Recommendation Systems and Consumer Satisfaction Online:
Moderating Effects of Consumer Product Awareness

Ling-Ling Wu
llwu@management.ntu.edu.tw
Dept. of Information Management, National Taiwan University, Taipei, Taiwan

Yuh-Izer Joung
joung@ntu.edu.tw

Jonglin Lee
fatfatfat@gmail.com

Abstract
Recommender systems are popular tools in e-commerce websites for helping people find the products that fit them. Many algorithms have been proposed, mostly to improve their accuracy. The underlying assumption is that accuracy will increase user satisfaction and ultimately lead to high purchase intention. However, past research has suggested that accuracy does not necessarily lead to satisfaction. As consumers usually being aware of at least some products when using recommenders, could this assumption depend on different types of recommender systems and consumer product awareness? This study examines the effects of popular recommenders such as collaborative-filtering and content-based systems to see if they have different effects on user satisfaction and willingness to purchase for consumers with different types of product awareness.

1. Introduction
Thanks to the success of the Internet, the number of products available online has increased exponentially in the past decades [1], along with numerous discussions and comments about them. This in turn has pushed more and more people to engage in online shopping activities. Too much information, however, has also made it difficult for consumers to locate suitable products [2]. Recommender systems are a popular technique to relieve this information overload problem, as they help users find products based on, say, their purchase histories, product ratings and/or personal profiles [3], [4].

As recommenders become more and more popular, measuring their effects becomes an important issue for e-commerce. Most of the previous researches on this problem focused on the efficiency and accuracy of the recommender algorithms. Techniques such as collaborative-filtering [5], content-based, [7] and hybrid recommenders [8] were developed and extensively researched upon, with one particular goal in mind: to improve the accuracy in predicting the purchase probability of the users. The underlying assumption is that optimal prediction ultimately leads to better user satisfaction and more purchase. However, previous researches have also shown that user satisfaction does not always correlate with high recommender accuracy [9], [10].

User satisfaction is widely considered a crucial factor of information system success [11], [12]. Several studies indicate that user satisfaction is one of, if not the most, important consumer reactions in B2C online environments [13]. Furthermore, user satisfaction has been linked to a number of important behavioral outcomes such as customer loyalty [14], repurchase intention [15], word of mouth [16] and willingness to purchase [17]. Among them, willingness to purchase holds particular importance. Whereas satisfaction is considered a cognitive and affective outcome [18], willingness to purchase is an indicator that directly influences a company’s financial performance. Therefore, the first purpose of this study is to examine how popular recommendation systems such as collaborative-filtering and content-based systems affect user satisfaction and willingness to purchase.

In addition, this study also considers the influence of a consumer characteristic, namely product awareness, on the effects of recommender satisfaction and willingness to purchase. Consumer product awareness refers to the set of products that the user is initially aware of before attending to the help of any recommender system [19]. It is important in that consumers are likely to compare recommended products against products that they are already aware of when making judgments. Furthermore, recommenders with a high rate of accuracy tend to recommend items already known to consumers, which would neither help consumers, nor increase their satisfaction. The rationale is: recommending to consumers new products similar to their product awareness is more likely to increase their satisfaction. Therefore, the second purpose of this study is to examine the role that product awareness plays in moderating the effects of recommenders on user satisfaction and willingness to purchase.

2. Literature Review
2.1. Effects of Recommender Systems: User Satisfaction and Willingness to Purchase
Two popular recommender strategies are collaborative-filtering (CF) and content-based (CB) systems respectively. CF systems recommend items by matching a user’s tastes to other similar users (or called peers) [5]. It collaboratively filters (hence the name) user data from multiple data resources such as agents,
viewpoints and ratings to match a consumer with peers that are highly correlated (i.e., similar), and then recommends most popular items among the peers to the consumer. CB systems, on the other hand, recommend products based on the product’s attributes, instead of other users’ purchase history or ratings [7]. CB systems attempt to match a user’s profile (i.e., personal preference) with the product’s attributes, and recommend items highly correlated with the user’s profile.

According to the past literature, CF systems recommend products that have been popularly rated or purchased among similar consumers such they tend to recommend popular or hit products. CB systems, on the other hand, recommend products according to personal preference. As preferences are largely diverse between individuals, CB systems tend to recommend more obscure and niche products.

User Satisfaction and willingness to purchase have been widely adopted as an important factor for information systems’ success[11],[12],[17]. The goal of recommenders as a whole is to help consumers locate preferable products from the vast number of products available. Hence, with the help of recommenders, the user should gain more satisfaction and be more willing to purchase products compared to contexts without recommenders. Previous researches that have evaluated the impact of recommenders on user satisfaction have adopted different definitions and dimensions as measures of satisfaction. For instance, Su, et al. [20] found that the presence of recommendation agents increased satisfaction through lowered perceived costs of the information search and heightened perceived benefits of the decision making process. Swearingen and Sinha [21] adopted perceived usefulness, satisfaction with recommended products and satisfaction with the recommendation system as dependent variables on the quality of recommenders. Their results indicate that effective recommenders inspire trust in the system, point users towards new, not-yet-experienced items and provide details about recommended items. That is, a satisfactory recommender needs to recommend new products that fit their preferences or tastes. However, recommendation systems typically recommend products preferred by consumers based on past behavior of consumers. Therefore, recommenders with a high rate of accuracy tend to recommend items already known to consumers, which would neither help consumers, nor increase their satisfaction. Previous researches have also shown that user satisfaction does not always correlate with high recommender accuracy [9], [10]. Balancing between accuracy and satisfaction hence becomes a challenging task for current recommendation systems. Following the previous discussion, taking consumer product awareness into account could be a possible way to achieve such balance.

### 2.2. Product Awareness

A user’s awareness set is the set of products that the user is initially aware of before attending to the help of recommender systems[19]. Consumers could acquire product information from being passively exposed to mass media or actively search for product information based on their own needs or preferences. Therefore, the variable of consumer product awareness could be formulated in terms of product exposure and personal preferences. Product exposure is defined as any opportunity for a reader, viewer or listener to see and/or hear an advertising or discussion message in a particular media vehicle [22]. Popular media vehicles include TV, radio, internet portals and Internet forums, etc. Consumers could also actively search for product information of their interests. As Anderson has noted [23], the growth of the Internet has allowed consumers to easily locate products that truly fit their needs. Although such products might not receive much exposure, with the help of search engines or other information tools, consumers can nonetheless easily locate them without much effort.

To sum up, product exposure and personal preference play crucial roles in the formulation of a user’s product awareness. Product exposure is more of a “push” factor, and is decided by the amount of product information the user unintentionally receives while doing unrelated activities, such as being shown products ads while watching TV or surfing the web. Personal preference, on the other hand, is more of a “pull” factor, and is decided by the amount of product information the user intentionally gets according to his own will, such as seeking information regarding a movie. If consumers rely mostly on the “push” mode for acquiring product information, the set of products in their awareness will likely consist of mostly popular items. On the other hand, if a consumer more relies on active search for products that fit his specific needs, then the set of products in his awareness will be more likely to include niche products that fit his tastes as preferences are largely diverse between individuals. We named the former consumer awareness as the “hit awareness”, while the latter as the “niche awareness”. Consumers with hit awareness are aware of relatively more hit products compared to niche products, and vice versa for their counterparts with niche awareness. Of course, most consumers exhibit both hit and niche awareness nature. However, to understand how the awareness type factor affects recommendation systems, we shall consider consumers of relatively hit-awareness oriented and consumers of relatively niche-awareness oriented. The dichotomy is a common practice for academic inquiry of such nature.

Past research has indicated that when consumers are faced with a purchase decision, high product familiarity will result in higher satisfaction and repurchase intention than those of low familiarity [24], [25]. Further, recommended products that have
According to the above discussion, we hypothesize the satisfaction and willingness to purchase levels as relevant and familiar, hence yielding higher deem recommendations produced by CF (CB) systems. Consumers of hit (niche) awareness will most likely recommend products according to personal preference. As preferences are largely diverse between individuals, CB systems tend to recommend more obscure and niche products. Therefore, consumers of hit (niche) awareness will most likely deem recommendations produced by CF (CB) systems as relevant and familiar, hence yielding higher satisfaction and willingness to purchase levels. According to the above discussion, we hypothesize the following:

H1: There is a significant interaction effect between recommendation system types and consumer product awareness on satisfaction with recommended products. Consumers with hit awareness are more satisfied with recommended products when aided by CF recommenders, whereas consumers with niche awareness are more satisfied with recommended products when aided by CB recommenders.

H2: There is a significant interaction effect between recommendation system types and consumer product awareness on willingness to purchase recommended products. Consumers with hit awareness are more willing to purchase recommended products when aided by CF recommenders, whereas consumers with niche awareness are more willing to purchase recommended products when aided by CB recommenders.

H3: There is a significant interaction effect between recommendation system types and consumer product awareness on numbers of recommended products chosen to be purchased. Consumers with hit awareness will accept more recommendations when aided by CF recommenders, whereas consumers with niche awareness will accept more recommendations when aided by CB recommenders.

H4: There is a significant interaction effect between recommendation system types and consumer product awareness on satisfaction towards website. Consumers with hit awareness are more satisfied with the website when aided by CF recommenders, whereas consumers with niche awareness are more satisfied with the website when aided by CB recommenders.

3. Experiment Design

The independent variables were product awareness (hit and niche) and recommender strategy (CF, CB and no recommender), with all of them being between-subject variables. Therefore, we adopted a 2 x 3 between-subjects factorial design for our experiment. We selected DVD movies as our experiment product of choice, and online DVD rental as our experiment context. Below we describe the experiment details.

3.1. Experiment Material

3.1.1. Choice of Product. A number of prerequisites were considered for the choice of the product set and experiment context. First of all, we would like the set to have a diverse catalog consisting of both hit and niche products to allow our awareness levels to take effect. In addition, the set needs to be both large in size and rich in content to increase the need for recommenders.

DVD movies and online DVD rental fit these needs due to the following reasons. First, DVD movies span a lot of genres, have numerous alternatives and are diverse enough to meet the needs of various kinds of consumers. Secondly, there are “best-sellers” or “new releases” present at all times that are highly exposed. Thirdly, several movie datasets are available online, allowing us to retrieve relevant information and provide recommendations. Finally, online DVD rental service is increasingly popular and is gradually replacing brick-and-mortar alternatives. Hopefully, our results in this study can provide some practical guidelines for building this important service in e-commerce.

3.1.2. Dataset. We used the Netflix Prize dataset as both our basis for recommendation calculation and DVD movie catalog for our simulated online retailer. The Netflix Prize dataset contains 100,480,507 ratings of 17,770 movies (released between 1927 and 2005) rated by 480,189 users. In order to facilitate recommendation calculation for both CF and CB recommenders, the users’ preference vectors and movies’ attributes vectors with matching dimensions need to be elicited from the data. As the Netflix Prize dataset lacks any attribute for the movies other than release year, we obtained another set of movie data from IMDB and matched the two datasets to obtain sufficient movie attributes for our recommenders. The IMDB dataset classifies movies into 23 genres, and each movie can belong to more than one genre.

In addition, we need movie information such as movie posters and movie descriptions to allow our participants to search and browse for movie information on our simulated online retailer. As our experiment was based in Taiwan where English is not the primary language, the additional movie information needs to be in Traditional Chinese. The IMDB dataset, however, contains information only in English. Therefore, we also resorted to a commercial movie website based in Taiwan, AtMovies, and retrieved information on our simulated online retailer.
Chinese information of a movie from the site. We retrieved 18,524 titles from AtMovies with Mandarin title, director, actor list, movie length, movie poster and movie description information. We then matched them with the movies from the Netflix and IMDB datasets, and obtained 4,021 movies for our experiment.

Several reasons can explain why there were only 4,021 matches between the AtMovies dataset (which contains 18,524 titles) and the Netflix+IMDB dataset (which contains 17,770 movies titles). First, the AtMovies dataset consisted of movies from both the Western and Eastern hemispheres, whereas the Netflix+IMDB dataset consisted of mostly Western movies. Second, a portion of movies in the AtMovies dataset (3,516, 18.9%) contained only English title and release year, making such movies impossible to be included in the experiment. Third, the AtMovies dataset also contained movies that were either released after 2005 (1,932, 10.4%) or was originated from Asia (3,327, 17.9%), which were not listed in the Netflix+IMDB dataset. Finally, the Netflix+IMDB dataset contained multiple versions of the same movie for some titles, usually in the form of “Bonus Material” or “Extended Edition”. Such multiple versions only counted as one movie in our experiment.

Although at the end we included only 4,021 movies of the original 17,770 movies from the Netflix+IMDB dataset, the former can still represent a valid sample of the latter because of the following reasons. First, the 4,021 (22.6%) movies still contained more than 72,778,413 (72.4%) ratings rated by 478,636 (99.7%) of the users, indicating that the 4,021 movies were mostly relevant movies that have been extensively rated by Netflix users. Second, they have similar cumulative ratings distributions. As shown in Fig. 1, our final 4,021-movie dataset still retained the long-tail aspect of the original Netflix dataset, fulfilling our requirement for both hit and niche products to be present.

3.1.3. Experiment Website. The experiment context is online DVD rental. We implemented a simulated online retailer, "Movie Rental King", for the study. The decision to design our own website instead of adopting a commercial website was to prevent any possible biases from participants’ past usage experiences with online DVD rental services. In addition, we would like to minimize the amount of influence from popular tools that are prominently used in practice (such as top-10 best-selling lists) to prevent them from interfering with the participants’ decision making process.

The homepage of the website consisted of a search bar to allow users to perform keyword search on our movie dataset. The users can also browse movies either by genre, title (both in English and Mandarin), release year or a combination of the four attributes. We kept the website design simple so as to prevent possible biases induced by unnecessary manipulations and to let participants concentrate solely on the tasks the experiment expects them to do.

3.2. Independent Variables

3.2.1. Product awareness. A user’s initial awareness set is the set of products that the user is initially aware of before attending to the help of any recommender system. A user is categorized as either hit or niche awareness by the proportions of hit and niche products in the set. A user that is aware of relatively more hit products than niche products were considered hit awareness, and vice versa for niche awareness.

To distinguish these two types of users, we first need to define “niche products”. We followed the Pareto principle, and categorized the bottom 80% (3,217 movies) of our DVD movie catalog (a total of 4,021 movies) ranked by relative popularity (number of ratings incurred by each movie) as niche products.

In the beginning of the experiment, participants were presented a random list of 48 movies (with their titles and posters) chosen from our movie catalog of 4,021 movies. They were instructed to choose the movies they have previously known or heard of from the list. The percentage of niche products out of every chosen movie from the list were then calculated to form the basis in grouping the participant as either hit awareness, niche awareness or neither of the two (hybrid). Those that scored higher than 0.50 (at least 50% of selected movies were niche movies) were grouped as niche awareness, and those that scored lower than 0.33 (no more than 33% of selected movies were niche movies) were grouped as hit awareness. The boundaries of 0.50 and 0.33 were chosen in a pre-test of 70 subjects to ensure that only the top 1/3 and bottom 1/3 portions of the total sample were regarded as valid subjects. In addition, to ensure the strength of the effect of the awareness measure, outliers that were aware of too few movies (less than 5) were considered invalid and neglected from our analyses.
3.2.2. Recommender strategy. We consider two types of recommendation strategies: CF and CB. The detailed algorithms for implementing the strategies will be presented in Section 3.4.

3.3. Dependent Variables

The participants were requested to perform the following tasks: pre-purchase information seeking, purchase decision, and evaluate the website after purchase. Therefore the dependent variables were measured in the two stages: pre-purchase stage and post-purchase stage. On the pre-purchase stage, satisfaction towards recommended products, willingness to rent recommended products, were measured by using scales developed in previous researches [27, 14] [21]. For satisfaction towards recommended products, we adopted single-item measures that specifically asked how satisfied the participants were with each of the recommended products presented to the participants. Measures were obtained using a semantic difference ranging from 1 to 5, with 1 being “Extremely Unsatisfied” and 5 being “Extreme Satisfied”. For willingness to rent recommended product, we also adopted single-item measures that specifically asked the users, on a scale of 1 to 5, how willing they were to rent each of the recommended products. In the pre-purchase stage, participants were instructed to choose 3 DVDs, either from their own browse or from recommendations. Therefore, number of products chosen from recommendation was also calculated to evaluate the effects of recommendation systems. This variable refers to the number of recommended products that were actually chosen for purchase by the participants in the experiment. As participants were instructed to select a total of 3 products, the range of this dependent variable was between 0 and 3.

In the post-purchase stage, three dependent variables were measured: satisfaction towards the 3 chosen products, willingness to rent the 3 chosen products, and satisfaction towards website in general. Given that the participants were instructed to choose 3 DVDs, their satisfaction and willingness to rent these chosen DVDs could vary. Similar to the construct of satisfaction towards recommended products, these two dependent variables were also measured with 5-point semantic single-item measures. We adopted an overall evaluation of the e-commerce system. Finally, for satisfaction towards the website, we adopted the 21-item measure developed by McKinney, Yoon and Zahedi’s dimensions for web satisfaction [28]. Their research developed six dimensions to measure web satisfaction that involve information quality (understandability, reliability and usefulness) and system quality (access, usability and navigation). All items were measured using a 5-point Likert-type scale.

3.4. Recommendation Algorithms

To implement CF and CB recommendation strategies, we need information about movies and users. For movies, we use the genres a movie belongs to represent the movie. Let \(x.Attr[i]\) be the Boolean value indicating if a movie \(x\) is of genre \(i\). The vector \(x.Attr\) then represents a movie’s attributes. For users, we use \(u_PREF[i]\) to represent a user \(u\)’s preference level (on a scale of 1 to 5) towards genre \(i\). The vector \(u_PREF\) then represents \(u\)’s preference.

For each subject \(u\) in the experiment, we collected \(u_PREF\) as follows. At the beginning of the experiment participants were presented a questionnaire containing a list of 23 genres (Action, Adventure, Animation, Comedy, Crime, Documentary, Drama, Fantasy, Family, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fi, Short, Thriller, War, Western, History, Sport, Biography, Music) adopted from IMDB, and were instructed to rate each genre according to their preference levels on a 5-point semantic difference. The results of the questionnaire were then used to obtain the participants’ preferences.

Next, we discuss the recommendation algorithm. For CF recommendation, we first calculated similarity between a participant \(p\) and each movie \(x\) as follows:

\[
\text{sim}(p, x) = \frac{\sum_{i} pROAD[i] \times x.ATTR[i]}{\sum_{i} x.ATTR[i]}
\]

Suppose \(x_m\) was the movie that had the highest similarity with the participant \(p\). Then, \(p\)’s candidate set of interested movies was defined as \(\{x \in X \mid \text{sim}(p, x) \geq 0.8 \times \text{sim}(p, x_m)\}\) where \(X\) is the set of movies. After \(p\)’s candidate set has been formed, the average ratings of each movie in the candidate set rated by every Netflix user were calculated to form a score for each movie. The movies were then ranked according to their respective scores, and the top 10 movies were recommended to the participant.

For collaborative filtering, we need to find for each participant \(p\) the set of users (called \(p\)’s neighbor set) that have a similar preference with \(p\), and then recommend most popular items among the neighbor set to \(p\). There are some difficulties, however. First, for the recommendation to be accurate, the process requires some amount of data on user preferences as well as on their purchases history (DVD renting in our context). The number of participants and the number of DVDs they intended to rent in the experiment are simply not enough. To overcome this problem, we resort to the Netflix dataset we screened in Section 3.1.1. As discussed in the section, we obtained 4,021 movies from the dataset, along with 72,778,413 ratings of them by 478,636 users. We used these Netflix users to calculate a participant’s neighbor set as follows: Let \(u\) represent a Netflix user. Suppose \(u\) has rated \(x_1, \ldots, x_n\)

\(^{1}\) For all 5-point scale in this study, 1 represents the lowest value while 5 represents the highest.
movies, with \( r_1, \ldots, r_n \) being the ratings, then \( u \)'s preference \( u.\text{Pref} \) was calculated by

\[
\text{u.\text{Pref}}[i] = \frac{\sum_{j=1}^{n} x_j \cdot \text{Attr}_j[i] \cdot r_i}{n}
\]

With the preferences from Netflix users, we used Pearson’s correlation to measure the similarity between a participant in our experiment \( p \) and each Netflix user \( u \):

\[
\text{sim}(p, u) = \frac{\sum_{j=1}^{n} p.\text{Pref}_j \cdot u.\text{Pref}_j - \left( \frac{\sum_{j=1}^{n} p.\text{Pref}_j \cdot \sum_{j=1}^{n} u.\text{Pref}_j}{n} \right) \cdot \left( \frac{\sum_{j=1}^{n} (p.\text{Pref}_j)^2 \cdot (u.\text{Pref}_j)^2}{n} \right)}{\sqrt{\left( \frac{\sum_{j=1}^{n} (p.\text{Pref}_j)^2 \cdot (u.\text{Pref}_j)^2}{n} \right) - \left( \frac{\sum_{j=1}^{n} p.\text{Pref}_i \cdot \sum_{j=1}^{n} u.\text{Pref}_j}{n} \right)^2}}
\]

Let \( u_m \) be the Netflix user that has the highest similarity with the participant \( p \). Then, \( p \)'s neighbor set was defined as \( p.\text{Neighbor} = \{ u \in U | \text{sim}(p, u) \geq 0.8 \times \text{sim}(p, u_m) \} \), where \( U \) is the set of Netflix users.

After finding a participant \( p \)'s neighbor set, we determined the so called “popular” movies in the set, and recommended the movies to \( p \). There are a number of ways to define “popularity”. In our preliminary test we evaluated some of them, and found that using “Sum of Ratings” to rank movies is a simple yet quite accurate approach, and thus adopted it in the experiment. In this ranking, a movie \( x \) is considered more popular than another movie \( y \) if the total rating scores \( x \) has garnered from the neighbor set is higher than \( y \) has. Based on this ranking, our CF algorithm recommended the top 10 most popular movies from a participant’s neighbor set to the participant.

### 3.5. Experiment Procedure

Participants were invited to our lab to conduct the experiment. Each of them filled out a consent form before the experiment, and was then asked to be seated in front of a computer where all instructions and experiment materials were presented. The experiment was divided into three stages. The first stage consisted of the preference collection step mentioned in the previous section, as well as the grouping of participants according to their awareness types. In this stage, participants were first presented the aforementioned preference questionnaire to obtain their preferences vectors. Next, to determine which awareness (hit or niche) groups they belong to, participants were presented measurement materials for the product awareness factors. Only participants that were successfully grouped to one of the two possible conditions (hit/niche awareness) were considered valid samples in our experiment. All participants were instructed to finish the entire experiment, although only data from the valid samples were used in our empirical analyses. After the participants have finished taking the preference questionnaire and the grouping test, they were then randomly assigned to a recommender strategy (CF, CB or no recommender) and were led into the second stage of our experiment. Note that participants were not aware of their measured awareness types, as well as the recommender condition assigned to them.

In the second stage of our experiment, participants were presented with our online DVD rental website “Movie Rental King”, and instructed to rent 3 DVD movies from the website. Participants can search or browse for DVD movies. The results were ranked and presented according to the movies’ average Netflix ratings, with 10 results shown per page. For participants assigned to the CF and CB recommenders, in addition to the 10 search or browse results, 10 DVD movies recommended by their respective recommenders were also presented. Please note that participants were instructed to choose 3 DVDs and those 3 DVDs could be chosen either from the movies they browsed or searched on their own, or from the movies recommended by the systems.

After the participants have decided on 3 final DVDs to rent in their shopping cart, they were then taken to the third stage of the experiment and were asked to complete a questionnaire to obtain their decision on how satisfied and willing to rent the final 3 chosen DVD movies they were. In addition, they were also told to report their satisfactory levels towards the website. Participants assigned to the CF and CB recommender conditions were also asked to report satisfaction and willingness to rent levels for the 10 recommended results presented to them. The entire experiment took roughly 30 minutes to complete, and each participant was given NT$100 as reward for their participation. ANOVA was used to test all our hypotheses.

### 3.6. Participants

A total of 383 participants were recruited from Taipei, Taiwan to attend the experiment. Out of the 383 participants, 180 (47%) were valid (successfully grouped as hit or niche awareness) samples, with 30 samples in each of the 6 experiment conditions. Among them, there were 91 (50.5%) male and 89 (49.5%) female; 155 (86.1%) were students, and 25 (13.9%) were non-students. Forty (22.2%) were aged 19 years or below, 121 (67.2%) between 20 and 24 years old, and 19 (10.6%) aged 25 years or above.

### 4. Experiment Results

For the pre-purchase information seeking and decision stage, the effects of recommendation systems and consumer awareness will be assessed in terms of satisfaction towards the recommended products, willingness to rent them, and number of products chosen from recommendation by the participants. Furthermore, on the post-purchase stage, three variables will also be assessed: satisfaction towards the 3 chosen products and willingness to rent them, and
satisfaction towards the website. Specifically, we will test the interaction effects proposed in the hypotheses.

4.1. Pre-purchase information seeking and decision making stage

The empirical results of this stage demonstrate a consistent pattern: the main effects of recommendation systems and product awareness are not significant whereas the interaction effect of these two variables is statistically significant for three different dependent variables (satisfaction towards recommended products, willingness to rent recommended products, and number of products chosen from recommendations). Please note that the condition of NO recommender is not applicable at this stage because all the dependent variables are regarding recommended products.

The main effects of recommendation systems and consumer product awareness are not significant for satisfaction towards recommended products (recommendation systems: CF: 3.42 vs. CB: 3.28, F(1, 30.5) = 1.54, MSe = 3.92, p = .22; product awareness: F(1,42.2) = 3.11, MSe = 3.18, p = .08). Nonetheless, as hypothesized, the interaction effect of these two variables is highly significant (F(1,55.9) = 7.54, MSe = 2.655, p < .01). As shown in Fig. 2, the interaction pattern between the two factors was consistent with our hypothesis. Specifically, participants assigned to the CF recommender were no more satisfied with the recommended products when they were of hit awareness (3.46, SD = 1.02) than niche awareness (3.38, SD = 1.09; t(598) = 0.82, p = .41). On the other hand, participants assigned to the CB recommender were significantly more satisfied when they were niche (3.50, SD = 0.97) than hit awareness (3.06, SD = 1.05; t(598) = 5.34, p < .0001). Thus, Hypothesis 1 was strongly supported by the empirical results.

As shown in Fig. 3, participants assigned to the CF recommender were more willing to rent the recommended products when they were of hit awareness (3.13, SD = 1.16) compared to niche awareness (2.87, SD = 1.36, t(598) = 2.45, p < .05). On the other hand, participants assigned to the CB recommender were significantly more satisfied when they were niche (3.34, SD = 1.10) than hit awareness (2.87, SD = 1.27, t(598) = 4.81, p < .0001). Thus, Hypothesis 2 was also strongly supported.

4.2 Post purchase stage

The empirical results of willingness to rent recommended products are similar to that of satisfaction towards recommended products. The main effects of recommendation systems and product awareness are not significant (recommendation systems: CF: 3.00 vs. CB: 3.10, F(1, 34.1) = 0.47, MSe = 6.832, p = .50; product awareness: F(1,56.5) = 0.62, MSe = 5.147, p = .43). However, the interaction effect of these two variables is, again, highly significant (F(1,55.3) = 7.60, MSe = 5.210, p < .01).

As shown in Fig. 3, participants assigned to the CF recommender were more willing to rent the recommended products when they were of hit awareness (3.13, SD = 1.16) compared to niche awareness (2.87, SD = 1.36, t(598) = 2.45, p < .05). On the other hand, participants assigned to the CB recommender were significantly more satisfied when they were niche (3.34, SD = 1.10) than hit awareness (2.87, SD = 1.27, t(598) = 4.81, p < .0001). Thus, Hypothesis 2 was also strongly supported.

Fig. 2: Satisfaction towards recommended products

Fig. 3: Willingness to rent recommended products.

For the results of number of products chosen from recommendations, the main effects of recommendation systems and product awareness are not significant either (recommendation systems: CF: 1.08 vs. CB: 1.10, F(1,112) = 0.01, MSe = 1.013, p = .93; product awareness: F(1,112) = 0.21, MSe = 1.013, p = .65). Once again, the interaction effect between the awareness and recommendation factors was significant although the significance was marginal (F(1,112) = 2.97, MSe = 1.013, p = .09). As demonstrated in Fig. 4, participants assigned to the CF recommender chose more products from the recommendations list when they were of hit awareness (1.20, SD = 0.92) compared to niche awareness (0.97, SD = 1.00; t(58) = 0.94, p = .35). In contrast, participants assigned to the CB recommender chose more recommended product when they were niche (1.30, SD = 1.06) than hit awareness (0.90, SD = 1.12; t(58) = 1.42, p = .16). The pair-wise comparisons of the interaction are not significant even though the overall interaction effect is. Thus, Hypothesis 3 was partially supported.

Fig. 4: Number of products chosen from recommendation.
In this stage, we measured participants’ satisfaction towards the 3 chosen products, their willingness to rent recommended products, and their overall level of satisfaction towards the website. For the variable of satisfaction towards the chosen products, differences between the CF, CB and no recommendation conditions were statistically significant ($F(2,36.4) = 4.53, MSe = 0.934, p < .05$). As shown in Table 1, both the CF ($4.49, SD = 0.64, t(358) = 3.80, p < .001$) and CB ($4.43, SD = 0.71, t(358) = 2.94, p < .01$) groups resulted in significantly higher satisfaction than the no recommender group ($4.20, SD = 0.79$). In addition, participants assigned to the CF and CB groups showed no significant difference ($t(358) = 0.78, p = .44$) in their satisfaction towards final chosen products. These results suggest that both CF and CB successfully increase the satisfaction level towards the chosen products than their counterparts in the NO recommender conditions.

Table 1: Satisfaction towards the 3 chosen products

<table>
<thead>
<tr>
<th>Awareness</th>
<th>CF</th>
<th>CB</th>
<th>NO</th>
<th>Average (Hit vs. Niche)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit</td>
<td>4.42 (0.65)</td>
<td>4.51 (0.62)</td>
<td>4.27 (0.80)</td>
<td>4.40 (0.70)</td>
</tr>
<tr>
<td>Niche</td>
<td>4.56 (0.62)</td>
<td>4.36 (0.78)</td>
<td>4.13 (0.78)</td>
<td>4.35 (0.75)</td>
</tr>
<tr>
<td>Average (CF vs CB vs. NO)</td>
<td>4.49 (0.64)</td>
<td>4.43 (0.71)</td>
<td>4.20 (0.79)</td>
<td></td>
</tr>
</tbody>
</table>

It could be argued that satisfaction level boosted in both recommendation groups might not be necessarily resulted from quality recommendations. Instead, consumer might actively search more due to the “priming” of recommendations. In our experiment, the 3 final products chosen by participants may be either from our recommendation list, or selected from the participants’ own browsing and search activities. Below we examine if these two sources may contribute to consumer satisfaction towards the chosen products and willingness to rent them, as shown in Table 2.

Table 2: Satisfaction towards the 3 chosen products and willingness to rent them

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Source</th>
<th>CF</th>
<th>CB</th>
<th>NO</th>
<th>Average (Hit vs. Niche)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction towards final chosen products</td>
<td>Recommended</td>
<td>4.63 (0.63)</td>
<td>4.25 (0.87)</td>
<td>4.20 (0.79)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Self-selected</td>
<td>4.41 (0.63)</td>
<td>4.53 (0.58)</td>
<td>(Self)</td>
<td></td>
</tr>
<tr>
<td>Willingness to rent final chosen products</td>
<td>Recommended</td>
<td>4.58 (0.61)</td>
<td>4.25 (0.87)</td>
<td>4.03 (1.00)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Self-selected</td>
<td>4.26 (0.83)</td>
<td>4.58 (0.70)</td>
<td>(Self)</td>
<td></td>
</tr>
</tbody>
</table>

Please note that product information source in the No recommender condition must be self selected given no recommendations. We compare products that were selected from the recommender’s list and products that were self-selected, under the same recommender condition. As seen in Table 2, under the CF condition, both satisfaction ($4.63, SD = 0.63$) and willingness to rent ($4.58, SD = 0.61$) recommended products garnered significantly higher ratings than their self-selected satisfaction ($4.41, SD = 0.63, t(178) = 2.27, p < .05$) and willingness to rent ($4.26, SD = 0.83 t(178) = 2.76, p < .01$) counterparts. However, under CB conditions, the result is opposite: self-selected products garnered significantly higher satisfaction ($4.53, SD = 0.58$) and willingness to rent ($4.58, SD = 0.70$) than the satisfaction ($4.25, SD = 0.87, t(178) = 2.62, p < .01$) and willingness to rent ($4.25, SD = 0.87, t(178) = 2.75, p < .01$) of recommended products. The results suggest that CF directly increased satisfaction and willingness to purchase through their recommended products, whereas CB recommender might expand the search scope of consumers through priming new products that might interest them.

Lastly, we assess the results of satisfaction towards website in general. As shown in Table 3, The satisfaction levels in the three recommendation condition do not significantly differ from each other ($F(2,168) = 0.09, MSe = 0.337, p = .91$). Furthermore, participants under the hit awareness and niche awareness conditions showed no difference in their satisfaction levels towards the website ($F(1,112) = 0.10, MSe = 0.382, p = .75$). Finally, no interaction effect was observed between awareness and recommendation on satisfaction towards website ($F(1,112) = 0.93, MSe = 0.382, p = .34$). The results suggest that the recommenders do not affect consumer’s overall evaluation of the website even though they do affect consumer satisfaction towards the recommended products and willingness to rent them. It could be because that there are so many components in websites and recommender is only one of them. Therefore, albeit their significant influences on consumer purchase decision, recommenders do not affect user satisfaction level of the website. H4 was not supported by our results.

Table 3: Satisfaction towards the website

<table>
<thead>
<tr>
<th>Awareness</th>
<th>CF</th>
<th>CB</th>
<th>NO</th>
<th>Average (Hit vs. Niche)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit</td>
<td>3.65 (0.67)</td>
<td>3.72 (0.62)</td>
<td>3.67 (0.66)</td>
<td>3.68 (0.64)</td>
</tr>
<tr>
<td>Niche</td>
<td>3.74 (0.68)</td>
<td>3.60 (0.57)</td>
<td>3.64 (0.51)</td>
<td>3.66 (0.59)</td>
</tr>
<tr>
<td>Average (CF vs CB vs. NO)</td>
<td>3.70 (0.67)</td>
<td>3.66 (0.59)</td>
<td>3.66 (0.58)</td>
<td></td>
</tr>
</tbody>
</table>

The empirical results strongly support H1, H2, and at least partially support H3. Although no main effects existed for both awareness and recommendation on the
dependent variables, the interaction effects are highly significant. The results show that participants of hit awareness would favor CF systems and those of niche awareness would favor CB systems. However, for satisfaction towards website, no main effects or interaction effect existed.

5. Conclusions, Managerial Implications, Limitations and Future Research

5.1. Conclusions
Recommender systems have been popularly used in e-commerce websites to help people find their products. Abundant research has been conducted to evaluate their effects, mostly on their accuracy. However, accuracy alone is not enough given that accuracy does not necessarily lead to increase in user satisfaction and purchase intention. This study tries to fill up this gap and proposed that consumer awareness should be considered for CF and CB systems to satisfy online consumers and increase sales.

To conclude, this study contributes to the ongoing research on recommender impact by providing empirical evidence that the presence of recommenders increases satisfaction and willingness to purchase. First of all, results of the study suggest that product awareness plays an important role in moderating the impact of recommenders. Consumers of hit awareness will be more satisfied and willing to purchase products when they are helped by CF systems, whereas ones of niche awareness will favor CB systems more.

The results indicate that the presence of recommenders increased user satisfaction and willingness to purchase at the product level, but caused no difference on satisfaction towards website. This implies that in the e-commerce context, consumers being aided by recommenders will be more satisfied and willing to purchase products compared to ones that were not aided by recommenders. However, the presence of recommenders alone will not cause a significant difference on consumers’ satisfaction towards website. Perhaps other factors exist and should also be considered when measuring such a multi-dimensional concept. Finally, the CF and CB groups did not differ each other in both their satisfactory and willingness to purchase levels. This was consistent with the literature which stated that there was no clear advantage between these two popular recommenders.

Further analysis of the source of the products chosen by the participants revealed mixed results for CF and CB systems. Although products selected from both the recommended list and through self-selection garnered higher satisfactory and willingness to purchase results than that of the no recommender condition, participants under the CF condition felt more satisfied and willing to rent recommended products than self-selected products. However, participants under the CB condition felt the opposite, in that self-selected products garnered higher satisfaction and willingness to purchase levels than recommended products.

5.2. Managerial Implications
This research provides evidence that the presence of recommenders has a positive effect on user satisfaction and willingness to purchase. In addition, consumer product awareness was found to moderate the impact of recommenders. Therefore, not only should e-commerce providers seek to utilize the prowess of recommendations to increase satisfaction and willingness to purchase their products, they are further encouraged to identify consumer product awareness and adjust their recommendations accordingly to maximize their recommender’s impacts.

First of all, e-commerce providers should take into consideration individual differences on the product awareness dimension. Particularly, if a consumer had a relatively niche awareness set, chances are that CB systems would garner more positive responses on the satisfaction and willingness to purchase dimensions from that consumer. On the other hand, consumers with a more hit awareness set should be targeted with CF systems instead. The measurement of awareness proposed by this study can easily be extended to practice. For example, e-commerce providers could implicitly analyze their consumer’s browsing behaviors and count the number of niche products that the consumer has browsed.

5.3. Limitations and Future Research
There are a number of limitations in this study. First, the scale and method utilized in this study to measure awareness is new and has not been rigorously proved in the literature. More thorough investigation of its definition and measurement might be needed in the future.

Secondly, this study adopted the Netflix, IMDB and Atmovies datasets as a basis for both the recommendation and the website catalog. In addition, DVD movie was chosen as the experiment product and online DVD rental retailer as the experiment context. Although results of this research supported our hypotheses, the same results might not be obtained using a different dataset (i.e. the MovieLens dataset) or a different product context altogether. Future research could replicate the study by using other datasets and different contexts and products to generalize the results to other realms.

Third, consumer characteristics other than the ones chosen in this study might also have an effect on recommendation impact. For example, product awareness may be correlated with personal expertise or enduring involvement. Indeed, consumers that are experts or are enduringly involved with a particular product domain might be aware of a wider range of
niche products. In addition, trust and conformity may also have a relationship with the recommendation susceptibility. Therefore, future research could adopt the model and procedures proposed in this study to examine the effects other consumer characteristics have on recommendation impact.

Finally, the algorithms used by both the CF and CB recommenders in this study were just basic algorithms used in literature, as the focus here was not on the algorithms, but rather on the impacts of them. Still, we see that the CB recommender adopted in our study produced mixed results on their effect on satisfaction and willingness to purchase, as compared to the CF recommender we used. This might imply that the CB recommender algorithm we used has room to improve. Future research can adopt more sophisticated algorithms and examine their impacts.

Acknowledgment We thank the anonymous reviewers for their valuable comments and suggestions.

6. References