

Effects of Online Recommendations on Consumers' Willingness to Pay

Gediminas Adomavicius
University of Minnesota
Minneapolis, MN
gedas@umn.edu

Jesse Bockstedt
University of Arizona
Tucson, AZ
bockstedt@email.arizona.edu

Shawn Curley
University of Minnesota
Minneapolis, MN
curley@umn.edu

Jingjing Zhang
Indiana University
Bloomington, IN
jjzhang@indiana.edu

ABSTRACT

We present the results of two controlled behavioral studies on the effects of online recommendations on consumers' economic behavior. In the first study, we found strong evidence that participants' willingness to pay was significantly affected by randomly assigned song recommendations, even when controlling for participants' preferences and demographics. In the second study, we presented participants with actual system-generated recommendations that were intentionally perturbed (i.e., significant error was introduced) and observed similar effects on willingness to pay. The results have significant implications for the design and application of recommender systems as well as for e-commerce practice.

1. INTRODUCTION

Recommender systems have become commonplace in online purchasing environments. Much research in information systems and computer science has focused on algorithmic design and improving recommender systems' performance (see Adomavicius & Tuzhilin 2005 for a review). However, little research has explored the impact of recommender systems on consumer behavior from an economic or decision-making perspective. Considering how important recommender systems have become in helping consumers reduce search costs to make purchase decisions, it is necessary to understand how online recommender systems influence purchases.

In this paper, we investigate the relationship between recommender systems and consumers' economic behavior. Drawing on theory from behavioral economics, judgment and decision-making, and marketing, we hypothesize that online recommendations¹ significantly pull a consumer's willingness to pay in the direction of the recommendation. We test our hypotheses using two controlled behavioral experiments on the recommendation and sale of digital songs. In the first study, we find strong evidence that randomly generated recommendations (i.e., not based on user preferences) significantly impact consumers' willingness to pay, even when we control for user preferences for the song, demographic and consumption-related factors, and individual level heterogeneity. In the second study,

¹ In this paper, for ease of exposition, we use the term "recommendations" in a broad sense. Any rating that the consumer receives purportedly from a recommendation system, even if negative (e.g., 1 star on a five-star scale), is termed a recommendation of the system.

Paper presented at the 2012 Decisions@RecSys workshop in conjunction with the 6th ACM conference on Recommender Systems. Copyright © 2012 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

we extend these results and find strong evidence that these effects still exist with real recommendations generated by a live real-time recommender system. The results of the second study demonstrate that errors in recommendation, a common feature of live recommender systems, can significantly impact the economic behaviors of consumers toward the recommended products.

2. LITERATURE REVIEW AND HYPOTHESES

Behavioral research has indicated that judgments can be constructed upon request and, consequently, are often influenced by elements of the environment. One such influence arises from the use of an anchoring-and-adjustment heuristic (Tversky and Kahneman 1974; see review by Chapman and Johnson 2002), the focus of the current study. Using this heuristic, the decision maker begins with an initial value and adjusts it as needed to arrive at the final judgment. A systematic bias has been observed with this process in that decision makers tend to arrive at a judgment that is skewed toward the initial anchor.

Past studies have largely been performed using tasks for which a verifiable outcome is being judged, leading to a bias measured against an objective performance standard (e.g., see review by Chapman and Johnson 2002). In the recommendation setting, the judgment is a subjective preference and is not verifiable against an objective standard. This aspect of the recommendation setting is one of the task elements illustrated in Figure 1, where accuracy is measured as a comparison between the rating prediction and the consumer's actual rating, a subjective outcome. Also illustrated in Figure 1 is the feedback system involved in the use of recommender systems. Predicted ratings (recommendations) are systematically tied to the consumer's perceptions of products. Therefore, providing consumers with a predicted "system rating" can potentially introduce anchoring biases that significantly influence their subsequent ratings of items.

One of the few papers identified in the mainstream anchoring literature that has looked directly at anchoring effects in preference construction is that of Schkade and Johnson (1989). However, their work studied preferences between abstract, stylized, simple (two-outcome) lotteries. This preference situation is far removed from the more realistic situation that we address in this work. More similar to our setting, Ariely et al. (2003) observed anchoring in bids provided by students participating in auctions of consumer products (e.g., wine, books, chocolates) in a classroom setting. However, participants were not allowed to sample the goods, an issue we address in this study.

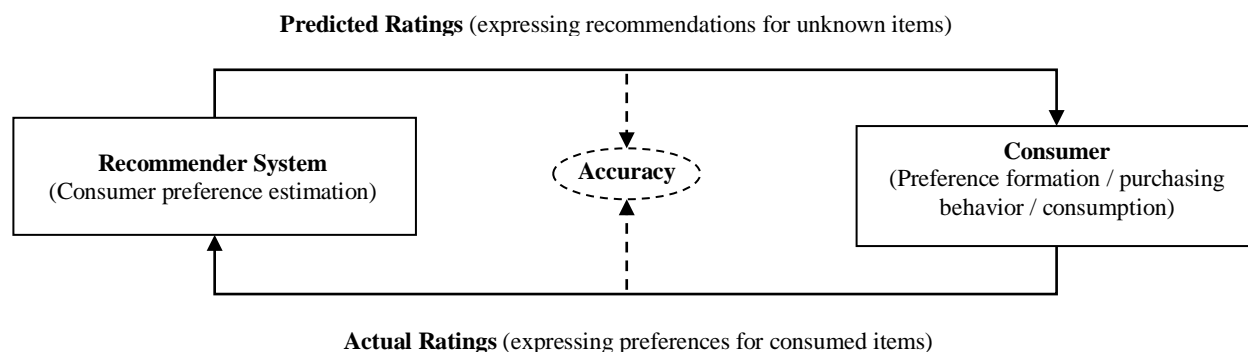


Figure 1. Ratings as part of a feedback loop in consumer-recommender interactions.

Very little research has explored how the cues provided by recommender systems influence online consumer behavior. Cosley et al. (2003) dealt with a related but significantly different anchoring phenomenon in the context of recommender systems. They explored the effects of system-generated recommendations on user re-ratings of movies. They found that users showed high test-retest consistency when being asked to re-rate a movie with no prediction provided. However, when users were asked to re-rate a movie while being shown a “predicted” rating that was altered upward or downward from their original rating by a single fixed amount of one rating point (providing a high or a low anchor), users tended to give higher or lower ratings, respectively (compared to a control group receiving accurate original ratings). This showed that anchoring could affect consumers’ ratings based on preference recall, for movies seen in the past and now being evaluated.

Adomavicius et al. (2011) looked at a similar effect in an even more controlled setting, in which the consumer preference ratings for items were elicited at the time of item consumption. Even without a delay between consumption and elicited preference, anchoring effects were observed. The predicted ratings, when perturbed to be higher or lower, affected the consumer ratings to move in the same direction. The effects on consumer ratings are potentially important for a number of reasons, e.g., as identified by Cosley et al. (2003): (1) Biases can contaminate the inputs of the recommender system, reducing its effectiveness. (2) Biases can artificially improve the resulting accuracy, providing a distorted view of the system’s performance. (3) Biases might allow agents to manipulate the system so that it operates in their favor. Therefore, it is an important and open research question as to the direct effects of recommendations on consumer behavior.

However, in addition to the preference formation and consumption issues, there is also the purchasing decision of the consumer, as mentioned in Figure 1. Aside from the effects on ratings, there is the important question of the possibility of anchoring effects on economic behavior. Hence, the primary focus of this research is to determine how anchoring effects created by online recommendations impact consumers’ economic behavior as measured by their willingness to pay. Based on the prior research, we expect there to be similar effects on economic behavior as observed with consumer ratings and preferences. Specifically, we first hypothesize that recommendations will significantly impact consumers’ economic behavior by pulling their willingness to pay in the direction of the recommendation, regardless of the accuracy of the recommendation.

Hypothesis 1: Participants exposed to randomly generated artificially high (low) recommendations for a product will exhibit a higher (lower) willingness to pay for that product.

A common issue for recommender systems is error (often measured by RMSE) in predicted ratings. This is evidenced by Netflix’s recent competition for a better recommendation algorithm with the goal of reducing prediction error by 10% (Bennet and Lanning 2007). If anchoring biases can be generated by recommendations, then accuracy of recommender systems becomes all the more important. Therefore, we wish to explore the potential anchoring effects introduced when real recommendations (i.e., based on the state-of-the-art recommender systems algorithms) are erroneous. We hypothesize that significant errors in real recommendations can have similar effects on consumers’ behavior as captured by their willingness to pay for products.

Hypothesis 2: Participants exposed to a recommendation that contains significant error in an upward (downward) direction will exhibit a higher (lower) willingness to pay for the product.

We test these hypotheses with two controlled behavioral studies, discussed next.

3. STUDY 1: RECOMMENDATIONS AND WILLINGNESS-TO-PAY

Study 1 was designed to test Hypothesis 1 and establish whether or not randomly generated recommendations could significantly impact a consumer’s willingness to pay.

3.1. Procedure

Both studies presented in this paper were conducted using the same behavioral research lab at a large public North American university, and participants were recruited from the university’s research participant pool. Participants were paid a \$10 fee plus a \$5 endowment that was used in the experimental procedure (discussed below). Summary statistics on the participant pool for both Study 1 and Study 2 are presented in Table 1. Seven participants were dropped from Study 1 because of response issues, leaving data on 42 participants for analysis.

The experimental procedure for Study 1 consisted of three main tasks, all of which were conducted on a web-based application using personal computers with headphones and dividers between

participants. In the first task, participants were asked to provide ratings for at least 50 popular music songs on a scale from one to five stars with half-star increments. The songs presented for the initial rating task were randomly selected from a pool of 200 popular songs, which was generated by taking the songs ranked in the bottom half of the year-end Billboard 100 charts from 2006 and 2009.² For each song, the artist name(s), song title, duration, album name, and a 30-second sample were provided. The objective of the song-rating task was to capture music preferences from the participants so that recommendations could later be generated using a recommendation algorithm (in Study 2 and post-hoc analysis of Study 1, as discussed later).

Table 1 Participant summary statistics.

| | Study 1 | Study 2 |
|---|-----------------------|-----------------------------------|
| # of Participants (n) | 42 | 55 |
| Average Age (years) | 21.5 (1.95) | 22.9 (2.44) |
| Gender | 28 Female, 14 Male | 31 Female, 24 Male |
| Prior experience with recommender systems | 50% (21/42) | 47.3% (26/55) |
| Student Level | 36 undergrad, 6 grad | 27 undergrad, 25 grad, 3 other |
| Buy new music at least once a month | 66.7% (28/42) | 63.6% (35/55) |
| Own more than 1000 songs | 50% (21/42) | 47.3% (26/55) |

In the second task, a different list of songs was presented (with the same information for each song as in the first task) from the same set of 200 songs. For each song, the participant was asked whether or not they owned the song. Songs that were owned were excluded from the third task, in which willingness-to-pay judgments were obtained. When the participants identified at least 40 songs that they did not own, the third task was initiated.

In the third main task of Study 1, participants completed a within-subjects experiment where the treatment was the star rating of the song recommendation and the dependent variable was willingness to pay for the songs. In the experiment, participants were presented with 40 songs that they did not own, which included a star rating recommendation, artist name(s), song title, duration, album name, and a 30 second sample for each song. Ten of the 40 songs were presented with a randomly generated low recommendation between one and two stars (drawn from a uniform distribution; all recommendations were presented with a one decimal place precision, e.g., 1.3 stars), ten were presented with a randomly generated high recommendation between four and five stars, ten were presented with a randomly generated mid-range recommendation between 2.5 and 3.5 stars, and ten were presented with no recommendation to act as a control. The 30 songs presented with recommendations were randomly ordered, and the 10 control songs were presented last.

To capture willingness to pay, we employed the incentive-compatible Becker-DeGroot-Marschack method (BDM) commonly used in experimental economics (Becker et al. 1984). For each song presented during the third task of the study, participants were asked to declare a price they were willing to pay between zero and 99 cents. Participants were informed that five songs selected at random at the end of the study would be assigned random prices, based on a uniform distribution, between one and 99 cents. For each of these five songs, the participant was required to purchase the song using money from their \$5 endowment at the randomly assigned price if it was equal to or below their declared willingness to pay. Participants were presented with a detailed explanation of the BDM method so that they understood that the procedure incentivizes accurate reporting of their prices, and were required to take a short quiz on the method and endowment distribution before starting the study.

At the conclusion of the study, they completed a short survey collecting demographic and other individual information for use in the analyses. The participation fee and the endowment minus fees paid for the required purchases were distributed to participants in cash. MP3 versions of the songs purchased by participants were “gifted” to them through Amazon.com approximately within 12 hours after the study was concluded.

3.2. Analysis and Results

We start by presenting a plot of the aggregate means of willingness to pay for each of the treatment groups in Figure 2. Note that, although there were three treatment groups, the actual ratings shown to the participants were randomly assigned star ratings from within the corresponding treatment group range (low: 1.0-2.0 stars, mid: 2.5-3.5 stars, high: 4.0-5.0 stars).

As an initial analysis, we performed a repeated measure ANOVA, as shown in Table 2, demonstrating a statistically significant effect of the shown rating on willingness to pay. Since the overall treatment effect was significant, we followed with pair-wise contrasts using t-tests across treatment levels and against the control group as shown in Table 3. All three treatment conditions significantly differed, showing a clear, positive effect of the treatment on economic behavior.

To provide additional depth for our analysis, we used a panel data regression model to explore the relationship between the shown star rating (continuous variable) and willingness to pay, while controlling for participant level factors. A Hausman test was conducted, and a random effects model was deemed appropriate, which also allowed us to account for participant level covariates in the analysis. The dependent variable, i.e., willingness to pay, was measured on an integer scale between 0 and 99 and skewed toward the lower end of the scale. This is representative of typical count data; therefore, a Poisson regression was used (overdispersion was not an issue). The main independent variable was the shown star rating of the recommendation, which was continuous between one and five stars. Control variables for several demographic and consumer-related factors were included, which were captured in the survey at the end of the study. Additionally, we controlled for the participants’ preferences by calculating an actual predicted star rating recommendation for each song (on a 5 star scale with one decimal precision), post hoc, using the popular and widely-used item-based collaborative

² The Billboard 100 provides a list of popular songs released in each year. The top half of each year’s list was not used to reduce the number of songs in our database that participants would already own.

filtering algorithm (IBCF) (Sarwar et al. 2001).³ By including this predicted rating (which was not shown to the participant during the study) in the analysis, we are able to determine if the random recommendations had an impact on willingness to pay above and beyond the participant's predicted preferences.

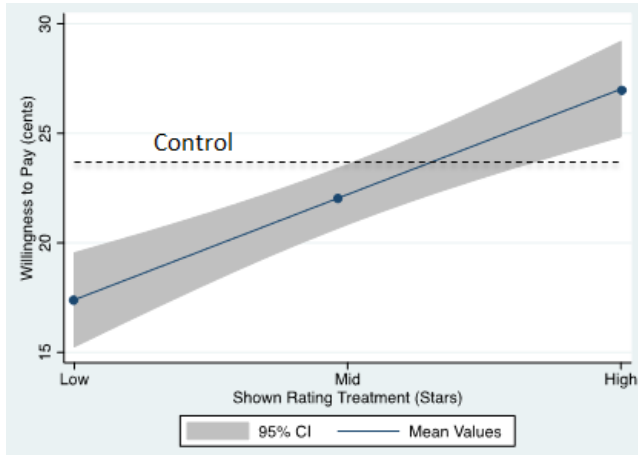


Figure 2. Study 1 treatment means.

Table 2. Study 1 repeated measures ANOVA.

| | Partial Sum of Squares | Degrees of Freedom | Mean Square | F Statistic | P value |
|-----------------|------------------------|--------------------|-------------|-------------|---------|
| Participant | 396744.78 | 41 | 9676.70 | | |
| Treatment Level | 24469.41 | 2 | 12234.70 | 42.27 | <0.000 |
| Residual | 346142.41 | 1196 | 289.42 | | |
| Total | 762747.40 | 1239 | 615.62 | | |

Table 3. Comparison of aggregate treatment group means with *t*-tests.

| | Control | Low | Mid |
|--------------------|----------|----------|---------|
| Low (1-2 Star) | 4.436*** | | |
| Mid (2.5-3.5 Star) | 0.555 | 4.075*** | |
| High (4-5 Star) | 1.138 | 5.501*** | 1.723** |

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
2-tailed *t*-test for *Control* vs. *Mid*, all else 1-tailed.

The resulting Poisson regression model is shown below, where WTP_{ij} is the reported willingness to pay for participant i on song j , $ShownRating_{ij}$ is the recommendation star rating shown to participant i for song j , $PredictedRating_{ij}$ is the predicted recommendation star rating for participant i on song j , and **Controls_{*i*}** is a vector of demographic and consumer-related variables for participant i . The controls included in the model were gender (binary), age (integer), school level (undergrad yes/no binary), whether they have prior experience with recommendation systems (yes/no binary), preference ratings

(interval five point scale) for the music genres country, rock, hip hop, and pop, the number of songs owned (interval five point scale), frequency of music purchases (interval five point scale), whether they thought recommendations in the study were accurate (interval five point scale), and whether they thought the recommendations were useful (interval five point scale). The composite error term ($u_i + \varepsilon_{ij}$) includes the individual participant effect u_i and the standard disturbance term ε_{ij} .

$$\log(WTP_{ij}) = b_0 + b_1(ShownRating_{ij}) + b_2(PredictedRating_{ij}) + \mathbf{b}_3(\mathbf{Controls}_i) + u_i + \varepsilon_{ij}$$

The results of the regression are shown in Table 4. Note that the control observations were not included, since they had null values for the main dependent variable *ShownRating*.

The results of our analysis for Study 1 provide strong support for Hypothesis 1 and demonstrate clearly that there is a significant effect of recommendations on consumers' economic behavior. Specifically, we have shown that even randomly generated recommendations with no basis on user preferences can impact consumers' perceptions of a product and, thus, their willingness to pay. The regression analysis goes further and controls for participant level factors and, most importantly, the participant's predicted preferences for the product being recommended. As can be seen in Table 4, after controlling for all these factors, a one unit change in the shown rating results in a 0.168 change (in the same direction) in the log of the expected willingness to pay (in cents). As an example, assuming a consumer has a willingness to pay of \$0.50 for a specific song and is given a recommendation, increasing the recommendation star rating by one star would increase the consumer's willingness to pay to \$0.59.

Table 4. Study 1 regression results

| Dependent Variable: log(Willingness to Pay) | | |
|---|-------------|------------|
| Variable | Coefficient | Std. Error |
| ShownRating | 0.168*** | 0.004 |
| PredictedRating | 0.323*** | 0.015 |
| Controls | | |
| male | -0.636** | 0.289 |
| undergrad | -0.142 | 0.642 |
| age | -0.105 | 0.119 |
| usedRecSys | -0.836** | 0.319 |
| country | 0.103 | 0.108 |
| rock | 0.125 | 0.157 |
| hiphop | 0.152 | 0.132 |
| pop | 0.157 | 0.156 |
| recomUseful | -0.374 | 0.255 |
| recomAccurate | 0.414* | 0.217 |
| buyingFreq | -0.180 | 0.175 |
| songsOwned | -0.407* | 0.223 |
| constant | 4.437 | 3.414 |
| Number of Obs. | 1240 | |
| Number of Participants | 42 | |
| Log-likelihood | -9983.3312 | |
| Wald Chi-Square Statistic | 1566.34 | |
| (p-value) | (0.0000) | |

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4. STUDY 2: ERRORS IN RECOMMENDATIONS

The goal of Study 2 was to extend the results of Study 1 by testing Hypothesis 2 and exploring the impact of significant error in true

³ Several recommendation algorithms were evaluated based on the Study 1 training data, and IBCF was selected for us in both studies because it had the highest predictive accuracy.

recommendations on consumers' willingness to pay. As discussed below, the design of this study is intended to test for similar effects as Study 1, but in a more realistic setting with recommender-system-generated, real-time recommendations.

4.1. Procedure

Participants in Study 2 used the same facilities and were recruited from the same pool as in Study 1; however, there was no overlap in participants across the two studies. The same participation fee and endowment used in Study 1 was provided to participants in Study 2. 15 participants were removed from the analysis in Study 2 because of issues in their responses, leaving data on 55 participants for analysis.

Study 2 was also a within-subjects design with perturbation of the recommendation star rating as the treatment and willingness to pay as the dependent variable. The main tasks for Study 2 were virtually identical to those in Study 1. The only differences between the studies were the treatments and the process for assigning stimuli to the participants in the recommendation task of the study. In Study 2, all participants completed the initial song-rating and song ownership tasks as in Study 1. Next, real song recommendations were calculated based on the participants' preferences, which were then perturbed (i.e., excess error was introduced to each recommendation) to generate the shown recommendation ratings. In other words, unlike Study 1 in which random recommendations were presented to participants, in Study 2 participants were presented with perturbed versions of their actual personalized recommendations. Perturbations of -1.5 stars, -1 star, -0.5 stars, 0 stars, +0.5 stars, +1 star, and +1.5 stars were added to the actual recommendations, representing seven treatment levels. The perturbed recommendation shown to the participant was constrained to be between one and five stars, therefore perturbations were pseudo-randomly assigned to ensure that the sum of the actual recommendation and the perturbation would fit within the allowed rating scale. The recommendations were calculated using the item-based collaborative filtering (IBCF) algorithm (Sarwar et al. 2001), and the ratings data from Study 1 was used as training data.

Each participant was presented with 35 perturbed, personalized song recommendations, five from each of the seven treatment levels. The song recommendations were presented in a random order. Participants were asked to provide their willingness to pay for each song, which was captured using the same BDM technique as in Study 1. The final survey, payouts, and song distribution were also conducted in the same manner as in Study 1.

4.2. Analysis and Results

For Study 2, we focus on the regression analysis to determine the relationship between error in a recommendation and willingness to pay. We follow a similar approach as in Study 1 and model this relationship using a Poisson random effects regression model. The distribution of willingness to pay data in Study 2 was similar to that of Study 1, overdispersion was not an issue, and the results of a Hausman test for fixed versus random effects suggested that a random effects model was appropriate. We control for the participants' preferences using the predicted rating for each song in the study (i.e., the recommendation rating prior to perturbation), which was calculated using the IBCF algorithm. Furthermore, the same set of control variables used in Study 1 was included in our regression model for Study 2. The resulting regression model is presented below, where the main difference

from the model used in Study 1 is the inclusion of $Perturbation_{ij}$ (i.e., the error introduced for the recommendation of song j to participant i) as the main independent variable. The results are presented in Table 5.

$$\log(WTP_{ij}) = b_0 + b_1(Perturbation_{ij}) + b_2(PredictedRating_{ij}) + b_3(\mathbf{Controls}_i) + u_i + \varepsilon_{ij}$$

The results of Study 2 provide strong support for Hypothesis 2 and extend the results of Study 1 in two important ways. First, Study 2 provides more realism to the analysis, since it utilizes real recommendations generated using an actual real-time recommender system. Second, rather than randomly assigning recommendations as in Study 1, in Study 2 the recommendations presented to participants were calculated based on their preferences and then perturbed to introduce realistic levels of system error. Thus, considering the fact that all recommender systems have some level of error in their recommendations, Study 2 contributes by demonstrating the potential impact of these errors. As seen in Table 5, while controlling for preferences and other factors, a one unit perturbation in the actual rating results in a 0.115 change in the log of the participant's willingness to pay. As an example, assuming a consumer has a willingness to pay of \$0.50 for a given song, perturbing the system's recommendation positively by one star would increase the consumer's willingness to pay to \$0.56.

Table 5. Study 2 regression results.

| Dependent Variable: log(Willingness to Pay) | | |
|---|-------------|------------|
| Variable | Coefficient | Std. Error |
| Perturbation | 0.115*** | 0.005 |
| PredictedRating | 0.483*** | 0.012 |
| Controls | | |
| male | -0.045 | 0.254 |
| undergrad | -0.092 | 0.293 |
| age | -0.002 | 0.053 |
| usedRecSys | 0.379 | 0.253 |
| country | -0.056 | 0.129 |
| rock | -0.132 | 0.112 |
| hiphop | 0.0137 | 0.108 |
| pop | -0.035 | 0.124 |
| recomUseful | 0.203* | 0.112 |
| recomAccurate | 0.060 | 0.161 |
| buyingFreq | 0.276** | 0.128 |
| songsOwned | -0.036 | 0.156 |
| constant | 0.548 | 1.623 |
| Number of Obs. | 1925 | |
| Number of Participants | 55 | |
| Log-likelihood | -16630.547 | |
| Wald Chi-Square Statistic | 2374.72 | |
| (p-value) | (0.0000) | |

* p<0.1, ** p<0.05, *** p<0.01

5. CONCLUSIONS

Study 1 provided strong evidence that willingness to pay can be affected by online recommendations through a randomized trial design. Participants presented with random recommendations were influenced even when controlling for demographic factors and general preferences. Study 2 extended these results to demonstrate that the same effects exist for real recommendations that contain errors, which were calculated using the state-of-the-art recommendation algorithms used in practice.

There are significant implications of the results presented. First, the results raise new issues on the design of recommender systems. If recommender systems can generate biases in consumer decision-making, do the algorithms need to be adjusted to compensate for such biases? Furthermore, since recommender systems use a feedback loop based on consumer purchase decisions, do recommender systems need to be calibrated to handle biased input? Second, biases in decision-making based on online recommendations can potentially be used to the advantage of e-commerce companies, and retailers can potentially become more strategic in their use of recommender systems as a means of increasing profit and marketing to consumers. Third, consumers may need to become more cognizant of the potential decision making biases introduced through online recommendations. Just as savvy consumers understand the impacts of advertising, discounting, and pricing strategies, they may also need to consider the potential impact of recommendations on their purchasing decisions.

6. ACKNOWLEDGMENT

This work is supported in part by the National Science Foundation grant IIS-0546443.

REFERENCES

- [1] Adomavicius, G., Bockstedt, J., Curley, S., and Zhang, J. 2011. Recommender Systems, Consumer Preferences, and Anchoring Effects. *Proceedings of the RecSys 2011 Workshop on Human Decision Making in Recommender Systems (Decisions@RecSys 2011)*, Chicago IL, October 27, pp. 35-42.
- [2] Adomavicius, G. and Tuzhilin, A. 2005. Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17 (6) pp. 734-749.
- [3] Ariely, D., Lewenstein, G., and Prelec, D. 2003. "Coherent arbitrariness": Stable demand curves without stable preferences. *Quarterly Journal of Economics* (118), pp. 73-105.
- [4] Becker G.M., DeGroot M.H., and Marschak J. 1964. Measuring utility by a single-response sequential method. *Behavioral Science*, 9 (3) pp. 226-32.
- [5] Bennet, J. and Lanning, S. 2007. The Netflix Prize. *KDD Cup and Workshop*. [www.netflixprize.com].
- [6] Chapman, G. and Johnson, E. 2002. Incorporating the irrelevant: anchors in judgments of belief and value. *Heuristics and Biases: The Psychology of Intuitive Judgment*, T. Gilovich, D. Griffin and D. Kahneman (eds.), Cambridge University Press, Cambridge, pp. 120-138.
- [7] Cosley, D., Lam, S., Albert, I., Konstan, J.A., and Riedl, J. 2003. Is seeing believing? How recommender interfaces affect users' opinions. *CHI 2003 Conference*, Fort Lauderdale FL.
- [8] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. 2001. Item-based collaborative filtering algorithms. *10th Annual World Wide Web Conference (WWW10)*, May 1-5, Hong Kong.
- [9] Schkade, D.A. and Johnson, E.J. 1989. Cognitive processes in preference reversals. *Organizational Behavior and Human Decision Processes*, (44), pp. 203-231.
- [10] Tversky, A., and Kahneman, D. 1974. Judgment under uncertainty: Heuristics and biases. *Science*, (185), pp. 1124-1131.