Interaction Design for Recommender Systems

Kirsten Swearingen, Rashmi Sinha
School of Information Management & Systems, University of California, Berkeley, CA[1]

ABSTRACT

Recommender systems act as personalized decision guides for users, aiding them in decision making about matters related to personal taste. Research has focused mostly on the algorithms that drive the system, with little understanding of design issues from the user's perspective. The goal of our research is to study users' interactions with recommender systems in order to develop general design guidelines. We have studied users' interactions with 11 online recommender systems. Our studies have highlighted the role of transparency (understanding of system logic), familiar recommendations, and information about recommended items in the user's interaction with the system. Our results also indicate that there are multiple models for successful recommender systems.

Keywords

Evaluation, Information Retrieval, Usability Studies, User Studies, World Wide Web

INTRODUCTION

In everyday life, people must often rely on incomplete information when deciding which books to read, movies to watch or music to purchase. When presented with a number of unfamiliar alternatives, people tend to seek out recommendations from friends or expert reviews in newspapers and magazines to aid them in decision-making. In recent years, online recommender systems have begun providing a technological proxy for this social recommendation process. Most recommender systems work by asking users to rate some sample items. Collaborative filtering algorithms, which often form the backbone of such systems, use this input to match the current user with others who share similar tastes. Recommender systems have gained increasing popularity on the web, both in research systems (e.g. GroupLens [1] and MovieLens [2]) and online commerce sites (e.g. Amazon.com and CDNow.com), that offer recommender systems as one way for consumers to find products they might like to purchase.

Typically the effectiveness of recommender systems has been indexed by statistical accuracy metrics such as Mean Absolute Error (MAE) [3]. However, satisfaction with a recommender system is only partly determined by the accuracy of the algorithm behind it [2]. What factors lead to satisfaction with a recommender system? What encourages users to reveal their tastes to online systems, and act upon the recommendations provided by such systems? While there is a lot of research on the accuracy of recommender system algorithms, there is little focus on interaction design for recommender systems.

To design an effective interaction, one must consider two questions: (1) what user needs are satisfied by interacting with the system; and (2) what specific system features lead to satisfaction of those needs. Our research studies have attempted to answer both of these questions. Below is a brief overview of our study methodology and main findings. Subsequently we discuss the results in greater detail and offer design guidelines based on those results.

OVERVIEW OF OUR RESEARCH PROGRAM

More than 20 different book, movie and music recommender systems are currently available online. Though the basic interaction paradigm is similar (user provides some input and the system processes that information to generate a list of recommendations), recommender systems differ in the specifics of the interaction (e.g., amount and type of input user is required to give, familiarity of recommendations, transparency of system logic, number of recommendations). Our approach has been to sample a variety of interaction models in order to identify best practices and generate guidelines for designers of Recommender Systems. For 6 of the 11 systems tested, we also compared user's liking for systems' recommendations with liking for recommendations provided by their friends.

[1] kirstens@sims.berkeley.edu, sinha@sims.berkeley.edu

Figure 1: User can choose between Online recommender systems and social recommendations (from friends)

Input from user

Output (Recommendations)

Social Recommendations

Online Recommender System
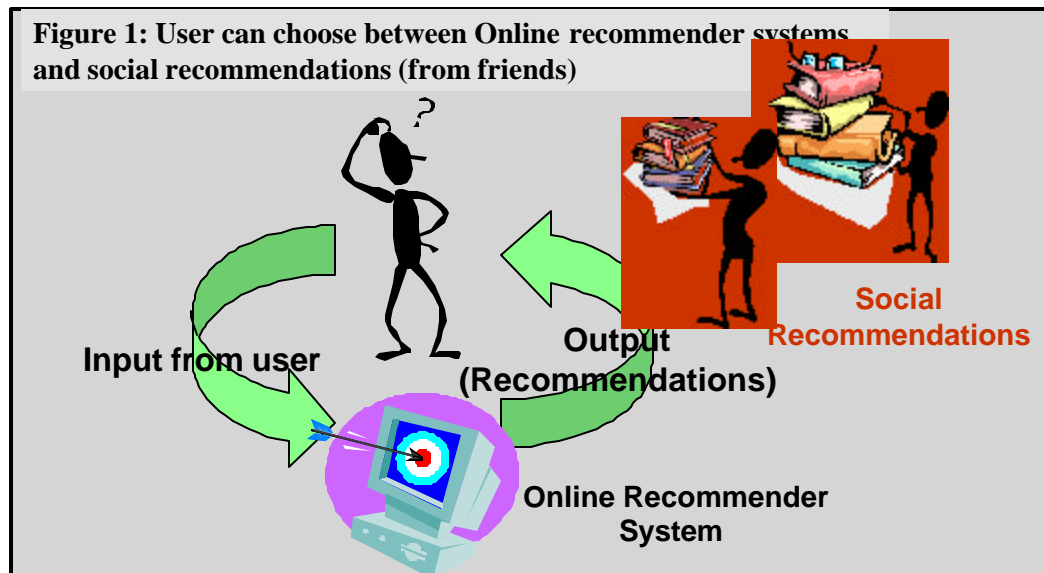
Our study methodology incorporates a mix of quantitative and qualitative techniques. For both of our studies we asked users to interact with several recommender systems, presented in random order. Users provided input to the systems and received a set of recommendations. We then asked users to rate 10 recommendations from each system, evaluating aspects such as: liking, action towards item (would they buy it / download it / do nothing); transparency (if they understood why system recommended that item); and familiarity (any previous experience of the item). Users were also asked to rate the system as a whole on a number of dimensions: usefulness, trustworthiness, and ease of use. For Study 1, we also asked users to evaluate recommendations provided by three of their friends using similar criteria. We recorded user behaviour and comments while they interacted with each system. At the end of each session, we asked users to name the system they preferred and explain their reasoning. Study 1 involved 20 participants and Study 2 involved 12. All participants were regular Internet users, and ranged in age from 19 to 44 years. Below, we describe our research studies in greater detail.

Study 1, Part 1: What user needs do recommender systems satisfy that a friend cannot?

Since the goal of most recommender systems is to replace (or at least augment) the social recommendation process (also called word-of-mouth), we began by directly comparing the two ways of receiving recommendations (friends and online recommender systems—see Figure 1) [4]. Do users like receiving recommendations from an online system? How do the recommendations provided by online systems differ from those provided by a user's friends? The results of our study indicated that users preferred recommendations made by their friends to those made by online systems. Though users preferred recommendations

made by friends, they expressed a high level of overall satisfaction with the online recommenders and indicated that they found the systems useful and intended to use them again [5]. This seemed to be due in part to the ability of recommender systems to suggest items that users had not previously heard of. In the words of one user, "I'm impressed with the types of movies that came back-- there were movies I hadn't seen—more interesting, more obscure. The system pulls from a large database—no one person can know about all the movies I might like."

The results of this study offer insight into the popularity of recommender systems. While users are happy with the age-old ways of getting recommendations, they like the breadth that online systems offer. Recommender systems allow users a unique opportunity to explore their tastes, and learn about new items.

Study 1, Part 2: Interface Analysis of Book and Movie Recommender Systems

The next question we asked was: What constitutes a satisfying interaction with recommender systems? To address this question, we conducted an exploratory study examining the interface of three book and three movie recommender systems:

- Amazon.com (books and movies)
- RatingZone's QuickPicks (books)
- Sleeper (books)
- Moviecritic.com (movies)
- Reel.com (movies)

A recommender system may take input from users implicitly or explicitly, or a combination of the two [6]; our study focused on systems that relied upon explicit input. Within this subset of recommenders, we chose systems that offered a wide variety of interaction paradigms to the user: differences in interfaces such as layout, navigation, color,
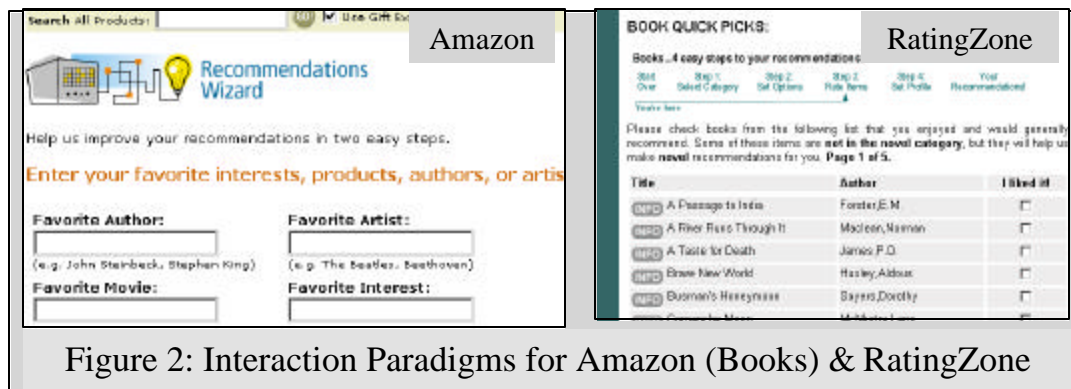
Figure 2: Interaction Paradigms for Amazon (Books) & RatingZone

graphics, and user instructions, types of input required, and information displayed with recommendations (see Figure 2 for illustration, and Appendix for full system comparison chart).

Our findings in this study suggested that, from a user's perspective, an effective recommender system inspires trust in the system; has system logic that is at least somewhat transparent; points users towards new, not-yet-experienced items; provides details about recommended items, including pictures and community ratings; and finally, provides ways to refine recommendations by including or excluding particular genres. Users expressed willingness to provide more input to the system in return for more effective recommendations.

Study 2: Interface Analysis of Music Recommender Systems

The goal of our second study was to verify the findings from Study 1, and extend them to another recommendation domain—that of music. In Study 1 we had focused on specific aspects of the interface (number of input items, number of results etc.). In Study 2 we considered the systems more holistically, seeking in particular to answer the question "what leads a user to trust the system's recommendations?"

In this study, we chose to examine music recommender systems, for two reasons. First, with the increasing availability and usage of online music, we anticipate that music recommender systems will increase in popularity. Second, and more importantly, music recommenders allow users to sample the item recommended—most systems provide access to a 30 second audio sample. This gave us the unique opportunity to evaluate the efficacy of recommendations in the lab setting. Users could sample the audio clip during the test session. Thus, their evaluations of the recommended items are based upon direct experience rather than an abstract estimate of liking.

We examined five music recommender systems:

- Amazon's Recommendations Explorer
- CDNow
- Mood Logic Filters Browser

- Song Explorer
- Media Unbound (5-minute version)

From this study, we found that trust was affected by several aspects of the users' interactions with the systems, in addition to the accuracy of the recommendations themselves: transparency of system logic, familiarity of the items recommended, and the process for receiving recommendations.

Interaction Design for Recommender Systems

Our analysis of recommender systems is divided into three parts. User interaction with such systems typically involves some input to the system; the system processes this input; and the user receives the output or recommendations. First we take recommender systems apart and analyse the input and output phases. What characteristics of these two phases distinguish recommender systems? Which of these design options do users prefer and why?

User interaction with recommender systems can also be conceptualised on a more gestalt or holistic level. What overall system features lead to satisfaction with recommendations? How do users decide whether to trust recommendations? What kinds of recommendations do they find the most useful? For each of these questions, we describe pertinent study results (both quantitative and qualitative); and suggest design options.

1) TAKING THINGS APART: INPUT TO THE SYSTEM

Recommender systems differ widely in terms of the type and amount of input users must provide in order to generate recommendations. Some recommender systems use an open-ended technique, asking users to indicate their favorite author, musician, or actor. Other systems ask users to rate a series of given items (books, songs, or movies) on a Likert Scale, while still others use a hybrid technique first asking general questions about taste (e.g., what phrase best indicates how you feel about FM radio?) followed by ratings of individual items, followed by item comparisons (e.g. do you like this song more or less than this other song?).
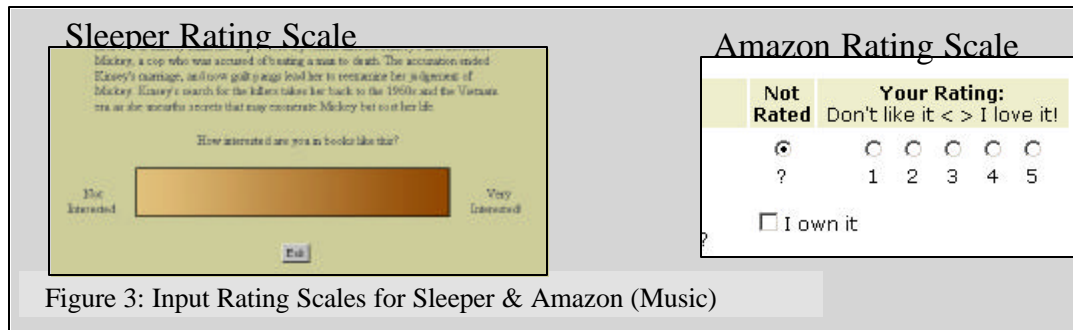
Figure 3: Input Rating Scales for Sleeper & Amazon (Music)

How many Items to Rate?

A few systems ask the user to enter only 1 piece of information to receive recommendations, while others require a minimum commitment of at least 30 ratings. Our quantitative and qualitative results indicate that users do not mind giving a little more input to the system in order to receive more accurate suggestions. Across both of our studies 39% of the users felt that the input required by systems was not enough, in contrast to only 9.4 % of our users who thought that the input required was too much. Table 1 shows users' opinions regarding the amount of input for music recommender systems (Study 2). Even for a system like MediaUnbound that required answers to 34 questions, only 8% of users regarded this as too much. Users indicated that their opinion of required input was influenced by the kind of recommendations they received. For systems whose recommendations were perceived as too simplistic (Amazon), or inaccurate (SongExplorer), most (>50%) users thought that input was not enough.

System	No. of Input Ratings	How users felt about number of ratings		
		Not Enough	Just Right	Too Much
Amazon	4-20	67%	33%	0.0%
CDNow	3	67%	33%	0.0%
MoodLogic	~4	45%	55%	0.0%
SongExplorer	20	58%	25%	8.3%
MediaUnbound	34	17%	75%	8.3%
Table 1: Input Ratings (From Study 2)				

Design Suggestion: Designers of recommender systems are often faced with a choice between enhancing ease of use (by asking users to rate fewer items) or enhancing the accuracy of the algorithms (by asking users to provide more ratings). Our suggestion is that it is fine to ask the users for a few more ratings if that leads to substantial increases in accuracy. Users dislike bad recommendations more than they dislike providing a few additional ratings.

What kind of rating process?

In the systems we studied, there were four types of rating input formats: (a) Open-ended: Name an artist / writer you like. When asked to name one "favorite" artist, some users found themselves stumped. With only one opportunity to provide input to the system, they felt pressure to choose with extreme caution. (b) Ratings on Likert Scale: Users were asked to rate items on 5-10 point scale ranging from Like to Dislike. This could become repetitive and boring. At SongExplorer, MovieCritic, and RatingZone users expressed irritation at having to page through lists of items in order to provide the requisite number of ratings. Another manifestation of a Likert scale was a continuous rating bar ranging from Like to Dislike. Users liked the rating bar since they could click anywhere to indicate degree of liking for an item. The Sleeper system used such a scale (see Figure 3). (c) Binary Liking: For this type of question, users were simply asked to check a box if they liked an item. This was simple to do, but could become repetitive and boring as well. (d) Hybrid Rating Process: Such systems incorporated features from all the above types of questions as appropriate. MediaUnbound used such a process and also provided continuous feedback to the user, keeping him / her engaged. Another aspect of the input process was the set of items that was rated. Often users had little or no experience of the item, leading to frustration with the rating process. One user commented at RatingZone "I'm worried because I haven't read many of these—I don't know what I'm going to get back," while at SongExplorer, another user observed "The [items to be rated] are all so obvious. I feel like I'm more sophisticated than the system is going to give me credit for."

Design Suggestion: It is important to design an easy and engaging process that keeps users from getting bored or frustrated. A mix of different types of questions, and continuous feedback during the input phase can help achieve this goal.

Filtering by Genre

Several recommender systems ask users whether they want recommendations from a particular genre. For example, MovieCritic allows users to set a variety of genre filters. Without being asked, almost all of the users volunteered favorable comments on these filters—they liked being able to quickly set the "include" and "exclude" options on a list of about 20 genres. However, we discovered two possible problems with genre filtering. Several users commented that
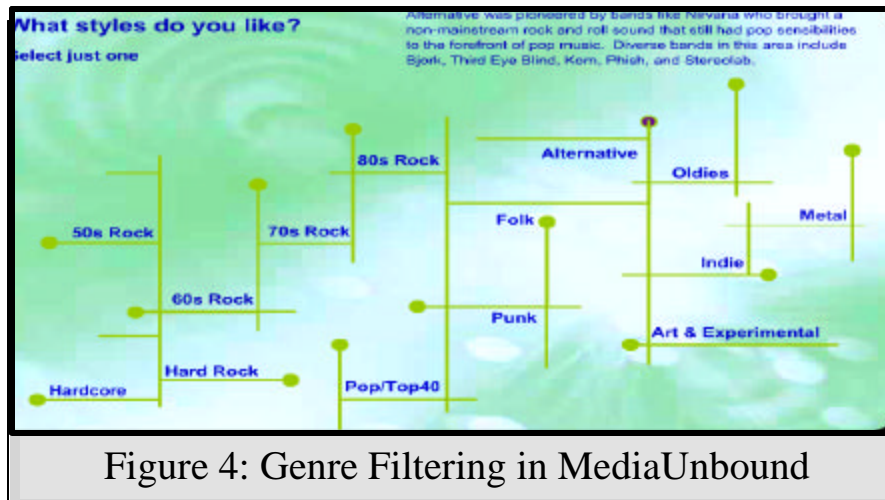
Figure 4: Genre Filtering in MediaUnbound

they did not like being forced to name the single genre they preferred, feeling that their tastes bridged several genres. Other users were unsure what exactly what kinds of music the genre represented since the system's categorization into genres did not map to their mental models. MediaUnbound and SongExplorer, two of the music recommender systems, faced such genre filtering problems (see Figure 4).

Genre is a tricky thing in recommendations. On the one hand recommender systems offer a way for users to move beyond genre-based book / movie / music exploration. On the other hand, genres do work well as shorthand for a lot of likes and dislikes of the user, and therefore help focus the recommendations. Over the course of the past year, we have observed that nearly all the major recommender systems have added a question about genre preferences.

Design Suggestion: Our design suggestion is to offer filter-like controls over genres, but to make them as simple and self-explanatory as possible. Users should be given the choice of choosing more than one genre. Also a few lines of explanation of each genre should be provided. This will allow users to understand what kind of music / books / movies the genre label represents.

2) TAKING THINGS APART: OUTPUT FROM THE SYSTEM
Ease of Getting More Recommendations

Recommender systems vary in the number of recommendations they generate. Amazon suggests 15 items in the initial set, while other sites show 10 items per screen, for as many screens as the user wishes to view. Users appear to be sensitive to the number of recommendations. However, the sheer number is less important than the ease of generating additional sets of recommendations. Some systems permit users to modify their recommendations simply by rating additional items. Other systems, however, require the user to repeat the entire rating process to see new recommendations. Users perceive the system as easier to use

if they can easily generate new sets of recommendations without a lot of effort.

Design Suggestion: Users should not perceive the recommendation set as a dead end. This is important regardless of whether they like the recommendations or not. If they like the recommendations, then they might be interested in looking at more; if they dislike the recommendations, they might be interested in refining their ratings in order to generate new recommendation sets.

Information about Recommended Items

The presence of longer descriptions of individual items correlates positively with both the perceived usefulness and ease of use of the recommender system (Study 1). This indicates that users like to have more information about the recommended item (book / movie description, author / actor / musician, plot summary, genre information, reviews by other users).

This finding was reinforced by the difference between the two versions of Rating Zone. The first version of RatingZone's Quick Picks showed only the book title and author name in the list of recommendations; user evaluations were almost wholly negative as a result. The second version of RatingZone changed this situation very simply: by providing a link to item-specific information at Amazon.com. Figure 5 shows the difference in perceived usefulness between both versions of the same systems. (Note: Error bars in figures 5 - 9 represent standard errors.) A different problem occurred at MovieCritic, where detailed information was offered but users had trouble finding it. This was because the item information was located several mouse clicks away and the site had poor navigation design. We have noticed that users find several types of information key in making up their minds. We use music systems as an example to describe the type of information users found useful.

Basic Item Information: This includes song, album, artists name, genre information, when album was released. Users

also like to look at the album cover. This often serves as a visual reminder for any previous experience with the item (e.g., they had seen that album in the store or at a friend's house).

Expert and Community Ratings: Reviews and ratings by other users seemed to be especially important. Several users indicated that ratings and reviews by other users helped them

Fig. 5: % Useful Recs. for Both Versions of RatingZone (From Study 2)
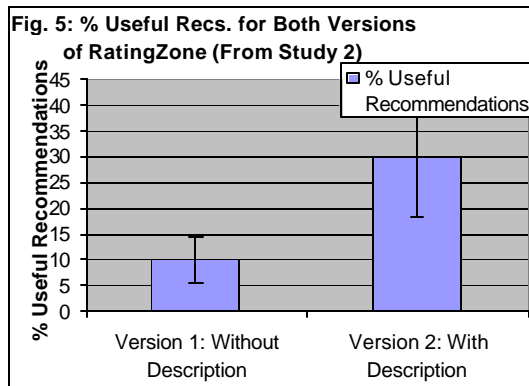
in their decision-making. In Study 2, 75% of the users indicated that community ratings in Amazon were helpful in deciding whether to trust the recommendations.

Item Sample: Users indicated that this was very helpful in making up their minds about the recommended songs. In the case of SongExplorer, one of the reasons users were dissatisfied with the system was that it was difficult to find the audio clip.

- "Of limited use, because no description of the books."(Comment about RatingZone, Version 1)

- "Red dots [Predicted ratings] don't tell me anything. I want to know what the movie's about."(Comment about MovieCritic)

- "I liked seeing cover of box in initial list of result… The image helps."(Comment about Amazon)

Design Suggestion: We recommend providing clear paths to detailed item information, validated through user testing. Simple changes to the navigational structure can have a large impact on user satisfaction. If the designer does not have access to lots of detailed item information (e.g. reviews by critics, plot synopses), offering some kind of a community forum for users to post comments can be a relatively easy way to dramatically increase the system's efficacy.

3) THE GESTALT VIEW: WHAT MAKES GOOD RECOMMENDER SYSTEMS?

Earlier we identified specific aspects of the interface that can affect the success of recommender systems. We focused upon mostly concrete dimensions of the user's interaction with the system: number of input items required, rating scales and recommendations. Next, we consider more holistic questions about what makes recommender systems work. What leads to trust in a system's recommendations, and what kind of systems do users prefer? How do users

decide if they should act upon the system's recommendations (e.g., buy / download the music, read the book or watch the movie)? Two factors emerged as strongly affecting levels of user trust: familiarity with recommended items and transparency of system logic.

The Advantages and Disadvantages of Familiar Recommendations

Recommender systems differ in the proportions of recommendations that have been previously experienced by users. For some systems, a large proportion of recommended items are familiar to the user, while other systems recommend mostly unfamiliar items. For example, 72% of Amazon's, 60% of MediaUnbound's, and 45% of MoodLogic's recommendations were familiar (Study 2).

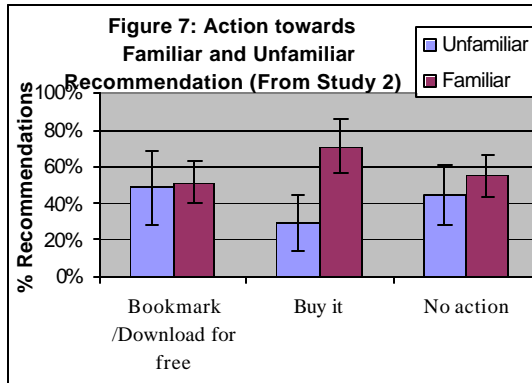*Users like and prefer to buy previously familiar recommendations*: In our first study we found preliminary indications that the presence of already-known items reinforces trust in the recommender system. We examined this issue in greater depth in Study 2, and found that mean

Figure 6: Mean Liking for Familiar & Not Familiar Recommendations (From Study 2)
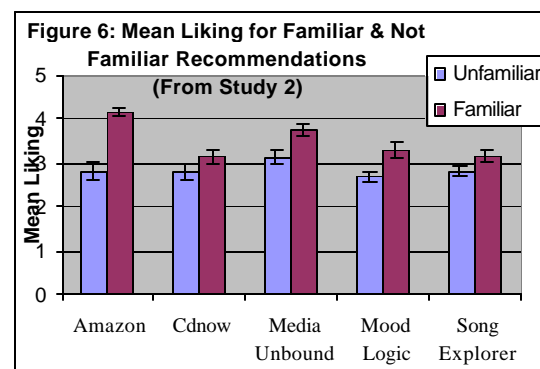
liking for familiar recommendations was higher than that for unfamiliar recommendations (Figure 6). The pairwise differences were significant for all systems except for CDNow [all ts' > 1.8; all p's <.05].

Familiar items appear to play a crucial role in establishing trust in the system. Previous positive experience with a recommended item increases trust in the system while previous negative experience causes trust in the system to decrease. Most of our users agreed that the inclusion of previously liked items in the recommendation set increased their trust. We also asked users whether they would be interested in buying, downloading for free, or bookmarking a recommended item. Figure 7 shows that users expressed greater willingness to buy familiar than unfamiliar recommended items (Note: Error Bars in figure represent standard errors). Most (70%) of the items that users expressed an interest in buying were familiar items. This makes sense since a familiar item is a less risky purchase decision.

Does too much familiarity breed contempt? While users did show a preference for familiar items, they did not like recommendations that were too directly related to their input

Figure 7: Action towards Familiar and Unfamiliar Recommendation (From Study 2)

Legend: Unfamiliar, Familiar

Y-axis: % Recommendations (0% to 100%)
X-axis: Bookmark/Download for free, Buy it, No action

ratings. For example, many of our users were frustrated by Amazon's recommendations that were albums by the same artists that the users had input into the system. "They're just going to give me things with this guy [same artist he named]?" one user commented. So while Amazon recommendations might remind users about a favorite song not heard recently, they did not help users expand their tastes in new directions. This perception was also reflected in the mean useful ratings for various music systems. Users in our study thought that MediaUnbound was a more useful system than Amazon because it introduced them to new items they liked and thereby allowed them to broaden their musical tastes.

Design Suggestion: A recommender system needs to understand user needs with relation to familiarity. Users differ in the degree of familiarity they want from their recommendations. The system might ask users about how familiar they would like their recommendation set to be. This would help systems cater to user needs more effectively. MediaUnbound, for example, includes a slider bar for users to indicate how familiar the music suggested should be. During the evaluation session, several users stated that they liked this option.

Our investigation into the effects of item familiarity on user satisfaction led us to some broader conclusions about RS design. We observed that two users with the same musical tastes often differ widely in what they expect and need from a recommender system. The range of user recommendation needs we have identified includes:
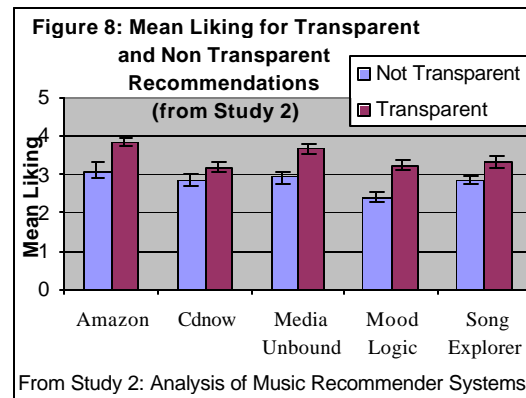
- Reminder recommendations, mostly from within the same genre ("I was planning to read this anyway, it's my typical kind of item")

- "More like this" recommendations, from within genre, similar to a particular item ("I am in the mood for a movie similar to GoodFellas")

- New items, within a particular genre, just released, that they / their friends do not know about

- "Broaden my horizon" recommendations (might be from other genres)

A user who is looking to discover new music might be frustrated by a system that keeps recommending artists
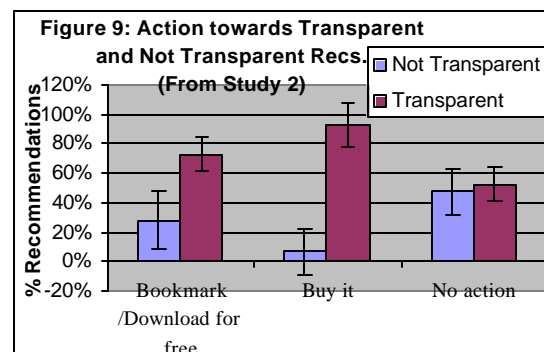
whose names it input into the system. As noted above, several of our users complained about this aspect while using Amazon.com. On the other hand, a user who is seeking "more like this" recommendations may feel thwarted by a system that does not return items similar to the ones he or she rates highly during the input step.

System Transparency

We were interested in exploring whether users perceive recommender system logic to be transparent, or whether they feel that they lack insight into why an item has been recommended. Is perceived transparency related to a greater liking for the system's recommendations? Results showed that users perceived systems to be very different on transparency. For Amazon, users thought they understood system logic 92% of the time, for MediaUnbound 76% of the time, and for MoodLogic 67% of the time (Study 2). Also users liked transparent recommendations more than not transparent recommendations (Figure 8) for all five systems. Mean liking was significantly higher for transparent than

Figure 8: Mean Liking for Transparent and Non Transparent Recommendations (from Study 2)

Legend: Not Transparent, Transparent

Y-axis: Mean Liking (0 to 5)
X-axis: Amazon, Cdnow, Media Unbound, Mood Logic, Song Explorer

From Study 2: Analysis of Music Recommender Systems

non-transparent recommendations for all systems except CDNow [all t's > 1.7; all p's<.05]. Furthermore, users more frequently indicated they would acquire a transparent recommendation (by buying or downloading it) than for a not transparent recommendation. (see Figure 9).

Figure 9: Action towards Transparent and Not Transparent Recs. (From Study 2)

Legend: Not Transparent, Transparent

Y-axis: % Recommendations (-20% to 120%)
X-axis: Bookmark/Download for free, Buy it, No action

Design Suggestions: This is an important finding from the perspective of system designers. A good CF algorithm that generates accurate recommendations is not enough to constitute a useful system from the users' perspective. The

system needs to convey to the user its inner logic and why a particular recommendation is suitable for them.

Users like the reasoning of recommender systems to be at least somewhat transparent. Herlocker et al., (2000) suggest that there are many ways to for the system to convey its inner logic to the user: (a) an explanation (e.g. "this item was recommended to you because you rated 'x' positively"), (b) predicted ratings (e.g. "we think you'll give this item an 8.5 out of 10") (c) including a few familiar recommendations (by artists or writers who are very close to input items) (d) community opinions (both reviews and numerical ratings) are all effective ways to provide more information about the recommendation.

ANALYSIS OF INTERACTION STYLE OF TWO RECOMMENDER SYSTEMS

In the preceding sections we have described some dimensions of user interactions with recommender systems. We have described our study findings and offered design suggestions based on those findings. Below we analyze the interaction style of two very different music recommender systems, in order to illustrate different models of recommendation success. Our analysis should also help illustrate the design guidelines identified above.

Results of Study 2 showed that mean liking for Amazon (Mean = 3.78; Standard Error=.11) was higher than for MediaUnbound (Mean = 3.49; Standard Error = .09). Users also indicated a greater willingness to buy Amazon recommendations (20% of items) as compared to MediaUnbound (7% of items). However in terms of overall system perception, MediaUnbound was rated as more useful (Mean = 1.5; Standard Error = .15) than Amazon (Mean = 1.16; Standard Error = .2). MediaUnbound was also rated as the system that understood users' tastes best, and most likely to be used again.

In general, our results suggest that a recommender system that allows users to explore their tastes and expand their musical horizons might be liked and used. But it might not influence buying decisions to the same degree as a system that merely reminds people of music to which they have previously been exposed. This paradox is further illustrated in our analysis of the different styles of recommending music offered by Amazon and MediaUnbound.

Recommendations by Amazon: Conservative Recommendations, Trustworthy System

Amazon's Recommendations Explorer performed well when examined in terms of recommendations that users liked the most, or were willing to spend the most resources on. Amazon recommended items that were very close to the user's input items. Therefore there was a high probability that users had directly or indirectly experienced these items previously. Many of the recommended items were simply albums by the same artist named by the user.

This conservative approach to recommendations had a number of effects. It led to high *system transparency*. Users understood why items had been recommended and could clearly see the link between the input and their output. Because users had previously experienced and liked so many of the recommended items, they perceived that the system understood their tastes and were inclined to trust it more. Amazon also provided users with *detailed information* about the item (pictures, expert reviews), as well as *community ratings* that further aided users in decision making. In addition, Amazon provided *sound clips* for most recommendations, allowing users to experience the item, and make their own judgments. Finally, the unit of recommendation was the album, rather than artist or song. This made it easier for users to think in terms of buying the recommendation.

Did Amazon succeed as a recommender system? If the purpose of a recommender system is to allow users to explore their tastes, then Amazon had only limited success. Users did not learn many new things about their tastes. But Amazon did succeed as an e-commerce system. It successfully guided users to items that they expressed an interest in buying.

Recommendations by MediaUnbound: Helping Users Explore Their Tastes

When users were asked about the system they found the most useful, and the one they thought best understood their musical preferences, the unanimous choice was MediaUnbound. Also, users seemed to enjoy the *recommendation process* with MediaUnbound. They liked the easy interaction with the FlashPlayer audio samples, the varied and humorous questions during the input process, and the overall look of the site. As one user commented, "[Media Unbound] entertains you with the process, the way you interact with the system. It felt like I was building a little pyramid--feels like the process you'd go through yourself naturally as a human being." The rating process itself seemed to inspire trust in the system and users liked the system's recommendations.

However the profile of items recommended by MediaUnbound was very different than that for Amazon. Users understood why an item was recommended for only 76% of the items as compared to 92% for Amazon. Users had previous experience with 60% of recommendations at MediaUnbound, in contrast to 72% of Amazon recommendations.

Users expressed a willingness to buy only 7% of the items recommended by MediaUnbound. This discrepancy between liking for the system and action towards its recommendations might be explained by the fact that a large percentage of items recommended by MediaUnbound were new to the users. While users enjoy being introduced to new

items that suit their tastes, they are not immediately willing to commit any resources. In addition, MediaUnbound presents only a list of individual songs, rather than complete albums, and does not offer the means for acquiring the item (e.g. a link to an e-commerce site). Therefore the user may perceive more of a psychological barrier to acquiring the item.

Finally, the user had only limited time to interact with the system during the course of our study. It is possible that in more realistic settings, users might have more time to explore MediaUnbound's recommendations and be willing to commit to purchasing recommended items. Recall that all of our users had indicated that they would use MediaUnbound in the future. Currently, we are following up with our study participants to find out if they have been using MediaUnbound as they had indicated, and whether they have bought any of the music recommended to them.

CONCLUSIONS

Both Amazon and MediaUnbound inspired trust in the user (albeit for different reasons). However, Amazon is a successful model of a recommender system integrated into an online commerce engine. In contrast, MediaUnbound offers users the chance to learn more about their musical tastes. Users liked both the systems but for different purposes.

Our suggestion to designers is to determine the purported role of the system—its primary purpose. A system may be designed very differently depending on the system's goals. It might also be possible to build some kind of a hybrid system that guides people to items they would be interested in buying immediately, but also allows them to explore and develop their tastes in the future.

REFERENCES

1. Konstan, J.A., Miller, B.N, Maltz, D., Herlocker, J.L., Gordon, L.R., and Riedl, J. GroupLens: Applying Collaborative Filtering to Usenet News. *Commun. ACM 40*, 3 (77-87).

2. Herlocker, J., Konstan, J.A., Riedl, J. Explaining Collaborative Filtering Recommendations. ACM 2000 Conference on Computer-Supported Collaborative Work.

3. Breese, J., Heckerman, D., and Kadie, C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, 1998 (43-52).

4. Resnick, P, and Varian, H.R. Recommender Systems. 1997 *Commun. ACM 40*, 3 (56-58).

5. Sinha, R. and Swearingen, K. Comparing Recommendations made by Online Systems and Friends. *Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries,* 2001.

6. Schafer, J.B., Konstan, J.A., and Riedl, J. Recommender Systems in E-Commerce. *Proceedings of the ACM Conference on Electronic Commerce*, November 1999.

Appendix. Recommender System Comparison Chart

	System type	No. of Input Items	Items to rate generated by system or individual?	Rating scale	Use of Genres or Categories	Info about recommended items	Methods for expanding on recommend-ations
Amazon (www.amazon.com)	Books, Movies, Music (part of e-commerce site)	1 item, then a list of 15 items with opportunity to refine	First by individual, then system	0 (don't like it) to 5 (love it!)	Not used explicitly (at the time of study)	Cover, title, artist/author/actors/directors, synopsis, audio clips, critics' & customer reviews	Give feedback on rec set (I own it/ 5 point rating scale/ Not interested)
Sleeper (www.pmetrics.com)	Books	12 books	System	Shaded bar	Not used explicitly	Cover, title, author, synopsis, number of pages	Give feedback on rec set. 1 item displayed at a time.
RatingZone QuickPicks (www.ratingzone.com)	Books, Movies, Music, TV		System	Like it or not (checkbox)	Must choose 1 genre before beginning	Books -Version1:Title, author & year only Books -Version2:Added link to info at Amazon.	Repeat rating process
MovieCritic (www.moviecritic.com)	Movies	20 movies	System or Individual	1 to 9	Optional genre filters	Title and year only, with link to Internet Movie Database	Give feedback on rec set
Reel (www.reel.com)	Movies (part of e-commerce site)	1 movie	Individual	None	Not used explicitly	Cover, title, year, director, actors, brief synopsis	Repeat rating process
Media Unbound (www.mediaunbound.com)	Music	8 audio clip ratings, & 5-6 other questions, and feedback on 2 items from each set of recs	System	Scale differs depending on question.	Must choose 1 genre at the beginning. Can choose sub-genre later	Song title, artist, and audio clip	Rate items on playlist or repeat rating process from the start
Song Explorer (www.songexplorer.com)	Music	20 songs	Initial set is System generated	0 (lowest) to 10 (highest), with no opinion option	Must choose 1 or more genres at the beginning.	Cover, artist name, album title, audio clip (sometimes)	Give feedback on rec set.
CDNow (www.cdnow.com)	Music (part of e-commerce site)	3 artists	Individual	None	Not used explicitly	Cover, artist name, album title, audio clip (sometimes), critics' reviews, customer ratings	Repeat rating process
Mood Logic (www.moodlogic.com)		1 to 9 filters	Individual	None	Option to choose genre as filter.	Cover, artist name, album title, audio clip (sometimes)	Repeat rating process