Accounting for Taste: Using Profile Similarity to Improve Recommender Systems

Philip Bonhard†, Clare Harries‡, John McCarthy†, M. Angela Sasse†

†Department of Computer Science
University College London
Gower Street
London WC1E 6BT, UK
{p.bonhard, j.mccarthy, a.sasse}@cs.ucl.ac.uk

‡Department of Psychology
University College London
Gower Street
London WC1E 6BT, UK
c.harries@ucl.ac.uk

ABSTRACT

Recommender systems have been developed to address the abundance of choice we face in taste domains (films, music, restaurants) when shopping or going out. However, consumers currently struggle to evaluate the appropriateness of recommendations offered. With collaborative filtering, recommendations are based on people's ratings of items. In this paper, we propose that the usefulness of recommender systems can be improved by including more information about recommenders. We conducted a laboratory online experiment with 100 participants simulating a movie recommender system to determine how *familiarity* of the recommender, *profile similarity* between decision-maker and recommender, and *rating overlap* with a particular recommender influence the choices of decision-makers in such a context. While *familiarity* in this experiment did not affect the participants' choices, *profile similarity* and *rating overlap* had a significant influence. These results help us understand the decision-making processes in an online context and form the basis for user-centered social recommender system design.

Author Keywords

Recommender systems, online advice-seeking, social networking, decision-making

ACM Classification Keywords

H.5.3. Information interfaces and presentation (e.g., HCI): Group and Organization Interfaces - Collaborative Computing.

INTRODUCTION

Recommender systems (RS) have been developed as a solution to the abundance of choice people face in many situations today. In effect, choice has become a burden [19]. Although recommender systems aim to filter choices according to the consumers' preferences, they usually do not provide any or very limited explanations as to where these filtered results come from.

For instance, consider the following scenario. Mary has just come home from a long week of work and wants to relax at home with a DVD from her local DVD rental shop. Unfortunately, she does not have a clear idea of what she wants to watch, and her DVD shop has about 3000 titles on offer from all kinds of genres, directors and years. Normally she would ask her best friend Jane for a recommendation, because Jane likes the same films and goes to the cinema more often. But what if Jane is not around? Mary's DVD shop offers a movie recommender system that asks her to rate a couple of films and then generates recommendations based on that. Mary rates 40 films, which takes her about 10 minutes, and in the end receives … 316 recommendations!

Helping consumers make choices – are we helping?

Collaborative filter RS seek to solve the problem of choice overload by emulating the social process of seeking recommendations from friends, while drawing on the data in a large user database [16]. But merely providing filtered lists of available choices is not the same as receiving a recommendation from a friend [2]. Without any explanation or justification, the user of a RS struggles to convert this data into meaningful information. This is a problem a RS should solve rather than create. Similarly, since people inherently know which of their friends to trust for a particular recommendation, RS should imitate this naturally occurring evaluation and decision-making process. It is therefore necessary to understand how people arrive at their decisions in such domains and which recommender characteristics are important to them.

The background section of this paper presents a brief overview of RS. It then examines the psychology literature of decision-making and advice-seeking and how it relates to RS design, which forms the basis for our

CHI 2006, April 22–28, 2006, Montréal, Québec, Canada.

experiment. The recommender experiment section then details the experimental method and procedure of our study with 100 participants. Our results show that profile similarity, in combination with rating overlap, had a significant impact on the participants' choices. We proceed to use these results to show how user profile similarity and rating overlap visualizations can aid decision-making and present how basing recommendation explanations on profiles can help consumers judge the appropriateness of recommendations they receive. This is revelatory about online decision-making processes and opens interesting new research avenues for social system design.

BACKGROUND

RS have been deployed in various e-commerce contexts, such as book or music shopping like Amazon (*www.amazon.com*), general rating sites (*www.ratingzone.com),* and specific rating sites such as MovieLens (*http://movielens.umn.edu*).

RS aggregate the information received and redirect it to the appropriate recipients. Depending on the recommendation strategies or algorithms employed, users are presented with options based on their personal preference profiles identified through ratings, keywords, usage history, content analysis, item to item or user to user correlation, to name just a few.

While many strategies for computing recommendations have been explored such as item-based collaborative filtering [18], Bayesian networks [3] and factor analysis [4], user-user collaborative filtering (CF) comes closest to emulating real world recommendations because they are based on user rather than item matching. Recommendations are generated for a given user by comparing her existing ratings to those of all other users in the database. In doing so, a neighborhood of similar users is established, and based on that rating predictions are computed for items that the user has not yet rated, but her closest neighbors have.

The assumption behind this strategy is that, if Mary and John have rated several items in a similar fashion, chances are that they have a taste overlap. Thus items that John has rated highly, but which Mary has not yet acquired might be of interest to her.

RS research to date has predominantly focused on designing algorithms for more effective and efficient computation of rating predictions [3,9,12,17]. Precision effectiveness is tested through existing rating data sets, where part of the rating set is deleted, and the prediction results from algorithms are compared against the real ratings. Prediction efficiency is concerned with the computational cost in terms of time and resources for calculating these predictions. Even trust has been examined from an algorithmic perspective [11,13]. HCI approaches to RS research have been limited to examining existing RS in order to establish interaction design guidelines [10,20,21].

The research approach presented in this paper, however, is to take a step back from the above evaluation approaches. Instead, we examine the psychology literature of advice-seeking and decision-making to establish the purpose and process of seeking recommendations as a basis for designing RS.

Psychology of decision-making & advice-seeking

The advice-seeking literature to date has predominantly focused on objective domains where advice can be classified as right or wrong [8,22,23,24]. Generally, advice-seeking is seen as a problem of weighting of different information sources to come to a final conclusion. One of the key questions is how decision makers arrive at their weightings. The classic approach to this problem argues that *"decision makers are exquisitely rational beings in solving judgment and choice problems"* [14]. Studies have shown that advice-seekers tend to place a greater weight on their own opinion than that of the advisor (*egocentric discounting*) [24]. The assumption is that advice-seekers can assess what they know and the strength of their opinions, which they are not able to do for the opinions of others. Related to that is that the confidence in one's opinion depends on how much evidence can be retrieved that supports it [22].

Whereas the benefit of combining opinions in matters of fact is well understood, the same is not true for matters of taste [22]. Simple aggregation of opinions in taste domains raises conceptual difficulties as people are entitled to different opinions on movies, music or restaurants. Yaniv [22] argues that the most promising strategy to examine these factors in taste domains is to consider the *personal match* between decision maker and recommender, assuming that the greater the similarity between them, the greater the impact and benefit of advice received.

Advice-seeking in taste domains

A recent qualitative study into advice-seeking in taste domains [2] has shown that decision-makers tend to consult familiar advisors, for a number of reasons. Applied to the scenario in the introduction, these reasons are:

1. Mary knows where her tastes overlap with those of Jane, and where not. Therefore she can judge whether Jane can give appropriate advice or not.
2. Mary might have previously received good advice from Jane. If this is the case,
3. … the risk of receiving bad advice is already reduced for Mary.

4. Mary does not have to expend a lot of effort in evaluating the advice Jane gives her because of the above. She can simply *rely* on it.
5. Even if Jane did not like the same movies as Mary, being a close friend of Mary's, Jane would probably know Mary's tastes sufficiently well to be able to give her good advice.

Bonhard & Sasse [2] suggest that matching people according to their profiles in terms of hobbies and interests - rather than item ratings or user neighborhood-statistical data alone – and explaining recommendations terms of similarity, would make it easier for users to judge the appropriateness of a recommendation.

So you're like me? Principles of attraction in first-time encounters

Social systems like Friendster (*www.friendster.com*) or Orkut (*www.orkut.com*) are based on people revealing information about themselves such as demographic data, hobbies and interests. RS, on the other hand, almost exclusively use item ratings to define the user profile for the system.

We propose that RS can be made more effective and usable by appropriating some functionality from such social systems. Perugini *et al*. pointed out that RS have an inherently social element and ultimately bring people together [15]. They believe that in a brick-and-mortar setting, the process of recommendation is intrinsically dependent on a mutual knowledge of each others' tastes. A mutually reinforcing dynamic ensues where the recommender's knowledge of the decision maker's tastes are incorporated into the recommendation process. On the other hand, the decision-maker's personal knowledge and past experience of the advice from a particular recommender helps her evaluate the appropriateness of the recommendation. RS therefore need to model the user, which involves the representation of her preferences and interests.

Social psychology has shown that people like others who are, among other things, familiar, similar to themselves and with whom they have a history of interaction [1]. Gigerenzer also suggested that when choosing from multiple options people generally prefer familiar ones (*recognition heuristic*) [7]. Cosley *et al*. examined similarity in online task-focused interactions [5], in an experiment that simulated a TV quiz show (Family Feud). They found that participants preferred to cooperate with people with a similar demographic background, yet interest similarity did not have an effect on cooperation. They suggest that a social recommender system should take demographic data in combination with the interest profile into account, when generating recommendations.

It is worth noting that Cosley *et al*.'s experimental scenario was task-focused, as opposed to an advice-seeking situation, where there might not even be any direct communication between the decision-maker and the advisor/recommender. The aim of the study presented in this paper was to represent similarity in a RS context, and see if *familiarity* and visualizations of *profile similarity* and *rating overlap* would influence the choice of recommendations and their perceived usefulness.

RECOMMENDER EXPERIMENT

The experiment aimed to explore what makes people trust recommenders in an online context. More specifically, what combination of *familiarity*, *profile similarity* and *rating overlap* would have an influence on the choices people make in a RS context? Would a visualization of *profile similarity* between the decision-maker and recommender influence the decision-maker's choice as suggested before [1,2]? Following Perugini *et al*.'s idea of modeling the user [15] and representing her preferences and interests, we aimed to visualize a recommender in a way to help the decision-maker judge the appropriateness of a recommendation.

Experiment Scenario

Since every participant would be different in terms of demographic data, interests and tastes, we had to create an experiment that would adapt to each individual participant, while conceptually remaining consistent for everyone. To do this we devised a film festival scenario where participants receive fictitious movie recommendations from recommenders (generated on the fly) that were *familiar* or *unfamiliar*, *similar* or *dissimilar*, and either had the *same or different film tastes* (in terms of rating overlap). The individual characteristics of the recommenders were adapted on the fly to each participant.

Participants & Procedure

A total of 100 participants completed the study, which lasted 30-45 minutes, in a computer lab. The age range varied from 18-44 with a variety of backgrounds including students and professionals.

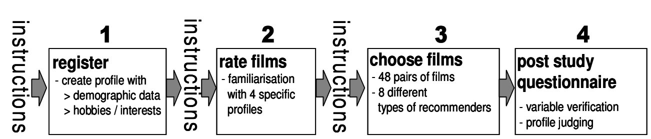

Figure 1 - Experiment overview

Each participant encountered 4 phases (see Figure 0 for an overview). In the first phase, each participant was required to provide a basic profile, consisting of demographic data such as age, gender and profession, but also preferences and interests such as preferred film genres, hobbies, leisure activities and music tastes. In phase 2, participants rated 20-30 films, which would serve as a basis for generating their recommendations.

In phase 3, participants chose from a series of 48 pairs of films recommended by people who – the scenario made out - had already seen those films.

Phase 4 consisted of a post study questionnaire, where participants rated some of the profiles they had previously seen in terms of *familiarity, profile similarity, rating overlap* and *trust*., and provided qualitative data (in text fields) about the decision reasoning.

Independent variables – recommender profile characteristics

Familiarity

Considering that people consult known sources for recommendations, we aimed to simulate familiarity with a recommender through repeated exposure as suggested in social psychology [1] and the advice-seeking and decision-making literature [7]. Exposing participants to a limited number of profiles before actually receiving recommendations simulates the process of *getting to know* that particular profile in relation to one's own. This variable had two levels, *familiar* and *unfamiliar*. Thus during the rating phase 2 of the experiment (phase 2, see Figure 0), after each rating, participants would see one of four "buddy" profiles and be told whether these buddies had rated those films the same way or not (see Figure 2 - Profile familiarization).

Figure 2 - Profile familiarization

Profile similarity

Profile similarity was based on the demographic data, film genre preferences and interests and had two levels, *similar* or *dissimilar*. Thus a similar profile would be the same gender, similar age (+-2 years), same profession and have a significant overlap in their film genre preferences, hobbies and interests. *Profile similarity* was visualized through highlighting of the interests or preferences the participant and recommender had in common (see Figure 2).

Figure 3 – Similar profile with high rating overlap (text visualization)

A dissimilar profile on the other would show little or no overlap in terms of interests and preferences and significantly differ in terms of demographic data.

Rating Overlap

Rating overlap was based on the film ratings participants had previously supplied (phase 2). Thus recommenders could either have *high* or *low rating overlap*, which was visualized in one of two ways (consistent for each participant).

The *text visualization*, similar to the *profile similarity* visualization, showed explicitly highlighted, which films the participant and the recommender had rated in a similar fashion (see Figure 2).

The *symbol visualization* showed a Venn diagram with two overlapping circles. A large middle circle would represent a large *rating overlap* (see Figure 6) and a small middle circle would represent a small *rating overlap*.

Film Recommender Profiles

With these three independent variables, each with two levels (e.g. familiar *vs.* unfamiliar) this experiment was a 2x2x2 design. All possible combinations of these resulted in eight recommender profiles outlined below:

Profile 1: - familiar - similar prof. - high RO	**Profile 2:** - unfamiliar - similar prof. - high RO	**Profile 3:** - familiar - similar prof. - low RO	**Profile 4:** - unfamiliar - similar prof. - low RO
Profile 5: - familiar - dissimilar - high RO	**Profile 6:** - unfamiliar - dissimilar - high RO	**Profile 7:** - familiar - dissimilar - low RO	**Profile 8:** - unfamiliar - dissimilar - low RO

Figure 4 - Recommender profiles

Conditions

There were three condition types (*familiarity*, *profile similarity* & *rating overlap*). Within each of these there were four pairs of choices, where the recommenders differed in each of the independent variables while keeping the other two constant, resulting in 12 different conditions on total (see Figure 5 below).

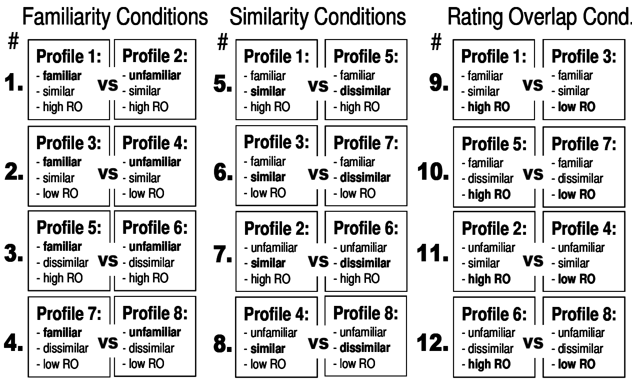

Figure 5 - Experiment conditions

Each individual pair was repeated four times to allow within-participant consistency checking and position counter balancing (left / right – see Figure 6). Thus, there were a total of 48 recommendation pairs. Participants saw every recommender profile the same number of times (12).

Figure 6 below shows an example of a choice between two recommenders that are both familiar (seen during the rating phase 2) and have similar profiles, but differ in rating overlap.

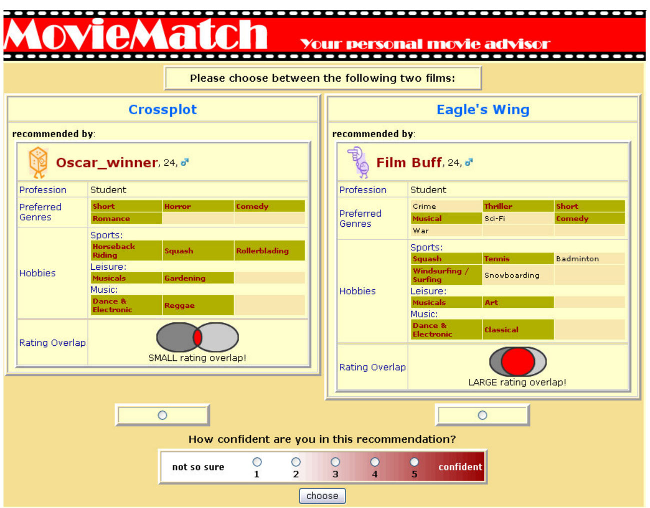

Figure 6 - Film recommendation

Dependent variable – choice

In phase 3 (phase 3 – see Figure 0), participants saw the titles or completely fictitious films and profile information of the recommender. The reason for not providing any fictitious reviews or synopses was to minimize noise through a possible participant bias towards any film based on the film's properties.

Each recommendation pair was a forced choice, thus the dependent variable was which film the participant would choose and in turn which recommender they would trust.

In addition to the actual choice, we recorded a confidence rating on Likert scale from 1 to 5 for each choice.

From the recommendation session (phase 3), the resulting data consisted of the number identifying the profile chosen, a confidence rating for each choice, as well as the following values:

- **0** = if the participant chose a familiar, similar or high rating overlap recommender
- **1** = if the participant chose an unfamiliar, dissimilar or low rating overlap recommender

Since each condition was repeated four times, we were able to conduct a within participant consistency check. The average (*consistency average*) of those four values allowed us to see if the participant consistently chose a particular recommender:

- **Average = 0 - 0.25** Participant tended to select familiar, similar profile or high rating overlap recommenders
- **Average = 0.5** Participant chose recommenders randomly
- **Average = 0.75 – 1.00** Participant tended to select unfamiliar, dissimilar profile or low rating overlap recommenders.

Further, we examined the ratio of number of times a particular profile was chosen and the number of times a profile was seen (*choice ratio*).

Hypothesis & Analysis

Hypothesis 1 – Familiar recommenders will be preferred
We wanted to test the prediction of [2] that in taste domains, people prefer recommendations from familiar over those from unfamiliar recommenders [2]. This is also in line with the idea in social psychology, that familiar people are judged to be safe and unlikely to cause harm [1].

Even though participants would not be actual friends with the recommenders they encountered during the phase 2, the idea was that merely having been exposed to their profiles and film ratings might have an influence on the participants' choices.

Hypothesis 2 – Similar recommenders will be preferred
We wanted to examine if, in an online RS context, similarity in terms of demographic data, interests and hobbies would result in decision makers preferring films from recommenders who are like them.

While Cosley *et al.* examined similarity in a task-focused online context [5], we were interested in an advice-seeking context, notably in a taste domain like films. The decision-making process is different in such a context because the advice taken does not require any expertise from the recommender, but the right match between advice seeker and recommender. This match is based on *profile similarity* and *rating overlap*.

Hypothesis 3 – Recommenders with high rating overlap will be preferred

While this is the underlying assumption behind user collaborative filter RS, it has not yet been tested as a separate independent variable in a quantitative fashion. Our aim therefore was to confirm that decision-makers in taste domains actually do prefer recommendations from people with a high rating overlap, given other profile characteristics such as familiarity and similarity.

Analysis

We analyzed the data from phase 2 in three ways:

1. The choice ratios were analyzed using a 2x2x2 repeated measure ANOVA.
2. The profile choices for each of the three condition types, were analyzed separately with an independent one sample t-test on the within participant *consistency average* (see Figure 5 above)
3. The confidence ratings were analyzed within each condition type with a 2x2 repeated measures ANOVA.

RESULTS

Overall Profiles Chosen

Figure 7 below shows the overall trend of *choice ratios*, as well as the how these profiles were rated in the end in terms of trust.

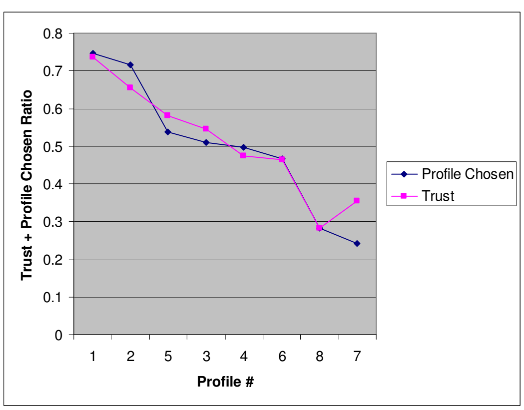

Figure 7 - Profile choice ratios / Trust rating

Trust was elicited through a Likert scale (1-5) in phase 4 in the post-study questionnaire. It is interesting to note the trend in above graph. When mapped against profile characteristics, it points to a significant influence of profile similarity and rating overlap. Figure 8 below shows the above trend mapped out with the individual profile characteristics.

1	2	5	3	4	6	8	7
F	UF	F	F	UF	UF	UF	F
SP	*SP*	*DSP*	*SP*	*SP*	*DSP*	*DSP*	*DSP*
HRO	HRO	HRO	LRO	LRO	HRO	LRO	LRO

F = familiar
UF = unfamiliar
SP = similar profile
DSP = dissimilar profile
HRO = high rating overlap
LRO = low rating overlap

Figure 8 - Chosen profiles with ratios & characteristics

Overall *profile similarity* and *high rating overlap* had an effect on the choice of which recommender to trust.

Participants rated the same profiles in terms of *trust* in phase 4 of the experiment. Interestingly, the trend for the *trust ratings* is the same as the one for the profiles chosen, except for the profiles 7 and 8. Here participants trusted the familiar profile more, whereas they chose the unfamiliar one more often.

Individual Variable Analysis

Since in each condition recommenders differed in one characteristic, while the others were kept constant it is worth examining the independent variables individually within each condition type. The results are summarized in Figure 9 below.

#	Familiarity Conditions		#	Similarity Conditions		#	Rating Overlap Cond.	
1.	Profile 1: 55.75% familiar	vs Profile 2: 44.25% unfamiliar	5.	Profile 1: 83.75% similar	vs Profile 5: 16.25% dissimilar	9.	Profile 1: 84.5% high RO	vs Profile 3: 15.5% low RO
2.	Profile 3: 50% familiar	vs Profile 4: 50% unfamiliar	6.	Profile 3: 87.5% similar	vs Profile 7: 12.5% dissimilar	10.	Profile 5: 88.5% high RO	vs Profile 7: 11.5% low RO
3.	Profile 5: 56.75% familiar	vs Profile 6: 43.25% unfamiliar	7.	Profile 2: 84.75% similar	vs Profile 6: 15.25% dissimilar	11.	Profile 2: 85.5% high RO	vs Profile 4: 14.5% low RO
4.	Profile 7: 48.75% familiar	vs Profile 8: 51.25% unfamiliar	8.	Profile 4: 84.75% similar	vs Profile 8: 15.25% dissimilar	12.	Profile 6: 81.5% high RO	vs Profile 8: 18.5% low RO

Figure 9 - Profiles chosen – individual percentages per condition

Familiarity

In ANOVA analysis the effect of *familiarity* in this experiment was not statistically significant ($F(1,99)=3.85$, $p > 0.05$). The t-tests on the consistency averages revealed that only condition 3 was significant ($t= -2.0152$, $p= .047$). All other conditions were not statistically significant ($p > .05$). Thus overall our hypothesis 1 was not supported, which is in contrast to Bonhard & Sasse's predictions [2].

Profile Similarity

Profile similarity had a significant impact on the recommender choice both in considering consistency average t-tests (all $p < 0.01$, t (cond 5)= -13.38, t (cond 6) = -17.23, t(cond 7) = -16.1, t(cond 8) = -14.28) as well as the ANOVA analysis of the overall profile chosen ratios

(F (1,99) = 474.512, p < 0.01). Hypothesis 2 was therefore supported.

Rating Overlap

Rating overlap also had a significant impact on the recommender choice as revealed in the consistency average t-tests (all p <.01, t(cond 9) = -14.35, t(cond 10) = -17.04, t(cond 11) = -15.57, t(cond 12) = -13.27) and the ANOVA (F (1,99) = 405.305), p<0.01). Thus participants generally chose films from recommenders with whom they had a high rating overlap, which supports hypothesis 3.

Examining the different visualizations of *rating overlap*, *text* & *symbol,* it is worth noting that there was a higher tendency to choose films from recommenders with high rating overlap in the symbol conditions (87.6 %) than in the text conditions (82.5 %).

Interaction effects

Interestingly, the ANOVA analysis revealed an interesting interaction between familiarity and rating overlap. When rating overlap was high there was a small but significant effect of familiarity (F (1,99) = 7.511), p<0.01). No other interaction effects were statistically significant.

Choice Confidence

For every choice participants provided a confidence rating on a Likert scale from 1 to 5. We considered a significant deviation of the mean from 3 (p<0.05) as examinable. With the exception of condition 10, all rating means significantly deviated from 3 (see Figure 10).

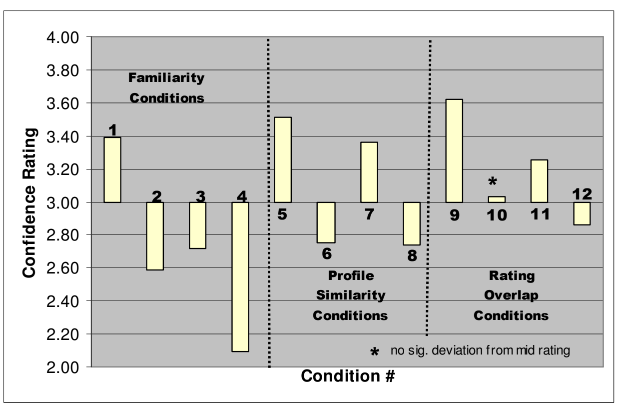

Figure 10 – Choice confidence rating means (across subjects & independent of choice made)

Familiarity conditions

The only condition where participants were confident in their choices was condition 1, where the choice was between a familiar and unfamiliar recommender that was similar and had high rating overlap with the participant.

Profile similarity conditions

In those conditions it was interesting to observe that participants were more confident in their choices in conditions 5 and 7 where recommenders had a *high rating overlap*, than in conditions 6 and 8 where recommenders had a *low rating overlap* with the participant. The ANOVA analysis of the confidence ratings confirmed a significant influence of the *rating overlap* variable (F(1,399)=236.966, p<0.01).

Rating Overlap Conditions

Similar to the *profile similarity* conditions, participants were more confident in their choices in conditions 9 and 11, where the recommenders had a *similar profile*, compared to conditions 10 and 12, where the recommender profile was *dissimilar*. The ANOVA analysis of the confidence ratings confirmed a significant influence of *profile similarity* on the confidence ratings (F(1,399)=192.698, p<0.01).

Post-study questionnaire

Verifying the effectiveness of the variable implementation

In phase 4 participants were asked to rate profiles 1 through 8 on the profile characteristics *familiarity*, *profile similarity*, *rating overlap* and *trust* on a Likert scale from 1-5 (e.g. "*How similar do you think you are to this recommender*?") The aim was to check whether participants actually understood that they are *familiar/unfamiliar* with, *similar/dissimilar* to and have a *high* or *low rating overlap* with a particular recommender.

Examining the ratio of the difference of mean ratings over the maximum rating (=5) for the two recommender profiles for each condition was indicative of whether the variables were understood correctly. The differences for each variable in each condition type significantly differed from 0 (p<0.01), which means that participants understood the different characteristics correctly.

Recommender characteristics

We also asked participants to rate how important the following are (Likert scale, 1-5) when making a movie choice decision in such an online recommendation context as well as when they are seeking advice from friends (see Figure 11).

1. Film title,
2. Recommender
3. Age
4. Gender
5. Profession
6. Hobbies
7. Music tastes
8. Rating / film taste overlap

All the ratings significantly deviated from the mid rating (3, p<0.05) except for the importance of the *film title* in the online context (mean = 3.2, p>0.05). For the real-world context, the film title was rated as important (mean=3.29).

In both contexts, participants placed importance on *who* recommends the film. *Age*, *gender* and *profession* were not rated as particularly important.

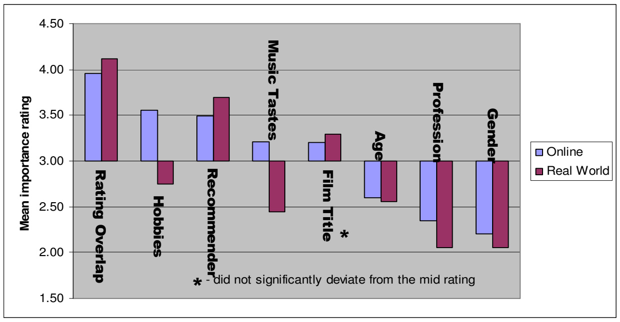

Figure 11 - Importance ratings for recommender characteristics (ordered by effect size)

In the online context, *hobbies* were rated as more important (mean = 3.55) than in a real-world context where they were not important (mean = 2.75, sig, difference, $p < 0.01$). This observation is the same for *music tastes* (online mean = 3.21, real world mean = 2.45, sig, difference, $p < 0.01$). *Rating/film taste overlap* were rated important in both contexts.

Qualitative data

Participants could also include comments about reasons for their decisions in this online context, as well as when seeking recommendations in the real world. A total of 151 comments were examined using content analysis [6].

In an online context, while there was a universal consensus that rating and film taste overlap was very important; it was revelatory to see that a large number of participants judged *profile similarity* as an important factor in their decision-making. Many of those actually mentioned profile similarity and rating overlap together as an important factor ("*… similarity of recommender in terms of hobbies and other films we both enjoyed.*").

In a real-world context, participants stated that it mattered how well they knew the recommender ("*… how close a recommender is as a friend.*", "*… how close this person is to me, e.g. my boyfriend would influence me more than a distant acquaintance.")*. Further, whether participants had had past experience with recommendations from a particular person played an important role ("*… the recommender's history of good or bad recommendations.*").

DISCUSSION

The results provide a number of insights into decision-making processes in taste domains in an online context. Our hypotheses and the results were as follows:

1. **Familiar recommenders will be preferred** – this hypothesis was not supported. Participants did not prefer to choose films recommended by people they had seen before
2. **Similar recommenders will be preferred** – this was overwhelmingly supported. Participants predominantly chose films from recommenders with *similar profiles*. This was also confirmed in the choice confidence ratings, the post study questionnaire ratings and qualitative data.
3. **Recommenders with high rating overlap will be preferred** – this hypothesis was also supported. Participants predominantly chose films recommended by people with whom they had a *large rating overlap*. This was also confirmed in the post study questionnaire ratings and qualitative data.

Thus, combinations of different profile characteristics had a definitive influence on participants' choices, which lends itself to a new view on recommender systems design.

Social Recommender Systems – more than a database

Friends from whom we seek recommendations are not just a source of information for us: we know their tastes, views and they provide not only recommendations, but also justification and explanations for them. The results of our study suggest that RS should be more than intelligent databases that filter the information according to our taste profile. Seeking and receiving a recommendation is a social activity that often involves the discussion of a particular item. Why did the recommender like it? Would she want to experience/buy it again? Will the experience change after a while?

Revealing recommenders' profile information and supporting communication between users, brings back this social element. Thus, integrating recommender with social systems capitalizes on that in that the focus is not limited to recommending items, but also bringing like-minded people together. This idea goes back to Perugini *et al.*'s [15] idea of how user preferences and interests can be represented in such a system.

Familiarity

Though participants stated in the post study-questionnaire that their relation to the recommender is important in a real-world context, in our experiment, *familiarity* did not have an effect on the participants' choices. We believe that this was based on a failure to operationalize the variable successfully. Brief exposure to profiles in the early phases of the experiment was not sufficient to induce familiarity, even though these were repeated online encounters. With hindsight, to properly examine the effect of familiarity on decision-making in an online context, profiles of friends or people actually known to participants need to be used.

Profile Similarity

Interests and preferences and their visualizations, especially with regards to showing similarity are at the core of any social networking application. Yet RS have to date ignored the potential of visualization, and just displayed rating scores. Our results demonstrate that recommender characteristics, such as interests and preferences, and not just ratings, are important to decision-makers when choosing taste domain items. Both the confidence in their choices when *high rating overlap* combined with a *similar profile*, as well as the importance rating of hobbies and music tastes in an online context confirm that this information helps consumers judge the appropriateness of a recommendation. We are not arguing that profile similarity should replace ratings as the core of a recommendation strategy, but it has huge potential to complement rating functionality. Basing recommendation explanations on people, rather than items relations (as Amazon.com does) or rating statistics used by collaborative filtering, makes it easier for consumers to judge the appropriateness of a recommendation.

Rating Overlap

While rating overlap is the *raison d'etre* of collaborative filter-based RS, this is the first experiment to employ a visualization of rating overlap between a *single* decision maker and a *single* recommender.

We explored one symbol and one text visualization. While both had a significant effect on the participants' choices, the symbol visualization had a slightly stronger effect. In a social recommender system that allows more information to be revealed about a recommender, a decision-maker could choose to see more details about what films exactly she and the recommender agree on.

Our results show that *rating overlap* in combination with *profile similarity* can be a powerful source of information for a decision-maker when judging the validity of a recommendation. Further, these results have shown that participants were more confident in their choices when the recommender had a *high rating overlap* with them in combination with a *similar profile*. This was further confirmed by the post study questionnaire ratings as well as the qualitative data.

CONCLUSIONS

Based on the results of this study, we propose a new breed of RS, which integrates functionality from social systems such as Orkut (www.orkut.com) and Friendster (www.friendster.com) and classic RS.

Our results have shown that decision-makers trust recommenders more when they have not only *high rating overlap*, but also a *similar profile*. Finding like-minded people in these social systems at the moment is done manually, by browsing profiles in networks of friends and interest groups. In communities of thousands of users this can become quite cumbersome. RS already match users in order to generate recommendations, but they do not use this information in any other form. Revealing profile information of a particular user's neighborhood of similar users could not only help the user judge the appropriateness of recommendations, but also allow communication within that neighborhood. By doing so, a social recommender system would effectively bring back the inherently social element of the recommendation process as suggested by Perugini et al. [15].

To apply this to the scenario introduced in the introduction, if Mary and Jane were both registered with a social movie recommender system, Mary could easily find out about films that Jane has seen lately and liked. Not only that, but she could also find other people who might be similar to Jane and herself, and thus extend her network of sources of potentially useful recommendations.

Further research should investigate how much actual familiarity with the recommender (either in the *real world* or through repeated online interaction) influences the decision-making and trust. Which specific characteristics of a profile are particularly important to consumers and why? How can profile similarity be incorporated into the recommendation strategy? There is also the question of privacy, i.e. whether users of such systems would be willing to reveal personal information such as interests and preferences. Also, how can the combination of rating overlap and profile similarity information be used to combat malicious attacks on or manipulation of RS?

Accounting for taste therefore is not merely about considering item ratings. Since taste is an intricate phenomenon, we believe that incorporating profile similarity into social recommender system and algorithm design has the potential of improving the user experience of such systems. We especially hope that this paper will give algorithm designers useful input in order to tailor recommender algorithms to better suit users' decision making and advice seeking behavior. We encourage researchers to consider our results for more detailed explorations of *familiarity*, *profile similarity* and *rating overlap*, as they can be instructive on a theoretical and practical level for social recommender system design.

REFERENCES

1. E. Berscheid & H. T. Reis. Attraction and close relationships. D. T. Gilbert, Fiske, S. T, and Linzey, G., The Handbook of Social Psychology Vol. 2, 4th[22], Oxford University Press, (1998), p. 193-281

2. P. Bonhard & M. A. Sasse. I thought it was terrible and everyone else loved it" - A New Perspective for Effective Recommender System Design. Springer Verlag, In *Proc.*

of the 19th British HCI Group Annual Conference, Napier University, Edinburgh, UK 5-9 September 2005 (2005), p. 251-265

3. J. S. Breese, D. Heckerman, C. Kadie. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In *Proc. of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence* (1998), p. 43-52

4. J. F. Canny. Collaborative Filtering with Privacy via Factor Analysis. In *Proc. of SIGIR 2002, Tampere, Finland* (2002), p. 238-245

5. D. Cosley, P. J. Ludford, L. Terveen. Studying the effect of similarity in online task-focused interactions. ACM Press, In *Proc. of the 2003 international ACM SIGGROUP conference on Supporting group work* (2003), p. 321-329

6. U. Flick. An Introduction to Qualitative Research. London, Sage Publications, (1998)

7. G. Gigerenzer. Fast and Frugal Heuristics: The Tools of Bounded Rationality. N. Harvey & Koehler, Derek, Blackwell Handbook of Judgment and Decision Making, [4]Oxford, Blackwell, (2004), p. 62-88

8. N. Harvey & I. Fischer. Taking Advice: Accepting Help, Improving Judgement, and Sharing Responsibility. In *Organizational Behavior and Human Decision Processes* 70[2](1997), p. 117-133

9. J. L. Herlocker, J. A. Konstan, A. Borchers, J. Riedl. An Algorithmic Framework for Performing Collaborative Filtering. In *Proc. of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (1999), p. 230-237

10. J. L. Herlocker, J. A. Konstan, A. Borchers, J. Riedl. Explaining Collaborative Filtering Recommendations. In *Proc. of the ACM 2000 Conference on Computer Supported Cooperative Work* (2000), p. 241-250

11. P. Massa & P. Avesani. Trust-aware Collaborative Filtering for Recommender Systems. In *Proc. of The International Conference on Cooperative Information Systems (CoopIS), 25 - 29 October 2004, Larnaca, Cyprus* (2004),

12. M. R. McLaughlin & J. L. Herlocker. A collaborative filtering algorithm and evaluation metric that accurately model the user experience. In *Proc. of the 27th annual international ACM SIGIR conference on research and development in information retrieval* (2004), p. 329-336

13. J. O'Donovan & B. Smyth. Trust in Recommender Systems. In *Proc. of The International Conference on Intelligent User Interfaces, San Diego, California, Jan 9-12, 2005* (2005), p. 167-174

14. J. W. Payne, J. R. Bettman, M. F. Luce. Behavioural Decision Research: An Overview. Measurement, Judgement and Decision Making, [5]London, Academic Press, (1998), p. 303-359

15. S. Perugini, M. A. Goncalves, E. A. Fox. Recommender Systems Research: A Connection-Centric Survey. In *Journal of Intelligent Information Systems* 23[2](2004), p. 107-143

16. P. Resnick & H. R. Varian. Recommender Systems. In *Communications of the ACM* 40[3](1997), p. 56-58

17. B. M. Sarwar, G. Karypis, J. A. Konstan, J. Riedl. Analysis of Recommendation Algorithms for E-Commerce. In *Proc. of the ACM E-Commerce '00 Conference. Minneapolis, MN* (2000), p. 158-167

18. B. M. Sarwar, G. Karypis, J. A. Konstan, J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proc. of the Tenth International World Wide Web Conference* (2001), p. 285-295

19. B. Schwartz. The Paradox of Choice. Harper Collins, (2005)

20. K. Swearingen & R. Sinha. Beyond algorithms: An HCI perspective on recommender systems. In *Proc. of ACM SIGIR 2001 Workshop on Recommender Systems, New Orleans, Lousiana* (2001), p. 24-33

21. K. Swearingen & R. Sinha. Interaction design for recommender systems. http://www.rashmisinha.com/articles/musicDIS.pdf (last accessed 2006-1-15)

22. I. Yaniv. Receiving other peoples' advice: Influence and Benefit. In *Organizational Behavior and Human Decision Processes* 93[1](2004), p. 1-13

23. I. Yaniv. The Benefit of Additional Opinions. In *American Psychological Society* 13[2](2004), p. 75-78

24. I. Yaniv & E. Kleinberger. Advice Taking in Decision Making: Egocentric Discounting and Reputation Formation. In *Organizational Behavior and Human Decision Processes* 83[2](2000), p. 260-281