference Elicitation Methods on User Satisfaction of a Recommender System

Bart P. Knijnenburg
Eindhoven University of Technology,
Human-Technology Interaction group
P.O. Box 513, 5600 MB Eindhoven
+31 40 247 8420
B.P.Knijnenburg@tue.nl

Martijn C. Willemsen
Eindhoven University of Technology,
Human-Technology Interaction group
P.O. Box 513, 5600 MB Eindhoven
+31 40 247 2561
M.C.Willemsen@tue.nl

ABSTRACT
In a recommender system that suggests options based on user attribute weights, the method of preference elicitation (PE) employed by a recommender system can influence users' satisfaction with the system, as well as the perceived usefulness and the understandability of the system. Specifically, we hypothesize that users with different levels of domain knowledge prefer different types of PE. While domain experts reported higher satisfaction and perceived usefulness with attribute-based PE (i.e., indicating preference levels for the domain-related attributes), novices preferred case-based PE (i.e., indicating the preference for specific examples, from which attribute-preferences can then be implicitly calculated). The paper discusses the decision-theoretical principles that are believed to lead to this distinction, as well as an experiment that provides substantial evidence for the hypothesis. Consequently, we introduce the idea of adapting the method of PE to users' domain knowledge on the fly using click stream data.

Categories and Subject Descriptors
H.1.2. **[Models and principles]**: User/Machine Systems–*software psychology*; H.4.2. **[Information Systems Applications]**: Types of Systems–*decision support*; H.5.2 **[Information Interfaces and Presentation]**: User Interfaces–*evaluation/methodology, interaction styles, user centered design*

General Terms
Measurement, Design, Experimentation, Human Factors.

Keywords
Attribute-based versus Case-based Preference Elicitation, User Interfaces, Satisfaction, Perceived Usefulness.

1. INTRODUCTION
Multi-Attribute Utility Theory (MAUT) is one of the methods recommender systems can employ to decide what to recommend [1]. For each user, the utility of a certain choice option is calculated by multiplying the values of each of its attributes with the user's weight of that attribute. Using MAUT requires that the system somehow discovers the attribute weights that optimize the recommendations for each user. This process is called *preference elicitation* (PE). The most straightforward way to do PE is by letting users explicitly assign attribute weights (attribute-based PE, [2,3]), but there are also ways to derive these weights from users' critique on entire choice options (case-based PE [4,5,6,7]).

Although MAUT has a universal applicability, we believe that the preferred method of PE critically depends on the type of person using the system. Letting go of the 'one-size-fits-all' solution commonly used in recommender systems [8], this paper argues that the preference elicitation method of a recommender system should be tailored (or even dynamically adapted) to the level of user *domain knowledge*. The remainder of this paper develops this argument in more detail and discusses the results of an experiment that provides evidence for its validity.

2. THEORY
In choice situations, 'domain knowledge' or 'expertise' can be described as the knowledge that is instrumental (or even required) to make adequate decisions. This includes knowledge about the attributes, their values, and their implications on product quality, as well as common trade-offs in making choices in the current domain. As domain experts and novices differ strongly in both the amount and detail of their domain knowledge, different ways of PE might be optimal for experts and novices [9].

2.1 Attribute-based PE for experts
In the most-used PE method, 'attribute-weight selection', users directly indicate the importance of each of the attributes with which choice options are described. Although several studies [2,3] find a general increase in decision quality and satisfaction by using attribute-based PE, we predict that this PE method works best for domain experts, because they are more familiar with these attributes [10], better understand the value of each of them [11], and are more capable of making trade-offs between them [12]. With such in-depth knowledge about the attributes, it is natural to let expert users assign weights to these attributes directly [13]. On the other hand, novices often do not possess detailed knowledge of attributes to be able to directly assign weights to them [10,14,12] and therefore might need a different PE method.

2.2 Case-based PE for novices
Instead of letting users explicitly assign weights to attributes, 'Case-based Recommendation' calculates attribute weights by analyzing the users' positive or negative evaluation of exemplary choice options [15,16]. Although recent studies [6,7] show that

RecSys'09, October 23–25, 2009, New York, New York, USA.

this case-based PE results in better decisions and higher satisfaction across the board, we expect that experts prefer a direct assignment of attribute weights over a latent one. Novice decision-makers, on the other hand, lack the knowledge to decide which attributes are important [4,5,11,17,10], and they don't know what combinations of attribute values to expect in the choice set [18]. Because of this, novices prefer to evaluate entire choice options holistically and determine their preference incrementally [14]. Case-based PE matches these needs, as it allows users to evaluate entire choice options, and gives the user feedforward by showing what type of options would show up if they would change their preference accordingly.

3. USER EXPERIMENT

3.1 Hypothesis

Xiao and Benbasat noted that no study has investigated the effect of the PE method on users' perceptions of how easy it is to use a recommender system [10]. Based on the reasoning provided in our theory section, we therefore state the following hypothesis to be investigated in our user experiment:

H1. Novices have a higher satisfaction and perceive the system as more useful when they use the case-based PE method (compared to the attribute-based PE method), while experts have a higher satisfaction and perceive the system as more useful when they use the attribute-based PE method (compared to the case-based PE method).

3.2 System, participants and task

We developed an online MAUT-based recommender system and provided it with a wide variety of 80 energy-saving measures. The domain of energy-saving was used because of its relevance in today's society. Also, in this domain it is natural to make multiple choices, and to devise a realistic experiment without having to sell our participants anything. More importantly, literature on energy-saving behavior shows that domain knowledge and commitment to save energy are two main decision-maker characteristics [19,20]. The energy-saving domain thus provides a wide variety of experts and novices, with various levels of commitment.

We developed two versions of the system: one with attribute-based PE and one with case-based PE (see Figure 1). The top part of the interface contains the PE interface. The middle part shows recommendations (based on attribute weights extracted by the PE in the top part). The bottom part contains two lists of chosen measures ("what I want to do", and "what I am doing already") and their total amount of energy/cost savings. The attribute-based PE version shows each attribute with buttons to increase or decrease their weight. The case-based PE version shows a list of *trade-off recommendations* that match the attribute weights except for one attribute, which has been made more important (i.e. its weight is increased at the expense of all others). The set adheres to the 'look ahead' principle [6,7]: the options show the possible outcomes of the compromises the user can make to the current recommendations. This also reduces the risk of turning preference elicitation into a self-fulfilling prophecy [21]. A positive (negative) evaluation in the case-based system can be treated similar to the increase (decrease) of the weight of the 'important' attribute in the attribute-based system. This way, the same MAUT-based recommendation engine is used for both systems.

Participants were recruited via posts on Internet forums and weblogs, some energy-related and some general interest. They were asked to participate to "help make further improvements to the system", and were promised a small financial reward. 145

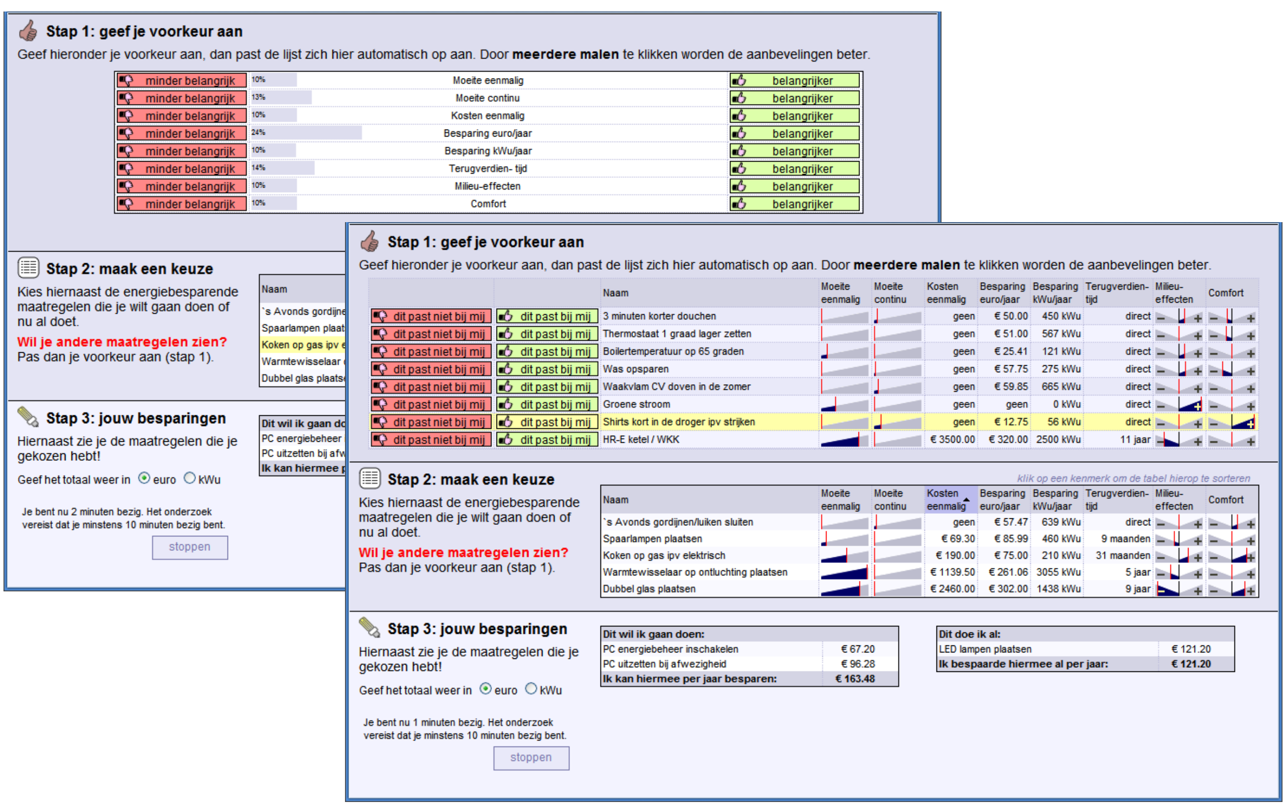

Figure 1. Our recommender system, in the two varieties of case-based PE (top) and attribute-based PE (bottom).

participants finished the pre-experimental questionnaires. 89 of them also finished the rest of the experiment*. The resulting sample was biased towards males (34 female, 111 male), but had good distribution of age (M = 35.7, SD = 11.6), education (12 high school, 23 intermediate, 59 higher, 52 university education) and occupation (27 students, 104 employed, 14 retired).

Participants were given several pre-experimental questionnaires and an explanation of the system. Participants were instructed that the goal of the interaction was to "find new saving measures that match your preference and at the same time catalogue saving measures that you are doing already." They were then randomly assigned to one of two PE-methods and routed to the actual system, which they were required to use for at least 10 minutes. Finally, they were given several post-experimental questionnaires.

3.3 Measures

Before interaction with the system, 31 five-point scale questions were asked to measure domain knowledge and commitment. These questions were entered in an exploratory factor analysis, using Generalized Least Squares extraction and Varimax rotation. Two factors were extracted, together explaining 36% of the variance. These factors divided the items in one factor with the expertise items and one with the commitment items.

After interaction with the system, satisfaction with the system was measured using the five general items of the QUIS†. The 9-point scaled items were summed to obtain a single satisfaction score (Chronbach's $\alpha = .83$, M = 26.0, SD = 8.06).

The post-experimental questionnaires also included 21 five-point scale questions covering other aspects related to satisfaction. These questions were entered in an exploratory factor analysis, using Maximum Likelihood extraction and Oblimin rotation ($\delta = -.5$). Three factors were extracted that together explained 47% of the variance. The factors were interpreted to entail the concepts 'perceived usefulness of the system', 'understandability of the interaction' and 'satisfaction with the chosen measures'.

4. RESULTS

The hypothesis was tested by performing linear regressions using PE-method (attribute-based versus case-based), expertise and commitment as predictors, and satisfaction with the system, perceived usefulness, understandability, and satisfaction with the chosen measures as dependent variables.

4.1 Satisfaction with the system

Table 1 presents the results of the regression on satisfaction. First of all, we found a strong significant main effect of commitment. Committed individuals were more satisfied with the system than are less committed individuals. Furthermore, a significant interaction of expertise with PE-method supported our main hypothesis with a medium-sized effect. The lack of main effects indicates a double dissociation, which means that neither of the systems was preferred by both experts and novices. Instead, experts were more satisfied with attribute-based PE than with case-based PE, while novices were more satisfied with case-based PE than with attribute-based PE.

* Our best guess for stopped sessions is a lack of interest or time. We found no significant predictors for why people prematurely ended the experiment, nor did it influence our analyses.

† See http://hcibib.org/perlman/question.cgi?form=QUIS. We excluded item 4, because it raised questions during pretesting.

Table 1. Predicting satisfaction (adjusted $R^2 = .165$)

	B (std. err)	t	partial η^2
intercept	25.36 (0.82)	31.11***	0.923
PE-method	0.73 (0.82)	0.89	0.010
expertise	-1.16 (0.86)	-1.34	0.022
commitment	3.24 (0.87)	3.74***	0.147
Expertise*PE-method	1.99 (0.86)	2.30*	0.061
Commitment*PE-method	-1.15 (0.87)	-1.33	0.021

* $p < .05$, ** $p < .01$, *** $p < .001$

4.2 Perceived usefulness

The regression on perceived usefulness is presented in Table 2. Again, a strong significant effect was found for commitment, showing that committed individuals perceived the system as more useful than less committed individuals. Our main hypothesis was again supported by a significant interaction between expertise and PE-method. This medium-sized significant effect, showed a similar double dissociation as for satisfaction.

Table 2. Predicting perceived usefulness (adjusted $R^2 = .265$):

	B (std. err)	t	partial η^2
intercept	-0.07 (0.092)	-0.75	0.007
PE-method	-0.15 (0.092)	-1.63	0.033
expertise	-0.05 (0.096)	-0.55	0.004
commitment	0.43 (0.097)	4.42***	0.200
Expertise*PE-method	0.30 (0.096)	3.13**	0.112
Commitment*PE-method	-0.14 (0.097)	-1.41	0.025

4.3 Understandability

The regression on understandability is presented in Table 3. Predicting understandability, we found a medium-sized significant effect of PE-method. In contrast to earlier findings [6,7], attribute-based PE is on average more understandable than case-based PE. The attribute-based PE is a straightforward specification of attribute weights and gives an unambiguous display of the user's preference. The case-based PE method, however, obscures the attribute weight specification in a less understandable critiquing of examples. Thus, although novices rated the case-based PE-method as more *satisfying* and *useful*, novices as well as experts rated the attribute-based PE-method as more *understandable*.

Table 3. Predicting understandability (adjusted $R^2 = 0.051$):

	B (std. err)	t	partial η^2
intercept	-0.03 (0.102)	-0.31	0.923
PE-method	0.27 (0.102)	2.61*	0.010
expertise	-0.12 (0.107)	-1.10	0.022
commitment	0.16 (0.108)	1.46	0.147
Expertise*PE-method	0.05 (0.107)	0.47	0.061
Commitment*PE-method	-0.07 (0.108)	-0.66	0.021

4.4 Satisfaction with the chosen measures

There was no direct significant effect of our predictors on the measure 'satisfaction with the chosen measures'. Consequently, we predicted that an effect of PE-method, expertise and/or commitment on this measure could be mediated by the one of the other dependent variables. Further analysis indeed revealed a medium-sized significant effect of ‘satisfaction with the system’ on ‘satisfaction with the chosen measures’ ($p < .01$, partial $\eta^2 = .078$). This is an interesting result which suggests that, in general, the satisfaction with a recommender system can reflect on the items chosen/purchased using the system.

5. CONCLUSION AND FUTURE WORK

Our experiment supports our main hypothesis that novices prefer case-based PE while experts prefer attribute-based PE. A match between the user’s domain knowledge and the PE method employed significantly increases satisfaction with the system and perceived usefulness of the system. Satisfaction with the system, in turn increases the satisfaction with the chosen measures, which of course is a primary goal of a recommender system.

If the right PE-method is employed, a user might benefit from faster and higher quality recommendations, as it is easier to explicate preferences to the system. This suggests that it would be beneficial to create a recommender system that tailors the PE-method to the user’s domain knowledge [9]. In the current experiment we used extensive questionnaires to measure this ‘user trait’, but in real life implementations of recommender systems, this would be inconvenient. Alternatively, one could use process data (click streams) to measure user traits *during* the interaction, and adapt the PE-method to the user on the fly. Inspired by similar attempts at creating an adaptive recommender system [22], we are currently running experiments comparing several variants of an adaptive version of our system (i.e. with or without explanations, with or without an anthropomorphic agent) to the ‘static’ variant. Process data gathered in the experiment reported above provided us click stream predictors of domain knowledge and commitment, which informed the user model of this adaptive system.

We acknowledge that a combination decision-theoretic principles and interface design can improve the usability of recommender systems, thereby increasing user satisfaction. We believe that a sound understanding of these fields will play a key role in the future research and development of recommender systems.

6. ACKNOWLEDGEMENTS

We thank Benedict Dellaert for feedback during our research, Steven Langerwerf for help implementing our system, and Evelien Matthijssen for help with the energy-saving use-case.

7. REFERENCES

[1] R. H. Guttman and P. Maes, "Agent-Mediated Integrative Negotiation for Retail Electronic Commerce," in *First International Workshop on Agent Mediated Electronic Trading*, Minneapolis, MN, USA, 1998, pp. 70-90.

[2] G. Haubl and V. Trifts, "Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids," *Market. Sci.*, vol. 19, no. 1, pp. 4-21, 2000.

[3] E. L. Olson and R. E. Widing II, "Are Interactive Decision Aids Better than Passive Decision Aids?," *Journal of Interactive Marketing*, vol. 16, no. 2, pp. 22-33, 2002.

[4] P. H. Z. Pu and P. Kumar, "Evaluating Example-based Search Tools," in *ACM Conference on Electronic Commerce*, New York, NY, 2004, pp. 208-217.

[5] P. H. Z. Pu and L. Chen, "Integrating Tradeoff Support in Product Search Tools for E-Commerce Sites," in *ACM Conference on Electronic Commerce*, Vancouver, British Columbia, Canada, 2005, pp. 269-278.

[6] P. Viappiani, B. Faltings, and P. Pu, "Preference-based Search using Example-Critiquing with Suggestions," *J. of Artificial Intelligence Research*, vol. 27, pp. 465-503, 2006.

[7] P. Viappiani, P. H. Z. Pu, and B. Faltings, "Conversational Recommenders with Adaptive Suggestions," in *Proc. ACM RecSys*, Minneapolis, MN, USA, 2007, pp. 89-96.

[8] S. Spiekermann and C. Paraschiv, "Motivating Human–Agent Interaction: Transferring Insights from Behavioral Marketing to Interface Design," *Electronic Commerce Research*, pp. 255-285, 2002.

[9] T. Randall, C. Terwiesch, and K. Ulrich, "User Design of Customized Products," *Market.Sci.*, vol. 26, pp. 268-280, 2007.

[10] B. Xiao and I. Benbasat, "E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact," *MIS Quarterly*, vol. 31, no. 1, pp. 137-209, Mar. 2007.

[11] J. W. Alba and J. W. Hutchinson, "Dimensions of Consumer Expertise," *J. Consum. Res.*, vol. 13, pp. 411-454, Mar. 1987.

[12] J. R. Bettman, M. F. Luce, and J. W. Payne, "Constructive Consumer Choice Processes," *J. of Consumer Research*, vol. 25, no. 3, pp. 187-217, Dec. 1998.

[13] S. Spiekermann, "Online Information Search with Electronic Agents: Drivers, Impediments, and Privacy Issues," PhD Thesis, Humboldt University, Berlin, 2001.

[14] R. H. Guttman, "Merchant Differentiation through Integrative Negotiation in Agent-mediated Electronic Commerce," 1998.

[15] B. Smyth, "Case-Based Recommendation," in *The Adaptive Web*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin: Springer Berlin / Heidelberg, 2007, pp. 342-376.

[16] L. McGinty and B. Smyth, "Comparison-Based Recommendation," in *Advances in Case-Based Reasoning, ECCBR 6*, Aberdeen, Scotland, UK, 2002, pp. 575-589.

[17] E. Coupey, J. R. Irwin, and J. W. Payne, "Product Category Familiarity and Preference Construction," *J. of Consumer Research*, vol. 24, pp. 459-468, Mar. 1998.

[18] A. Chernev, "When More Is Less and Less Is More: The Role of Ideal Point Availability and Assortment in Consumer Choice," *J. Consum. Res.*, vol. 30, pp. 170-183, Sep. 2003.

[19] S. Darby, "Making sense of energy advice," in *ECEEE 2003 proceedings*, Saint-Raphaël, France, 2003, pp. 1217-1226.

[20] R. Parnell and O. Popovic Larsen, "Informing the Development of Domestic Energy Efficiency Initiatives," *Environment and Behavior*, vol. 37, pp. 787-807, 2005.

[21] B. Smyth and P. McClave, "Similarity vs. Diversity," in *ICCBR 4*, Vancouver, BC, Canada, 2001, pp. 347-361.

[22] J. Hauser, G. Urban, G. Liberali, and M. Braun, "Website Morphing," *Market. Sci.*, vol. 28, pp. 202-223, 2009.