# Trust Building with Explanation Interfaces

Pearl Pu and Li Chen
Human Computer Interaction Group, School of Computer and Communication Sciences
Swiss Federal Institute of Technology in Lausanne (EPFL)
CH-1015, Lausanne, Switzerland
{pearl.pu, li.chen}@epfl.ch

## ABSTRACT
Based on our recent work on the development of a trust model for recommender agents and a qualitative survey, we explore the potential of building users' trust with explanation interfaces. We present the major results from the survey, which provided a roadmap identifying the most promising areas for investigating design issues for trust-inducing interfaces. We then describe a set of general principles derived from an in-depth examination of various design dimensions for constructing explanation interfaces, which most contribute to trust formation. We present results of a significant-scale user study, which indicate that the organization-based explanation is highly effective in building users' trust in the recommendation interface, with the benefit of increasing users' intention to return to the agent and save cognitive effort.

## Categories and Subject Descriptors
H.1.2 [**Models and Principles**]: User/Machine Systems – *human factors, software psychology*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *evaluation/methodology, graphical user interfaces (GUI), user-centered design*.

## General Terms
Design, Experimentation, Human Factors, Algorithms.

## Keywords
Explanation interfaces, trust building, recommender agents, tradeoff assistance.

## 1. INTRODUCTION
The importance of explanation interfaces in providing system transparency and thus increasing user acceptance has been well recognized in a number of fields: expert systems [9], medical decision support systems [2], intelligent tutoring systems [23], and data exploration systems [4].

Being able to effectively explain results is especially important for product recommender systems. When users face the difficulty of choosing the right product to purchase, the ability to convince them to buy a proposed item is an important goal of any recommender system in e-commerce environments. A number of researchers have started exploring the potential benefits of explanation interfaces in a number of directions.

Case-based reasoning recommender systems that can explain their recommendations include ExpertClerk [21], Dynamic critiquing systems [10], FirstCase and TopCase [14, 15]. ExpertClerk explained the selling point of each sample, in terms of its difference from the other two contrasting samples. In a similar way, FirstCase can explain why one case is more highly recommended than another by highlighting the benefits it offers and also the compromises it involves with respect to the user's preferences. In TopCase, the relevance of any question the user is asked can also be explained in terms of its ability to discriminate between competing cases.

McCarthy et al [10] proposes to educate users about product knowledge by explaining what products do exist instead of justifying why the system failed to produce a satisfactory outcome. This is similar to the goal of resolving users' preference conflict by providing them with partially satisfied solutions [19].

A number of researchers also reported evaluation of explanation interfaces with real users. Herlocker et al addressed explanation interfaces for ACF (automated collaborative filtering) systems, and demonstrated that the histogram with grouping of neighbor ratings was the most compelling explanation component among the users that they studied [8]. They also showed that providing explanations can improve the acceptance of ACF systems and potentially improve users' filtering performance. Sinha and Swearingen [22] found that users like and feel more confident about recommendations that they perceive as transparent.

Some consumer decision support systems with explanation interfaces can be found on commercial websites such as Logical Decisions (www.logicaldecisions.com), Active Decisions (www.activedecisions.com), and SmartSort (shopping.yahoo.com/smartsort).

So far, previous work on explanation interfaces has not explored the potential of using explanation interfaces for building users' trust, which is a long term relationship between a user and the organization that the recommender system represents.

Trust issues are critical to study for recommender systems used in e-commerce where the traditional salesperson is replaced by a product recommender agent. Studies show that customer trust is positively associated with customers' intention to transact, purchase a product, and return to the website [7]. These results have mainly been derived from online shops' ability to ensure security, privacy and reputation, i.e. the integrity and benevolence aspects of trust constructs, and less on a system's competence such as a recommender system's ability to explain its result. The contribution of our work is that we both investigate the inherent benefits of using explanation for trust building and examine whether such interface features provide the same trust-related benefits as other trust


*IUI'06*, January 29-February 1, 2006, Sydney, Australia.

constructs. We primarily consider the competence perception and its essential contribution to trust-induced benefits. See [16] for example regarding other trust-related issues based on reputation in recommender systems.

This paper is organized as follows: section 2 presents a trust model for recommender systems and some results from a carefully constructed survey, identifying explanation interfaces as one of the most promising areas to investigate design issues for trust-inducing interfaces; section 3 describes a set of general principles derived from an in-depth examination of various design dimensions for constructing explanation interfaces; section 4 presents a significant-scale empirical study; section 5 presents results from that study, which indicate that the organization-based explanation is highly effective to build users' trust in the recommendation interface, with the benefit of increasing their intention to return to the agent and save their cognitive effort; section 6 discusses the implication of this work to related work in this area, followed by the conclusion section.

# 2. TRUST MODEL AND TRUSTING INTENTIONS

We summarize our recent work on building a trust model for recommender systems which was reported in a workshop [5]. The results highly influence our current work and therefore are presented here as an integral part of our investigation of trust building with explanation interfaces.

## 2.1 Trust Model for Recommender Systems

We have conceptualized a general trust model for recommender agents. It consists of three components: system features, trustworthiness of the agents, and trusting intentions. The system features mainly deal with those design aspects of a recommender agent that can contribute to the promotion of its trustworthiness. We classified them into three groups: the interface display techniques, the algorithms that are used to propose recommendations, and user-system interaction models such as how an agent elicits users' preferences.

The agent trustworthiness is a trust formation process based on the users' perception of the agent's competence, reputation, integrity, and benevolence. It has been regarded as the main positive influence on the trusting intentions [7, 13]. In this paper, we primarily consider the competence perception and its essential contribution to trust-induced benefits.

The trusting intentions are the benefits expected from users once trust has been established by the recommender agents. The trusting intentions include the intention to purchase a recommended item, return to the store for more information on products or purchase more recommended products, and save effort. The intention to save effort is of particular interest to us because it examines whether upon establishing a certain trust level with the agent, users will likely spend less cognitive effort or actual time in selecting the recommended items.

## 2.2 Trust Building with Explanation Interfaces

As a first step, we primarily consider trust building by the different design dimensions of interface display techniques such as content selection, explanation generation, and recommendation algorithms. We investigate the modality of explanation, e.g., the use of graphics vs. text, the amount of information used to explain, e.g., whether long or short text is more trust inspiring, and most importantly whether alternative explanation techniques exist that are more effective in trust building than the simple "why" construct currently used in most e-commerce websites.

The explanation generation comprises the steps of content selection and organization, media allocation, and media realization and coordination [4]. Content selection determines what information should be included in the explanations. Once the content is selected, we must know how to organize and display it. The simplest strategy is to display the recommendation content in a rank ordered list with a "why" tool tip explaining the computational reasoning behind it.

As an alternative and potentially more effective technique, we have designed an organization-based explanation interface where the best matching item is displayed at the top of the interface along with several categories of tradeoff alternatives. Each category is labeled with a title explaining the characteristics of the items the respective category contains (see Figure 2).

## 2.3 Qualitative Survey and Results

We have conducted a survey with 53 users in order to understand the interaction among the three components of our trust model: the effect of an agent's competence in building users' trust, the influence of trust on users' problem solving efficiency and other trusting intentions, and the effective means to build trust using explanation-based interfaces. Nine hypotheses (see [5] for details) were established, each of which is a statement for which the participants indicated their level of agreement.

Results indicate that the competence of recommender agents would not be the only contribution to users' trust formation process, but it is positively correlated with the trusting intention to return. In other words, if users possess a high perception of the recommender agent's competence, they would be more inclined to return to the agent for other product information and recommendations, but they would not necessarily intend to buy the product from the website where the agent was found. Post-survey discussion indicated that they would visit more websites to compare the product's price before making a purchase. The website's security, reputation, delivery service and privacy policy were also important considerations in buying a product.

Users positively responded that explanation can be an effective means to achieve users' trust, and the organization interface is a more effective explanation technique than the simple "why" construct. On the other hand, the modality and richness of an explanation interface did not seem to contribute to the effectiveness of the interface. From the participants' viewpoints, these two aspects were mostly dependent on the concrete product domain. Users would prefer a short and concise conversational sentence for the so-called low-risk products such as movies and books, but if they were selecting products which carry a high level of financial and emotional risks such as cars and houses, a more detailed and reasonable explanation would be favored. In addition, people from different educational backgrounds seemed to have different preferences on the media richness.

Based on the trust model and results from the qualitative survey, we have decided to focus our attention on explanation interfaces and the related design issues for building users' trust.

# 3. ORGANIZATION-BASED EXPLANATION INTERFACES

Traditional product search and recommender systems present a set of top-k alternatives to users. We call this style of displaying the results the k-best interface. Because these alternatives are calculated based on users' revealed preferences (directly or indirectly), these top-k items may not provide for diversity. Recently the need to include more diversified items in the result list has been recognized. Methods have been developed to address users' potentially unstated preferences [6,17], to cover topic diversity [24], to propose possible tradeoffs a user may be prepared to accept [14], and to allow faster navigation to the target choice by critiquing the proposed items [3,11,20]. These related works have led us to developing an organization-based explanation interface, combining the ideas of diversity, tradeoff reasoning, and explanation. Here we review a set of design principles that show promise for the design of such interfaces.

## 3.1 Design Principles

We have implemented more than 13 paper prototypes of the organization-based interface, exploring all design dimensions such as how to generate categories, whether to use short or long text for category titles, how many tradeoff dimensions to include, whether to include example products in the categories or just the category titles, etc. We have derived 5 principles based the results of testing these prototypes with real users in the form of pilot studies and interviews.

*Principle 1: Categorize remaining recommendations according to their similar tradeoff properties relative to the top candidate*

We consider the case where the explanation interface is used in the early stage of the entire interaction cycle between a user and a recommender agent. We assume that users are unlikely to have stated all of their preferences. Consequently, they have not considered tradeoff alternatives of the product currently being considered. According to [18, 20], integrating tradeoff support in a product search tool can improve users' decision accuracy by up to 57%. Thus, this principle suggests displaying tradeoff alternatives in addition to the top candidate. Each category comprises a set of similar items having the same tradeoff properties. For example, one category contains the recommendations of notebooks that are cheaper but heavier than the top candidate, and another category's notebooks are lighter but more expensive. Each category indicates a tradeoff direction where users would potentially navigate to for achieving their final decision goals.

*Principle 2: Propose improvements and compromises in the category title using conversational language; keep the number of tradeoff attributes under five to avoid information overload*

Here we consider designing a category's title in terms of its format and richness. After surveying some users, we found that most of them preferred the category title displayed in natural and conversational language because that makes them feel at ease. For example, the title "these notebooks have a lower price and faster processor speed, but heavier weight" is preferred to the title "cheaper and faster processor speed and heavier." Moreover, the former title is also preferred to the title "they have a lower price and faster processor speed and bigger memory, but heavier weight and larger display size" which includes too many tradeoff properties. Many users indicate that they cannot handle tradeoff with more than three attributes.

*Principle 3: Eliminate dominated categories, and diversify the categories in terms of their titles and contained recommendations*

The third principle proposes to provide the most beneficial but diverse categories to users. If one category is too similar with, or dominated by, another one in terms of their tradeoff properties, it would not provide potential recommendation power to the user. Therefore, it is better to exclude the dominated categories and diversify the returned categories. In addition, the pilot study on category design showed that the number of total displayed categories is more effective when less than four since too many categories cause information overload and confusion.

*Principle 4: Include actual products in a recommended category*

When we compared two interface designs where one displays only category titles versus one displaying both category titles and a few actual products in each category, users indicated a strong preference in favor of the latter design, mainly due to the fact that they were able to find their target choice much faster. Given the limitation of the display size and users' cognitive effort, a designer can choose up to 6 items to display in each category.

*Principle 5: Rank recommendations within each category by exchange rate rather than similarity measure*

We have also performed a pilot study to compare the effects of two ranking strategies for the recommendations within the category. The *similarity* strategy is broadly used by early case-based and preference-based reasoning systems (CBR), which rank items according to the similarity relative to a user's current query. We propose another strategy based on the *exchange rate* of an item relative to the top candidate, i.e. its potential gains versus losses compared with the top candidate (the detail formula for exchange rate calculation will be shown shortly). The study showed that users could more quickly find their target choice when the recommended items within each category were sorted by the exchange rate rather than by similarity.

## 3.2 Organization Algorithm

The organization algorithm was designed and implemented optimizing the overall objectives of the five principles. The top level of the algorithm can be described in four steps: generate all possible category titles by the Apriori algorithm [1]; exclude dominated categories; select a few prominent categories not only with longer tradeoff distance with the top candidate but also with higher diversity degree between each other; rank the recommended items within each category by their exchange rates relative to the top candidate. A resulting example based on the organization algorithm can be seen in Figure 2.

### Step 1: Generate all possible categories

We generate the categories using the method presented in [11]. A slight modification is that we represent each recommendation as a tradeoff vector comprising a set of (*attribute*, *tradeoff*) pairs (the pair is also called an item in the algorithm). Each *tradeoff* vector indicates whether the *attribute* of the recommendation is *improved* (denoted as ↑) or *compromised* (denoted as ↓) compared to the same attribute of the top candidate. An example of a notebook recommendation is denoted by a tradeoff vector {(price, ↑), (processor speed, ↓), (memory, ↓), (hard drive size, ↑), (display size, ↑), (weight, ↓)}, indicating that this notebook has a lower price, more hard drive size, and larger display size, but heavier weight, slower processor speed, and less memory relative to the top

recommended notebook. Thus a tradeoff vector describes how the current product compared to the top candidate in terms of its advantages and disadvantages, rather than the simple equality comparison used in dynamic critiquing (bigger, smaller, equal, different, etc.). After all tradeoff vectors are used as input to the Apriori algorithm, we obtain the frequent item sets in terms of their tradeoff potentials underlying all the recommendations.

In order to limit the number of attributes involved in the category title (principle 2), we set the Apriori's option "maximal number of items per set" as 3.

The recommendations whose tradeoff vectors contain the same subset of items are grouped in the same category. Indeed, a recommendation can belong to more than one category given that it has different subsets of items shared by other groups of recommendations, thus leading to a large amount of categories potentially produced by Apriori. In the following steps, we concentrated on how to select the most beneficial categories to present to users based on our design principles.

### Step 2: Exclude dominated categories

If one category is strictly dominated by another category in terms of the item sets they contain in the titles, we will not show it to the user. Formally, a category title $C_1$ is dominated by another category title $C_2$ if $C_1$ is a subset of $C_2$ in terms of the items (*attribute, tradeoff*) contained in the title or the two titles have the same number of items (i.e. $|C_1| = |C_2|$), but

$\forall$ item $T_i \in C_1$, $\exists$ $T_j \in C_2$:

where $T_i.attribute = T_j.attribute$ (with equal attribute name)

and $T_i.tradeoff \leq T_j.tradeoff$ ( with equal or less preferred tradeoff property, i.e. "↓" $\prec$ "↑")

and $\exists$ $T_p \in C_1$ and $T_q \in C_2$

where $T_p.attribute = T_q.attribute$

and $T_p.tradeoff \prec T_q.tradeoff$ (at least one item is with less preferred tradeoff property)

### Step 3: Select prominent categories with longer tradeoff distance and higher diversity degree

This is where we depart from the dynamic critiquing method [11], which uses the low support value to select categories. We use two criteria to select up to four categories: the maximal tradeoff distance with the top candidate and maximal diversity among each other in terms of their titles and contained recommendations (principle 3). The tradeoff distance of each category is defined as the average sum of the exchange rate of all recommendations which are contained in the category:

$$TradeoffDistance(C_i, TC) = \frac{1}{|SR(C_i)|} \sum_{R \in SR(C_i)}^{|SR(C_i)|} ExRate(R, TC)$$

where TC is the top candidate, $SR(C_i)$ is the set of recommendations contained in the category $C_i$, and ExRate(R, TC) is the exchange rate of the recommendation R compared to the top candidate (see the ExRate formula in Step 4). Intuitively, a higher tradeoff distance indicates that a category provides the highest overall tradeoff benefits to users (more gains than losses).

During the selection process, the category with the longest tradeoff distance will be initially selected as the first category. The second category will be selected if it has the biggest value of $F(C_i)$ in the remaining non-selected categories according to the following formula:

$$F(C_i) = TradeoffDistance(C_i, TC) \times Diversity(C_i, SC)$$

where $C_i$ is the current considered category in the remaining set, TC is the top candidate, and SC denotes the set of categories so far selected. $F(C_i)$ is the combination of the category's tradeoff distance and diversity degree with respect to the categories selected so far. The subsequent categories are selected according to the same rule. The selection process will end when the desired *k* categories have been selected.

The global diversity of $C_i$ with SC is the average sum of its local diversity with each category in the SC set. The local diversity of two categories is further determined by two factors: the title diversity and recommendation diversity.

$$Diversity(C_i, SC) = \frac{1}{|SC|} \sum_{C_j \in SC}^{|SC|} TitleDiv(C_i, C_j) \times \text{Re}comDiv(C_i, C_j)$$

The title diversity determines the degree of difference between the two item sets ($C_i$ and $C_j$) respectively representing the two compared categories' titles:

$$TitleDiv(C_i, C_j) = 1 - \frac{|C_i \cap C_j|}{|C_i|}$$

The recommendation diversity measures the different amount of recommendations contained in the two compared categories:

$$\text{Re}comDiv(C_i, C_j) = 1 - \frac{|SR(C_i) \cap SR(C_j)|}{|SR(C_i)|}$$

where $SR(C_i)$ represents the set of recommendations included in category $C_i$.

### Step 4: Rank recommendations within a given category by exchange rate

The global exchange rate for each recommendation R is formulated as:

$$ExRate\ (R, TC) = \sum_{i=1}^{p} w_i exrate\ (v_{r,i}, v_{tc,i})$$

where *p* is the number of attributes, $w_i$ is the weight of attribute i, and *exrate* is the local exchange rate computed for each attribute ($v_{r,i}$ and $v_{tc,i}$ are the values of the $i^{th}$ attribute of R and TC respectively).

For numeric attributes, $exrate\ (v_i, v_j) = q \times \frac{v_i - v_j}{range}$. The parameter $q$=1 if the attribute $i$ is in increasing order (i.e. the more, the better), and $q$=−1 if $i$ in decreasing order (i.e. the less, the better). For symbolic attributes, $exrate\ (v_i, v_j) = 1$ if $v_i \neq v_j$ and $v_i$ is preferred to $v_j$, or $-1$ if contrarily, or 0 if $v_i = v_j$.

Therefore, the exchange rate motivates a user to consider alternative choices. A positive and higher exchange rate means that there are potentially more gains than losses of an alternative product compared to the top candidate.

# 4. USER EVALUATION

In order to understand whether the tradeoff-based organization interface can be an alternative and more effective way to explain recommendations, we conducted a significant-scale empirical study during April-June 2005 that compared the organized view with the traditional "why" interface in a within-subjects design. The main objective is to measure the difference of users' trust in the two interfaces, from their perceived trustworthiness of the interface in terms of the competence construct and two trusting intentions, the intention to return and save effort.

## 4.1 Materials

In order to avoid any carryover effects due to the within-subjects design, we developed four (2 x 2) experiment conditions. A total of 72 participants were randomly assigned to one of the four experiment conditions, resulting in a sample size of 18 subjects for each condition cell. Each condition has a different order of appeared interfaces and a different product domain associated with the interface. For example, the 18 users in one experiment condition evaluated the ranked list interface with "why" explanations for finding a digital camera (similar to Figure 1 but with digital cameras as the product domain), and then the organization interface for finding a notebook (Figure 2).

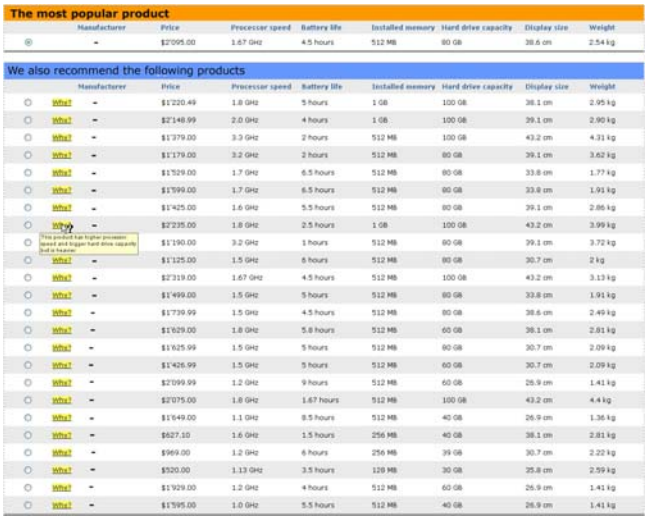


**Figure 1. The "why" interface used in the user study.**

Both product domains comprise 25 up-to-date items, where each notebook has 8 attributes (manufacturer, price, processor speed, battery life, etc.) and each digital camera contains 9 attributes (manufacturer, price, megapixels, optical zooms, etc.). To prevent the brand of products from influencing users' choice, we replaced them by manufacturers which do not exist (masked out in the figures).

To minimize these behavior differences, we considered asking users to select an item out of the top 25 most popular products from a commercial website (www.pricegrabber.com) in this user study. The top candidate is the most popular item in both interfaces (Figure 1 and 2). In the "why" interface the remaining 24 products are sorted by their exchange rates relative to the top candidate, where the "why" tool-tip explains how one product compares to the most popular item (Figure 1). In the organization interface, the remaining items are grouped in four ($k$=4) categories generated based on our organization selection and ranking algorithms (Figure 2). The radio button alongside with each item is used by participants to select the product that they are prepared to purchase. Since the most popular candidates in both interfaces are based on the website's opinion, rather than the evaluators' own opinions, we judged that the respondent is likely to view the other 24 products and consult the explanations. As it turned out, it was indeed the case since less than 11.3% of users selected the top candidate in the "why" interface, and only 8.3% in the case of the organization interface.

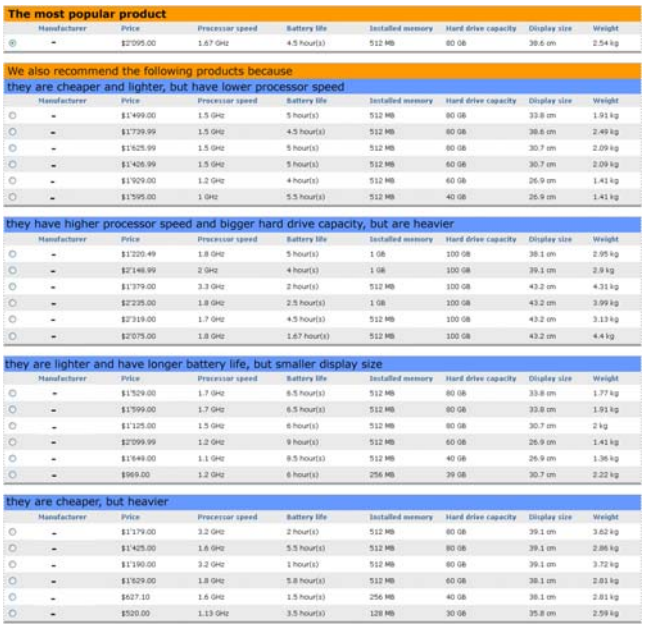


**Figure 2. The organization interface used in the user study.**

## 4.2 Participants

A total of 72 volunteers were recruited as participants in the user study. They are from 16 different countries and have different professions (student, professor, research assistant, engineer, secretary, sales clerk and manager) and educational backgrounds (high school, bachelor, master and doctor). Table 1 shows some of their demographic characteristics.

**Table 1. Demographic characteristics of participants (total 72)**

| | | | |
|---|---|---|---|
| Gender | Female | Male | |
| | 19 (26.4%) | 53 (73.6%) | |
| Education | High school, Bachelor, Master, Doctor | | |
| Nationality | 16 countries (Spain, Canada, China, etc.) | | |
| Age | 20-30 | 30-40 | >40 |
| | 64 (88.9%) | 4 (5.56%) | 4 (5.56%) |
| Online shopping experience | Yes | No | |
| | 62 (86.1%) | 10 (13.9%) | |

Among the participants, 54 had bought a notebook in the past two years, and 59 users had bought a digital camera. Most of all the participants intend to purchase a new notebook (57 users) and digital camera (60 users) in the near future.

## 4.3 Procedure

The user study was conducted at places convenient for the participants (office, home, cafeteria, etc.) with the help of a provided notebook or desktop computer. An online procedure containing the instructions, evaluated interfaces and questionnaires was implemented so that users can easily follow, and also for us to record all of their actions in a log file. There was also an administrator present in each user study to answer any of the user's questions in addition to taking notes.

The online experiment was prepared in two versions, English and French, since these are the participants' native languages. At the beginning of each session, the participants were first asked to choose the language that they prefer, and then they were debriefed on the objective of the experiment and the upcoming tasks. In particular, they were asked to evaluate two graphical recommendation interfaces and to determine which interface is more helpful in recommending products to users. Thereafter, a short questionnaire was to be filled out about their demographics, e-commerce experience and product knowledge.

Participants would then start evaluating the two interfaces one by one corresponding to the order defined in the assigned experiment condition. For each interface, the main user task was to select a product the participant would purchase if given the opportunity, followed by a set of 6 questions about his/her overall opinions regarding the interface. Users were also encouraged to provide any comment on the interface.

## 4.4 Hypothesis and Measured Data

Our main hypothesis was that users would build more trust in the organization-based explanation interface than the simple "why" construct in the list view. This was mainly assessed by the three trust constructs in our trust model: perceived competence, the intention to return, and the intention to save effort. The intention to save effort is further measured by the cognitive effort and actual completion time consumed.

# 5. RESULTS ANALYSIS

Results were analyzed for each measured variable using paired samples t-test.

## 5.1 Perceived Competence

Users' subjective perception of the competence in the interface was mainly measured by their perception of the interface's ease of use and efficiency in comparing products. Each is asked by one item (or question) in the post-questionnaire marked on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Table 2 indicated participants' mean responses to each item for the two interfaces, and the Cronbach's alpha value representing how well the two items are related and unified to the construct "perceived competence".

Both items were responded to be on average higher for the organization interface, which showed that most users regarded the organization-based explanation interface more comfortable to use and perceived it to be more efficient in making product comparisons. The overall level of perceived competence of the organization interface was thus higher than that provided by the "why" interface (mean=3.31, SD=1.05, vs. mean=2.75, SD=1.20 for the "why" interface, t=3.74, $p$<0.001, see Figure 3; median=3.5 vs. 3; mode=4 vs. 3.5).

**Table 2. Perceived-competence construct**

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

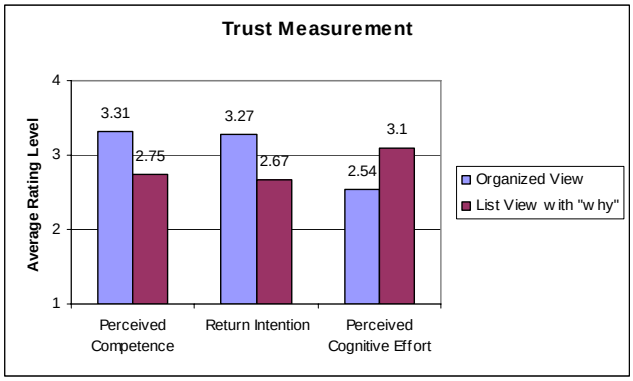


**Figure 3. Mean difference of participants' trust formation for the two interfaces.**

## 5.2 Intention to Return

As demonstrated in our previous work [5], the most remarkable benefit of the competence-inspired trust was its positive influence on users' intention to return. Accordingly, we regard the "intention to return" as an important criterion to judge the trust achievement of explanation-based recommendation interfaces. In our user study, it was assessed by two interrelated post-questions (still using the 5-point Likert scale), which asked participants, positively then negatively, about their genuine intention to use the interface again for future shopping (see Table 3).

**Table 3. Intention-to-return construct**

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

The results showed that most of participants had stronger intention of returning to the organization-based explanation interface in the future, than the simple "why" list view. The difference in overall mean value proved to be highly significant (mean=3.27, SD=1.11

for the organization vs. mean=2.67, SD=1.24 for the "why" interface, t=4.58, $p<0.001$, see Figure 3; median=3.5 vs. 2.5; mode=4 vs. 1).

## 5.3 Intention to Save Effort

### *5.3.1 Perceived Cognitive Effort*

The cognitive effort refers to the psychological costs users perceived to obtain and process information that enable them to arrive at a decision. Like the other constructs, it was also made up of two items (or questions) respectively responded on a 5-point Likert scale (see Table 4 for the items and their mean responses).

**Table 4. Cognitive-effort construct**

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

The lower mean rate represents a less cognitive cost the average user experienced during the interaction with the corresponding interface. As a result, the overall cognitive effort was perceived significantly lower (t=-3.89, $p<0.001$) on the organization-based explanation interface (mean=2.54, SD=0.96, vs. mean=3.10, SD=1.13 for the "why" interface, see Figure 3; median=2.5 vs. 3; mode=2 vs. 3.5).

### *5.3.2 Actual Completion Time*

The completion time was defined as the amount of time a participant accomplished the task of locating a desired product in the interface. No significant difference was found between the two interfaces in terms of task completion time (mean=2.60 minutes, SD=1.74 vs. mean=2.62 minutes, SD=1.67 for the organization interface, t=0.13, $p=0.45$). Users took slightly less time to complete the task using the organization interface when compared by the median time (median=2.13 vs. 2.18 minutes for the "why" interface).

## 5.4 Discussion

Further investigating the correlation between the above three trust constructs (see Table 5), we found that the perceived competence is actually highly positively correlated with the trusting intention to return and save cognitive effort ($p<0.001$). This suggests an important concept: if users perceive an interface to be more competent, they are more willing to return to it for more product recommendations and are also more likely to save their cognitive effort consumed on the interface. The actual completion time, however, has no significant correlation with the other variables ($p>0.1$). Thus, even though less task time is spent on the interface, it does not predict that users perceive less cognitive effort and have the intention to use it again.

From users' comments, the reasons that the organization interface was subjectively preferred to the simple "why" list by the majority of participants are quite clear. As a matter of fact, many users considered it well structured and easier to use for comparing products from different categories or in one category. Some users thought it was a little surprising at the beginning, but they soon got used to it and found it to be useful. It was also accepted as a good idea to label each category to distinguish it from others. In another word, the grouping allowed them to locate a product matching their needs more quickly than the ungrouped display.

**Table 5. Correlations among perceived competence, intention to return, intention to save cognitive effort and completion time (Pearson Correlation)**

| | Perceived Competence | Intention to Return | Cognitive Effort | Completion Time |
|---|---|---|---|---|
| Perceived Competence | 1 | .778** (.000) | -.826 ** (.000) | -.018 (.830) |
| Intention to Return | .778** (.000) | 1 | -.675** (.000) | -.042 (.619) |
| Cognitive Effort | -.826 ** (.000) | -.675** (.000) | 1 | .069 (.414) |
| Completion Time | -.018 (.830) | -.042 (.619) | .069 (.414) | 1 |

** Correlation is significant at the 0.01 level (2-tailed).

# 6. IMPLICATION TO RELATED WORK

Results from our empirical study strongly support a current trend in displaying a diverse set of recommendations rather than the k-best matching ones. McGinty and Smyth [12] maintain that showing diverse items can reduce the recommendation cycles. McSherry [14] advocates that the displayed items should cover all possible tradeoffs that the user may be prepared to accept. Faltings et al [6] propose to show products that can be potentially acceptable to users had they stated all of their preferences. In the same spirit, Price and Messinger [17] propose to generate the display set taking into account users' preference uncertainty. Our work demonstrates that displaying a diverse set of results in an organization-based interface more effectively enables users' trust formation compared to the simple k-best interface even after the "why" enhancement. We believe that similar trust-related benefits can be obtained for these diversity-driven interfaces proposed by other researchers in this field.

# 7. CONCLUSION

Based on our recent work on the development of a trust model for recommender agents, we have shown that explanation interfaces have the greatest potential to build a competence-inspired trust relationship with its users. A carefully designed survey indicated that a recommender agent's competence is positively correlated with users' intention to return, but not necessarily with their intention to purchase. It also showed that an organization-based explanation interface is likely to be more effective than the simple "why" interface, since most participants felt that it would be easier for them to compare different products and make a quicker decision.

We proposed a set of five principles for the design of organization interfaces and an algorithm for generating the content of such interfaces. We reported a significant-scale comparative study to further quantify users' trust formation and trusting intentions. Results show that the organization interface significantly increases users' perception of the interface's competence, resulting in their higher intention to use the interface again and save their cognitive effort. Moreover, we found that the actual time spent looking for a product did not have significant impact on users' subjective emotions. This indicates that less time spent on the interface does not predict that users would subjectively experience a smaller

amount of decision effort, nor does it predict that users will form more intention to return to the website.

## 8. REFERENCES


[1] Agrawal, R., Imielinski, T. and Swami, A. Mining association rules between sets of tems in large databases. In *Proceedings of the International ACM SIGMOD Conference*, Washington DC, USA, 1993, 207–216.

[2] Armengol, E., Palaudàries, A., Plaza, E. Individual prognosis of diabetes long-term risks: a CBR approach. *Methods of Information in Medicine* 40**,** 2001, 46-51.

[3] Burke, R., Hammond, K. and Young, B. The FindMe approach to assisted browsing. *Journal of IEEE Expert*, 12(4), 1997, 32–40.

[4] Carenini, G. and Moore, J. Multimedia explanations in IDEA decision support system. *Working Notes of the AAAI Spring Symposium on Interactive and Mixed-Initiative Decision Theoretic Systems*, 1998.

[5] Chen, L. and Pu, P. Trust building in recommender agents. In *Proceedings of the Workshop on Web Personalization, Recommender Systems and Intelligent User Interfaces at the 2nd International Conference on E-Business and Telecommunication Networks (ICETE'02)*, 2005.

[6] Faltings, B., Pu, P., Torrens, M. and Viappiani, P. Designing example-critiquing interaction. *International Conference on Intelligent User Interfaces (IUI'04)*, 2004, 22-29.

[7] Grabner-Kräuter, S. and Kaluscha, E.A. Empirical research in on-line trust: a review and critical assessment. *International Journal of Human-Computer Studies 58*, 2003.

[8] Herlocker, J.L., Konstan, J.A. and Riedl, J. Explaining collaborative filtering recommendations. In *ACM Conference on Computer Supported Cooperative Work*, 2000.

[9] Klein, D.A. and Shortliffe, E.H. A framework for explaining decision-theoretic advice. *Artificial Intelligence* 67, 1994, 201-243.

[10] McCarthy, K., Reilly, J., McGinty, L. and Smyth, B. Thinking positively – explanatory feedback for conversational recommender systems. In *Proceedings of the Workshop on Explanation in CBR at the Seventh European Conference on Case-Based Reasoning (ECCB'04)*, 2004, 115-124.

[11] McCarthy, K., Reilly, J., McGinty, L. and Smyth, B. Experiments in dynamic critiquing. In *Proceedings of the International Conference on Intelligent User Interfaces (IUI'05)*, 2005, 175-182.

[12] McGinty, L. and Smyth, B. On the role of diversity in conversational recommender systems. In *Proceedings of the Fifth International Conference on Case-Based Reasoning (ICCBR'03)*, 2003, 276-290.

[13] McKnight, D.H., and Chervany, N.L. What trust means in e-commerce customer relationships: conceptual typology. *International Journal of Electronic Commerce*, 2002, 35-59.

[14] McSherry, D. Similarity and compromise. In *Proceedings of the International Conference on Case-Based Reasoning Research and Development (ICCBR'03),* 2003, 291-305.

[15] McSherry, D. Explanation in recommender systems. In *Workshop Proceedings of the 7th European Conference on Case-Based Reasoning*, 2004, 125-134.

[16] O'Donovan, J. and Smyth, B. Trust in recommender systems. In *Proceedings of the 10th International Conference on Intelligent User Interfaces (IUI'05)*, 2005, 167–174.

[17] Price, B. and Messinger, P.R. Optimal recommendation sets: covering uncertainty over user preferences. In *National Conference on Artificial Intelligence (AAAI'05)*, 2005.

[18] Pu, P. and Chen, L. Integrating tradeoff support in product search tools for e-commerce sites. In *Proceeding of the ACM Conference on Electronic Commerce (EC'05)*, 2005, 269-278.

[19] Pu, P., Faltings, B. and Torrens, M. Effective interaction principles for online product search environments. In *Proceedings of the IEEE/WIC/ACM International Joint Conference on Intelligent Agent Technology and Web Intelligence,* 2004, 724-727.

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

[21] Shimazu, H. ExpertClerk: a conversational case-based reasoning tool for developing salesclerk agents in e-commerce webshops. *Artificial Intelligence Review* 18, 2002, 223-244.

[22] Sinha, R. and Swearingen, K. The role of transparency in recommender Systems. In *Extended Abstracts of Conference on Human Factors in Computing Systems (CHI'02)*, 2002.

[23] Sørmo, F. and Aamodt, A. Knowledge communication and CBR. In *Proceedings of the ECCBR-02 Workshop on Case-Based Reasoning for Education and Training*, 2002, 47-59.

[24] Ziegler, C.N., McNee, S.M., Konstan, J.A. and Lausen, G. Improving recommendation lists through topic diversification. In *Proceedings of the 14th International World Wide Web Conference (WWW'05*), 2005.