Integrating Tradeoff Support in Product Search Tools for E-Commerce Sites

Pearl Pu
Human Computer Interaction Group
School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
+41-21-6936081

pearl.pu@epfl.ch

Li Chen
Human Computer Interaction Group
School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
+41-21-6931246

li.chen@epfl.ch

ABSTRACT

In a previously reported user study, we found that users were able to perform decision tradeoff tasks more efficiently and commit considerably fewer errors with the example critiquing interface than with the ranked list. We concluded that example-based search tools were likely to be useful particularly for extending the scope of consumer e-commerce to more complex products where decision making is critical. This paper presents results from a follow-up user study quantifying the benefits of tradeoff support. Users were able to refine the quality of their preference structures and improve decision accuracy by up to 57% after performing tradeoff tasks. Tradeoff support also significantly increased users' confidence in their choices. Together, these two studies show that example critiquing enables users to more accurately find what they want and be confident in their choices, while only requiring a level of effort that is comparable to the ranked list interface.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces–*evaluation/methodology, interaction styles, user-centered design*; H.4.2. [**Information Systems Applications**]: Types of Systems–*decision support*; I.3.6. [**Computer Graphics**]: Methodology and Techniques–*interaction techniques*.

General Terms

Performance, Experimentation, Human Factors.

Keywords

Consumer e-commerce, decision support, example critiquing interactions, SmartClient systems, empirical user study.

1. INTRODUCTION

When the example critiquing interface was initially developed, we were interested in identifying usability requirements for preference elicitation in product search tools. Pu and Kumar provided details of such a requirements catalog in the previous

ACM E-Commerce Conference [17]. While conducting a series of user studies to validate some of the requirements, we discovered that example critiquing enabled users to perform tradeoff tasks more efficiently with considerably fewer errors than the ranked list interface. We concluded that such tools were likely to be useful particularly for extending the scope of consumer e-commerce to more complex products where decision making is critical. However, we did not know the exact benefit of tradeoff tasks.

This paper presents evidence based on a second user study showing that participants were able to improve decision accuracy by up to 57% using the tradeoff support (called tradeoff navigation) provided by our example critiquing interface. Moreover, the same tradeoff support helped users refine the quality of their preference structures and increase their confidence in the choices that they have made. Together, these two studies show that example critiquing enables users to make highly accurate decisions while only requiring a level of effort that is comparable to the ranked list interface.

1.1 Summary of the First User Study

We compared example critiquing with the commonly used ranked list interface and measured task performance and error rate as participants were instructed to perform tradeoff navigations [17]. There were two main reasons for choosing the ranked list interface as the baseline. First, it implements one of the heuristic decision strategies, known as the lexicographical ordering strategy (LEX). LEX is said to be a very efficient decision strategy, although it does not achieve 100% decision accuracy [13]. Second, a ranked list is the most commonly used interface in consumer e-commerce websites to display search results, thus serving well as a baseline model.

22 participants (7 females) were instructed to use the example critiquing and ranked list interfaces in two evenly divided groups to perform tradeoff navigations. The first group evaluated the example critiquing first and then the ranked list interface, while the second one evaluated the interfaces in the opposite order. Counterbalance measures were taken to eliminate order and learning effects as much as possible. The set of user tasks (tradeoff navigation) were further divided into identifying simple and complex tradeoff alternatives. Despite the fact that a ranked list was much more familiar to the participants, the first study showed that example critiquing interfaces were comparable to the ranked list on simple tradeoff tasks both in terms of task time and error rate. For complex tasks, users performed 15% faster using

example critiquing interfaces, and made 75% fewer errors compared to the ranked list. In addition, we reviewed three other example-based systems, such as FindMe [1], ATA [10] and Apt Decision [18], along the dimensions of ease of use and the complexities of tradeoff tasks that they could support. We concluded that example-based search tools were likely to overtake the popularity so far enjoyed by the ranked list as consumer e-commerce is extending its scope to more complex products where making judicious decisions is increasingly critical.

1.2 Contribution of Our Current Work

The first user study motivated us to emphasize decision support in search tools. Based on this finding, we made three important contributions in our current work. We have identified decision accuracy as the main objective for decision support. Therefore we investigated whether users actually improved their decision accuracy after performing tradeoff tasks with the help of the example critiquing interface. Secondly, we modified the interface to more actively guide the users to benefit from the tradeoff support. We further verified in our second study that users' confidence in the choices that they have made in the example critiquing interface did increase after they had performed tradeoff navigation. Finally, our current user study was the first one to detail the amount of accuracy that tradeoff analysis was able to achieve, even though many researchers believe that accurate decisions could be produced by compensatory decision strategies.

The rest of this paper is organized as follows. We first compare our work with other studies regarding the effect of decision support elements on consumer decision making. We then describe the most recent version of the SmartClient system, a preference-based search tool which implements the example critiquing interface. We present our user study in detail by describing the main hypotheses, the material and data sets used, the design of user tasks and the experimental procedure. We then analyze the results of the experiment and outline the main findings. We discuss our future plans to improve the example critiquing interface based on feedback collected from the current user study, followed by our concluding remarks.

2. RELATED WORK

According to behavior decision theorists, accurate decisions are made via compensatory strategies where a decision maker must diligently examine all relevant alternatives and compare their prospective pros and cons attribute by attribute [7,13]. Even though most decision makers want to be as accurate as possible, the effort required to process information and perform tradeoffs can simply be too overwhelming for the average user. Therefore, they rely on an array of non-compensatory decision strategies, which can lead to serious decision errors.

Our work aims to understand how the tradeoff analysis provided by the example critiquing method affects consumers' decision making. Two related works also studied the effects of interface technology on consumers. One investigated the effect of decision technology on consumers' choice of decision strategies [8], and the other studied the effect of interactive decision aids on the quality and efficiency of purchase decisions [6].

According to [6,13], individuals tend to use two-stage processes to reach their decisions in complex environments, where the depth of information processing varies by stages. Correspondingly, two kinds of interactive tools were designed to assist consumers in performing tasks of the two stages: a recommendation agent (RA) for the initial screening of available alternatives to identify a subset of the most promising alternatives, and a comparison matrix (CM) for the in-depth comparison of selected options to help users make actual decisions. In detail, the RA generates a personalized list of recommended alternatives based on consumers' self-defined information about attribute importance weights and minimum acceptable attribute levels, and the CM enables consumers to compare products more efficiently and accurately in an alternatives × attributes matrix.

Jedetski and Adelman [8] investigated whether consumers adapted their decision strategies on e-commerce Web sites to the presence of decision technology. They compared two web sites: CompareNet (compare.net) and Jango (jango.com). At the time of their experiment, CompareNet used a comparison matrix to display products side by side based on a set of attributes, and Jango simply presented the alternatives in a list without a comparison matrix. As demonstrated by their experiment, consumers employed more compensatory decision strategies when using CompareNet, and they were also more satisfied with it than with Jango. Another premise they proved was that the number of alternatives had a significant effect on decision strategies. Consumers use more compensatory strategies with a smaller number of alternatives (fewer than 30). Since the more compensatory decision strategies consumers use is directly related to making more accurate decisions, the authors suggested that Web site designers use decision technology to support alternative comparison (CM) and reduce the appearance of a large number of product alternatives.

Haubl and Trifts [6] focused on proving their hypotheses pertaining to the effects of RA and CM on various aspects of consumer decision making. They demonstrated that the use of RA could reduce consumers' search effort for product information, decrease the size but increase the quality of their consideration sets, and improve the quality of their purchase decisions. The use of CM could also lead to a decrease in the size but an increase in the quality of consumers' consideration sets, and has a favorable effect on objective decision quality. They concluded that RA and CM might have strong favorable effects on both the quality and the efficiency of purchasing decisions.

These two related works pointed out that a comparison matrix is likely to augment a user's decision accuracy. However, an important missing element is that neither of these two works described a method that guides users to *identify* the set of alternatives to be included in the comparison matrix. It seems that the user is left alone to complete this task, as is the case with CompareNet. A crucial contribution of the example critiquing system is that this set of alternatives (called tradeoff alternatives) is automatically recommended by the tradeoff component.

3. SMARTCLIENT AND EXAMPLE CRITIQUING INTERFACE

The example critiquing interaction model (see Figure 1) was initially used in ATP [22]. Later on, ATP became SmartClient, an online preference-based search tool for finding flights. This method was subsequently applied to catalogs of vacation

packages, insurance policies, and apartments. Example critiquing has been employed in a variety of SmartClient systems [15,16,21] that we have built to perform multi-attribute product search in such domains as planning trips, choosing insurance policies, finding restaurants, and searching for apartments.
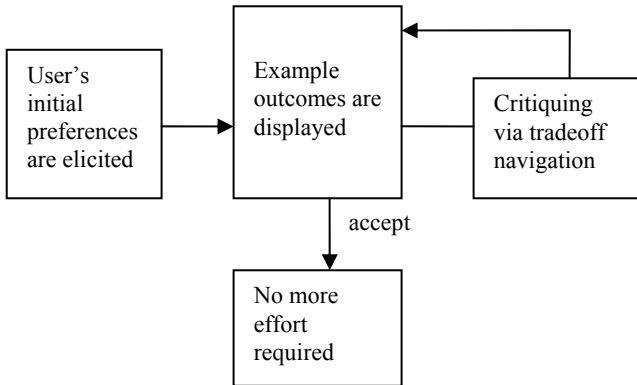

Figure 1. Example critiquing interaction model.

Given the users' comments on the example critiquing interface in the first study [17], we have improved the interface, especially modifying the weight input area so that user can more easily enter the weight value of each participating attribute in the tradeoff process. The example critiquing interface consists of three panels: the "Search" panel for the user to enter the search criteria, the "Search Results" panel for displaying search results, and the "Basket" panel for storing a set of finalists. Consider an apartment search scenario where a student has specified three preferences in the query area (see Figure 8): a shared apartment with a private bathroom, which is located 10 minutes from the university where the user is a student. Notice that if the student does not have any preferences on a particular attribute, the value can be set to "no opinion." Each preference is composed of one acceptable attribute value and the corresponding degree of importance (weight) of that attribute. The weight ranges over four values: least important, somewhat important, important, and most important. If a user expresses a preference, but does not know at this point how important this preference is, the default value is set to "somewhat important". A preference structure is a set of (attribute value, weight) pairs of all participating attributes.

The search engine then uses this preference structure to filter the product space to a smaller set with 7 items and displays it in the "Search Results" panel (see [5] for the optimal number of solutions to display). This set may change as users revise their preference models (see [16] for further details on how preference revision is handled). However, as users converge on a possible solution, known as the near-target, they can activate the tradeoff navigation by clicking on the "compare" button (see Figure 8). The user will be guided to query the system whether tradeoff alternatives of this near-target solution exist. Tradeoff alternatives are improvements of the current solution on some attributes, and compromises on other attributes. Tradeoff alternatives can be found if users post critiques to the near-target solution based on their desire to trade off more of one valued attribute for less of another valued attribute. For example, suppose that the user has selected apartment 91 as the starting point of the navigation process (see Figure 8). The "Search" panel has become the

"Compare" panel, where the user can specify the desire for a cheaper apartment by selecting "cheaper" option (see Figure 9). However, knowing something must be sacrificed for a cheaper apartment, s/he specifies "compromise" for both the "Bathroom" and "Kitchen" attributes. Once a set of critiques has been composed, the system will show another set of matching examples. Apartment 83 might be a good match for the student because it is 100 Swiss francs (CHF) cheaper with a shared bathroom but without a kitchen. Since the student is likely to eat at the university cafeteria most of the time, the 100 CHF savings seems to be a good tradeoff for the lack of a kitchen (see Figure 9). Notice that the system does not resolve tradeoffs for users, but only provides information for them to better understand the decision context. The final choice is left to the user. This query/critiquing completes one cycle of interaction, and it continues as long as users want to further refine the results.

When finding the tradeoff alternatives, the system applies the weighted additive sum rule, called the WADD rule (see [9] for further details). Since WADD is a compensatory decision strategy [13], the system achieves the highest decision accuracy possible in identifying the set of tradeoff alternatives.

4. DECISION IMPROVEMENT USER STUDY

4.1 Hypotheses Development

Our main objective was to investigate whether tradeoff navigation augments a user's decision accuracy. Secondly, we would like to understand whether users subsequently change their preference structures and if so, how they change them.

According to [13,20,23], decision makers do not have a set of innate preferences. Instead, they construct them in a way that is both adaptive to the decision task (how many options there are) and the decision environment (the options shown to them). Furthermore, according to [4,12], consumers with less familiarity of product catalogs tend to adopt an equal weighting scheme, while consumers with more familiarity tend to weigh the most important attribute more heavily. We were therefore interested in measuring the following parameters of the preference structure in regards to our example critiquing interface:

- Preference enumeration: do users express preferences on more attributes after tradeoff navigation?

- Revision of preference values: how many attribute values does a user modify after tradeoff navigation?

- Weight modification: how many attribute weights does a user modify after tradeoff navigation?

For self evaluation, we intended to ask users to directly state how certain they feel about their preference structure both before and after tradeoff navigation.

4.2 Hypotheses

There were three categories of hypotheses in this experiment. Here we discuss the concrete steps that we would take to prove each one.

4.2.1 Hypothesis 1: Choice Improvement

We assumed that an item existed in the database that was the most suitable choice for a given user. We called it the target choice. If a user would eventually find it, then we would say that s/he had achieved 100% decision accuracy. According to decision theories, this target can be identified if and only if a decision maker views all options in the collection and compares the options attribute by attribute to evaluate their pros and cons. This decision strategy is the normative method, known as the weighted added sum strategy [13]. A user is said to improve decision accuracy if s/he gradually moves toward the target choice. To measure improvement of accuracy, we would first record a user's choice (choice 1), which would be identified after an initial search using the example critiquing interface. Then the user would be instructed to perform a series of tradeoff navigation tasks and indicate a new choice (choice 2) if the latter was an improvement on choice 1 in her/his opinion. To evaluate whether the second choice was better than the initial one, we would instruct the user to review all apartments (100 apartments in this case) and tell us whether choice 1, choice 2, or a completely different one truly seemed best. If users would stand by their first choice, it would indicate that they had reached 100% accuracy without explicit tradeoff analysis. If users would stand by their second choice, it would indicate that they had reached their 100% accuracy with the help of explicit tradeoff analysis. If users chose yet another item, it would indicate that they had not reached 100% accuracy even though they performed tradeoff analysis. We expected that very few users would achieve 100% accuracy without explicit tradeoff analysis, but that many users would achieve 100% accuracy after explicit tradeoff analysis. We were curious to know the percentage of users who did not achieve 100% accuracy using explicit tradeoff analysis.

Here we should make the difference between error rate and decision accuracy. The error rate recorded in the first user study measured whether users found the correct answers to the tradeoff tasks assigned. The decision accuracy notion in this paper is defined as the fraction of users who find their target choices after performing tradeoff navigation.

4.2.2 Hypothesis 2: Preference Structure Improvement

The second hypothesis was that the explicit tradeoff navigation would help users refine their preference structures. We would compare a user's final preferences after tradeoff navigation with the initial preferences and analyze whether any improvement had occurred. More concretely, we would measure the enumeration of a user's preference structure and the number of modifications the user made to the preference values and weights. Furthermore, we would ask our participants to explicitly indicate their preference certainty levels before and after tradeoff navigation.

4.2.3 Hypothesis 3: Improvement of Users' Confidence in their choices

In addition to user's choice and preference structure improvement, we also hypothesized that users would increase their confidence in their choices after performing tradeoff analysis. To prove it, we would measure whether a user felt more confident about the choice that s/he has made after performing tradeoff navigation.

4.3 Material, Data Sets and Participants Used

The SmartClient system and example critiquing interface were provided in this user study. As in the previous study, the domain still dealt with apartment search scenarios [17]. However, we augmented the data set to contain as many as 100 apartments to provide more tradeoff alternatives. The user's preferences can be specified on a total of six attributes: type (room in a house or shared apartment), price (from 300 to 900 CHF), area (from 10 to 30 square meters), bathroom (private or shared), kitchen (private or shared), and distance between apartment and work place (from 5 to 60 minutes).

28 volunteers (10 females) were recruited as participants in the user study. In order to make the group as diverse as possible, the participants were selected from a variety of 10 nationalities, different levels of educational backgrounds, and professions (student, research assistant, engineer, broker, hotelier, sales clerk, and accountant). Table 1 shows some of their demographic characteristics.

Table 1. Demographic characteristics of participants

Characteristics	Gender		Age		
	Male	Female	10s	20s	30s
Participants	18	10	2	21	5
Total	28		28		

Characteristics	Education			Familiarity with online apartment search	
	High school	BS	MS/ PhD	Yes	No
Participants	8	7	13	11	17
Total	28			28	

4.4 User Tasks

4.4.1 Tradeoff Navigation

The tradeoff navigation involves finding products that have more optimal values on one attribute, while accepting compromised values for one or two other attributes. As in the first study [17], the participants in this study were also explicitly instructed to perform four tradeoff navigation tasks. Two of these tasks dealt with multi-attribute tradeoffs that allow only one attribute to be improved and one to be compromised. The other two tasks dealt with making tradeoffs when users improved values on one attribute and sacrificed values on up to two attributes.

The tradeoff tasks were adaptively chosen in reaction to the user's initial choice at the search stage. This was to ensure that correct answers existed for all tradeoff tasks. For example, provided that the user initially chose a 500 CHF apartment, we would ask her/him to improve the price attribute by finding a cheaper apartment. This task scenario would not have been possible if the user had chosen a 300 CHF apartment since that is the minimum available price. For this case, s/he was likely to be asked to improve on the distance attribute if it was greater than 20 minutes.

4.4.2 User Tasks Design

A total of 12 tasks have been designed which were given to each user in three steps.

Step 1 - User task: "Find your favorite apartment".

The goal was to let the participant find her/his favorite apartment by freely interacting with the apartment finder interface, where the example critiquing function (the "compare" button) was disabled. The answer to this task gave the participant a starting point for subsequent tradeoff analyses.

After a participant had made her/his initial choice, a measure of choice confidence level ("Are you confident that what you have found is the best choice?") and preference certainty level ("Are you certain about your current preferences?") were obtained. The confidence varied from 0% (not confident at all) to 100% (extremely confident), and the preference certainty varied from -5 (not certain at all) to +5 (extremely certain). This step thus contained three user tasks.

Step 2 - User tasks: perform a set of tradeoff tasks by posting critiques to the initially found apartment in step 1.

The second step was to instruct the participant to perform four tradeoff tasks with the example critiquing function (the "compare" button) enabled in the interface. For each task, a participant was required to find an apartment satisfying the instructed task condition.

For instance, if the apartment found in step 1 was a "shared apartment, 500 CHF, 20 square meters, private bathroom, shared kitchen, 20 minutes to work place" (called A1), the participant would be asked to accomplish the following four tradeoff tasks:

1. "Find an apartment which is cheaper than A1. You can compromise on only one attribute."

2. "Find an apartment which is bigger than A1. You can compromise on only one attribute." (see Figure 10)

3. "Find an apartment which is 100 CHF cheaper than A1. You can compromise up to two attributes." (see Figure 11)

4. "Find an apartment which is 5 square meters bigger than A1. You can compromise up to two attributes."

In [17], the first two tradeoff tasks were defined as (1, 1) tradeoffs (optimizing one attribute and compromising another attribute), whereas tasks 3 and 4 were called complex (1, 2) tradeoff cases (optimizing one attribute and compromising up to two attributes).

At the completion of the above tradeoff tasks, each participant was asked to:

1. "Select your most preferred apartment from the apartments that you have chosen as answers to the tradeoff tasks, together with the apartment found initially." (see Figure 12)

2. "Specify your current preferences and their degrees of importance."

At the end of this step, the questions which measure the choice confidence level and preference certainty level were asked again to each participant. This step thus contained 8 user tasks.

Step 3 - User task: "Do you still think the choice made at the end of step 2 is the best choice after you have reviewed all apartments?"

The final step was to ask the participant to review all apartments in our database. If the answer was "No", the user would be asked to point out the apartment that s/he thought was the best choice.

The choice made after all apartments had been reviewed is called the participant's target choice. This step contained one user task.

4.5 Experimental Procedure

We designed an online procedure to record most data in log files. These data, such as user preferences, choices and critiquing actions, were needed for proving the hypotheses. In this online procedure, a set of user interfaces was developed to guide participants to finish all of the tasks step by step (see Figure 10, 11 & 12).

Before each user study, we explained to each participant the experiment's objective, the main functions of our apartment finder interface, and the meaning of labels on the interface. We also told them that their interactions with the interface would be automatically recorded in a log file.

5. RESULTS ANALYSIS

5.1 Choice Improvement (Hypothesis 1)

Each participant's initial choice, the second choice made after tradeoff navigation, and the final choice found in the list of all alternatives were recorded and compared. 18% of the participants found the target choice initially (in step 1) since they did not waver from their first choice after tradeoff navigation and after reviewing all apartments (see Figure 2). 57% of participants discovered their target choice when they finished the four tradeoff tasks because they still thought the choice they made at the end of step 2 was the best choice after reviewing all apartments. Among these 16 participants, 10 participants' target choices were found after performing (1,1) tradeoffs, and the remaining 6 participants' target choices were found after (1,2) tradeoff tasks.
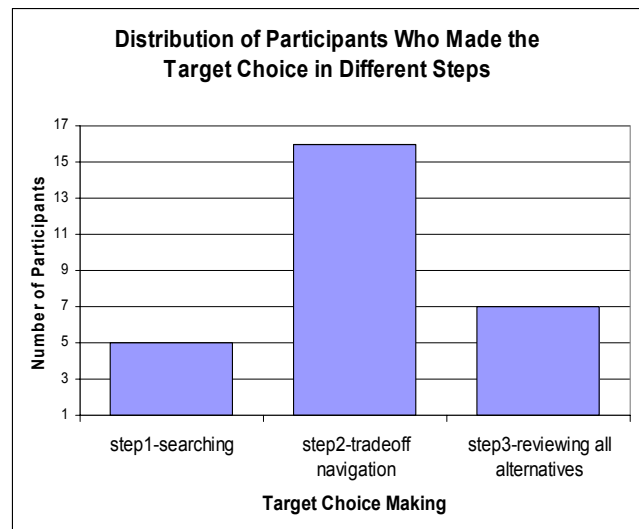

Figure 2. Distribution of participants who made the target choice in different steps of the experiment.

Therefore, due to the effect of explicit tradeoff navigation, the percent of users who found their target choices by the end of Step 2 increased from 18% to 75% (see Figure 3), which represents an increase of over 400% in decision accuracy. This effect is furthermore significant ($p<0.001$) according to the McNemar test, a test that allows us to know whether a process has a significant

influence on an established condition. The remaining 25% of participants identified a completely different choice when we revealed all apartments to them.
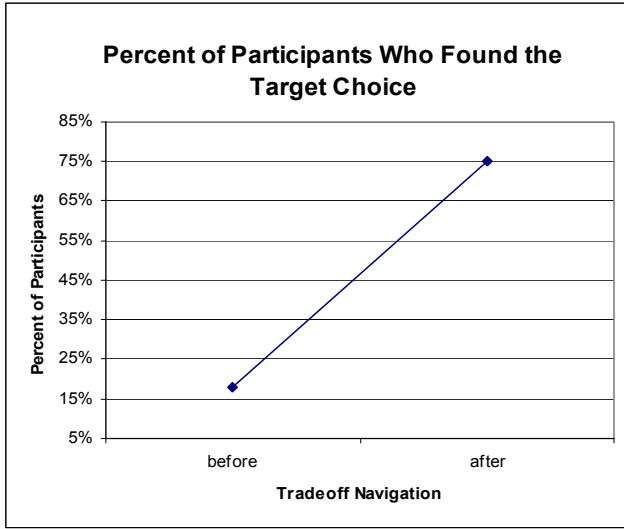
Percent of Participants Who Found the Target Choice

Figure 3. Effect of tradeoff navigation on finding the target choice.

Together with the experiment described in [17], we can reach the conclusion that the example critiquing interface not only enables users to find tradeoff alternatives more quickly, but also helps them achieve a higher level of decision accuracy via tradeoff navigation.

5.2 Preference Structure Improvement (Hypothesis 2)

Distribution of Participants Who Changed Attribute Values after Tradeoff Navigation
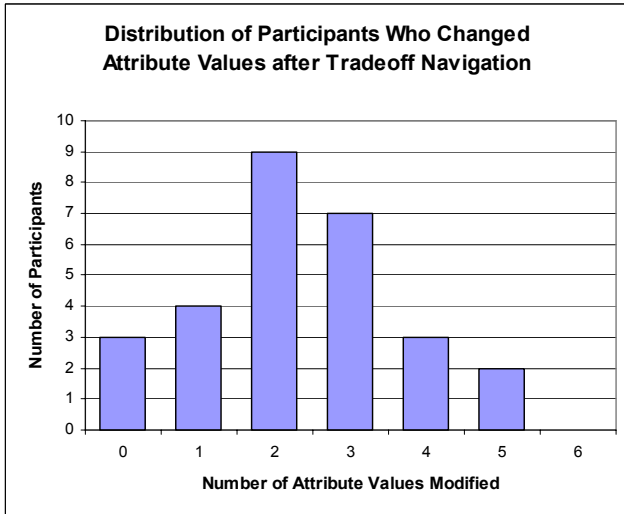
Figure 4. Effect of tradeoff navigation on the modification of attribute acceptable values.

To test the hypothesis regarding the improvement of users' preference structure, we collected and compared all participants' initial preferences with the preferences specified after tradeoff navigation. The mean number of preference enumeration increased from 5.25 to 5.5, and the average weight of all

preferences increased from 6.52 to 6.78. However, these phenomena were not highly significant (t=-1.491, p=0.148 and t=-0.993, p=0,329 respectively by the paired samples t-test). We believe that this may be due to the fact that most participants were so familiar with the apartment search scenario that they were likely to have strong preferences from the beginning.

The experiment results show that 100% of participants modified their preferences on at least one attribute value or weight after tradeoff analyses. Therefore the effect of the tradeoff navigation on preference modification is highly significant (p<0.001). Moreover, we found that the majority of participants made corrections at least on two attributes' acceptable values or weights (see Figure 4 & 5). This implies that most users quite actively modified their preferences. Simultaneously, the mean preference certainty level of all participants increased from 2.8 to 3.6 in a significant way (t=-2.556, p<0.05, see also Figure 6).
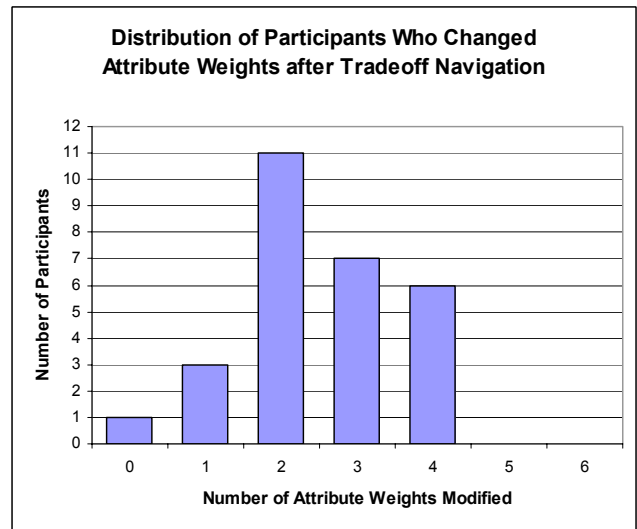
Distribution of Participants Who Changed Attribute Weights after Tradeoff Navigation

Figure 5. Effect of tradeoff navigation on the modification of attribute weights.

Preference Certainty Level
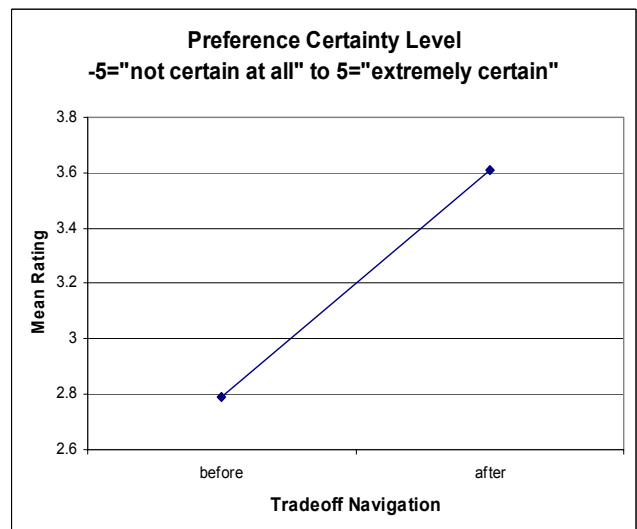**-5="not certain at all" to 5="extremely certain"**

Figure 6. Effect of tradeoff navigation on users' preference certainty levels.

The results show that the use of the tradeoff analysis by example critiquing method is an efficient approach for users to adaptively refine their preference structures while examining more tradeoff alternatives. This is even true for those users who have strong preferences initially.

5.3 Choice Confidence Improvement (Hypothesis 3)

The mean confidence level of all participants increased from 68.6% to 77.1% after performing tradeoff analysis (see Figure 7). This difference is significant by the paired samples t-test ($t=-2.175$, $p<0.05$). That is, participants were more certain about the accuracy of their choices after the tradeoff navigation.
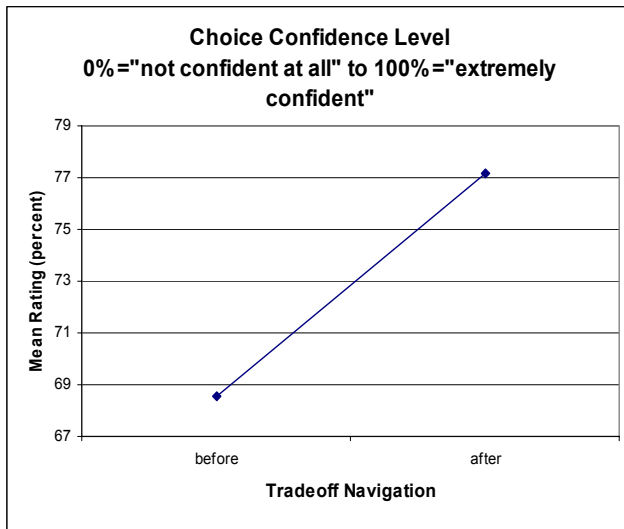
Choice Confidence Level
0%="not confident at all" to 100%="extremely confident"

Figure 7. Effect of tradeoff navigation on users' confidence in their choices.

5.4 Discussion

The experiment results support most of our hypotheses related to decision accuracy improvement via the tradeoff navigation provided by our example critiquing interface. More specifically, 57% of users found a better choice after tradeoff navigation. This is a significant improvement, especially given the fact that most users already achieved a fairly high accuracy before tradeoff navigation began. 25% of users did not find their target choice after using the tradeoff navigation. This indicates that example critiquing based on one selected item was not sufficient to enable users to achieve 100% accuracy.

Along with improved choice, the users' preference structures were refined simultaneously via tradeoff navigation. After comparing users' preferences specified before and after the tradeoff process, we can see that most preferences on acceptable attribute values and degrees of importance were modified. The users themselves also felt more certain about their final preferences. Therefore, the tradeoff navigation has a favorable effect on improving users' preferences by prompting them to learn more about alternatives and refine their initial uncertain preferences accordingly. However, this experiment did not provide enough evidence that most users increased their

preference enumeration in a significant way, contrary to our belief that initial preferences were scarce.

Other valuable evidence that we acquired from this study showed that users became more confident of their choices after the interface provided tradeoff assistance for searching products. This could be due to the fact that users were able to examine more tradeoff alternatives and achieve higher decision accuracies by performing tradeoff analyses with our interface.

6. FUTURE WORK

The findings of our previous and current user studies provided rather positive evaluations of the example critiquing interface. However, some improvements are still necessary. One main goal for the future is to integrate an interactive element in the example critiquing interface that can "teach" users to effectively make tradeoffs. This involves several trial designs of the interface. One solution we propose is to use a dialog box that pops up the very first time users intend to put a search result in the basket. In that dialog box called "compare", users will be invited to compare the current selection with alternatives which are improvements on one or several attributes. Users will also be asked to name the attributes that they are willing to sacrifice. If users do not want to compare their initial selection with other tradeoff alternatives, they can simply click the "cancel" button and the initial selection will be stored in the basket. This popup dialog box only appears once. In subsequent interactions, the "compare" popup window is accessible via the "compare" button located on the main interface. In the future, it remains to be seen if this new design effectively guides users to click the "compare" button (i.e. making tradeoff) more frequently.

Another main work is to integrate an explanation facility in the interface to explain how to match the tradeoff alternatives to their specified tradeoff criteria. We believe that the explanation would play an important role in further improving user's decision accuracy, search performance, and perception of the interface.

Furthermore, we also plan to extend the findings from this experiment towards designing a more general framework for complex and unfamiliar product domains where users tend to have fewer initial preferences. We plan to evaluate whether the tradeoff analysis would have a more prominent effect on decision improvement in those domains. In addition, we will further increase the number of available products and the number of values for each attribute to make the tradeoff task more challenging.

7. CONCLUSION

Our previous user study showed that the example critiquing interface enabled users to perform tradeoff navigation tasks much faster with considerably fewer errors than the commonly used ranked list interface. The user study described in this paper showed that tradeoff navigation allows users to significantly improve their decision accuracy by up to 57%. The participants also adaptively modified their preferences during the tradeoff navigation process and became more certain of their preference structures afterwards. Consequently, participants' confidence in their choices was significantly improved. These findings provide empirical evidence that example critiquing with its tradeoff support enables consumers to more accurately find what they

want and be confident in their choices, while requiring a level of cognitive effort that is comparable to a ranked list. According to [2,11], a good interface design allowing product comparison can increase traffic and sales. This suggests that current product search tools used in most e-commerce websites can significantly increase conversion rates by integrating tradeoff support and using example critiquing like interfaces.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] Burke, R., Hammond, K., and Young, R. The FindMe approach to assisted browsing. In *IEEE Expert: Intelligent Systems and Their Applications*, volume 12(4):32-40, 1997.

[2] Callahan, E. and Koenemann, J. A comparative usability evaluation of user interfaces for online product catalogs. In *Proceedings of ACM Conference on Electronic Commerce (EC'00)*, 2000, 197-206.

[3] Carenini, G. and Poole, D. Constructed preferences and value-focused thinking: implications for AI research on preference elicitation. *AAAI-02 Workshop on Preferences in AI and CP: symbolic approaches*, 2002.

[4] Coupey, E. Irwin, J.R. and Payne, J.W. Product category familiarity and preference construction. *Journal of Consumer Research*, Vol. 24, March 1998, 459-467.

[5] Faltings, B., Torrens, M. and Pu, P. Solution generation with qualitative models of preferences. *International Journal of Computational Intelligence and Applications*, Volume 20, Issue 2, 2004, 246-264.

[6] Haubl, G. and Trifts, V. Consumer decision making in online shopping environments: the effects of interactive decision aids. *Marketing Science*, Vol. 19, No. 1, 2000, 4-21.

[7] Hogarth, R. *Judgment and Choice*. John Wiley and Sons, New York, NY, United States, 1987.

[8] Jedetski, J. Adelman, L. and Yeo, C. How web site decision technology affects consumers. *IEEE Internet Computing*, March-April 2002, 72-79.

[9] Keeney, R. and Raiffa, H. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge University Press, 1976.

[10] Linden, G., Hanks, S. and Lesh, N. Interactive assessment of user preference models: the automated travel assistant. In *Proceedings of User Modeling'97*, 1997, 67-78.

[11] Lohse, G.L. and Spiller, P. Quantifying the effects of user interface design features on cyberstore traffic and sales. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI' 98)*, 1998, 211-218.

[12] Park, C.W. The effect of individual and situation-related factors on consumer selection of judgmental models. *Journal of Marketing Research*, 13(May), 1976, 144-151.

[13] Payne, J.W., Bettman, J.R. and Johnson, E.J. *The Adaptive Decision Maker*. Cambridge University Press, 1993.

[14] Payne, J.W., Bettman, J.R. and Schkade, D.A. Measuring constructed preferences: towards a building code. *Journal of Risk and Uncertainty*, 19, 1-3 (1999), 243-270.

[15] Pu, P. and Faltings, B. Enriching buyers' experiences: the SmartClient approach. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI' 00)*, ACM Press, 2000, 289-296.

[16] Pu, P. and Faltings, B. Decision tradeoff using example critiquing and constraint programming. *Special Issue on User-Interaction in Constraint Satisfaction, CONSTRAINT: an International Journal,* Kluwer Academic Publishers, Vol. 9, Issue 4, 2004.

[17] Pu, P. and Kumar, P. Evaluating example-based search tools. In *Proceedings of ACM Conference on Electronic Commerce (EC'04)*, 2004, 208-217.

[18] Shearin, S. and Lieberman, H. Intelligent profiling by example. In *Proceedings of International Conference on Intelligent User Interfaces (IUI '01),* ACM Press, 2001, 145-151.

[19] Shepard, R.N. On subjectively optimum selection among multiattribute alternatives. In Maynard W. Shelley and Glenn L. Bryan (eds.), *Human Judgments and Optimality*, New York, 1964.

[20] Slovic, P. The construction of preference. *American Psychologist*, 50 (August), 1995, 364-371.

[21] Torrens, M., Faltings, B. and Pu, P. SmartClients: constraint satisfaction as a paradigm for scaleable intelligent information systems. *International Journal of Constraints*, Jan. 2002, 49-69.

[22] Torrens, M., Weigel, R. and Faltings, B. Java constraint library: bringing constraints technology on the Internet using the Java language. In *workshop notes, Constraints and Agents, AAAI'97*, 1997, 10-15.

[23] Tversky, A. and Simonson I. Context-dependent preferences. *Management Science*, Vol. 19, No.10, October 1993.
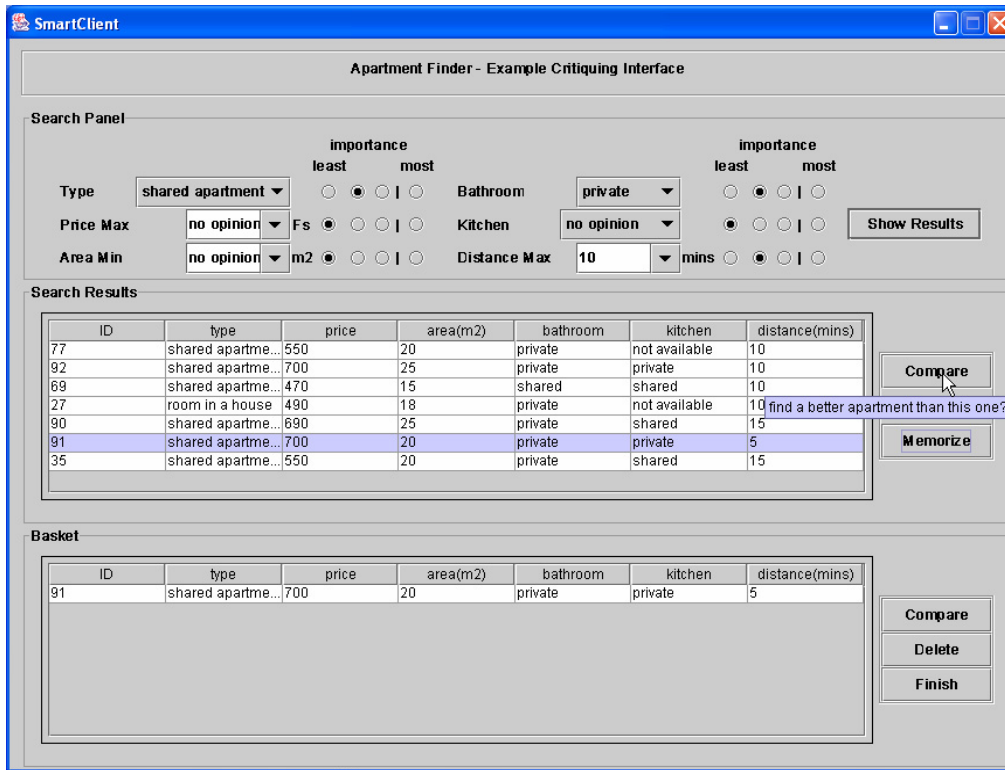
Figure 8. Step one of the example critiquing interface: the user enters the initial query based on her/his preferences, and the system returns seven results.
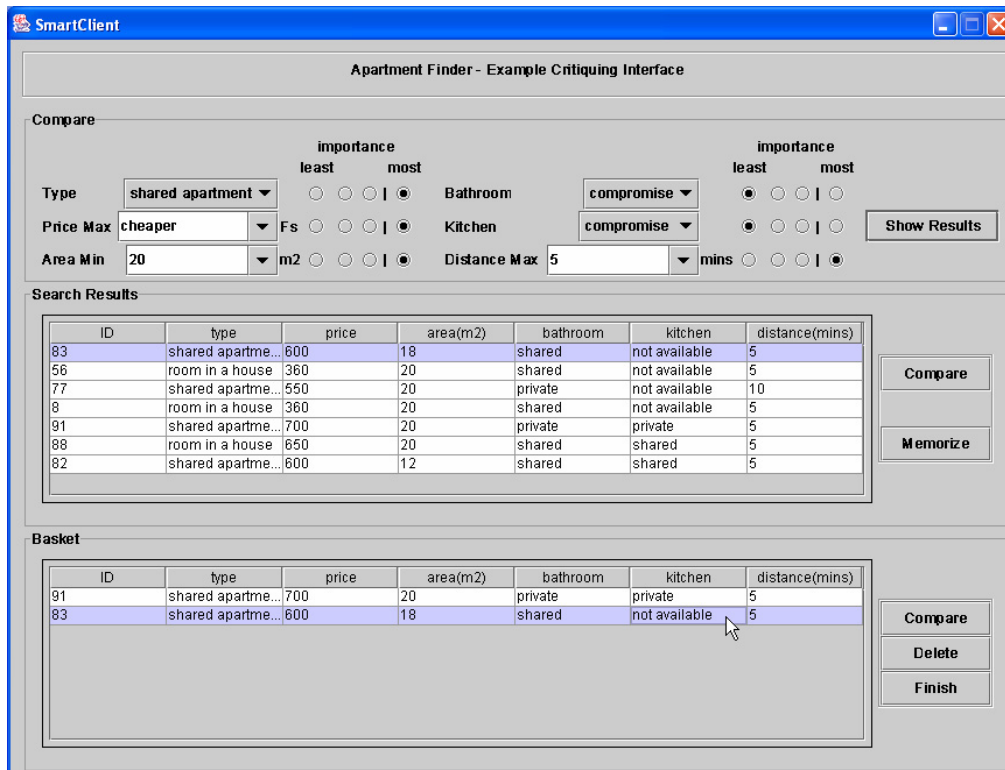

Figure 9. Step two of the example critiquing interface: the user finds a set of tradeoff alternatives using the critiquing module accessible via the "compare" button.
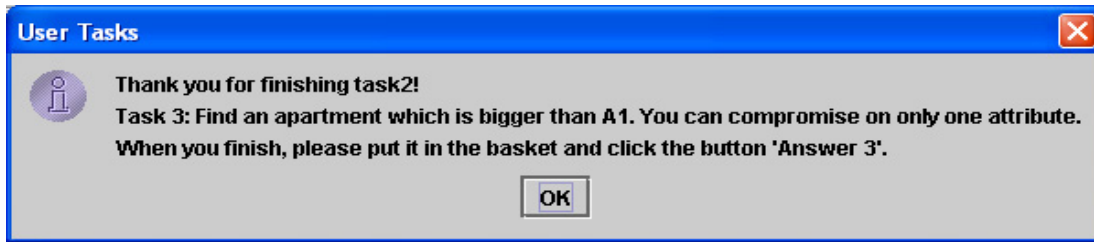
Figure 10. One of the (1,1) tradeoff tasks asking a user to improve on the area attribute while compromising the value of another attribute. The tradeoff tasks were adaptively chosen based on the user's initial choice, i.e. A1.
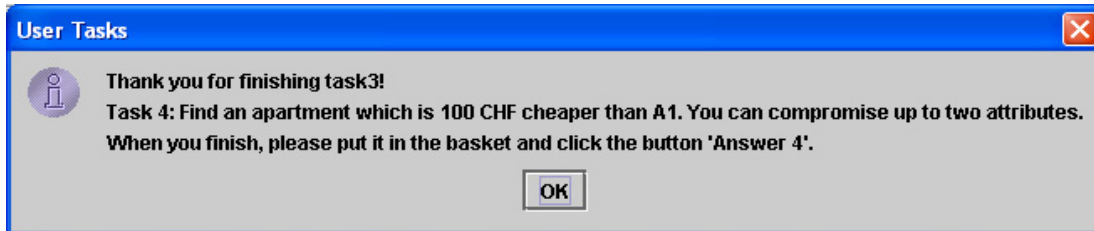

Figure 11. One of the (1,2) tradeoff tasks asking the user to improve on the price attribute (100 CHF cheaper) while compromising the values of up to two attributes.
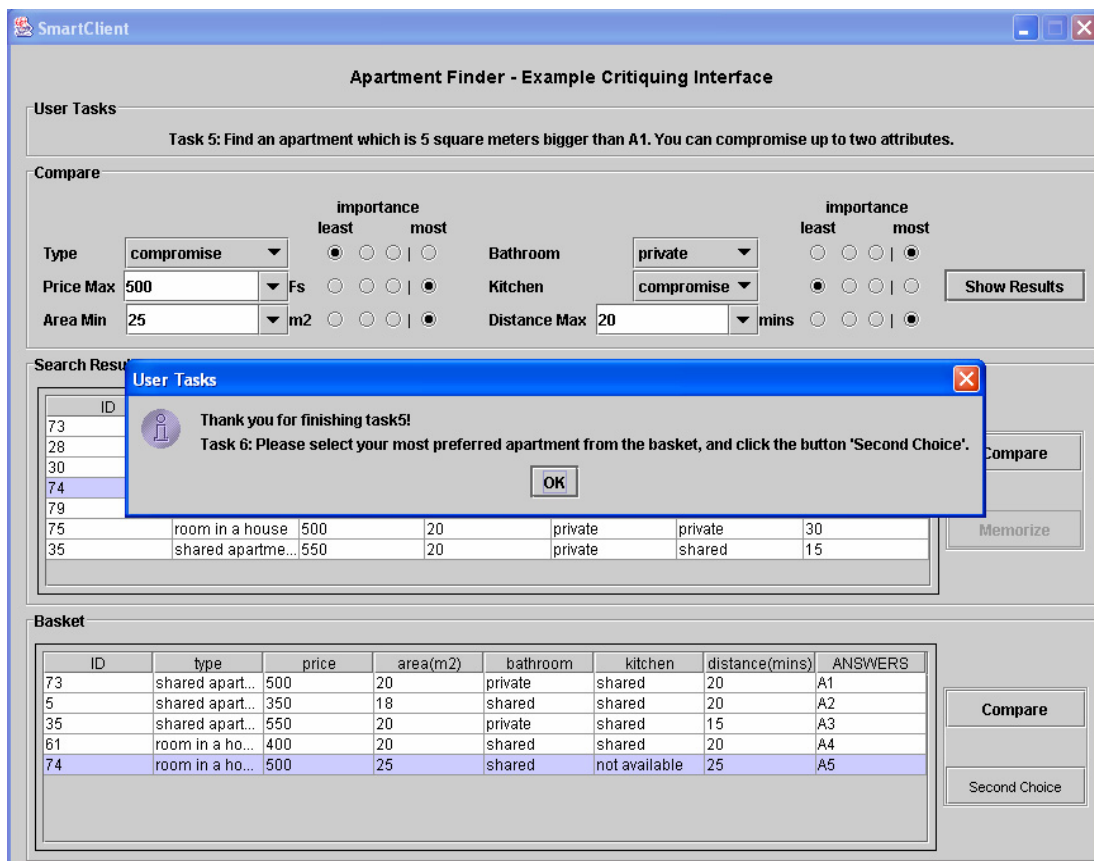

Figure 12. When a user finished tradeoff analyses, s/he would be asked to select the most preferred apartment from the basket, which includes the initial choice (A1) and the four answers (A2, A3, A4, A5) to the tradeoff tasks.