Predictors of User Perceptions of Web Recommender Systems:

How the Basis for Generating Experience and Search Product Recommendations Affects User Responses

Paloma Ochi[a], Shailendra Rao[a], Leila Takayama[b], Clifford Nass[a]

[a]Stanford University

Building 120, 450 Serra Mall

Stanford, CA 94305-2050

USA

[b]Willow Garage

68 Willow Road

Menlo Park, CA 94025

USA

{pochi, shailo, nass}@stanford.edu, takayama@willowgarage.com

Please direct printed correspondence to:

Professor Clifford Nass

Department of Communication

Stanford University

Building 120, 450 Serra Mall

Stanford, CA 94305-2050

USA

1-650-804-1733

Fax: 1-650-725-2472

Predictors of User Perceptions of Web Recommender Systems:

How the Basis for Generating Experience and Search Product Recommendations Affects User Responses

Abstract

One critical question suggested by Web 2.0 is: When is it better to leverage the knowledge of other users versus rely on the product-characteristic-based metrics for online product recommenders? Three recent and notable changes of recommender systems have been: (1) a shift from characteristic-based recommendation algorithms to social-based recommendation algorithms; (2) an increase in the number of dimensions on which algorithms are based; and (3) availability of products that cannot be examined for quality before purchase. The combination of these elements is affecting users' perceptions and attitudes regarding the recommender systems and the products recommended by them, but the psychological effects of these trends remain unexplored. The current study empirically examines the effects of these elements, using a 2 (Recommendation Approach: Content-based vs. Collaborative-based, within) x 2 (Dimensions Used to Generate Recommendations: 6 vs. 30, between) x 2 (Product Type: Experience Products (fragrances) vs. Search Products (rugs), between) web-based study (*N*=80). Participants were told that they would use two recommender systems distinguished by recommendation approach (in fact, the recommendations were identical). There were no substantive main effects, but all three variables exhibited two-way interactions, indicating that design strategies must be grounded in a multidimensional understanding of these variables. The implications of this research for the psychology and design of recommender systems are presented.

Keywords: recommender systems; content-based vs. social-based recommendations; number of dimensions for the basis of recommendation

1. Introduction

In an era of advanced Web technologies and a seemingly infinite number of products available online, recommender systems are not simply a luxury add-on feature for ecommerce websites. Instead, recommender systems have become a necessary and critical component of the online shopping experience for both retailers and consumers. Retailers can use online recommender systems to efficiently and effectively market their product selection to consumers, by both predicting whether an individual user will like a particular item, as well as identifying a set of items that will be of interest to an individual user (Degemmis et al., 2004; Sarwar et al., 2002). Similarly, consumers can take advantage of recommender systems to find exactly what they are looking for, thus bypassing the time-consuming and daunting task of browsing through vast product offerings or discovering the appropriate search terms and then deciding which product best matches their own unique needs. Additionally, when retailers properly harness the power of these recommenders, consumers can opportunistically discover items that they were not specifically looking for at that particular moment in time, but would like to purchase.

Recommender systems have three main components (Burke, 2002): (1) background data, which is already in the system before the commencement of the recommendation process, (2) input data, which the user gives the system in order to elicit a recommendation, and (3) an algorithm that combines background and input data to generate recommendations. This study focuses on the user's perspective of the third component of recommender systems, specifically on what type of recommendation algorithm is perceived to be most effective for generating product recommendations, while considering the genre of product and the number of data points or dimensions available to make these suggestions.

First generation ecommerce recommender systems used similarities between the user's preferences and associated characteristics of products as the basis for their product recommendations (e.g., book genres). In this product characteristic-based approach to generating recommendations, the system suggested items that had characteristics that the user was assumed to find desirable. For example, if a person ordered brightly colored clothes and woolen clothes, the system could recommend a bright red woolen sweater. This approach was oriented to the preferences of each individual user (Degemmis et al., 2004).

There are three types of characteristic-based recommender systems: content-based, utility-based, and knowledge-based (Burke, 2002; Degemmis et al., 2004; Resnick and Varian, 1997; Schafer, Konstan and Riedl, 1999; Terveen and Hill, 2001). Content-based recommendation creates each user's profile based on the associated features of products that the user has rated, and recommends items based on their matching associated features (Burke, 2002; Degemmis et al., 2004). Utility-based recommender systems can account for non-product attributes (such as vendor reliability and product availability), and generate recommendations based on a computation of the utility of each product for each user (Burke, 2002). Knowledge-based recommender systems have knowledge about how particular items meet particular user needs, and employ this knowledge to recommend items based on inferences about each user's needs and preferences (Burke, 2002).

Second-generation recommender systems focus on a more social approach grounded in the preferences of other people rather than the characteristics of other products. The idea, called social-based approaches, is that rather than attempting to classify products based on their relevant characteristics, which may be unclear or ambiguous, one should ground recommendations in the opinions of users similar to a given user. In essence, the more similar a person is to the target user, the more likely the system will recommend products that the person liked to the target user (Degemmis et al., 2004).

There are two types of social-based recommender systems: collaborative and demographic (Burke, 2002; Degemmis et al., 2004; Resnick and Varian, 1997; Schafer, Konstan and Riedl, 1999; Terveen and Hill, 2001). Collaborative filtering aggregates users' ratings of products, recognizes correlations among users' product ratings, and employs the product ratings of similar users to recommend new items of interest to individual users (Burke, 2002; Degemmis et al., 2004). Demographic recommender systems collect information about users' personal attributes, employ this demographic information to categorize users into demographic classes, and make recommendations based on correlations among these groupings[1] (Burke, 2002).

1.1 Richness of Dataset for Recommendations

Although characteristic-based and social-based recommender systems differ in how they generate recommendations, both recommendation techniques are powered by background data, which populates the system's working memory before the commencement of the recommendation process (Burke, 2002). Recommender systems vary in the size of this database or the number of dimensions used to generate recommendations. Some recommender systems may use a more extensive number of dimensions to generate recommendations, while others may use a more limited number of dimensions to generate their suggestions. Although recommender systems depend on a large dataset to generate recommendations (Burke, 2002), too many product

[1] Additional system techniques are based on hybrid recommendation approaches which combine two or more recommendation techniques to gain better performance with fewer of the drawbacks of any individual technique (Burke, 2002; Degemmis et al., 2004).

choices can be demotivating, as people have been shown to prefer a limited number of choices (6) rather than a more extensive number of choices (24 or 30) (Iyengar & Lepper, 2000). This conflict between the presumptive benefit of more data leading to more accurate predictions versus the demonstrated dislike of too much data in a purchasing context suggests that the effects of the number of dimensions on acceptance of recommendations is an open and important question.

1.2 Search vs. Experience Products

In addition to the number of dimensions, another issue that could influence the selection of a recommendation algorithm is whether the product is a "search" product or an "experience" product. This distinction was first explored by Phillip Nelson (1970, 1974). Search goods, such as cookware, house furnishings, carpets, cameras, garden supplies, and clothing, are products for which full information can be determined by inspection prior to purchase. Experience goods, such as food, drugs, toiletries, and books, are products for which full information cannot be acquired prior to purchase and use of the product, or for which information search is more costly and/or difficult than merely examining the product (Klein, 1998; Nelson, 1970, 1974, 1976, 1981).

Recommender systems are used for both search and experience goods. Because the nature of product characteristics differ in the two cases—the characteristics are hidden with experience products and are clear for search products—perceptions of recommended products could be affected by using characteristic-based vs. social-based algorithms.

1.3 Motivation for the Current Experiments

One of the key hallmarks of Web 2.0 interfaces (Web 2.0, 2009) is the leveraging of a large number of users to influence the behavior of a web application. One of the best examples of this leveraging is social-based recommendation. Although social-based recommendations started emerging in the mid-1990s with such systems as GroupLens (Resnick, et al., 1994) and Ringo/Firefly in 1994 (Lambert, et al., 2005)[2], many of the other aspects of Web 2.0 only began to appear at the turn of the century (the term "Web 2.0" was coined in 2004). Social-based recommender systems are typical Web 2.0 applications.

Of course, one cannot assume that Web 2.0 approaches are always better than previous approaches. Therefore, it is important to ask when it is better to leverage the knowledge of other users versus rely on the characteristic-based metrics powering the initial instantiation of online product recommenders. This study investigates in what situations it is more appropriate to use a recommender system based on social data rather than characteristic similarities (collaborative filtering vs. content-based recommendation) with respect to both the size of the system database used to generate the recommendations (large or small number of dimensions) and product type (experience good vs. search good). All of these eight possible combinations of ecommerce recommender systems have already been deployed without a systematic controlled experiment to either verify or challenge their fundamental assumptions and design choices in terms of their impact on users' liking and purchasing attitudes toward the recommended products. Users' perceptions of the quality of these systems and how these systems make users feel also remains an unexplored territory. The current experimental research addresses these issues for the psychology and design of future product recommender systems.

---

[2] While the Information Tapestry project (Goldberg, et al., 1992) was arguably the first collaborative filtering system, it did not obtain significant visibility.

2. Related Work

2.1 Techniques Used to Generate Recommendations

Existing literature on and the deployment of these recommender systems primarily focuses on collaborative-based and content-based recommendation techniques (Balabanovic and Shoham, 1997; Degemmis et al., 2004; Pazzani, 1999; Shardanand & Maes, 1995). Collaborative filtering is the most well-known and widely used recommendation technique to date (Burke, 2002; Degemmis et al., 2004). The strengths of collaborative filtering systems lie in their adaptivity (ability to improve over time as they aggregate ratings of objects), ability to identify "cross-genre niches," functioning without knowledge of the product domain, and requiring only implicit feedback from users (Burke, 2002). However, collaborative filtering recommender systems also have drawbacks. Collaborative filtering exhibits both the new user ramp-up problem, in which new users who have only rated a few objects are difficult to categorize and compare to other users, and the new item ramp-up problem, in which new items that have not received many ratings cannot be easily recommended to users (Burke, 2002; Konstan et al., 1998). Additionally, using collaborative filtering can be problematic because it depends on a large historical dataset to ensure that there is overlap in ratings across users. Accordingly, collaborative filtering only works well when there are many users who rate a small and static set of items (Burke, 2002). Because collaborative filtering relies on overlap in users' tastes, this recommendation technique does not work well for "gray sheep" who straddle the fence between groups of users (Burke, 2002; Claypool et al., 1999). Examples of recommender systems that use collaborative filtering (Burke, 2002) include GroupLens/ Net Perceptions (Resnick et al., 1994), Ringo/Firefly (Shardanand & Maes, 1995), Tapestry (Goldberg et al., 1992) and Recommender (Hill et al., 1995).

Content-based recommender systems exhibit similar strengths as recommender systems that use collaborative filtering, except content-based recommender systems cannot identify cross-genre niches (Burke, 2002), and have no way of generating "serendipitous finds"– recommended products that the user has not necessarily seen nor indicated liking previously (Shardanand & Maes, 1995). However, content-based recommender systems are not prone to the new item ramp-up and "gray sheep" drawbacks of collaborative filtering (Burke, 2002). An example of a content-based recommender system is NewsWeeder (Lang, 1995), which is a newsgroup filtering system that uses the words in the texts as the associated features of the texts (Burke, 2002).

Both collaborative filtering and content-based recommendation techniques are widely used in various arenas, but the comparative research in this domain is devoid of a focus on users' purchasing behaviors, as well as their attitudes toward and perceptions of these systems during usage in differing contexts. Knowledge of the advantages and drawbacks between these two recommendation methodologies according to data mining and algorithm design has advanced with a multitude of research. However, there is still a lack of experimental psychological research tackling fundamental questions concerning when to use social or characteristic-based data and their implications on the user experience. It is critical to fill this dearth with empirically grounded research to inform the use of social or characteristic-based data in the design of these recommender systems.

2.2 Number of Characteristics Used to Generate Recommendations

The quality of recommendations generated by both collaborative filtering and content-based recommendation is dependent upon having a large dataset (Burke, 2002). Recommender systems with larger data sets can better match a user with similar users (as in the case of collaborative filtering), or better match a user's preferences with the associated features of a product (as in the case of content-based recommendation). For users, systems with a greater number of dimensions may be perceived as having a more thorough and nuanced understanding of the data, leading to perceptions of better performance.

On the other hand, a recommender system that uses a large number of dimensions may actually lead to negative perceptions of the recommendations that are provided. For example, an experiment conducted by Iyengar and Lepper (2000) revealed that people are more likely to purchase products or undertake optional assignments when they are offered a limited array of 6 choices as opposed to a more extensive array of 24 or 30 choices. Additionally, participants in this study also reported greater satisfaction with their selections when they received a limited set of options from which to choose. If the system's dimensions are perceived as criteria that the user, as well as the system, must process, a larger number of dimensions might lead to less satisfaction (although there is clearly a difference between the number of dimensions and the number of recommendations based on those dimensions). Furthermore, a recommender system that bases its recommendations on an extensive number of data points may raise user expectations, which in turn will make bad choices more salient than good choices (Reeves & Nass, 1996). Investigating these issues will increase the understanding of whether the amount of data used to construct recommendations based on either collaborative filtering or content similarities affects perceptions of the quality of the recommender system and its recommendations.

2.3 Product Type Recommended

In addition to investigating differences in uses of recommendation techniques and the number of dimensions used to generate recommendations, our research examines differences in recommending search goods as compared to experience goods.

Minimal research on the applicability of the search/experience framework to the information search process has been conducted (for exceptions, see e.g., Klein, 1998; Maute and Forrester, 1991). Nelson (1970) predicted that the recommendations of others would be used more for purchases of experience goods than search goods. However, with new search technology and recommender systems, it is unclear if this remains the case. Klein (1998) suggests that new media will provide information for search goods that is more accessible, more customizable, and less costly, while at the same time, new media will enable experience goods to "virtually" turn into search goods, by allowing consumers to obtain important product performance information prior to purchase.

This study investigates whether recent advances in recommendation techniques—collaborative filtering and content-based recommendation—as well as the number of dimensions used to generate recommendations, should be applied differently, depending on the type of products that the website offers. More practically, given the type of products that a website offers and given the amount of system data available, which type of recommendation system engine will lead to increased consumer purchases and satisfaction with the overall shopping experience?

3. Study Design

3.1 Overview of Design

We conducted a 2 (Recommendation Approach: Content-based vs. Collaborative-based, within) x 2 (Dimensions Used to Generate Recommendations: 6 vs. 30, between) x 2 (Product Type:

Experience Products (fragrances) vs. Search Products (rugs), between) web-based, mixed-design experiment.

Participants began the study by indicating their preferences via a set of 40 profile questions, with different questions for fragrance participants as compared to rug participants. After filling out the profile questions, participants were told that they would use two different recommender systems for either rugs or fragrances, with both systems based on 6 dimensions or both based on 30 dimensions. One system was (ostensibly) content-based and the other system was (ostensibly) collaborative-based. As each system presented each product recommendation with values on the appropriate number of dimensions, participants were asked to evaluate the quality of the recommendation. After evaluating all of the recommendations ostensibly generated by a given recommender system, participants were asked to provide overall evaluations of that system.

3.2 Participants

A total of 80 participants (48 female, 32 male, *mean* age = 21.06 years, *SD* = 2.16) volunteered and consented to participate in this study. Participants were recruited primarily through university mailing lists, and were granted class credit or paid $10 cash for their participation in this study. They were split evenly among the eight experimental conditions by random assignment.

3.3 Product Type

Two types of products were recommended by the recommender systems. Half of the participants used two systems recommending fragrances, an exemplar experience good whose qualities cannot be determined until after purchase and usage. The other half of the participants used two systems recommending rugs, a prototypical search good whose qualities (such as color and size) can be determined prior to their purchase on the Web.

3.4 Recommendation Approach

Each participant tested two recommender systems that were ostensibly powered by two different types of recommendation approaches. The Content-based recommendation approach (ostensibly) matched participant's preferences with similar product characteristics in the system's database. The Collaborative-based recommendation approach (ostensibly) matched the participant with people in the system's database who had similar preferences. The order of presentation was balanced for each between-participants condition.

3.5 Dimensions Used to Generate Recommendations

A set of either 6 (small) or 30 (large) data dimensions was ostensibly used to generate the product recommendations for both the Content-based and Collaborative-based engines. These amounts parallel the small and large number of product choices used in Iyengar and Lepper (2000). In addition to referencing the number of dimensions in the introductions to the system, the (ostensible) values of each of the dimensions were indicated for each of the recommended items (see below).

3.6 Procedure

In the study invitations, participants were told that the web-based study about recommender systems would require 30-45 minutes to complete and could be taken at their convenience. After signing up for the study, participants were directed to the study website, where they read and consented to the human subjects internal review board-certified terms of the study.

The participants began by answering profile questions. For fragrance (experience) participants, the profile question was, "How much would you enjoy the flavors of the following foods?” followed by a list of foods (e.g., “coffee,” “teriyaki steak,” “curry”). For rug (search) participants, the profile question was, “How much would you enjoy the look of the following furnishings in your living room?”, followed by a list of furnishings (e.g., “glass cabinet,” “snow globe,” “rocking chair”). We chose these questions because they were plausibly related to the products to be recommended, but not so close that the subsequent recommendations would be obviously right or wrong. For both sets of profile questions, the response scale involved five options: “Dislike Very Much,” “Dislike,” “Neutral,” “Like,” and “Like Very Much.” Participants’ responses were not considered in the remainder of the experiment (although participants believed that the recommendations were based on their responses).

After completing the profile, participants were directed to the first of the two recommender systems; order of the content-based vs. collaborative-based system was randomly assigned and balanced across between-participants conditions. When participants arrived at a given recommender system, they received the introduction explaining the recommendation approach and number of dimensions used to generate recommendations: this constituted one aspect of the manipulation.

The content-based system read:

> “We are matching your food preferences with similar [fragrance/rug] characteristics in our database. Based on the **[fragrances/rugs] that match your preferences along [6/30] dimensions**, your [fragrance/rug] recommendations are being generated. Please hit the ‘Next’ button when the progress bar reaches the end.” [emphasis in original]

After the participants hit the “Next” button, the system read:

"Here are male and female fragrances that **match your preferences along [6/30] dimensions**." [emphasis in original].

Conversely, the collaborative-based system read:

"We are matching you with people who have similar preferences in our database. Based on the **people who match you along [6/30] dimensions**, your [fragrance/rug] recommendations are being generated. Please hit the 'Next' button when the progress bar reaches the end." [emphasis in original]

After the participants hit the "Next" button, the system read:

"Here are [male and female fragrances/rugs] that are recommended by **people who match you along [6/30] dimensions**." [emphasis in original]

To emphasize the differences between the two systems beyond their description at the introduction to each system, the two systems were assigned a different color scheme and name (*Whiff-o-matic* and *Sniff-o-matic* for the fragrances; *Lush-o-matic* and *Plush-o-matic* for the rugs) to clearly differentiate them during usage and evaluation. To ensure that the differences of name and color would not have an effect on our results, these assignments were balanced and randomized within each condition.

Participants then received eight pre-selected fragrance or rug recommendations from each recommender system. The sixteen total fragrance or rug recommendations given by the two recommender systems were identical across participants, recommendation approach, and dimensions used to generate recommendations.

Each of the recommendations listed a product title, picture, and some accompanying fragrance or rug notes (see Figure 1). Additionally, each suggested product was presented with a percentage match with the participant (see Figure 2). Content-based participants saw “X% of the ***characteristics*** you like match this [fragrance, rug]”, while Collaborative-based participants saw “X% of ***people*** like you recommend this [fragrance, rug].” This sentence served as an additional instantiation of the recommendation approach manipulation. The percentage matches for all eight recommendations (94%, 91%, 87%, 80%, 77%, 73%, 63%, and 62%) were identical for all eight recommender system versions. Each recommendation percentage match was augmented with a radar graph with either 6 points for recommendations based on 6 dimensions or 30 points for recommendations based on 30 dimensions. These images served as an additional instantiation of the number of dimensions recommendation. In sum, for each system using a Content-based Recommendation Approach, participants saw the phrase “your preferences” three times (twice in the introduction to the experiment and once in the introduction to the recommendations), and “characteristics” nine times (once in the introduction and once for each of the recommendations), while there was no use of the term “people.” Conversely, for each system using a Collaborative-based Recommendation Approach, participants saw the term “people” 11 times (twice in the introduction to the experiment, once in the introduction to the recommendations, and once for each of the recommendations), while there was no use of the term “your preferences” or “characteristics.”

Similarly, the number of dimensions appeared once in the introduction to the experiment and once in the introduction to the recommendations. In addition, each time a recommendation was made, there was a pictorial representation of either 6 or 30 dimensions for each of the eight products (see Figure 2). Thus, there were 10 instantiations of the number of dimensions for each system.

After evaluating the eight product recommendations from each recommender system, participants were directed to a post-questionnaire where they evaluated that particular recommender system and how they felt while using it.

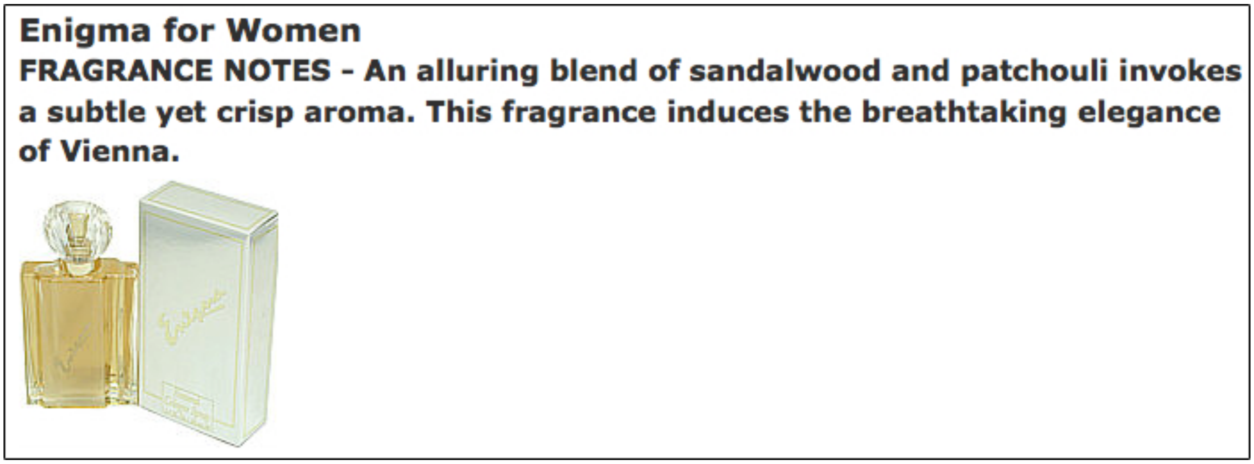

Figure 1: Sample recommendation displaying product title, picture, and notes.

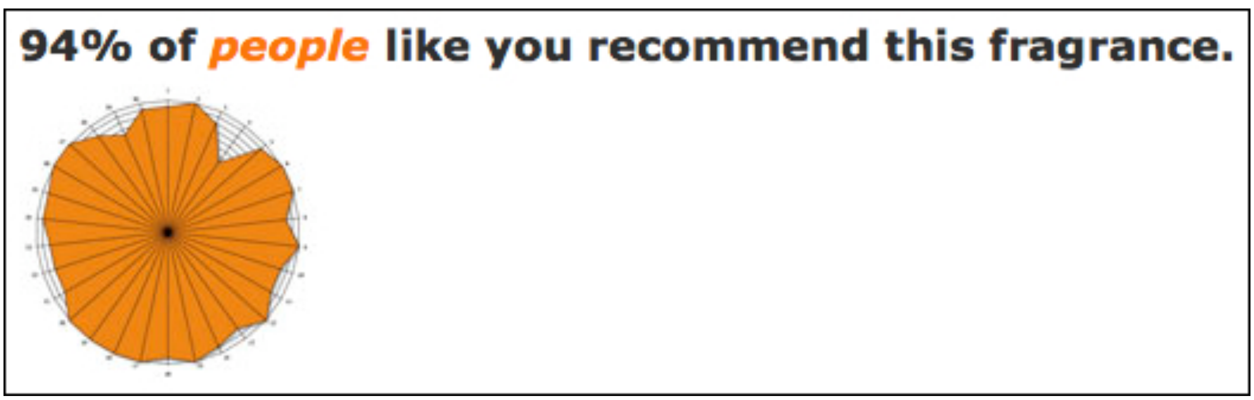

Figure 2: Sample percentage match indicator showing exact percentage by characteristics or people and radar graph with 30 points.

3.7 Measures

To examine our research questions, a set of dependent variables were derived from the questionnaire using Principal Components Analysis.

Liking of the Recommendations: For each fragrance and rug recommendation, participants answered: (1) “How much would you like this product?” (“Dislike Very Much” (=1) and “Like Very Much” (=10)), and (2) “How likely are you to buy this product?” (“Very Unlikely” (=1) and “Extremely Likely” (=10))—both on ten-point Likert scales. After removing the most strongly recommended and the most weakly recommend items (leaving the second through seventh items), the index was highly reliable (Cronbach’s α = .94).

User’s Positive Feelings: This index consists of four items from the post-questionnaire that assess to what degree the participant felt *upset* (reverse-scaled), *distressed* (reverse-scaled), *irritable* (reverse-scaled), and *afraid* (reverse-scaled) while using each system. Confirmatory factor analysis indicated a single factor for both the Content-based (*eigenvalue* = 2.55) and Collaborative-based (*eigenvalue* = 2.49) recommender systems; the items were combined into a very reliable index (α = .86).

Intelligence of the System: This index consists of six items from the post-questionnaire that assess to what degree the participant felt the system was *intelligent*, *accurate*, *competent*, *trustworthy*, *useless* (reverse-scaled), and *inaccurate* (reverse-scaled). These adjectives and the remaining measures were based on ten-point Likert scales (“Describes Very Poorly” (=1) and “Describes Very Well” (=10)). Confirmatory factor analysis indicated a single factor for both the Content-based (*eigenvalue* = 3.87) and Collaborative-based (*eigenvalue* = 3.66) recommender systems; the items were combined into a highly reliable index (α = .92).

4. Results

A series of 2 (Recommendation Approach, within) x 2 (Dimensions Used to Generate Recommendations, between) x 2 (Product Type, between) mixed-model Analysis of Variance (ANOVAs) were conducted to assess the effects of the three independent variables on each of the three dependent measures.

For *Liking of the Recommendations*, there was a significant interaction effect for Product Type by Recommendation Approach, $F(1,76) = 6.32$, $p < .02$, $\eta_{within}^2 = .08$ (see Figure 3).[3] For Search products, participants liked the recommendations more when they were powered by the Collaborative-based system as compared to the Content-based system; for Experience products, there was no significant difference.

There was also a significant cross-over interaction effect for Product Type by Dimensions Used to Generate Recommendations, $F(1,76) = 4.39$, $p < .05$, $\eta_{between}^2 = .06$ (see Figure 3), such that for Search products, participants liked the recommendations more when they were based on the smaller number of dimensions (6) compared to the larger (30), while for Experience products, participants liked the recommendations more when they were based on the larger number of dimensions (30) as compared to the smaller (6). One main effect was an artifact of an interaction: participants liked the recommendations more when they were generated with the Collaborative-based recommender system compared to the Content-based one, $F(1,76) = 14.11$, $p < .001$, $\eta_{between}^2 = .18$. The main effect for Product Type cannot be interpreted because the items and their descriptions are not comparable, $F(1,76) = 6.29$, $p < .02$, $\eta_{between}^2 = .08$. No other interaction or main effects were found.

---

[3] Although the error bars are presented, they are deceptive in mixed designs (Estes, 1997; Loftus & Masson,1994).

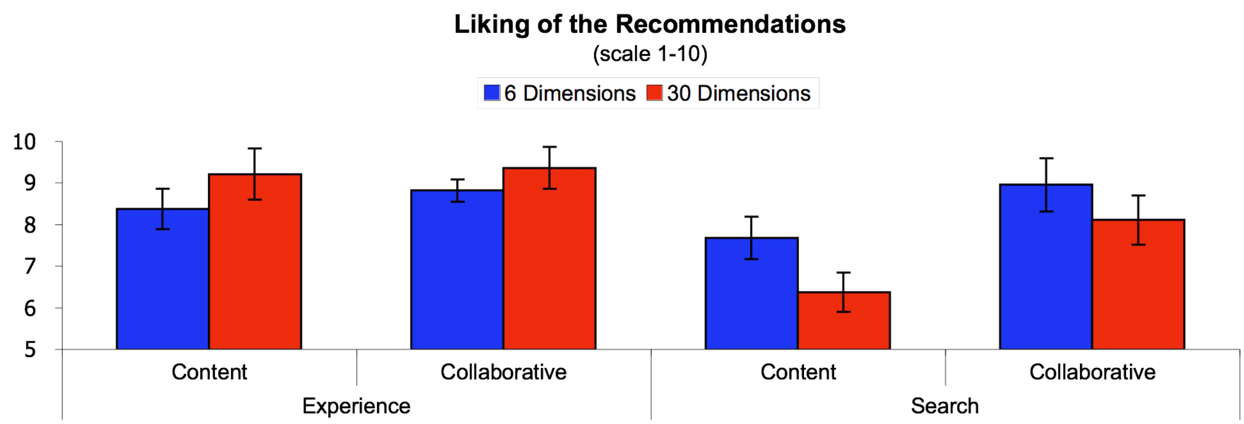

Figure 3: Mean Values for Each of the Eight Experiment Conditions Regarding User Liking of the Recommendations.

For *User's Positive Feelings,* the results were consistent with the results for liking. There was a significant cross-over interaction effect for Product Type by Recommendation Approach, $F(1,76) = 4.38$, $p < .04$, $\eta_{within}^2 = .06$ (see Figure 4). For Search products, participants tended to feel better about the recommendations when they came from the Collaborative-based system as compared to the Content-based system; for Experience products, participants tended to feel better about the recommendations when they came from the Content-based system as compared to the Collaborative-based system.

There was also a significant cross-over interaction effect for Product Type by Dimensions Used to Generate Recommendations, $F(1,76) = 11.33$, $p < .001$, $\eta_{between}^2 = .15$ (see Figure 4), such that for Search products, participants felt better when the recommendations were based on the larger number of dimensions (30) compared to the smaller (6), while for Experience products, participants felt better when the recommendations were based on the smaller number of dimensions (6) as compared to the larger (30). No other interactions and no main effects were found.

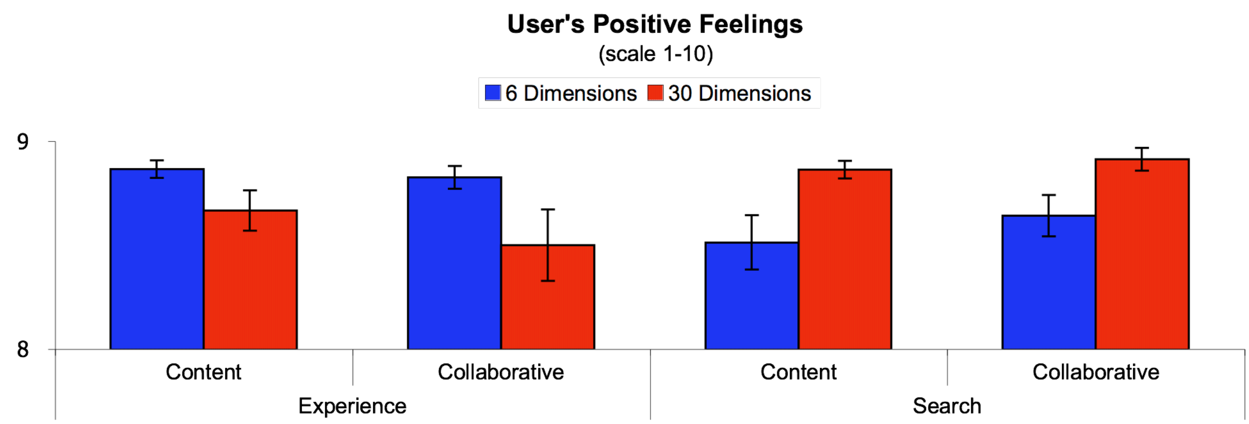

Figure 4: Mean Values for Each of the Eight Experiment Conditions Regarding Users' Positive Feelings in the Interaction.

The analysis for *Intelligence of the System* revealed a significant cross-over interaction effect for Product Type by Dimensions Used to Generate Recommendations, $F(1,76) = 5.91$, $p < .02$, $\eta_{between}^2 = .08$ (see Figure 5), such that for Experience products, participants thought the recommender system was more intelligent when it based its recommendations on the smaller number of dimensions (6) as compared to the larger (30); for Search products, there was no significant difference.

There was also a significant interaction for Dimensions Used to Generate Recommendations by Recommendation Approach, $F(1,76) = 10.88$, $p < .001$, $\eta_{within}^2 = .14$ (see Figure 5). Collaborative-based approaches with 30 dimensions were perceived as much less intelligent than the other three conditions.

The main effect for Product Type cannot be interpreted because the items and their descriptions are not comparable, $F(1,76) = 8.91$, $p < .01$, $\eta_{between}^2 = .12$. No other interaction or main effects were found.

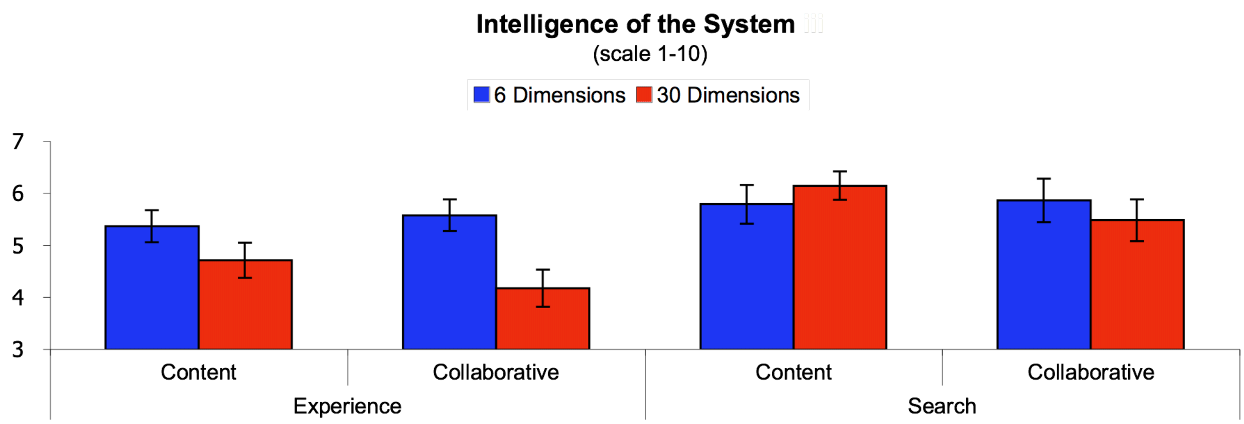

Figure 5: Mean Values for Each of the Eight Experiment Conditions Regarding Perceived Intelligence of the System.

5. Discussion

5.1 Summary of Results

One of the primary characteristics of the shift to Web 2.0 has been a dramatic increase in recommender systems that are social-based—recommending products that people who are similar to the user like—rather than characteristic-based—recommending products based on the characteristics of products that the user prefers. Similarly, as the size of recommender databases and the computational power of Web servers have grown, the number of dimensions on which recommendations can be based has also grown. These two dimensions have been viewed solely as a means to improve the accuracy of search product recommendations. Finally, while in the early days of the Web, the natural products to sell were those whose attributes could be easily verified by customers (i.e., search products), the developing trust in ecommerce now leads consumers to buy products that cannot be evaluated until after use (i.e., experience products).

The first key result of this paper is that these technological changes in the basis of recommender systems have psychological, as well as technical, impact. That is, although the recommendations were identical in all respects, participants believing that the recommendations were based on similar people rather than based on preferred product characteristics reacted very differently to the system. Similarly, although the use of 30 dimensions might seem *a priori* superior to 6 dimensions, the relationship between liking of the recommendations, positive feelings and even perceived intelligence of the system, and the number of dimensions was not simple. Even more complex, feelings about these dimensions were influenced by the types of products recommended, specifically whether the product characteristics were verifiable or not at the time of purchase.

The second key result of this paper is that there are no simple answers to whether the traditional or new versions of ecommerce sites are optimal. Instead, the key effects of all three of the variables only occur in two-way interactions. Notably and surprisingly, there was a clear tendency for positive outcomes when a first-generation aspect of the recommendation system was combined with a second-generation aspect. For example, people like the system better and feel better about using the system when the first generation of web products, i.e., search products, was coupled with the second generation of dimensions, i.e., 30, and while the second generation of products, i.e., experience products, was coupled with the simpler, first generation of dimensions, i.e., 6. (This result seems to be inconsistent with Maute and Forrester (1991), who found that people are more motivated to search for experience characteristics as compared to search characteristics, suggesting that experience products would be more appropriately described with a higher number of dimensions. Although future research should address this seeming inconsistency, one key difference is that the current research is focused on search vs. experience products as compared to search vs. experience characteristic *within* a product.) Similarly, for search products, people liked the system more when it used a collaborative, Web

2.0 approach. Finally, systems that combined collaboration with 30 dimensions were seen as less intelligent.

The bottom line is that a "one size fits all" approach for product recommender systems cannot satisfy users. For example, the recommendation technique must be adapted to the type of product. Similarly, while designers and marketers of recommender systems are frequently tempted to highlight the vastness of the data set and dimensions used to generate the recommendations, the results from this study caution against doing so without a consideration of the recommendation approach and the product type.

5.2 Open Questions and Extensions

A series of studies by Nass and colleagues has demonstrated that it works best to deploy multi-dimensional technologies with each dimension at the same level of technological development (Gong & Nass, 2007; Nass & Brave, 2005; Nass, Brave, Takayama, 2006; Nass & Lee, 2001; Isbister & Nass, 2000). For example, a synthetic face with a synthetic voice was clearly preferred to a synthetic face with a recorded voice or a video face with a synthetic voice. How can one reconcile these seeming benefits of inconsistency with respect to recommender systems with this other literature? One possibility is that what is advanced from a technological perspective may be less advanced from a psychological perspective. For example, while 30 dimensions may require much more complex and sophisticated processing from a technical perspective, it may seem to reflect a shot-gun approach that is less sophisticated than the more focused and carefully selected set of 6 dimensions. Similarly, while social-based approaches require more complex statistical and analytic models than content-based approaches, they may seem to reflect less "knowledge" of the products and less sophistication than the product-based approaches. Future research should determine whether there is in fact a discrepancy between actual and perceived sophistication. Of course, alternative explanations for the seeming preference for inconsistency should also be explored.

As recommender systems become ubiquitous as a result of the seemingly infinite collection of products and services available on the web, the selection of the psychologically-appropriate algorithm becomes critical. The current study selected one example of each of the two dominant approaches: (1) content-based, which conducts a deep analysis of characteristics and matches users with appropriate products and services, exemplifying characteristics-based approaches, and (2) collaborative-based, which leverages the wisdom of other similar people to suggest products, exemplifying social-based approaches. Future research should determine whether one can generalize from these exemplars to the other approaches that are available: utility-based and knowledge-based (characteristics-oriented), demographic (social-oriented), and the hybrid approaches. Of course, these approaches cannot be evaluated without reference to the number of dimensions used and the types of products to be recommended.

Another area for future research is assessing whether pre-questionnaires used to gather information about users play an important role in perceptions about the quality of recommendation output from various recommender engines with respect to the number of dimensions used and the type of products to be recommended. In this study, the questions differed for the two product domains, fragrances (Experience Products condition) and rugs (Search Products condition), but the same question sets were utilized by the two different recommendation techniques. Future work in this area can look at whether there are any side effects deriving from user attitudes about the appropriateness of these questions for different recommendation engines. For instance a question that seems intuitive for searching for a fragrance based on other people's preferences might not have the same desirable effect if the recommendations are to be purely based on characteristic matching by a computer.

While we selected products that were canonical exemplars of search and experience products, future research should use a wider range of products and services. Another interesting approach is to not differentiate search vs. experience by the nature of the products, but rather by whether the descriptions of the products and services highlight characteristics associated with search or experience qualities (Danielson, 2007). Similarly, while the choice of the number of dimensions was informed by the consumer research literature, it would be useful to examine a wider range of dimensions to uncover the critical cut-offs for differentiating user reactions. Also, we chose 6 and 30 dimensions based on studies distinguishing a "small" number of choices vs. a "large" number of choices. It will be important to determine if there are optimal numbers of dimensions based on type of product, recommendation approach, or other characteristics; for example, 30 dimensions may seem too many to be "searchable." Finally, it is important to determine whether different types and numbers of profile questions would be appropriate for content- vs. collaborative-based approaches, 6 vs. 30 dimensions, or search vs. experience products.

Another important domain of research involves the intersection between these high level descriptions of recommendation approaches and the more detailed explanations of the decision-making process in complex systems discussed by Herlocker (e.g., Herlocker, et al., 2000) and Tintarev (e.g., Tintarev and Masthoff, 2007), among others. This latter research is especially critical in the context of recommender systems, where understanding the relationship between product rating input and generated recommendation output can help users have predictable and efficient interaction with the recommender system (Johnson & Johnson, 1993; Nielsen, 1994;

Sinha & Swearingen, 2002). Specifically, having an explanation that provides transparency on how the recommender system works can benefit users in many ways: (1) explanations provide justification and reasoning for a recommendation, allowing users to decide how much confidence to place in the recommendation; (2) explanations increase user involvement, allowing users to complete the decision process with their own knowledge; (3) explanations educate users on the processes used to generate recommendations; and (4) explanations make the system's strengths and limitations, as well as justifications for suggestions, fully transparent, leading users to greater acceptance of the recommender system as a decision aid (Herlocker et al., 2000).

In our implementation, the system was presented to users as being extremely transparent, in that it clearly described and emphasized both the (ostensible) recommendation approach as well as the (ostensible) number of dimensions used in the approach. Future research should examine whether the benefits of transparency might be outweighed by the complexity of the interactions between the aspects of the search. Furthermore, depending on the amount of data powering their recommendations and the particular product or service type, different recommendation techniques may necessitate unique levels of transparency or feedback to the end user as well as distinctive framing. This is another important area for future research.

6. Conclusion

Recommender systems have become a necessity for any website that has a significant number of products or services to offer. The traditional—and appropriate—focus when deploying these systems has been finding the algorithm that will most accurately and efficiently present the optimal choices for consumers. The current research suggests that recommender systems must also recognize that the choice of algorithm can influence how optimal the recommendations *seem*, independent of their actual quality. Decisions concerning whether and when to use data-based or social-based approaches and whether, when it comes to dimensions, more is better or "less is more," must now be conditional on each other as well as the characteristics of the product or service. While the lack of simple answers might seem a problem, the rich interactions between these decisions present a powerful opportunity for differentiation.

References

Balabanovic, M., Shoham, Y., 1997. Fab: content-based, collaborative recommendation. Communications of ACM 40 (3), 66-72.

Burke, R., 2002. Hybrid recommender systems: survey and experiments. User Modeling and User-Adapted Interaction 12, 331-370.

Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., Sartin, M., 1999. Combining content-based and collaborative filters in an online newspaper. Proceedings of ACM SIGIR '99 Workshop on Recommender Systems: Algorithms and Evaluation.

Danielson, D., 2007. Source equivocality and claim verifiability in remote product evaluation. Doctoral dissertation, Stanford University.

Degemmis, M., Lops, P., Semeraro, G., Costabile, M., Guida, S., Licchelli, O., 2004. Improving collaborative recommender systems by means of user profiles, in: Karat, C-M., Blom, J. O., Karat, J. (Eds.), Designing Personalized User Experiences in eCommerce. Kluwer Academic Publishers, the Netherlands, pp. 253–274.

Estes, W. K., 1997. On the communication of information by displays of standard errors and confidence intervals. Psychonomic Bulletin & Review 4(3), 330-341.

Goldberg, D., Nichols, D., Oki, B. M., Terry, D., 1992. Using collaborative filtering to weave an information tapestry. Communications of the ACM 35(12), 61-70.

Gong, L., Nass, C., 2007. When a talking-face computer agent is half-human and half-humanoid: human identity and consistency preference. Human Communication Research 33, 163-193.

Herlocker, J. L., Konstan, J. A., Riedl, J., 2000. Explaining collaborative filtering recommendations. Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, 241-250.

Hill, W., Stead, L., Rosenstein, M., Furnas, G., 1995. Recommending and evaluating choices in a virtual community of use. CHI '95: Conference Proceedings on Human Factors in Computing Systems, 194-201.

Isbister, K., Nass, C., 2000. Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. International Journal of Human-Computer Studies 53(1), 251-267.

Iyengar, S. S., Lepper, M. R., 2000. When choice is demotivating: can one desire too much of a good thing? Journal of Personality and Social Psychology 79(6), 995-1006.

Johnson, H., Johnson, P., 1993. Explanation facilities and interactive systems. Proceedings of the 1st International Conference on Intelligent User Interfaces, 159-166.

Klein, L. R., 1998. Evaluating the potential of interactive media through a new lens: search versus experience goods. Journal of Business Research 41(3), 195-203.

Konstan, J. A., Riedl, J., Borchers, A., Herlocker, J. L., 1998. Recommender systems: a grouplens perspective. Recommender Systems: Papers from the 1998 Workshop (AAAI Technical Report WS-98-08), 60-64.

Lambert, L. Woodford, C., Poole, H., Moschovitis, C. J. P. (Eds.), 2005. The internet: a historical encyclopedia. ABC-CLIO, Santa Barbara, CA.

Lang, K., 1995. Newsweeder: learning to filter news. Proceedings of the 12th International Conference on Machine Learning, 331-339.

Loftus, G. R., Masson, M. E. J.,1994. Using confidence intervals in within-subject designs. Psychological Bulletin & Review 1(4), 476-490.

Maute, M. F., Forrester, W. R., 1991. The effect of attribute qualities on consumer decision making: a causal model of external information search, Journal of Economic Psychology 12, pp. 643–666.

Nass, C., Brave, S., 2005. Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship. MIT Press, Cambridge, MA.

Nass, C., Brave, S., Takayama, L., 2006. Socializing consistency: from technical homogeneity to human epitome, in: Zhang, P., Galletta. D. (Eds.), Human-Computer Interaction in Management Information Systems: Foundations. M. E. Sharpe, Armonk, pp. 373-391.

Nass, C., Lee, K., 2001. Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. Journal of Experimental Psychology: Applied 7(3), 171-181.

Nelson, P. J., 1970. Information and consumer behavior. Journal of Political Economy 78(2), 311-329.

Nelson, P. J., 1974. Advertising as information. Journal of Political Economy 82(4), 729-754.

Nelson, P. J., 1976. Economic value of advertising, in: Brozen, Y. (Ed.), Advertising and Society. New York University Press, New York, pp. 109-141.

Nelson, P. J., 1981. Consumer information and advertising, in: Galatin, M., Leiter, R. D. (Eds.), Economics of Information. N. Nijhoff Publishers, Boston, pp. 42-77.

Nielsen, J., 1994. Enhancing the explanatory power of usability heuristics. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence, 152-158.

Pazzani, M. J., 1999. A framework for collaborative, content-based and demographic filtering. Artificial Intelligence Review 13(5-6), 393-408.

Reeves, B., Nass, C., 1996. The Media Equation: How People Treat New Media Like Real People and Places. Cambridge University Press, New York.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J., 1994. GroupLens: an open architecture for collaborative filtering of netnews. Proceedings of the Conference on Computer Supported Cooperative Work, 175-186.

Resnick, P., Varian, H. R., 1997. Recommender systems. Communications of the ACM 40(3), 56-58.

Sarwar, B. M., Karypis, G., Konstan, J., Riedl, J., 2002. Recommender systems for large-scale e-commerce: scalable neighborhood formation using clustering. Proceedings of the Fifth International Conference on Computer and Information Technology. December 2002.

Schafer, J. B., Konstan, J., Riedl, J., 1999. Recommender systems in e-commerce. Electronic Commerce: Proceedings of the First ACM Conference on Electronic Commerce, 158-166.

Shardanand, U., Maes, P., 1995. Social information filtering: algorithms for automating "word of mouth". ACM Special Interest Group on Computer-Human Interaction, 210-217.

Sinha, R., Swearingen, K., 2002. The role of transparency in recommender systems. CHI '02 Extended Abstracts on Human Factors in Computing Systems, 830-831.

Terveen, L., Hill, W., 2001. Human-computer collaboration in recommender systems, in: Carroll, J. (Ed.), Human Computer Interaction in the New Millenium. Addison-Wesley, New York, pp. 223-242.

Tintarev, N., Masthoff, J., 2007. Effective explanations of recommendations: user-centered design. In ACM Recommender Systems '07, Minneapolis, MN.

Web 2.0, 2009. Wikipedia. http://en.wikipedia.org/wiki/Web_2.0.