Complements and Substitutes in Product Recommendations: The Differential Effects on Consumers' Willingness-to-pay

Mingyue Zhang
Tsinghua University
Beijing, China
zhangmy.12@sem.tsinghua.edu.cn

Jesse Bockstedt
Emory University
Atlanta, GA
bockstedt@emory.edu

ABSTRACT
Product recommendations have been shown to influence consumers' preferences and purchasing behavior. However, empirical evidence has yet to be found illustrating whether and how the recommendations of other products affect a consumers' economic behavior for the focal product. In many e-commerce websites, a product is presented with co-purchase and co-view recommendations which potentially contain complement and substitute products, respectively. Very little research has explored the differential effects of complementary and substitutable recommendations. In this study, we are interested in how the type of recommendations of other products impact the consumers' willingness-to-pay for the focal product, and additionally how the recommendations' price and the consumers' decision stage moderate this effect. We conducted a 2x2x2 randomized experiment to examine how the consumers' willingness-to-pay is affected by these factors. Experimental results provide evidence that there is no significant main effect difference between complementary and substitutable recommendations. But we observed a significant interaction effect between recommendation type and decision stage, which highlights the importance of timing in recommender systems. Other findings include that consumers are willing to pay more for a specific product when the price of a recommended product is high, as well as when they are in later decision stages. These findings have significant implications for the design and applications of recommender systems.

Keywords
recommender systems; complements; substitutes; willingness-to-pay; price; decision stage

1. INTRODUCTION
Recommender systems (RS) are becoming integral to how consumers discover new products and have a strong influence on what consumers buy and view. For instance, 60% of Netflix content consumption originates from recommendations, and 35% of Amazon sales are attributed to recommendations [11]. With their utmost importance for retailers, some studies have been conducted to explore the behavioral effects of recommender systems on consumers [1][2]. Specifically, prior studies found that consumers' preferences and willingness-to-pay for a product can be influenced by the values of personalized recommendations. This provides evidence that consumers' behavior is vulnerable toward recommendation agents. However, there is still space for researching the behavioral effects of detailed recommendation features. For example, when evaluating the information of a focal product on its own webpage, consumers are often exposed to additional relevant products as recommendations, such as on Amazon.com. It is an open research question whether consumer's purchase decisions such as willingness-to-pay for the focal product would be affected by the display of 'other products' and the type of information presented with these recommendations. Some work has studied this in offline settings [31], but little work addresses this issue in the online recommendation context.

Particularly, the types of 'other products' in a recommendation set may vary, but they can be generally categorized into substitutes and complements [19]. Substitutes are products that can be purchased instead of each other, while complements are products that experience joint demand. For example, when a user is evaluating a cellphone, it's reasonable to recommend other phones to better match his/her needs, but it also makes sense to recommend batteries, chargers, or cases, which commonly make up a bundle recommendation [34]. Research has shown that consumers factor into consideration the source or type of information when making their purchase decisions [26]. Economic theory suggests that complements increase demand for the focal product because of increasing the possibility of users finding added value for the focal product [31]. With the increased demand, the market price will increase accordingly, leading to higher individual willingness-to-pay. Whereas, substitutes decrease demand for the focal product due to competition, which leads to a lower market price and individual willingness-to-pay. Despite the extensive literature about complements and substitutes in economics, little research studied their differential effects in online recommendation settings.

In this study, we are interested in how the type of recommendations of other products impact the focal product, and additionally how the recommendations' price and consumers' decision stage moderate this effect. The price of a recommended product can be perceived as a contextual reference point. Along with the nature of cross-price elasticity of demands between complementary/ substitutable goods [20], the research question of how the price interacts with recommendation type is of great value. Additionally, when shopping online consumers tend to have a two-stage decision making process [9]: first, screening a large set of available products to identify a subset of the most promising alternatives; and second, evaluating the latter in more depth to make a final purchase decision. Zheng et al. (2009) [33] argue that consumers prefer different recommendations in each stage because they are driven by different goals in each stage, i.e., in stage 1 they are comparing alternatives, whereas in stage 2 they are reviewing candidates. For this reason, we are also interested in the moderation role of decision stages, which has not been previously studied.

In the following sections, we firstly introduce the theoretical underpinnings of this research, based on which five hypotheses are proposed. Then we discuss the design of a randomized experiment, which measures consumers' willingness-to-pay across different recommendation scenarios. We present the results of our analysis and discuss the implications for online retail practices and research involving recommender systems. The investigation provides a new angle for understanding the behavioral aspect of recommender systems, as well as guidelines to further improve the design of recommendation agents.

2. THEORETICAL FRAMEWORK

In this section, we discuss the relevant theoretical foundations for our research questions in terms of three dimensions: recommendation type, price, and consumers' decision stage. Based on our primary goal of uncovering the differential effects between different recommendation types, we firstly review research about complementary/substitutable goods in the marketing literature, as well as related empirical studies in the recommender system context. Second, since the basic relationship between complements/substitutes is their cross-price elasticity of demands, we discuss theories describing how one product's price might influence consumers' willingness-to-pay for another product. Finally, we discuss the related literature on consumer decision making processes and how this interacts with the recommender systems.

2.1 Complements and Substitutes in Recommendation

The study of complements and substitutes has long been a central subject in the marketing literature. Generally, products are considered complements (substitutes) if lowering (raising) the price of one product leads to an increase in sales of another [31]. Research shows that consumer choice is easily influenced by context and the set of alternatives available at the time of decision [24], thus, there is significant demand interrelationship among substitutable and complementary goods [20]. Generally, two moderate or strong substitutes should be offered separately, whereas two complements should be offered as a bundle [32], in order to maximize the profit. This is because the introduction of a complement may increase the possibility of buyers finding new uses or added value for existing products [31], whereas the substitutes can be consumed or used in place of one another.

In the online recommendation scenario, a focal product is often presented with several related items as the recommendations. Take Amazon.com as an example, each product is featured on its own designated webpage, along with additional relevant products as recommendations. Hence, a visible directed product network is created whereby products are explicitly connected by hyperlinks [16]. Some studies [22][23][5] have examined the behavioral impacts of recommendation networks, with these studies primarily focusing on the co-purchase recommendation network. Many e-commerce websites provide recommendations from two product networks: co-view and co-purchase product networks. Only recently have their differential effects been considered [16]. More interestingly, co-purchase and co-view networks can be used to implicitly represent two recommendation strategies, that is, recommending complementary and substitutable products, respectively. Although not always the case, it is common that co-purchased items contain complementary products while co-viewed products contain substitutable products. For example, a consumer buy a laptop computer may view several laptops, but purchase only one laptop along with a complementary mouse, software, screen protector or other accessories. Figure 1 and Figure 2 provide an example to this extent with the co-purchase and co-view recommendations on Amazon.com when the focal product is 'Dell Inspiron 15 i5558-5718SLV'.

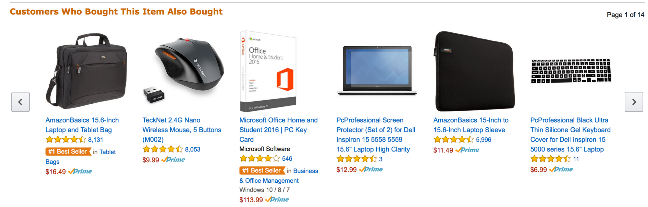

Figure 1. Amazon co-purchase product recommendation

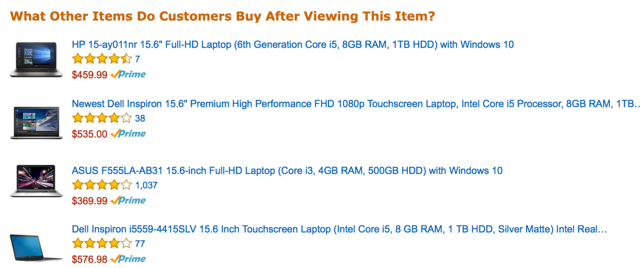

Figure 2. Amazon co-view product recommendation

By definition in microeconomics, if product A and B are complements, increased demand for product A should be associated with increased demand for product B [33]. This complementary product effect leads to the co-purchase network. On the other hand, substitute products have an inverse demand relationship. This leads to the co-view network because consumers tend to view and compare substitutes before making final purchase decisions. Given that recommendations of both complements and substitutes are often presented along with the focal product, it is of significant practical interest to understand their differential effects. Based on economic theory, complements increase demand for the focal product, and the market price will increase accordingly, leading to higher individual willingness-to-pay. Whereas, substitutes decrease demand for the focal product due to competition, which leads to a lower market price and individual willingness-to-pay. Thus, we put forth the following hypothesis:

H1: *Consumers tend to have higher willingness-to-pay for the focal product when it is displayed together with a complementary recommendation as compared to being displayed with a substitutable recommendation.*

Note that this initial hypothesis is intended to test a main effect of recommendation type and is price agnostic. We address the moderating effects of price in the next section.

2.2 Pricing

Researchers in marketing and economics have long recognized that pricing decisions sometimes incorporate more than one product. This is because consumers tend to respond to price relative to some reference price [25], such as the other prices in the store at the point of purchase. Both prospect theory and mental accounting suggest that consumers make decisions based on losses or gains relative to a reference point. When consumers compare the actual price of the focal product with other reference prices, incidental price learning [21] occurs without any explicit intention to memorize them. In offline physical stores, retailers can attempt to influence positively the degree to which the sales of one item affect sales of other items through in-store product locations and shelf space allocations, for example, locating two complements together. This is very similar to the online recommendations where other products are co-displayed with prices along with the focal product on the webpage. When consumers evaluate the focal product, the prices of recommended products are expected to affect their purchase-related decisions for the focal product. Contrast effect theory [30] suggests that the perceived value of the focal product's price is decreased (increased) when the recommendations presented along with have relatively higher (lower) price. That is, consumers are willing to pay more (less) when the product seems cheaper (more expensive) relative to other products. Thus, we have the following hypothesis about the main effect of *price* on willingness-to-pay:

H2: *The price of a recommended product has a positive influence on consumers' willingness-to-pay for the focal product, that is, consumers are willing to pay more when the price of the recommended product is higher than the price of focal product.*

There are also significant cross-relationships among the sales of substitutable and complementary products [20]. In particular, lower price or promotion of one product can stimulate sales of a complement, whereas supplant sales of other substitutes. That is to say, when the prices of complementary goods go up, the purchase likelihood for the other complementary good may go down, while if the price of one of the substitute goods goes up, the purchase likelihood for others will go up [12]. Furthermore, by influencing the demand of a complementary/ substitutive product through its price, the consumers' willingness-to-pay for that product will be influenced as well [18]. This cross-relationship nature of substitutes and complements indicates that the effect of price is stronger when the competition between two products is high, which is common among substitutes. Therefore, we hypothesize the following:

H2A: *There is an interaction effect between the price and type of recommendations, such that the positive influence of recommendation price on willingness-to-pay for the focal product is stronger when the recommendation type is substitute as compared to complement.*

2.3 Consumers' Two-Stage Decision Making

As illustrated in the previous literature [9][17][29], consumers are often not capable of evaluating all available alternatives in great depth, and this results in a two-stage decision making process. In the first stage, consumers usually browse a large set of available options and identify a small subset of candidates for further consideration. In the second stage, they tend to thoroughly evaluate the candidates and make a final purchase decision. In the second decision stage consumers' motivation and determination to make purchases are increased, thus having higher willingness-to-pay for selected alternatives. Hence, we hypothesize the following main effect of decision stage:

H3: *The decision stage has a positive influence on consumers' willingness-to-pay for the focal product, that is, consumers are willing to pay more when they are in the second stage.*

Given this multistage mental process, Ge et al. (2012) [8] argued that the manner in which information is processed differs systematically between the two decision stages. Their experimental results reveal that the timing of the presentation of specific pieces of information about an alternative across shopping stages has a great impact on consumers' choice. This difference can be attributed to the shopping goals theory [14] and construal level theory [15]. Specifically, consumers are less certain of their shopping goals in the first stage of a shopping process. Thus, their thinking is more abstract when in the first stage. Shopping goals become concrete when they are closer to the final purchase point in the second stage. Therefore, marketing promotions for similar products (i.e., substitutes) are more effective in influencing consumer's spending when their goals are less concrete [10][14]. Researchers have also studied the behavioral effects of recommendations beyond standard substitute recommendations. For instance, Zheng et al. (2009) [33] argued that customers prefer different types of recommendations in different purchase stages. In the first stage of an online purchase process, customers are navigating webpages to compare a large set of similar products. Whereas in the second stage, customers already have a clear candidate set through which to make a purchase decision. In the second stage, substitutive recommendations likely have little impact and recommendations of complement products may be preferred since they introduce items that can add value to the purchase of the focal product. Hence, we hypothesize the following interaction effect:

H3A: *There is an interaction effect between the stage and type of recommendations, such that the stage has a positive effect on willingness-to-pay for the focal product when the recommendation type is complement and it has a negative effect on willingness-to-pay for the focal product when recommendation type is substitute.*

3. EXPERIMENTS

Recommendations on Amazon.com and other platforms generally fall into the complement and substitute product types through the co-purchase and co-view lists. However, this is not always the case, and other contextual factors and user self-selection can impact the effect and content of these recommendations. Therefore, to eliminate these confounding factors and conditions that naturally occur in the field, we designed a randomized controlled experiment so that the recommendation type, recommendation price, and decision stage can be cleanly manipulated. This controlled and randomized treatment approach allowed us to test our hypotheses and make causal inferences.

3.1 Experiment Design and Participants

Our hypotheses express the main effects of each of three main factors (recommendation type, price of recommended product, and decision stage) as well as two two-way interaction effects (price x type and stage x type) on willingness-to-pay for a focal product. Since the focus of the study is on the effects of complementary versus substitutable recommendations, we did not hypothesize the interaction between recommendation price and decision stage. Additionally, since the three-way interaction among these factors is complex and no prior theory provides insights to this regard, this interaction was also not hypothesized.

A factorial experiment was used to test our hypotheses efficiently. Specifically, a 2 (types of recommendation: complements vs. substitutes) x 2 (recommended products' price relative to the focal product's price: low vs. high) x 2 (decision stage: stage 1 vs. stage 2) full-factorial design was used, which results in 8 treatment conditions. Although, the full factorial provides the opportunity to test the three-way interaction and the price x stage interaction, our analysis focuses only on the main effects and interactions identified in our hypotheses. The advantage of factorial experiment designs over randomized controlled trails (RCTs) is that they provide more statistical power with fewer participants. Generally, the objective of RCT is to compare the individual experimental conditions to each other directly, while in a factorial experiment the combinations of experimental conditions are compared, i.e., the main effects and interactions.

We manipulated the three factors between subjects, who were undergraduates from a business school in a large public university in North America. Subjects received extra credit for their participation in the experiment. We performed a power analysis with the assumption that the effect size of our model will be medium, i.e., $Cohen\ f^2 = 0.15$ (Cohen, 1988). To achieve power (1-β) of 0.80 and a medium effect size, as well as maintain a significance level (α) at 0.05, the minimum sample size for a model with three main effects and two interactions is 92 (the calculation was made by using the package 'pwr' in R).

We published a web link for our online experiment to a large undergraduate class containing approximately 400 students. 261 students clicked on the link to initiate the study. Participants were randomly assigned to one of the eight treatment conditions. The median time of completion is 12 minutes. We dropped observations for 126 participants for the following reasons: not completing the study, completing the experiments in an extremely short time (e.g., less than 4 minutes), completing the study in very long time (e.g., more than 4 hours indicating the study was started, stopped, and started again later), and not passing manipulation checks. Participants were informed that multiple manipulation checks would be used to determine if they took the study seriously, which would then impact whether they received extra credit for their participation. Since we collected the data as an online survey instead of bringing students to a laboratory, it is a common phenomenon that response and completion rates are relatively low [3]. As a result, 135 valid observations were left. The distribution of the valid observations across treatment groups is shown in Table 1. As can be seen in Table 1, the randomly assigned treatments are evenly distributed among participants.

Table 1. Experimental design and sample sizes per group

		complements	substitutes
Stage 1	Low	16	17
Stage 1	High	17	18
Stage 2	Low	17	16
Stage 2	High	17	17

Experiment participants were put in the scenario of purchasing a new computer mouse on an e-commerce site like Amazon.com. We chose a mouse because it is very common for consumers to purchase electronic products online, and a computer mouse has a large number of potential complements and substitutes with both low and high prices.

For the first manipulation factor (i.e., type of recommendation), participants were randomly assigned to one of two different shopping interfaces: the focal product page with recommendations of complementary products, or the focal product page with recommendations of substitute products. The pages included product descriptions directly from Amazon.com. We omitted brand information in the descriptions to eliminate any brand bias. Note that both the complementary and substitutable goods derived from real recommendations from the website. For the second manipulation factor, (i.e., price of the recommended products), participants were randomly assigned to either a high or low price condition. In the high price condition, the recommended products were higher in price than the focal product and in the low price condition the opposite was true. For the third manipulation factor (i.e., decision stage), we designed a two-stage shopping procedure (i.e., consider-then-choose) adapted directly from [8], which the participants followed prior to measuring dependent variables. Participants were randomly assigned to one of two decision stage manipulations: complete the first and second stage of the shopping procedure before being shown the focal product page or complete only the first stage of the shopping procedure before being shown the focal product page. In all treatment groups, the focal product and its posted price and description remained the same.

3.2 Stimuli and Procedures

Participants were first instructed with a cover story that they were participating in research focusing on consumers' preferences and purchase behavior. They were also told that there were no right or wrong answers. Following these initial instructions, participants were randomly assigned to one of the eight treatment groups.

Before the main task of the experiment, participants were asked to answer some basic questions about their opinions on electronic products and were asked to rate several different electronic product categories. Participants were told that their answers to these questions would be used later by our system to predict their preferences and make personalized recommendations for them. This pre-experiment task was used to eliminate their doubt about the basis of recommendations in later steps.

In the main task of the experiment, subjects were asked to shop for a computer mouse and make a purchase decision. We implemented the two-stage shopping decision process based on the methodology used in the marketing literature, (e.g., [8]). In the first stage process, participants were presented with descriptions of 12 computer mice as search results on the e-commerce store. They were asked to browse and evaluate all the product information presented. The 'Next Page' button appeared only after 30 seconds had elapsed, as means of preventing participants from moving ahead too quickly, without reviewing the stage 1 products. Manipulation check questions were also asked to check their impressions about these initial 12 mice. In the second stage, we narrowed down the choice set to 2 mice and participants were asked to evaluate the two items and pick one of them as their final purchase choice. Participants who were randomly assigned to the groups with condition 'stage 1' would only go through the first stage (i.e., browsing information). Comparably, those who were randomly assigned to groups with condition 'stage 2' would go through both the first and the second stage.

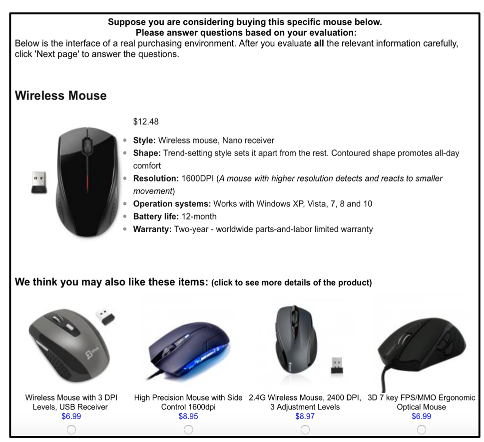

Figure 3. Example Screenshot for the experimental interface
(i.e., substitutes recommendation with low price)

After the stage manipulation, participants viewed a specific focal mouse product page. Along with the focal mouse, recommendations were presented according to the random treatment condition. Figure 3 provides example screenshots of the recommendation interface. In the groups with 'low (high) price' condition, all the recommendations' prices were slightly lower (higher) than the price of the focal mouse. Subjects could click on the recommendation to view a detailed description. The number of clicks and duration on each webpage were also recorded.

After viewing the focal product page based on their treatment condition, participants were asked to provide their willingness-to-pay for the focal product. Upon completing the shopping task, participants responded to a set of manipulation check questions and completed a short survey with demographic questions that we use as control variables in our analysis (age, gender, level of education, computer experience, web experience, e-commerce experience, familiarity with and attitudes toward recommender systems).

3.3 Dependent Measure

Willingness-to-pay is the maximum amount an individual is willing to sacrifice to procure a product. Here we adopted the method used by Rucker & Galinsky (2008), Rucker et al. (2014) and Kim & Gal (2014) [13][27][28] to measure willingness-to-pay. Participants indicated their willingness-to-pay using a sliding scale where they could choose from 0% to 120% of the retail price. The interval (i.e., 0%-120%) is used to reduce the amount of response variance and to guard against outliers. We are interested in relative changes in willingness-to-pay due to treatment effects and not in estimating point estimates of willingness-to-pay for specific products, thus the interval willingness-to-pay metric is sufficient. Furthermore, because the market price for the focal product was given, it's not realistic to use the Becker–DeGroot–Marschak approach [7] or second-price auctions to elicit willingness-to-pay. Figure 4 shows the interface for entering willingness-to-pay:

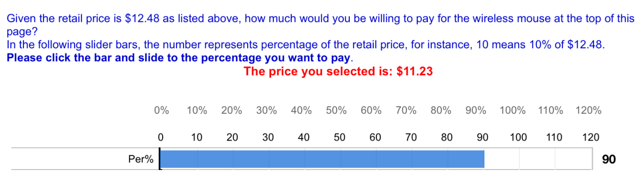

Figure 4. Entering willingness-to-pay

4. RESULTS

Table 2 provides summary statistics on the demographic items collected in our post-experiment survey.

Table 2. Demographic Summary Statistics

Control variables	Summary
Age	Mean: 21.6, SD: 3.35
Gender	50.37% --- female
Primary language	88.89% --- native English speaker
Experience with Internet	88.15% --- spend more than 4 hours per day on the Internet
Experience with e-commerce	77.04% --- browse e-commerce websites more frequently than once a week
Familiarity with RS	85.19% --- familiar with RS
Attitude toward RS	82.22% --- RS is helpful for finding relevant items

4.1 Manipulation Checks

In order to check the saliency of our recommendation type, two questions were asked of participants in the post-experiment survey: *(1) Do you think the products in the section titled 'We think you may also like these items' are complements to the mouse you evaluated?* and *(2) Do you think the products in the section titled 'We think you may also like these items' are substitutes to the mouse you evaluated?*. In terms of the decision stage manipulation check, we did not directly ask subjects' perceptions about *decision stage* because this may be an incomprehensible terminology. Instead, we asked them '*In the previous task you just finished, which procedure(s) have you been through?*', and provided the following possible responses: (1) *Evaluating a large set of alternative products as if you were gathering information in early stages of shopping* and (2) *Evaluating a small set of alternative products as if you were trying to choose a final one to purchase.* An additional question was used to check participants' perception about the relative price: '*What do you think of the price level of the mouse you just evaluated?*'.

First, to check if participants consciously distinguished between complement and substitute recommendations, we compared their responses toward the two manipulation check questions about recommendation type. They responded with the following 5 claims: "Definitely yes" (coded as 5), "Probably yes" (coded as 4), "Maybe" (coded as 3), "Probably not" (coded as 2), and "Definitely not" (coded as 1). As expected, participants in the complements group perceived the recommendations as complements ($M_{complements} = 4.15, M_{subsitutes} = 1.61,\ t(133) = 16.17, p < 0.001$), and not as substitutes ($M_{complements} = 1.42, M_{subsitutes} = 4.68, t(133) = -28.87,\ p < 0.001$). The extremely low p-values of these tests help guard against any potential multiple comparison issues. These results support the validity of our manipulation for recommendation types. Further, for the stage check question, participants in different stage conditions correctly perceived their decision stages ($M_{stage1} = 1.19, SD = 0.39, M_{stage2} = 1.97, SD = 0.17, t(133) = -14.81, p < 0.001$). Finally, a successful manipulation check was observed for the price. Due to the contrast effect, people in the high price recommendation condition felt the price of focal product is lower than those assigned in the low price recommendation condition ($M_{high} = 2.96, SD = 0.57, M_{low} = 3.33, SD = 0.68, t(133) = -3.45, p < 0.001$).

4.2 Main Results

Table 3 shows the mean and standard deviation of the willingness-to-pay, measured as a percentage (0%-120%) of the focal product's original price, in each of the eight treatment groups.

Table 3. Mean (SD) Willingness-to-pay (%) in Each Group

Decision stage	*Price*	*complements*	*substitutes*
Stage 1	Low	68.19 (15.86)	79.29 (17.66)
Stage 1	High	87.65 (13.50)	93.78 (21.81)
Stage 2	Low	88.29 (14.29)	81.63 (16.04)
Stage 2	High	93.94 (10.87)	89.53 (13.14)

To test the proposed hypotheses, we need to make comparisons between combinations of groups – the main and interaction effects. Since there are three manipulated factors in the experiment, we started by conducting a three-factor Analysis of Variance (ANOVA) and the results are presented in Table 4. The results reveal that the main effect of *stage* and *price*, as well as the interaction between *recommendation type* and *stage* are significant. We also conducted orthogonal contrast analysis, which provided consistent results and is omitted due to space constraints. Details can be obtained by contacting the authors directly.

Table 4. Results of Three-factor ANOVA

	Df	SSE	MSE	F value	Pr(>F)
Type	1	73	73	0.279	0.5981
Price	1	4670	4670	17.786	4.66e-05 ***
Stage	1	1200	1200	4.570	0.0345 *
Type : Price	1	9	9	0.036	0.8505
Type : Stage	1	1688	1688	6.430	0.0124 *
Price : Stage	1	872	872	3.321	0.0708 +
Residuals	127	33343	263		

Significance levels: + $p \leq 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Figure 5 displays the average willingness-to-pay under the combined conditions. Specifically, there is no significant difference between groups with complement and groups with substitute recommendations ($M_{complements} = 84.76$, $SD = 16.67, M_{substitutes} = 86.24, SD = 18.85, F = 0.279, p = 0.598$). However, when the prices of recommended products are relatively high, participants' willingness-to-pay is much higher than that in relatively low prices condition($M_{low} = 79.48$, $SD = 17.48, M_{high} = 91.26, SD = 15.75, F = 17.786, p < 0.001$), which supported H2. Similarly, the difference between conditions in stage 1 and stage 2 was in the expected directions ($M_{stage1} = 82.60, SD = 19.99, M_{high} = 88.45, SD = 14.26$, $F = 4.570, p < 0.05$), thus supporting the hypothesis that consumers are willing to pay more when they are in the second decision stage (H3). Further, we calculated Cohen's *d* to capture the effect size. Cohen's *d* is known as the difference of two population means and divided by the standard deviation from the data. The effect size for *price* and *stage* factors are 0.669 and 0.372, which indicate medium-to-large and small-to-medium effect, respectively.

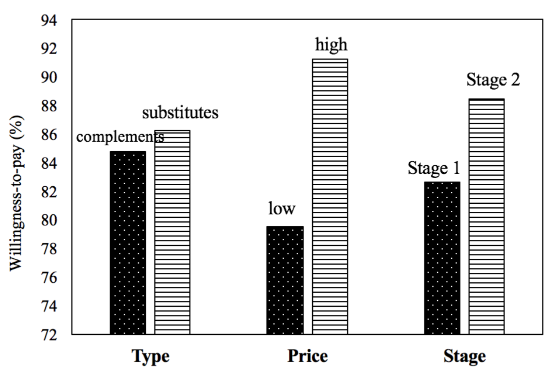

Figure 5. Average Willingness-to-pay in Combined Conditions

For the interaction effects, Figure 6 demonstrates the difference of mean values for complements and substitutes groups under different price levels and decision stages, respectively. The left figure shows no interaction between *type* and *price*, since both complements and substitutes groups have higher willingness-to-pay in high relative price conditions ($F = 0.036, p = 0.851$). The crossing lines in right figure indicate significant interaction effect between *type* and *stage* ($F = 6.430, p < 0.05$). Particularly, the effect of *decision stage* is much stronger when the recommendations are complements ($M_{stage1} = 78.21, SD = 17.62, M_{stage2} = 91.12, SD = 12.28, t(65) = -3.29, p < 0.001$), while it's not significant under substitute conditions ($M_{stage1} = 86.74, SD = 21.18, M_{stage2} = 85.70, SD = 15.14, t(66) = 0.23, p = 0.41$). Therefore, our hypothesis of interaction effect H3A is partially supported, and H2A is not supported.

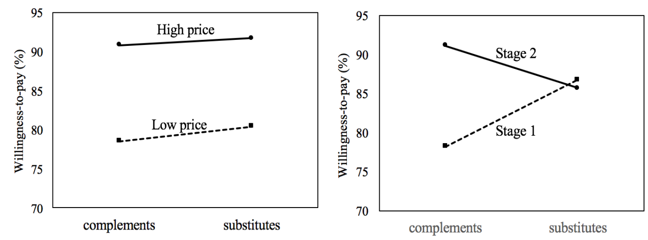

Figure 6. Interaction effects (Left: type × price; Right: type × stage)

Furthermore, to get the effect size of our model as well as coefficients of each factor, we estimated a sequential linear model. Firstly, we regressed the willingness-to-pay on a set of control variables, including gender, preference to the focal product, experience with e-commerce, familiarity with and attitude toward recommender systems. After that, we included the five independent variables of interests to the model. The *type* factor has two levels, either complements (0) or substitutes (1), the *price* factor is either low (0) or high (1), and the *stage* factor is either stage 1 (0) or stage 2 (1). The regression results of these two models are shown in Table 5, and the R-square increased 0.149 after including these five variables. Consistent with the ANOVA results, we got significant positive coefficients for *price* and *stage*, indicating that consumers have higher willingness-to-pay in the high price condition (compared to low price) and stage 2 condition (compared to stage 1). In addition, the interaction effect between type and stage is also marginally significant at level α = 0.1. The coefficients in the table suggest that consumers shown a recommended product with high price reported 10.353% higher willingness-to-pay in terms of the retail price. Similarly, consumers in the second stage reported 10.832% higher willingness-to-pay in terms of the retail price than those in the first stage.

Table 5. Results of the Linear Regression Models

	Dependent variable: Willingness-to-pay (%)					
	Model 1: Control ($R^2 = 0.1699$)			Model 2: Full model ($R^2 = 0.3189$)		
	Coefficient (SE)	T statistic	P-value	Coefficient (SE)	T statistic	P-value
Intercept	53.166 (10.450)	5.088	1.25e-06***	44.589 (9.862)	4.521	1.42e-05***
Type				5.572 (4.828)	1.154	0.251
Price				10.353 (3.769)	2.747	**0.007 ****
Stage				10.832 (4.008)	2.703	**0.008 ****
Type * Price				2.279 (5.336)	0.427	0.670
Type * Stage				-10.729 (5.691)	-1.885	**0.062 +**
Preference	7.141 (1.539)	4.640	8.45e-06***	6.232 (1.492)	4.178	5.50e-05***
Gender	3.255 (3.081)	1.056	0.292	3.803 (2.905)	1.309	0.193
Experience	-1.126 (1.616)	-0.697	0.487	-1.054 (1.527)	-0.690	0.492
Familiarity	0.600 (1.548)	0.388	0.699	1.412 (1.491)	0.947	0.345
Attitude	-6.362 (3.032)	-2.098	0.0378 *	-6.924 (2.807)	-2.467	0.015 *

Significance levels: + $p \leq 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

The effect size of our sequential multiple regression model is calculated by Cohen's f^2. It is defined as $f^2 = \frac{R_{AB}^2 - R_A^2}{1 - R_{AB}^2}$, where R_A^2 is the variance accounted for by a set of control variables A and R_{AB}^2

is the combined variance accounted by A and another set of independent variables of interest B. Here in our model, we have $R_A^2 = 0.1699$ and $R_{AB}^2 = 0.3189$, resulting in a medium-to-large effect size of $f_B^2 = 0.219$. We also conducted a post-hoc power analysis and with the 135 observations and the calculated effect size, the power of our model is 0.993 while maintaining the significance level at 0.05. This provides evidence that the null effects are true and not the result of a lack of power.

Since we measured the willingness-to-pay by restricting participants' choice from 0% to 120% of the stated retail price, it may result in censored and non-normal data. Therefore, we performed robustness check that removed the normality assumption. We conducted two non-parametric tests and ran a Tobit regression model using both dummy coding (0,1) and effect coding (-1,1). The coefficients and significance levels are consistent with our baseline analysis. Due to space constraints, details of the robustness check are omitted.

5. DISCUSSION AND CONCLUSIONS

5.1 Summary of Findings

In this paper, we conducted a randomized experiment to examine the impact of complement and substitute recommendations on consumers' willingness-to-pay for the focal product. Table 6 summarizes our findings which corresponds to the proposed hypotheses.

Table 6. Hypotheses and Results

Hypotheses	Results
H1: Consumers tend to have higher willingness-to-pay about the focal product when it is displayed together with a complementary recommendation as compared to being displayed with a substitutable recommendation.	Not Supported
H2: The price of a recommended product has a positive influence on consumers' willingness-to-pay for the focal product, that is, consumers are willing to pay more when the price of the recommended product is higher than the price of focal product.	**Supported**
H2A: There is an interaction effect between the *price* and *type* of recommendations, such that the positive influence of recommendation price on willingness-to-pay for the focal product is stronger when the recommendation type is substitute as compared to complement.	Not supported
H3: The decision stage has a positive influence on consumers' willingness-to-pay for the focal product, that is, consumers are willing to pay more when they are in the second stage.	**Supported**
H3A: There is an interaction effect between the *stage* and *type* of recommendations, such that the *stage* has a positive effect on willingness-to-pay for the focal product when the recommendation type is complement and it has a negative effect on willingness-to-pay for the focal product when recommendation type is substitute.	**Partially Supported**

Experimental results provide evidence that there is no significant main effect difference between complementary and substitutable recommendations on consumers' willingness-to-pay for the focal product. We further investigated two factors that commonly present with recommendations: decision stage and the price of recommended products. We found that consumers are willing to pay more for a specific product as decision stage increases. An interesting finding is the interaction between recommendation type and decision stage. The positive effect of stage vanished when the recommendation is substitute to the focal product, while it is very significant with complementary recommendations. This is consistent with previous findings that customers prefer different recommendations against different purchase stages, as well as highlighting the importance of timing in recommender systems. The price of recommended products was also found to have significant effects on willingness-to-pay. Serving as a reference point, the prices of recommendations may be compared with the retail price of the focal product, which could cause consumers to adjust their willingness-to-pay through incidental price learning. Under the condition with high recommendation prices, consumers tend to have higher willingness-to-pay for the focal product and vice versa. This positive effect is significant no matter the type of recommendation for other products.

5.2 Theoretical Contributions

Our research offers important theoretical contributions in the following ways. First, studies on product recommendations have focused on the consumers' different preferences and behaviors for one products in the presence of recommendations [1][2]. This paper extends the behavioral research on recommender systems by studying the question whether recommending 'other products' on the same webpage had an effect on consumers' willingness-to-pay for the focal product. Second, prior research has not paid much attention to different types of recommendations. Deriving from economics literature, two typical relationships between products are examined, that is, complements and substitutes. Third, our research is one of the few studies that examine the detailed recommendation features, i.e., price of recommended products as well as consumers' decision stage. Integrating the consumers' decision process, we have a better understanding of the behavioral aspects of recommendations in online purchases.

5.3 Implications for Practice

Beyond contributing to advancing the academic literature, our findings also have significant practical implications and may guide the platform's recommendation strategy. The vulnerability of consumers' willingness-to-pay indicates the importance of 'other products' in recommender systems. This suggests new possibilities for influencing product sales by manipulating the contextual information of recommendations. Another important implication is about the timing of recommendations, i.e., complementary recommendations should be delayed to the second decision stage.

5.4 Future Work

The main limitation of this study is that we are not observing real world purchases. In contrast, however, an advantage is that our controlled randomized experiment allows us to make causal inferences, and thus trading external validity for identification. Future research can be developed by exploring other factors associated with recommended products, such as average ratings, quality, pictures of complements/substitutes and so on. Additionally, we can use observational data to empirically validate the findings of our experiment. By examining the relationships between recommendation network properties and products' sales, we will have additional support for the influence of complementary and substitutable product recommendations on consumers' economic behavior from an aggregate level.

6. REFERENCES

[1] Adomavicius, G., Bockstedt, J.C., Curley, S.P. and Zhang, J. 2014. Suggest or Sway? Effects of Online Recommendations on Consumers' Willingness to Pay. (2014). *Working paper*.

[2] Adomavicius, G., Bockstedt, J.C., Curley, S.P. and Zhang, J. 2013. Do recommender systems manipulate consumer preferences? A study of anchoring effects. *Information Systems Research*. 24, 4 (2013), 956–975.

[3] Baruch, Y. and Holtom, B.C. 2008. Survey response rate levels and trends in organizational research. *Human Relations*. 61, 8 (2008), 1139–1160.

[4] Cohen, J. 1988. Statistical Power Analysis for the Behavioral Sciences (second ed.). Lawrence Erlbaum Associates.

[5] Dhar, V., Geva, T., Oestreicher-singer, G. and Sundararajan, A. 2014. Prediction in Economic Networks. *Information Systems Research*. 25, 2 (2014), 264–284.

[6] Donaldson, C., Jones, A.M., Mapp, T.J. and Olson, J.A. 1998. Limited dependent variables in willingness to pay studies: applications in health care. *Applied Economics*. 30, 5 (1998), 667–677.

[7] Frederick, S. 2012. Overestimating Others' Willingness to Pay. *Journal of Consumer Research*. 39, 1 (2012), 1–21.

[8] Ge, X., Häubl, G. and Elrod, T. 2012. What to Say When: Influencing Consumer Choice by Delaying the Presentation of Favorable Information. *Journal of Consumer Research*. 38, 6 (2012), 1004–1021.

[9] Häubl, G. and Trifts, V. 2000. Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing Science*. 19, 1 (2000), 4–21.

[10] Ho, S.Y., Bodoff, D. and Tam, K.Y. 2011. Timing of Adaptive Web Personalization and Its Effects on Online Consumer Behavior. *Information Systems Research*. 22, 3 (2011), 660–679.

[11] Hosanagar, K., Fleder, D., Lee, D. and Buja, A. 2014. Will the Global Village Fracture Into Tribes? Recommender Systems and Their Effects on Consumer Fragmentation. *Management Science*. 60, 4 (2014), 805–823.

[12] Jin, R.K.-X., Parkes, D.C. and Wolfe, P.J. 2007. Analysis of bidding networks in eBay: Aggregate preference identification through community detection. *AAAI Workshop - Technical Report*. WS-07-09, (2007), 66–73.

[13] Kim, S. and Gal, D. 2014. From Compensatory Consumption to Adaptive Consumption: The Role of Self- Acceptance in Resolving Self-Deficits. *Journal of Consumer Research*. 41, August (2014), 526–542.

[14] Lee, L. and Ariely, D. 2006. Shopping goals, goal concreteness, and Conditional Promotions. *Journal of Consumer Research*. 33, June (2006), 60–71.

[15] Liberman, N., Trope, Y. and Wakslak, C. 2007. Construal level theory and consumer behavior. *Journal of Consumer Psychology*. 17, 2 (2007), 113–117.

[16] Lin, Z., Goh, K.-Y. and Heng, C.-S. 2015. The demand effects of product recommendation networks: an empirical analysis of network diversity and stability. *MIS Quarterly*. (2015). *Forthcoming*.

[17] Liu, Q. and Arora, N. 2011. Efficient Choice Designs for a Consider-Then-Choose Model. *Marketing Science*. 30, 2 (2011), 321–338.

[18] Loomis, J., Gonzalez-Caban, A. and Gregory, R. 1994. Do Reminders of Substitutes Influence Contingent Valuation Estimates? *Land Economics*. 70, 4 (1994), 499–506.

[19] McAuley, J., Pandey, R. and Leskovec, J. 2015. Inferring Networks of Substitutable and Complementary Products. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2015).

[20] Mulhern, F.J. and Leone, R.P. 1991. Implicit Price Bundling of Retail Products: A Multiproduct Approach to Maximizing Store Profitability. *Journal of Marketing*. 55, 4 (1991), 63–76.

[21] Nunes, J.C. and Boatwright, P. 2004. Incidental Prices and Their Effect on Willingness to Pay. *Journal of Marketing Research*. 41, 4 (2004), 457–466.

[22] Oestreicher-Singer, G. and Sundararajan, A. 2012a. Recommendation networks and the long tail of electronic commerce. *MIS Quarterly*. 36, 1 (2012), 65–83.

[23] Oestreicher-Singer, G. and Sundararajan, A. 2012b. The visible hand? Demand effects of recommendation networks in electronic markets. *Management Science*. 58, 11 (2012), 1963–1981.

[24] Payne, J.W., Bettman, J.R. and Johnson, E.J. 1992. Behavioral decision research: a constructive processing perspective. *Annual Reviews of Psychology*. 43, 1 (1992), 87–131.

[25] Rajendran, K.N. and Tellis, G.J. 1994. Contextual and temporal components of reference price. *Journal of Marketing*. 58, 1 (1994), 22–34.

[26] Rao, A.R. and Sieben, W. a. 1992. The Effect of Prior Knowledge on Price Acceptability and the Type of Information Examined. *Journal of Consumer Research*. 19, 2 (1992), 256–270.

[27] Rucker, D.D. and Galinsky, A.D. 2008. Desire to Acquire: Powerlessness and Compensatory Consumption. *Journal of Consumer Research*. 35, 2 (2008), 257–267.

[28] Rucker, D.D., Hu, M. and Galinsky, A.D. 2014. The Experience versus the Expectations of Power: A Recipe for Altering the Effects of Power on Behavior. *Journal of Consumer Research*. 41, August (2014), 381–396.

[29] Russo, J.E. and Leclerc, F. 1994. An eye-fixation analysis of choice processes for consumer nondurables. *Journal of Consumer Research*. 21, September (1994), 274–290.

[30] Sherif, M. and Hovland, C.I. 1961. Social judgment: Assimilation and contrast effects in communication and attitude change. (1961).

[31] Shocker, A., Bayus, B. and Kim, N. 2004. Product Complements and Substitutes in the Real World: The Relevance of "Other Products." *Journal of Marketing*. 68, 1 (2004), 28–40.

[32] Venkatesh, R. and Kamakura, W. 2003. Optimal Bundling and Pricing under a Monopoly: Contrasting Complements and Substitutes from Independently Valued Products. *The Journal of Business*. 76, 2 (2003), 211–231.

[33] Zheng, J., Wu, X., Niu, J. and Bolivar, A. 2009. Substitutes or Complements: Another Step Forward in Recommendations. *Proceedings of the 10th ACM conference on Electronic commerce*. (2009), 139–145.

[34] Zhu, T., Harrington, P., Li, J. and Tang, L. 2014. Bundle recommendation in ecommerce. *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval - SIGIR '14*. (2014), 657–666.