Helping Users Perceive Recommendation Diversity

Rong Hu
Human Computer Interaction Group
Swiss Federal institute of Technology (EPFL)
CH-1015, Lausanne, Switzerland
rong.hu@epfl.ch

Pearl Pu
Human Computer Interaction Group
Swiss Federal institute of Technology (EPFL)
CH-1015, Lausanne, Switzerland
pearl.pu@epfl.ch

ABSTRACT
The recommendation diversity is increasingly being recognized as an important issue in satisfying users' needs for recommender systems. Various diversity-enhancing methods have been developed to increase diversity while making personalized recommendations to users. However, one crucial issue remains. Could the diversity, as system designers have carefully incorporated, be perceived by users and influence their interaction behaviors? In this paper, we try to investigate whether this issue can be addressed at the interface level. Our goal is to understand design issues that enhance users' perception of recommendation diversity and more importantly their satisfaction. A within-subject user study was conducted to compare an organization interface, which groups recommendations into categories, with a standard list interface. Our user study results show that the organization interface indeed effectively increased users' perceived diversity of recommendations, especially perceived categorical diversity. Correlation results reveal that the perceived categorical diversity in recommendation lists has a significant correlation with users' perceived ease of use of a system, perceived usefulness of the system and attitudes towards the system, thereby resulting in a positive effect on their intention to use the system. We conclude by proposing design guidelines based on our study observations.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – Evaluation/methodology.

General Terms
Design.

Keywords
Diversity, Recommender System, User Study, Interface Design, User Satisfaction.

1. INTRODUCTION
During a long period, prediction accuracy was considered as the sole criterion when evaluating recommender systems' quality. However, recent studies have increasingly indicated that accuracy is not enough for a satisfying recommender system, in particular from a user's point of view [8, 13]. Other criteria, such as diversity and serendipity, are emerging as important characteristics for consideration to generate *useful* recommendations [5, 8]. In this paper, we focus on recommendation diversity issues.

Diversity is an intrinsically desirable property for a recommender system. Firstly, users' needs are commonly uncertain beforehand [15, 16]. Varied options could broaden users' domain knowledge about the recommended items and help them clarify their requirements. Secondly, recommender systems are expected to help users explore and discover new items of interest [8]. For users, it is more valuable to obtain the recommendations that they would love, but are different from those which they have already purchased or used [9]. For e-commerce websites, recommending varied items has the potential to make more profits by increasing the sales diversity [7]. Thirdly, it is important for recommender systems to convince users that the recommended item is the best one for them. The existence of diversity in recommendations has the capability of decreasing the difficulty of making a choice and enhancing users' confidence in their choices by providing comparison among recommendations [2, 10].

Even though many diversity-enhancing algorithms haven been proposed in the literature [1, 12, 14, 21, 22, 23], few studies have investigated users' perception of recommendation diversity and how such a perception could influence their satisfaction and acceptance of a system. In [23], Ziegler et al. did a large scale online study, and their online experimental results show that users' overall satisfaction with recommendation lists not only depends on accuracy, but also on the range of reading interests covered. They also found that human perception can only capture a certain level of diversification inherent to a list. Beyond that point, it is difficult for users to notice the increasing diversity degree. Therefore, it is worth investigating how to help users overcome the cognitive limitation and be aware of the existence of diversity in recommendation lists, aimed at achieving a high level of satisfaction to a system.

Currently, the conventional ranked list interface is still a popular way of displaying search/recommendation results. However, this method is highly inefficient in some cases [3]. For example, the number of retrieved search results can be easily beyond the extent of human cognitive capability. Users tend to focus on the top of a list and items that are located farther down in the list would attract little attention. By nature, the ranked list interface is likely to impede users' perception of the diversification of recommendations. Therefore, we are considering whether alternative approaches, such as a proper interface layout design, could augment users' diversity perception.

In this paper, we conducted a within-subject user study, comparing an organization-based interface, which groups recommendations and displays them in a category style [3, 16], with a conventional list interface, while keeping the recommendations in the two systems identical. We utilized Amazon.com as our experimental platform due to its well-known reputation in the field of recommender systems. Its standard list interface for recommendations was replaced by an organization-

based interface with the help of a proxy program. In this study, we attempt to answer the following two research questions:

1) How can interface designs influence users' perceived diversity?

2) How does diversity perception affect users' satisfaction of a system?

The contributions of this paper include three aspects. Our results suggest that the organization-based interface indeed effectively increased users' perceived diversity of recommendations, especially perceived categorical diversity (i.e., users perceive that various kinds of items were recommended to them). In addition, we empirically explored the influence of perceived diversity on users' acceptance of a recommender system. Correlation results show that categorical diversity more significantly influences users' perceived usefulness of the recommender, their attitudes toward the system and their intentions to use the system. Finally, based on the findings in this study, we proposed specific design guidelines.

The reminder of the paper is organized as follows. We first provide an overview of related research work on diversity enhancing technologies and diversity-related user studies in recommender systems. In Section 3, we describe the organization-based interface design methods. In Section 4, we present a detailed description of our experiment, including experiment design, evaluation metrics, and dataset, followed by the experimental results, discussion and the derived design guidelines. Finally, we present the conclusions and future work.

2. RELATED WORK

Traditional diversity-enhancing methods are operated as a heuristic search. The bounded greedy algorithm proposed in [1, 21] is the first attempt to explicitly enhance the diversity of a recommendation list without significantly compromising their query similarity characteristics in case-based recommender systems. It first ranks all recommendable items according to their similarity to the current query. Then, it sequentially transfers items from this ranked list to a final recommendation list such that each selected item maximizes the product of its similarity to the target query and its diversity relative to the cases that have already been selected. Most diversity-enhancing methods follow this fundamental re-ordering strategy [14, 19].

The concept of diversity was also considered in the design of critiquing-based recommender systems. Pu and Chen [4, 16] proposed a dynamic compound critiques generation method, which takes diversity among critiques into account. McCarthy et al. [11] also proposed an idea of generating diverse compound critiques in the context of conversational recommender systems.

Zhang and Hurley [22] suggested presenting the competing concerns of similarity and diversity as constrained binary optimization problems. They applied their optimization strategy to the top-*N* prediction problem and achieved improvements on both diversity and accuracy compared to a standard item-based collaborative filtering algorithm.

McGinty and Smyth [12] highlighted the pitfalls of naively incorporating diversity-enhancing techniques into existing recommender systems and proposed an adaptive diversity-enhancing algorithm. They pointed out that diversity should be provided adaptively. When a recommender system appears to be close to the target case, diversity should be limited to avoid missing it. But when the recommender system is not correctly focused, diversity can be used to help refocus more effectively.

In [23], the authors proposed a topic diversification approach based on taxonomy-based similarity. They compared not only the accuracy measures in different levels of diversification for both user-based and item-based CF, but also subjective satisfaction results from a large scale user survey. Their results show that users' overall satisfaction of recommendation lists goes beyond accuracy and involves other factors, e.g., the users' perceived list diversity. Their work first shed light on the critical value of diversity from the perspective of users.

Castagnos et al. [2] investigated the impact of recommenders on users' product search patterns by observing their interaction behaviors with an online product retail website with an eye tracking system. They demonstrated that users' need for diversity led them to use the recommender systems, compared to the traditional information filtering tools. Furthermore, they found that the diversified recommendations could enhance users' confidence by providing the capability of comparison. To conclude their findings, they proposed a time-dependent satisfaction model which demonstrates the dynamic compromise between accuracy and diversity in recommender system. Our work is similar to theirs. Differently, we investigate the relations between perceived diversity and users' acceptance of the system in a within-subject user study by comparing the influence of two interface designs.

3. ORGANIZATION-BASED INTERFACE

The idea of organization-based interfaces was first proposed as an explanation interface, with the aim of inspiring users' trust in recommender systems [16]. Pu and Chen implemented more than 13 paper prototypes of organization-based interfaces to explore the design dimensions. Based on the results of testing these prototypes with real users in the form of pilot studies and interviews, they derived five design principles: 1) categorize remaining recommendations according to their similar tradeoff properties relative to the top candidate; 2) propose improvements and compromises in the category title using conversational language; keep the number of tradeoff attributes under five to avoid information overload, e.g., "these products are cheaper and lighter, but have slower processor speed"; 3) eliminate dominated categories, and diversify the categories in terms of their titles and contained recommendations; 4) include actual products in a recommended category; 5) rank recommendations within each category by exchange rate (i.e., the preference-based utility value relative to the top candidate) rather than similarity measure. Consequently, the organization-based interface design essentially considers the diversity issue both among categories and within each category.

Previous studies have indicated that organization-based interface designs are highly effective in building users' trust of a recommender system, with the benefit of increasing users' intention to return to the agent and saving users' cognitive effort [16]. More recently, Chen and Pu [3] performed a user study with an eye-tracker to compare the efficacy of two recommender interface designs, list-based and organization-based interfaces, in affecting users' decision making strategies through the observation of users' eye movements and product selection behavior. Their results showed that organization-based interfaces can significantly attract users' attentions to more items with the resulting benefit of enhancing their objective decision quality. Based on their findings, we assume that the organization interface

designs have the capability of assisting users in perceiving the diversity of recommendation lists. In our experiment, we utilized a variation of the conventional organization-based interface approach, Editorial Picked Critiques (EPC) technique, to generate categories for our organization-based interface. We will introduce EPC technique in detail in the following section.

3.1 Editorial Picked Critiques (EPC)

EPC was originally developed in the context of applying critiquing-based recommendation technology to public taste products such as music, films, perfumes, fashion goods and wine [18]. In contrast to high-involvement products such as PCs, digital cameras, users tend to spend less time choosing public taste goods and are more likely to rely on public opinions or experts' advice to make decisions [20]. EPC was designed to take into account the public opinions, popularity information and editorial suggestions, as well as the needs for personalization and diversity.

EPC first identifies five important unit critique categories that match users' attention and needs for public taste goods: price-driven critiques, popularity-based critiques, diversity-driven critiques, similarity-driven critiques, and special recommendation (similar to editorials special picks). Items in the similarity-driven critiques are those which are similar to the selected product and could be generated by recent similarity-based recommendation approaches, such as content-based or collaborative filtering methods. This category is titled as "people who like this may also like".

Compound critiquing categories are generated on the basis of these unit critiques. In [18], a set of five compound categories were proposed for perfume products: "more popular and cheaper", or "more popular but more expensive" in the case that the former category does not contain any products, "same brand and cheaper" or "same brand but more expensive", "just as popular and cheaper", "same price range and just as popular", and finally "people who like this also like". When generating recommendations for each category, users' preferences are taken into account.

In our experiment, we adopted these compound categories proposed in [18] as our classification categories for the organization interface. We remapped the recommendations from Amazon into these five categories, and we used Amazon's bestselling order and customers' ratings as a popularity measure. The items which cannot be categorized into any of the first four categories are put into the category "people who like this also like".

4. EXPERIMENT

4.1 Materials

A well-known commercial website, Amazon.com, was used as our experimental platform due to its high reputation in the field of recommender systems. Its standard list interface was used as the baseline. The organization version was achieved with the help of an open-source filtering HTTP proxy program, PAW [1]. The recommendation list we used was "Customer Who Viewed This Item Also Viewed" in the detailed information page for each product (perfume in our experiment). Unlike the organization-based interface designs in [2, 3, 16], the categories in this study were organized in a tab-based structure to better conform to the horizontal list style in that website. By clicking on each tab, users could see the recommendations in the corresponding category.

[1] http://paw-project.sourceforge.net

Table 1. Demographic characteristics of participants.

Gender	Male	Female
	10	10
Nationality	Chinese (10), Swiss(2), Indian(3), Romanian(1), Croatian(1), Portuguese(1), Iranian(1), Georgian(1)	
Education	Bachelor, Master, Doctor	
Profession	student, research assistant, engineer, interface designer	
Age	21-30	31-40
	19	1

The categories which had no products were not presented. A screenshot of the organization (ORG) interface is shown in Figure 1. The original list-view (LIST) interface used in the website was adapted to only show five products each time to remain consistent with the organization-based interface. A screenshot of the list interface is shown in Figure 2. In either interface, the number of displayed recommendations was restricted to be the same (five in the current study) and the "next" and "previous" buttons were used to explore more items in a list. In order to avoid confusion, we removed the recommendation list of "Customers Who Bought This Also Bought" from the page. In addition, we placed the section "Customers Who Viewed This Item Also Viewed" just beneath the selected product so that users could easily notice it.

4.2 Dataset and Participants

The dataset of perfumes used in this experiment was crawled from Amazon and updated just before launching the study to ensure that we had a dataset containing the most recent and popular fragrance products available on the market. In our experiment, 21,071 items were accessible, covering 13,246 items for women (6,281 Eau de Toilette, 689 Cologne and 6,276 Eau de Parfum) and 7,825 items for men (6,066 Eau de Toilette, 1,474 Cologne and 285 Eau de Parfum).

A total of 20 participants (10 females) were recruited in our user study. The incentive for the participants was a lottery: one out of the 20 users could win a 100 CHF gift voucher to purchase one of the perfumes the winner put in the basket during the study. These participants were from 8 different countries with various professions (student, research assistant, engineer, interface designer); their age ranged from 20 to 40, and they represented various educational backgrounds (from bachelor, master or Ph.D.). The details of their demographic characteristics are shown in Table 1. In addition, four background questions were asked in terms of users' previous computer knowledge, internet usage, perfume knowledge and experience with Amazon. All participants said that they were regular computer users and used the Internet frequently. 11 participants indicated "agree" to the statement "I have knowledge about perfume", 8 participants marked "neutral" and just one said "disagree". 17 of the participants had used Amazon before.

4.3 Evaluation Criteria

In order to evaluate users' perceived qualities of a recommender, we used a simplified version of a user-centric recommender evaluation model (***ResQue***) [17]. More specifically, two questions were designed to measure users' diversity perception of the recommendation lists. One referred to the difference among categories, querying whether "the items recommended to me are of various kinds" (called *categorical diversity*). The other

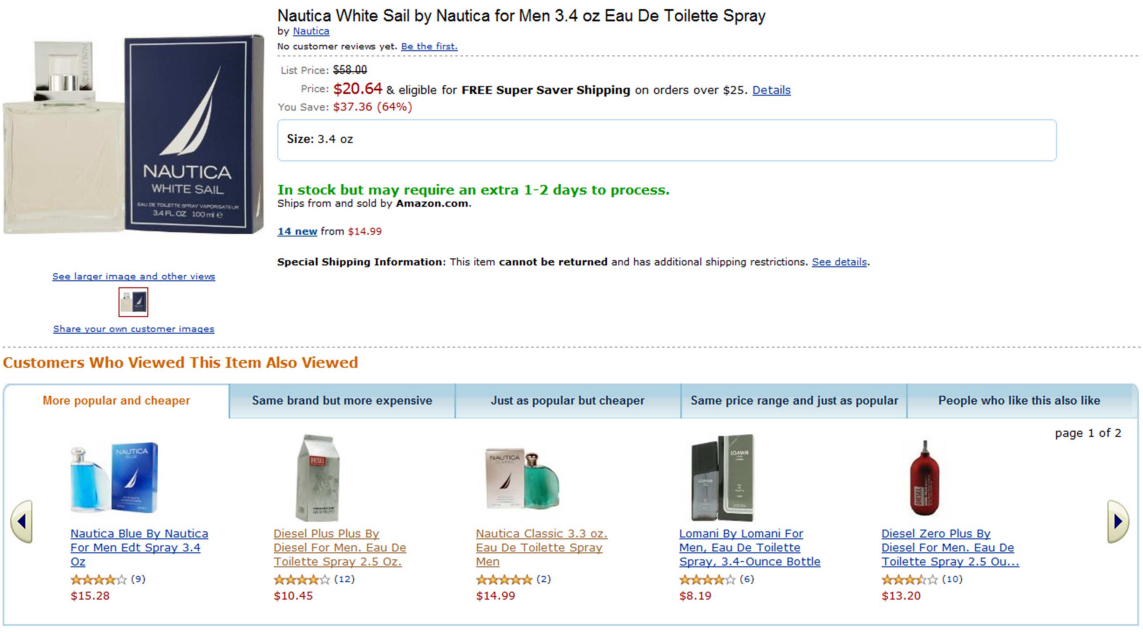

Figure 1. The simulated organization interface (content is identical to the recommendation results below).

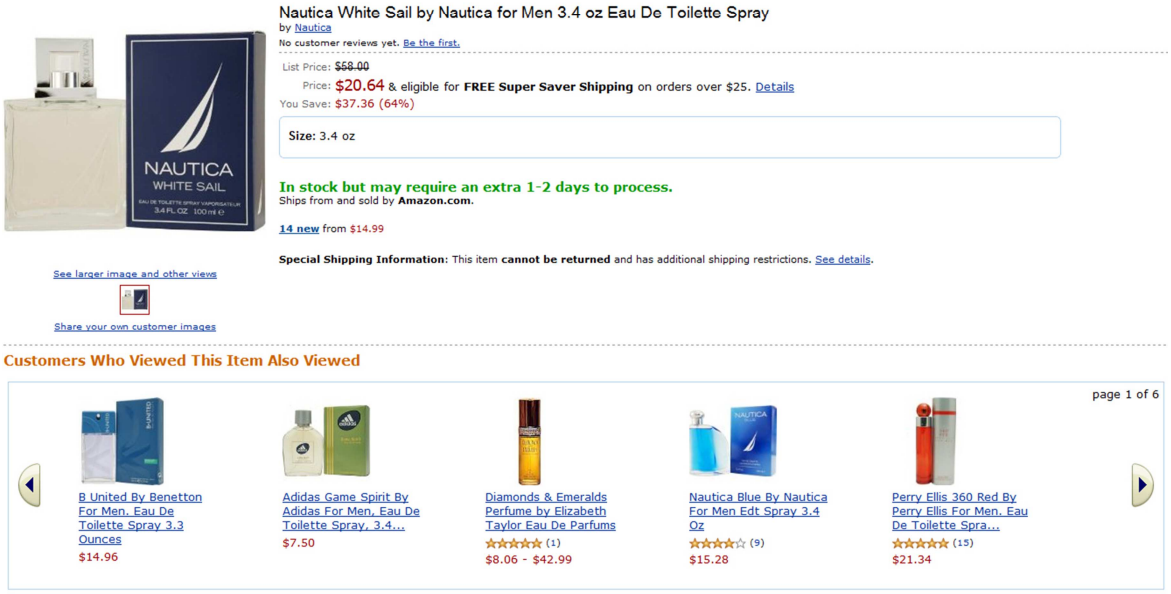

Figure 2. The standard list interface.

considers the difference among each item, asking whether "the items recommended to me are similar to each other" (also called *item-to-item diversity*). We also tried to investigate the influence of perceived diversity on users' acceptance of a recommender system. In our evaluation, we took into account perceived ease of use and useful ness of a system (facilitation, effectiveness, and supportiveness), users' attitudes towar ds the system (satisfaction, conviction, and confidence), and behavioral intentions to use it (intention to reuse, intention to tell friends, and intention to purchase). Besides, we measured users' perception on recommendation quality. Table 2 lists all of the questions as measures of these subjective variables. Each question was required to respond on a 5-point Likert scale from "strongly disagree" (1) to "strongly agree" (5).

4.4 Experiment Design and Procedure

Our user study was conducted in a within-subjects design. All participants used both interfaces, and then filled in a post-stage assessment questionnaire for the respective interface (see Table 2). In the end, they were asked to answer about their preferences on these two interfaces. All participants were randomly assigned to two experimental conditions, with a differing order in using the two interfaces. That is, 10 users in one condition evaluated the list view interface first and then the organization view interface; the

Table 2. Post-stage assessment questionnaire.

ID	Questions
Q1	I am interested in the items recommended to me.
Q2	The items recommended to me are of various kinds.
Q3	The items recommended to me are similar to each other. (reversal question)
Q4	Finding an item to buy with the help of the recommender is easy.
Q5	The recommended items effectively helped me find the ideal product.
Q6	I feel supported in selecting the items to buy with the help of the recommender.
Q7	Overall, I am satisfied with the recommender provided by this system.
Q8	I am convinced of the products recommended to me.
Q9	I am confident I will like the items recommended to me.
Q10	I will use this recommender again.
Q11	I will tell my friends about this recommender.
Q12	I would buy the items recommended, given the opportunity.

other condition had a reverse order. Counterbalance measures were taken to eliminate fatigue and learning effects as much as possible.

The user study was run at the office of an administrator who supervised the experiment and assisted participants to successfully complete all tasks, with the help of a desktop computer. Users' click behaviors were automatically recorded into log files. At the beginning, participants were asked to read a printed introduction and debriefed on the upcoming tasks. They then answered a series of background and demographic questions. In order to clarify the evaluated interfaces to the participants, two printed screenshots were shown and a brief description was given by the administrator. Then, they started using these two interfaces.

Participants were given specific tasks when using each interface. In the first interface, we asked a user to find up to three perfumes that he/she has never heard of or used before and would be willing to purchase for himself/herself given the opportunity and put them into the shopping cart. When using the second interface, the user was asked to search for three perfumes which he/she would be willing to purchase for someone of the opposite gender as a gift, in order to reduce the potential influence of users' familiarity with the product domain after using the first interface. After using each interface, the user was asked to fill in a post-stage assessment questionnaire to evaluate the interface he/she just tested. The questions are listed in Table 2.

Finally, all participants were asked to answer a questionnaire about their preferences on these two interfaces in terms of five aspects: general preference, informative, useful, good at recommending, and good at helping perceived diversity. These questions are listed in Table 3.

5. RESULTS ANALYSIS

5.1 Users' Subjective Evaluation

All responses for the post-stage questions were analyzed using paired sample t-tests. The results are shown in Figure 3. The questions marked with (**) denote that a significant difference

Table 3. Preference questionnaire.

ID	Questions
P1	Which recommendation interface did you prefer?
P2	Which recommendation interface did you find more informative?
P3	Which recommendation interface did you find more useful?
P4	Which recommendation interface was better at recommending perfumes you like?
P5	Which recommendation interface was better at helping perceive the diversity of recommendations?

were observed among users' responses. The detailed analysis is as follows. Users found the recommended items from both interfaces to be interesting (Q1) with a slight advantage for ORG ($p = 0.07$). It means that the subjective accuracy of the two interfaces is not significantly different.

With respect to users' perceived diversity in both interfaces, we asked two questions. One emphasizes the categorical difference (Q2). The other simply considers the general differences between each item (Q3). Interestingly, we could see from the results that the difference between the two interfaces was only significant with respect to the question Q2. That is, the level of perceived categorical diversity in the organization interface was significantly higher than that of the list interface (mean = 4.1, SD = 0.788 for ORG, vs. mean = 3.35, SD = 0.988 for LIST, $p < 0.05$, $t = 3.68$). However, no significant difference was measured on item-to-item diversity ($p = 0.186$). Users seemed to disagree that items were similar to each other in both interfaces (reverse scale of item-to-item diversity). Therefore, we conclude that the organization-based interface helped users' awareness of the diversity present by variety differences.

Perceived ease of use and usefulness of the system were evaluated in terms of three aspects: facilitation (Q4), effectiveness (Q5), and supportiveness (Q6). While users found ORG is more easy to use (Q4), the difference between ORG and LIST was slightly significant ($p = 0.09$). On the other hand, users thought that the recommended items were significantly more effective in helping them find the ideal product (Q5) in ORG (mean = 3.75, SD = 0.851, vs. mean = 3.2, SD = 1.005 for LIST, $p < 0.05$, $t = 2.773$). They also felt more supported in selecting the items to buy with the help of ORG (Q6, mean = 4.05, SD = 0.686, vs. mean = 3.4, SD = 1.095 for LIST, $p < 0.05$, $t = 2.371$).

In order to evaluate users' attitude towards the tested interfaces, three evaluation measures were considered: satisfaction (Q7), conviction (Q8), and confidence (Q9). Users expressed significantly higher satisfaction for ORG (Q7, mean = 3.9, SD = 0.912, vs. mean = 3.3, SD = 0.923 for LIST, $p < 0.05$, $p < 0.05$, $t = 3.559$). In addition, they seemed to be more confident that they would like the recommended items in ORG (Q9, mean = 3.65, SD = 0.875, vs. mean = 3.2, SD = 0.894 for LIST, $p < 0.05$, $t = 2.269$). Therefore, users had more positive attitudes towards the ORG interface.

Significant differences were also revealed on the measures of users' behavioral intentions to use a system. More specifically, users scored significantly higher for ORG on reusing the system (Q10, mean = 4.2, SD = 0.834, vs. mean = 3.5, SD = 0.827 for LIST, $p < 0.001$, $t = 4.273$), telling friends about it (Q11, mean =

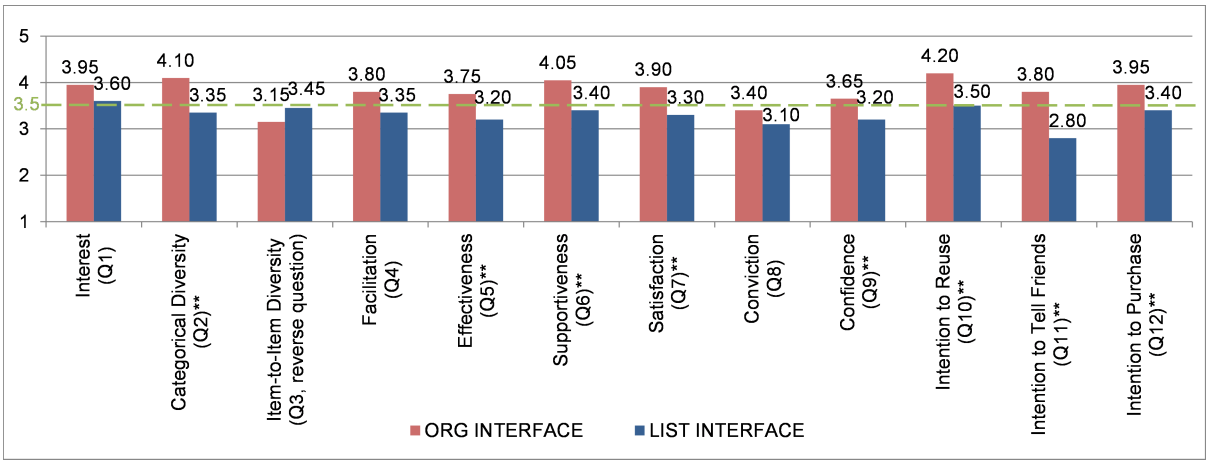

Figure 3. Usability and user satisfaction assessment results. A cut off value at 3.5 represents agreement on the 5-point Likert scale. ** is marked for significant differences at the 5% level (p-value < 0.05).

3.8, SD = 0.894, vs. mean = 2.8, SD = 0.894 for LIST, $p < 0.001$, $t = 4.359$) and purchasing the recommended items given the opportunity (Q12, mean = 3.95, SD = 0.686, vs. mean = 3.4, SD = 0.940 for LIST, $p < 0.05$, $t = 2.463$).

5.2 Final Preference

After evaluating two interfaces, users were asked to answer five questions regarding their preferences for these two interfaces. The results are shown in Figure 4. ORG got dominant preferences with more than 50% votes on all of the five questions. Particularly, 65% of users preferred the organization interface versus only 20% for the list interface, while 5% of them prefer both interfaces. More users thought that the organization-based interface was more informative (70% vs. only 10%), more useful (60% vs. 15%) and better at recommending items (50% vs. 10%). More importantly, 70% (vs. 15%) of users thought that the organization-based interface is better at helping them perceive the diversity of recommendations in contrast to the list interface.

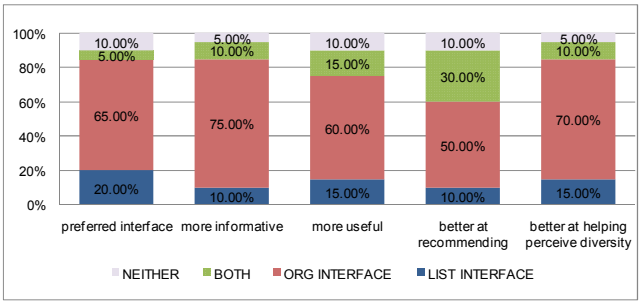

Figure 4. Preference Results.

5.3 Correlation Analysis

We did a correlation analysis between the perceived diversity (both categorical diversity and item-to-item diversity) and other subjective measures, aimed at understanding how perceived diversity influences users' acceptance of a recommender system. The results are shown in Table 4. All correlations presented in boldface and with the symbol (**) are statistical significant at the 0.05 level with two-tailed Pearson correlation coefficients.

More specifically, Table 4 shows the correlations between perceived categorical diversity and the other subjective measurements (perceived ease of use, perceived usefulness, attitudes and behaviors intentions to use). Perceived categorical diversity is highly positively related to the perceived ease of use (facilitation: $r = 0.405$, $p < 0.05$), and the perceived usefulness of the system (effectiveness: $r = 0.451$, $p < 0.01$, supportiveness: $r = 500$, $p < 0.01$). In addition, perceived categorical diversity is significantly positively correlated with satisfaction ($r = 0.576$, $p < 0.001$), conviction ($r = 0.456$, $p < 0.01$) and confidence ($r = 0.493$, $p < 0.01$). The same correlation is found with respect to behavioral intentions to use (intention to reuse: $r = 0.519$, $p < 0.01$, intention to tell friends: $r = 0.428$, $p < 0.006$, intention to purchase: $r = 0.386$, $p < 0.05$).

On the contrary, the item-to-item diversity has a weaker correlation to the three subjective measure factors. It only has a significantly correlation with facilitation ($r = -0.322$, $p < 0.05$) in the aspect of perceived usefulness. In the aspect of attitudes to the system, it is strongly related to conviction ($r = -0.390$, $p < 0.05$) and confidence ($r = -0.426$, $p < 0.01$). Furthermore, the item-to-item diversity is significantly correlated to intention to reuse ($r = -0.434$, $p < 0.01$).

5.4 Discussion and Design Guidelines

According to users' responses to the subjective questionnaires, we saw that users perceived more categorical diversity of recommendations in the organization interface compared to in the list interface. This suggests that the organization interface could indeed help users become aware of the diversity in recommendation lists, particularly the difference among categories which is difficult to perceive in the list view interface; there is a 22.4% increase. However, there is no significant statistical difference between the organization-based interface and the list-based interface with respect to the item-to-item diversity. The organization-based interface does not appear to be particularly advantageous in this case. After using two interfaces, users were asked to answer five questions regarding their preferences for these two interfaces. 70% (vs. 15%) of users thought that the organization-based interface is better at helping them perceive the diversity of recommendations in contrast to the list interface.

Table 4. Correlation results on categorical and item-to-item diversity (denotes statistical significance at the 0.05 level, i.e., p-value<0.05).**

Factors		Correlation (Sig.)	
		Categorical Diversity (Q2)	**Item-to-item Diversity (Q3)**
Ease of Use	Facilitation (Q4)	**0.405(0.01**)**	**-0.322(0.043**)**
Perceived Usefulness	Effectiveness (Q5)	**0.451(0.003**)**	-0.247(0.124)
	Supportiveness (Q6)	**0.500(0.001**)**	-0.247(0.124)
Attitudes	Satisfaction (Q7)	**0.576(0.000**)**	-0.263(0.101)
	Conviction (Q8)	**0.456(0.003**)**	**-0.390(0.013**)**
	Confidence (Q9)	**0.493(0.001**)**	**-0.426(0.006**)**
Behavioral Intentions	Intention to reuse (Q10)	**0.519(0.001**)**	**-0.434(0.005**)**
	Intention to tell friends (Q11)	**0.428(0.006**)**	-0.097(0.553)
	Intention to purchase (Q12)	**0.386(0.014**)**	-0.226(0.161)

Previous studies have shown that the diversity of recommendation lists influences users satisfaction [23]. However, it is still not well understood why and how such an impact occurs. Our correlation results reveal that categorical diversity in recommendation lists influences users' perceived ease of use of a system, perceived usefulness of the system and attitudes towards the system, thereby resulting in a positive effect on their intention to use the system. While the item-to-item diversity has an impact on users' acceptance to the system as well, the effect is not as strong as with categorical diversity. On the other hand, our results empirically demonstrate that perceived diversity is indeed one critical factor influencing users' adoption of a recommender system due to its strong correlation with the factors (perceived ease of use, perceived usefulness, attitudes, and behavioral intentions) which are considered in users' acceptance models, like TAM [6].

Furthermore, the correlation results show that perceived diversity plays a role in providing supporting information, which leads to increased user confidence in a system. In previous research about diversity-enhancing techniques, diversity has only been demonstrated to help users reduce interaction cycles and more efficiently find the target item [12, 22]. Our empirical results indicate that users obtained more supportive and convincing information when they perceive diversity, and thereafter they felt more confident about their decisions. In other words, diversity can not only make recommendations covering a wide range of users' interests, but can also provide supportive information to aid users make decisions.

The current study confirmed the critical role of diversity in a recommender's success. It further shows promising results that contribute to the field:

1) Even though a number of diversity-enhancing techniques have been proposed in the literature, interface design issues relative to diversity have been overlooked. Our study demonstrates that a simple reorganization of the results into a category layout could have a strong positive effect on users' perceived qualities of the system, especially their satisfaction and intention to use and purchase. This suggests a novel research direction on the issue of diversity-enhancing technology.

2) Our results show that perceived ***categorical*** diversity has an even stronger influence on users' positive perception and acceptance of a recommender system than item-to-item diversity. This highlights the critical role of categorical diversity on user experience of a recommender system. However, it doesn't mean the item-to-item diversity is trivial. According to users' responses, it is difficult for them to be aware of the item-to-item diversity in recommendations.

To conclude the findings of our study, we propose the following design guidelines.

Guideline 1: Take recommendation diversity into account when designing recommender systems.

Guideline 2: Make users aware of the diversity (both categorical diversity and item-to-item diversity) existed in recommendation lists by explaining the similarities and differences among the displayed items.

Guideline 3: Display recommendations in a category layout by adopting organization interface designs to enhance users' perception of the categorical diversity of the recommendations.

6. CONCLUSION AND FUTURE WORK

We conducted an in-depth user study to compare an organization-based interface with the standard list-based interface. Experimental results reveal that the ORG interface indeed influence the users' perception of the recommendation diversity. Users in the ORG interface had more strong perception of categorical diversity. Even though users found the recommended items to be interesting in both interfaces, ORG users were more satisfied with the recommender. While both interfaces were easy to use, ORG users indicated that the interface was more helpful for them in terms of locating the items they wanted to buy (decision support). Most importantly, ORG users are more likely to use the system again, tell their friends about it and buy the recommended items. Strong correlation has been found between perceived diversity and users' satisfaction.

Our future work includes validating our findings in other product domains, comprehensively investigating the influence of diversity on the success of a recommender system, exploring other formats of interface designs which can more effectively enhance users' experience with a recommender system.

7. ACKNOWLEDGMENTS

We thank EPFL, the Swiss National Science foundation, and the ministry of education of the People's Republic of China for supporting the reported research work. We are grateful to the participants of our user studies for their patience and time.

8. REFERENCES

[1] Bradley, K. and Barry, S. 2001. Improving Recommendation Diversity. In *Proceedings of the Twelfth Irish Conference on Artificial Intelligence and Cognitive Science* (Maynooth, Ireland, 2001), 85-94.

[2] Castagnos, S., Jones, N. and Pu, P. 2010. Eye-tracking product recommenders' usage. In *Proceedings of the fourth ACM conference on Recommender systems* (Barcelona, Spain, 2010). ACM, 1864717, 29-36.

[3] Chen, L. and Pu, P. 2010. Eye-Tracking Study of User Behavior in Recommender Interfaces.*User Modeling, Adaptation, and Personalization*, De Bra, P., Kobsa, A. and Chin, D., eds. LNCS 6075, Springer Berlin / Heidelberg, 375-380.

[4] Chen, L. and Pu, P. 2007. Preference-Based Organization Interfaces: Aiding User Critiques in Recommender Systems.*User Modeling 2007*, Conati, C., McCoy, K. and Paliouras, G., eds. Lecture Notes in Computer Science 4511, Springer Berlin / Heidelberg, 77-86.

[5] Cosley, D., Lawrence, S. and Pennock, D.M. 2002. REFEREE: an open framework for practical testing of recommender systems using ResearchIndex. In *Proceedings of the 28th international conference on Very Large Data Bases* (Hong Kong, China, 2002). VLDB Endowment, 1287374, 35-46.

[6] Davis, F.D., Bagozzi, R.P. and Warshaw, P.R. 1989. User acceptance of computer technology: a comparison of two theoretical models. *Manage. Sci. 35*, 8, 982-1003.

[7] Fleder, D.M. and Hosanagar, K. 2007. Recommender systems and their impact on sales diversity. In *Proceedings of the 8th ACM conference on Electronic commerce* (San Diego, California, USA, 2007). ACM, 1250939, 192-199.

[8] Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T. 2004. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst. 22*, 1, 5-53.

[9] Hijikata, Y., Shimizu, T. and Nishida, S. 2009. Discovery-oriented collaborative filtering for improving user satisfaction. In *Proceedings of the 14th international conference on Intelligent user interfaces* (Sanibel Island, Florida, USA, 2009). ACM, 1502663, 67-76.

[10] Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H. and Newell, C. 2011. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction Journal (UMUAI), Special Issue on User Interfaces for Recommender Systems* (Upcoming).

[11] Mccarthy, K., Reilly, J., Smyth, B. and Mcginty, L. 2005. Generating Diverse Compound Critiques. *Artif. Intell. Rev. 24*, 3-4, 339-357.

[12] McGinty, L. and Smyth, B. 2003. On the role of diversity in conversational recommender systems. In *Proceedings of the 5th international conference on Case-based reasoning: Research and Development* (Trondheim, Norway, 2003). Springer-Verlag, Berlin, 276-290.

[13] McNee, S.M., Riedl, J. and Konstan, J.A. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI '06 extended abstracts on Human factors in computing systems* (Montreal, Quebec, Canada, 2006). ACM, 1125659, 1097-1101.

[14] McSherry, D. 2002. Diversity-Conscious Retrieval. In *Proceedings of the 6th European Conference on Advances in Case-Based Reasoning* (2002). Springer-Verlag, 219-233.

[15] Mislevy, R.J. and Gitomer, D.H. 1996. The Role of Probability-Based Inference in an Intelligent Tutoring System. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*. *5*, 3-4, 253-282.

[16] Pu, P. and Chen, L. 2006. Trust building with explanation interfaces. In *Proceedings of the 11th international conference on Intelligent user interfaces* (Sydney, Australia, 2006). ACM, 1111475, 93-100.

[17] Pu, P. and Chen, L. 2010. A User-Centric Evaluation Framework of Recommender Systems. In *Proceedings of the ACM RecSys 2010 Workshop on User-Centric Evaluation of Recommender Systems and Their Interfaces (UCERSTI)* (Barcelona, Spain, 2010). CEUR-WS.org, 14-21.

[18] Pu, P., Zhou, M. and Castagnos, S. 2009. Critiquing recommenders for public taste products. In *Proceedings of the third ACM conference on Recommender systems* (New York, New York, USA, 2009). ACM, 1639760, 249-252.

[19] Shimazu, H. 2001. ExpertClerk: navigating shoppers' buying process with the combination of asking and proposing. In *Proceedings of the 17th international joint conference on Artificial intelligence - Volume 2* (Seattle, WA, USA, 2001). Morgan Kaufmann Publishers Inc., 1642287, 1443-1448.

[20] Smith, D., Menon, S. and Sivakumar, K. 2005. Online peer and editorial recommendations, trust, and choice in virtual markets. *Journal of Interactive Marketing*. *19*, 3, 15-37.

[21] Smyth, B. and McClave, P. 2001. Similarity vs. Diversity. In *Proceedings of the 4th International Conference on Case-Based Reasoning: Case-Based Reasoning Research and Development* (2001). Springer-Verlag, 758890, 347-361.

[22] Zhang, M. and Hurley, N. 2008. Avoiding monotony: improving the diversity of recommendation lists. In *Proceedings of the 2008 ACM conference on Recommender systems* (Lausanne, Switzerland, 2008). ACM, 1454030, 123-130.

[23] Ziegler, C.-N., McNee, S.M., Konstan, J.A. and Lausen, G. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web* (Chiba, Japan, 2005). ACM, 22-32.