When do Recommender Systems Work the Best? The Moderating Effects of Product Attributes and Consumer Reviews on Recommender Performance

Dokyun Lee
Carnegie Mellon University

Kartik Hosanagar
The Wharton School

Abstract

We investigate the moderating effect of product attributes and consumer reviews on the efficacy of a collaborative filtering recommender system on an e-commerce site. We run a randomized field experiment on a top North American retailer's website with 184,375 users split into a recommender-treated group and a control group with 37,215 unique products in the dataset. By augmenting the dataset with Amazon Mechanical Turk tagged product attributes and consumer reviews from the website, we study their moderating influence on recommenders in generating conversion.

We first confirm that the use of recommenders increases the baseline conversion rate by 5.9%. We find that recommenders act as substitutes for high average review ratings and review volumes with the effect of using recommenders increasing the conversion as much as about two additional average star ratings. Additionally, we find that positive impact on conversion from recommenders are greater for hedonic products compared to utilitarian products while search-experience quality did not have any impact. Lastly, we find that the higher the price, the lower the positive impact of recommenders, while providing more product descriptions increased the recommender effectiveness.

For managers, we 1) identify the products with which to use recommenders and 2) show how other product information sources on e-commerce sites interact with recommenders. From an academic standpoint, we provide insight into the underlying mechanism behind how recommenders cause consumers to purchase.

Keywords: E-Commerce, Recommender systems, Consumer Review, Item Attributes

1 Introduction

Recommender systems are now ubiquitous on the web. E-commerce sites regularly use such systems to guide consumers with prompts like "People who purchased this item also purchased..." to increase up-selling and cross-selling opportunities. Recommenders aid online shopping by reducing search cost (Anderson, 2008) and product uncertainty for consumers (Bergemann and Ozmen, 2006). As such, many existing studies have already shown that recommender systems increase revenue and profitability for firms (Anderson, 2008; Bodapati, 2008; Dias et al., 2008a; Das et al., 2007; De et al., 2010; Dias et al., 2008b; Fleder and Hosanagar, 2009; Hosanagar et al., 2014; Jannach and Hegelich, 2009; Monetate, 2013; Oestreicher-Singer and Sundararajan, 2012; Thompson, 2008). Consequently, according to a study by Econsultancy and Monetate (2013), 94% of e-commerce sites now consider recommendation systems to be critical competitive advantage to be implemented. At the same time however, the same study reveal that only about 15% of the company were getting good return on investment and 72% attributed failure to lack of knowledge on recommender systems. This is because recommenders almost always coexist with other factors and features on web that influence purchase decisions through product uncertainty levels[1]. For example, different products have different search cost (Hann and Terwiesch, 2003) and product uncertainty (Dimoka et al., 2012), while user-generated reviews reduce product uncertainty. As such, effective implementation of recommenders must account for complicated interaction with these factors. However, there is a lack of literature on how the impact of recommenders are moderated by other factors such as types of items sold, item attributes, and consumer-generated reviews. In this study, through a randomized field experiment, we investigate how factors that influence product uncertainty online, such as product attributes and consumer reviews, interact with a recommender system to affect conversion rate, defined as the percentage of product views that result in purchases.

Existing studies have shown that utilizing recommender systems in e-commerce settings lead to an increase in usage, revenue, and profitability - in short, an increase in sales volume (Anderson, 2008; Bodapati, 2008; Dias et al., 2008a; Das et al., 2007; De et al., 2010; Dias et al., 2008b; Fleder and Hosanagar, 2009; Hosanagar et al., 2014; Jannach and Hegelich, 2009; Monetate, 2013; Oestreicher-Singer and Sundararajan, 2012; Thompson, 2008). Other studies have investigated the

[1]Product uncertainty is defined as the consumer's difficulty in evaluating product attributes and predicting how a product will perform in the future (Hong and Pavlou, 2014).

impact of recommenders on sales diversity (Hinz and Eckert, 2010; Matt et al., 2013; Hosanagar et al., 2014; Fleder and Hosanagar, 2009; Oestreicher-Singer and Sundararajan, 2012; Jannach et al., 2013), in which the focus was to study how the use of recommender systems' influences the assortment of items viewed and purchased by consumers. While it is clear that the use of a recommender system generally leads to an increase in sales volume and influences sales diversity, there is a lack of investigation on how product-specific attributes or reviews influence the effectiveness of recommenders. Researchers and managers still don't know under what conditions and for what products a recommender system works well. Specifically, there is a lack of actual field studies that investigate the interaction between other factors that influence product purchase decisions (e.g., product-level attributes and review data) and the efficacy of a recommender system to generate conversion. How do certain item attributes increase or decrease the effectiveness of recommender systems in causing purchases? For example, are recommenders substitutes or complements for high review ratings and review volumes? Will a recommender system cause more or fewer purchases for highly priced items? How about for hedonic vs. utilitarian product or search vs. experience products? Many of these highly insightful and managerially impactful questions are not answered or are partially answered due to limited data. The lack of access to a field experiment setting covering a wide range of products and the sheer amount of resources required to content-code attributes of a large number of products are just a few reasons for this gap. Answers to the questions above can guide recommender implementation in e-commerce and provide insight into consumer purchase behavior in online settings.

Our study attempts to address these gaps by running a randomized field experiment on an e-commerce site of a top retailer in North America[2]. We run a randomized experiment with recommender treatment and control groups, then proceed to identify several key product attributes of more than 37,000 unique items viewed or purchased during the period of the field experiment. We utilize Amazon Mechanical Turk to efficiently content-code a large number of items and item attributes. After augmenting the dataset with the consumer review data pulled from APIs (Application Programming Interface), we run logistic regressions to tease out the moderating effects of product attributes in causing conversion under the use of recommenders.

[2]We are not allowed to disclose the identity of the company. But it is one of the biggest companies offline, also ranking top 5 in e-commerce revenue *worldwide*.

Briefly, our main results show the follow. We first confirm that the use of a recommender increases the conversion rate in general (by 5.9%), but this increase is highly moderated by product attributes. For example, the higher the price, the lower the positive influence of recommenders. We also find that while the baseline conversion rate is higher for utilitarian products online, benefit from recommenders is higher for hedonic products compared to utilitarian products. We find that contrary to conjectures from existing literature, the search-experience attribute does not influence the power of recommenders. Furthermore, we find that the use of a recommender increases conversion rates as much as approximately 2 additional stars out of 5 in average review ratings. While the higher review volume increases conversion rates, once recommenders are accounted for, the volume no longer had any effect on conversion. Essentially, recommenders act as substitutes for high average review ratings. Besides these, we have many more insights with more details in the results section.

Our results provide both broad and specific insights for understanding the moderating effects of product attributes on the power of recommender systems. This study makes several contributions. From an academic standpoint, ours is the first individualized field experiment study to look at the moderating effects of product attributes like price, hedonic-utilitarian quality, search-experience quality, and review data on a recommender with individual-level conversion field data. By working with a retailer that ranks top 5 in the world in e-commerce revenue and sells the most expansive list of product categories, we increase external validity. At the practice, our study has several managerial implications. First, managers can determine which specific products would be best served by recommenders and which would not. Second, managers will have insight into how other e-commerce features, such as product descriptions and user-generated reviews, interact with the power of recommenders. Managers can then optimize e-commerce sites appropriately and decide which features (e.g., reviews, more descriptions, recommenders) to implement in combination. Ultimately, we provide insight for better utilizing recommenders online for increased conversion rates.

2 Data

Our main dataset consists of complete individual-item level views and purchase transactional data from running a field experiment. The cooperating company that ran the experiment randomly assigned incoming new customers into either a treated group, in which the recommendation panel is shown, or a control group, in which the recommendation panel is not shown. We capture click-stream data as well as eventual conversion data. This dataset is augmented with 1) complete review data from the pages of all the products appearing in the dataset and 2) item attributes separately tagged via a survey instrument and workers on Amazon Mechanical Turk, an online marketplace for data tagging and cleaning.

2.1 Field Experiment & Data Description

With the cooperation of one of the top retailers in North America, we ran the field experiment on their e-commerce site for a two-week period in August 2013. The company has both an online and offline presence and is one of the top 3 in the North American region by size and revenue. It's e-commerce presence is ranked top 5 in the world with more than $10 billion in e-commerce revenue alone in 2014[3]. The company ran the field experiment using a state-of-the-art A/B/n testing platform. This platform implements a session tracking technology whereby each visitor's IP address is recorded and given a unique visitor ID. Then, visitors' behaviors are tracked over the period of the field experiment. This enables the website to track individuals' viewing logs and purchases over the period of field experiment duration. Whenever new visitors access the website for the first time, they are randomly chosen to be in the control group or in the treatment group. Upon clicking and viewing a particular item, the visitors assigned to the treated group are shown a recommender panel, as seen in Figure 1. Visitors in the control group do not see this panel. There are many types of recommender systems and it is infeasible to run all types of recommender systems in the field experiment setting due to the amount of resources required to implement and opportunity cost for the retailer. In order to increase the external validity, we utilize the most common type of recommender system used in the industry, a purchase-based collaborative filtering algorithm - "People who purchased this item also purchased" (Adomavicius and Tuzhilin, 2005)[4]. The specific

[3]https://www.internetretailer.com/top500/?cid=2014-IRAGP

[4]Within *Personalized Recommenders* systems, a broad taxonomy distinguishes three types of algorithms: Content-based, Collaborative Filtering, and Hybrid, which combines the first two. Content-based systems analyze product attributes to suggest products that are similar to those that a consumer bought or liked in the past. Collaborative filtering recommenders, unaware of product attributes, recommend products either purchased or liked by similar consumers, where similarity is measured by historical purchase (or like) data. We discovered through talking to a large e-business analytics firm, which implements recommenders for many clients, that out of about 300 firms, only 3 utilized content-based recommenders. The rest utilized purchase-based collaborative filtering. A majority of companies utilize collaborative filtering algorithm simply because content-based recommender systems require

algorithm used in the study is obtained from the most widely used open-source machine learning framework called the Apache Mahout (mahout.apache.org).

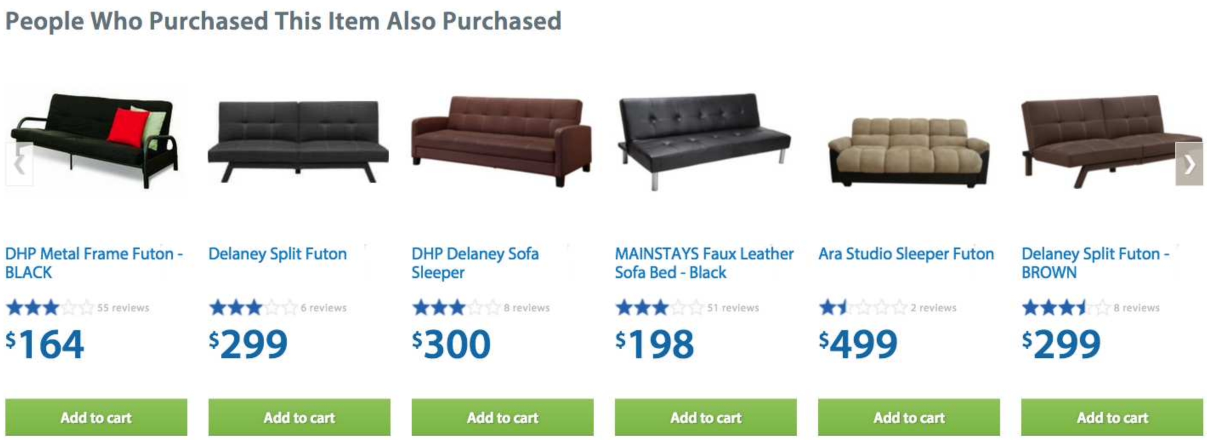

Table 1: **Recommendation Panel** : Example of a recommender shown to a consumer. Most commonly used recommender algorithm, “People who purchased this item also purchased”, is used.

The dataset, which spans 355,084 rows of individual-item transactional records, tracks 184,375 unique users split into 92,188 treated users and 92,187 control users. Users clicked and viewed details of 37,215 unique items and bought 3,642 unique items and a total of 9,761 items. In addition, we collected review data of all items appearing in the dataset, retailer’s description of the item, categorization including the subcategorization to the maximum depth, and more. Table 2 shows the top-level category appearance in the data and Table 3 gives the summary of the data. At the top level, the retailer has 18 categories including house appliances, automotive, electronics, movies, furniture, jewelry, and so on. We carefully chose the retailer with one of the most extensive coverage of SKUs and product categories to increases the external validity of the results.

2.2 Product Attribute Tagging on Amazon Mechanical Turk

Given the data from the field experiment, we still need to identify product attributes of interest. With more than 37,000 unique number of items, it is challenging to identify many product attributes at this scale. We have identified several product attributes motivated by extant literature to analyze for the products in our dataset. We discuss these attributes and relevant literature in Section 3. We

expensive attribute tagging and content analysis. One prominent exception is Pandora.com (a music genome project) that managed to content-code a large library of songs.

Products Appearance in Data by Categories as Classified by the Retailer - Top Level Categorization					
Appliances	Automotive	Baby	Clothing & Accessories	Electronics	Furniture
29545	5366	27843	7080	40733	39856
Grocery	Halloween	Health & Beauty	Holiday Gift Centre	Home & Pets	Jewelry & Watches
8422	6	28719	7621	50859	4015
Movies Music & Books	Office & Stationery	Outdoor Living	Sports & Rec	Toys	Video Games
26000	12352	6297	27681	20657	12032

Table 2: **Product Categories Occurring In the Dataset:** The first level product categorization as classified by the retailer online. There are in total 4 levels of depths and subcategories. 1st depth has 18 categories, 2nd -> 149, 3rd ->884, and 4th -> 492.

Variable	**Description**	**Source**	**Mean**	**SD**	**Min**	**Max**
REC	Recommender system treatment condition. 1 means the user was randomly selected to be shown recommendations.	Treatment	0.503	0.49	0	1
PRICE	Item price.	Site	85.94	120.69	0.01	998.00
DESLEN	Length of item description on the site.	Site	269.71	251.06	0	3882
AVGRATING	Average review star rating out of 5.	Site	2.44	2.22	0	5
RATINGNUMB	The number of reviews the item obtained.	Site	12.46	107.93	0	19407
BRAND	% of Amazon Mechanical Turkers who recognized the brand. Asked 5 Turkers per item.	AMT	0.53	0.35	0	0
DURABILITY	Durability of the item. Likert scale from 1-7 with 7 being the most durable.	AMT	4.97	1.37	1	7
UTILHEDO	Classification into utilitarian or hedonic product. 1 if utilitarian, 0 if hedonic.	AMT	Util	18529	Hed	18596
SEARCHEXP	Classification into search or experience product. 1 if search, 0 if experience.	AMT	Sea	15798	Exp	21327
Views	For a given user-item session, the number of times the user viewed the item.	Site	1.3	0.79	1	48
Quantity	The number ordered.	Site	0.02	0.32	0	48
			Number		Treated	Control
User ID	Unique user ID	Site	184375		92188	92187
Products Viewed	Unique products viewed by users	Site	37125			
Products Purchased	Unique products purchased by users	Site	3642			
	Total number of products purchased by users	Site	9762			
RATINGSEXIST	The number of items with existing reviews	Site	9631			
Product Category	Product category name classified by the retailer up to three level depth	Site				
Visit Start Time	User session start time	Site				
Transaction time	Transaction time	Site				

Table 3: **Variable Descriptions and Summary for Content-coded Data:**

now describe our methodology for identifying product attributes using Amazon Mechanical Turk (AMT). AMT is a crowd sourcing marketplace for simple tasks such as data collection, surveys, and photo and text analyses. To obtain product attributes for a given item, we create a survey instrument based on existing constructs, operating definitions, and measurement questions previously used in other studies. To ensure high-quality responses from the Turkers, we follow several best practices identified in literature (e.g., we obtain tags from at least 5 different Turkers choosing only those who are from the U.S., have more than 500 completed tasks, and an approval rate more than 98%. We also include an attention-verification question.) Please see the appendix for the measurement questions used and the complete list of strategies implemented to ensure output quality.

Ultimately, we achieve values greater than 0.8 for all the constructs in Krippendorff's Alpha, a inter-rater reliability measure in which any value above 0.8 is accepted in the literature as a satisfactory outcome[5]. We end up utilizing a several thousand unique AMT workers answering many questions about more than 37,000 unique items.

3 Product Attributes & Hypotheses

Extant literature in consumer economics, marketing, and information systems research have identified many product attributes that influence purchase decisions. Relating to products sold online on e-commerce sites, the literature has identified information uncertainty (Stigler, 1961; Arrow, 1963) related to product uncertainty and search cost (Degeratu et al., 2000; Bakos, 1997; Johnson et al., 2004; Wu et al., 2004; Hong and Pavlou, 2014; Kim and Krishnan, 2013; Lauraeus-Niinivaara et al., 2007) to be one of the main deterrents in product purchase decisions. Focusing on product-related uncertainty [6], the main aspects of product uncertainty online is description and performance uncertainty (Dimoka et al., 2012), defined *"as the buyer's difficulty in assessing the product's characteristics and predicting how the product will perform in the future."* Similarly, Liang and Huang (1998) have shown that that 1) different products do have different customer acceptance on the electronic market and 2) the customer acceptance is determined by the transaction cost, which is

[5]Another reliability measure, Cronbach's Alpha, produced the same result.

[6]We do not consider buyers experience and retailer uncertainty in this study. Buyer experience is not a concern since we randomize a large number of users into different groups. The retailer uncertainty is not a concern since our retailer is one of the most recognized retailers in the world. In fact, many company ranking lists rank our retailer as number one among US retailers.

in turn determined by the uncertainty and asset specificity. Hann and Terwiesch (2003) have also shown that different products have different search costs associated with them. Lastly, connecting product type and complexity to recommenders on e-commerce sites, Xiao and Benbasat (2007),Aggarwal and Vaidyanathan (2005), and Senecal and Nantel (2004) suggested that product type and complexity may influence users' acceptance of recommender systems. Thus in this paper, we analyze factors that influence product uncertainty in the online setting, which may influence recommender performance: product attributes and consumer-generated product reviews.

Product uncertainty can be ameliorated via product descriptions and reviews up to a certain point but this reduction also heavily depends on the type of product and the consumers' willingness to search. For example, Nelson's 1970s seminal work on economics of information and advertising (Nelson, 1970, 1974) classified products into search and experience goods. Search goods are dominated by characteristics and attributes that can be discerned prior to purchase and are often objective in nature. Experience goods are dominated by characteristics that can only be discerned by using the product or are subjective in nature. Nelson's search and experience framework has been used to explain how people react to advertising, search for different products online, and ultimately make purchases (Klein, 1998; Klein and Ford, 2003). Another product attribute that may influence purchase decision is the hedonic-utilitarian framework. Hedonic (pleasure-oriented consumption) or utilitarian (goal-oriented consumption) purpose related to a product (Dhar and Wertenbroch, 2000; Khan et al., 2005) has been shown to change the way consumers shop online. For example, this attribute interacts with uncertainty reducing mechanisms such as reviews and descriptions online. Sen and Lerman (2007) show that online consumers trust negative reviews more for utilitarian products. There are many other attributes that influence purchase decision via difference in information cost and product uncertainty. As such, we posit that these product attributes will also influence the effectiveness of recommender systems, commonly acknowledged as an electronic word-of-mouth or another source of information for awareness and product fit. In this paper, we look at the impact of these product attributes in an e-commerce setting in which recommenders are implemented.

Since it is infeasible to go through all of product attributes, we have focused our attention on identifying product attributes that 1) are shown in word-of-mouth and online review literature to influence consumers purchase behavior, 2) are clear and simple in concept for maximal managerial implication, 3) have strong theoretical background with existing and well-used operational definition and measurement survey questions. Following these criteria, we have identified several control variables as well as main variables of interest that may influence the effectiveness of a recommender. We next discuss each variable, related literature, how we tagged the attributes using extant operating definitions, and our hypotheses on how each will moderate the power of a recommender system. Details and sources of survey instruments for measuring product attributes are discussed in the Appendix.

3.1 Product Attributes

Hedonic VS. Utilitarian

A product characteristic often discussed and used to categorize products across industries is whether the product is dominantly a utilitarian product or a hedonic product (Dhar and Wertenbroch, 2000; Strahilevitz and Myers, 1998; Hirschman and Holbrook, 1982). The literature (Dhar and Wertenbroch, 2000; Strahilevitz and Myers, 1998; Hirschman and Holbrook, 1982) define utilitarian goods as those for which consumption is cognitively driven, instrumental, goal-oriented, and accomplishes a functional or practical task. Hedonic goods are defined as ones whose consumption is primarily characterized by an affective and sensory experience of aesthetic or sensual pleasure, fantasy, and fun. Broadly, the hedonic-utilitarian attribute has been shown to influence consumer product search behavior, purchase decisions, and even consumers' value of products (Hirschman and Holbrook, 1982; Bart et al., 2014; Khan et al., 2005).

Connecting to online shopping, studies have shown that consumers are more goal-oriented and utilitarian motivated online. Consumers with utilitarian motivation shop online for convenience, cost savings, and readily available information online (To et al., 2007). Since utilitarian goods dominantly consist of objective attributes that serve specific functions (e.g., hammer, memory card, and ink toners) and are apt for goal-oriented shopping, consumers may use online shopping for utilitarian products more than for hedonic products. As such, we posit that the baseline conversion rate is higher for utilitarian product.

Relating to recommender systems, extant literature have shown that the hedonic-utilitarian attribute moderates the trust and re-use intention of recommender systems. For example, Choi et al. (2011) suggests that consumers' trust for recommender systems and re-use intention is increased when the recommender provides a "social presence", defined as "*the extent to which a website allows users to experience others as psychologically present*". This increase in trust and re-use intention is greater for hedonic products compared to utilitarian products. Extending along these lines, we draw from past advertising literature to theorize how hedonic-utilitarian attributes may moderate the power of recommender systems in directly increasing conversion rates. Studies have shown that the effectiveness of product endorsement depends on whether the product is utilitarian or hedonic (Feick and Higie, 1992; Stafford et al., 2002). When consumers are shopping for a utilitarian product, the purchase decisions are guided by information about objective functional attributes. As such, consumers prefer expert endorsers. However, for hedonic products with many subjective attributes and high heterogeneity in preferences, it's been suggested that consumers prefer opinions of people who are more like them (Feick and Higie, 1992). The collaborative filtering algorithm implemented in our dataset provides recommendations to a consumer based on purchase histories of other consumers similar to the consumer and signal this clearly. Thus, we posit that conversion rates will be increased for hedonic products under the use of recommender systems since recommenders claim to reveal preferences of similar consumers. Thus, our hypotheses are as follows.

Hypothesis 1 *The base conversion rate for utilitarian goods will be higher in online settings.*

Hypothesis 2 *The increase in conversion rate under the use of a recommender will be higher for hedonic goods, compared to utilitarian goods.*

To measure and classify an item into a hedonic or a utilitarian product, we surveyed the extant literature and found several operating definitions and measurement questions. One measurement survey defines hedonic and utilitarian values and for each value, asks to rate the product on a 1 to 7 likert scale. This results in two separate measurements for utilitarian and hedonic quality. Another scale condenses this into one scale starting from purely utilitarian to purely hedonic in intervals. We asked all three as seen in Table 4 to at least five different Turkers, then took mean values. Finally, based on these three dimensions, the k-means clustering algorithm (Hartigan, 1975) was used to classify products into two clusters: utilitarian or hedonic. The cluster means for each product are shown in Table 4. The Appendix has the full list of questions used, question sources, and the inter-rater reliability measure.

Measurement Questions	Utilitarian Product Cluster Mean	Hedonic Product Cluster Mean
Given the above definition of hedonic and utilitarian value of a product, rate the product above in the scale below on hedonic value and utilitarian value.		
Hedonic Value [1 NOT AT ALL HEDONIC to 7 PURELY HEDONIC]	2.28	6.17
Utilitarian Value [1 NOT AT ALL UTILITARIAN to 7 PURELY UTILITARIAN]	5.98	1.95
Please give the scale on how much comparative utilitarian VS hedonic value the product offers. [1 PURELY UTILITARIAN to 7 PURELY HEDONIC]	2.19	5.97

Table 4: **Utilitarian VS. Hedonic Product Cluster Means**: Definition given is in the appendix.

Search VS. Experience

Philip Nelson's seminal work on economics of information and advertising (Nelson, 1970, 1974) classified products into search and experience goods. Search goods consist of attributes that can easily be discerned before purchase and are dominated by attributes with lower informational search cost and objective attributes, such as the speed and memory of a computer. In contrast, experience goods consist of attributes that cannot easily be discerned before purchase and are dominated by attributes with higher information search cost and subjective attributes like taste of wine or the entertainment value of movies. Nelson originally theorized and calculated the total cost of the product as the sum of the product cost and the consumers' search cost. Following this work, numerous studies in economics, marketing, and information systems have investigated how this search and experience classification of product influence consumers' search, consideration set, and purchase behavior (Klein, 1998; Klein and Ford, 2003; Girard and Dion, 2010; Huang et al., 2009; Hsieh et al., 2005; Krishnan and Hartline, 2001; Hong and Pavlou, 2014; Dimoka et al., 2012). Specifically, in online settings, product information uncertainty and higher search cost for experience goods has been shown to be a major hurdle and challenge for e-commerce managers (Hong and Pavlou, 2014; Dimoka et al., 2012; Weathers et al., 2007; Girard et al., 2003). While experience goods like wine, cosmetics, apparel, etc are increasingly sold on e-commerce sites, these sites still find it challenging to satisfy consumers' information needs to convert, or satisfy them enough to prevent high rates of return (Hong and Pavlou, 2014; Dimoka et al., 2012). A few studies have suggested several remedies like the use of search engines, multimedia product descriptions, and finally recommender systems to overcome high search costs (e.g., Hinz and Eckert (2010); De et al. (2013, 2010)). However, literature lacks studies on comparing search vs experience goods in the context of recommender systems. Traditionally, recommender systems were popularized on experience goods like movies, music, and books. However, now recommenders are being utilized for all types of products and we can compare the differential impact.

Nelson theorized that consumers' search for experience goods will be characterized by heavier reliance on word-of-mouth and experience of other consumers since the cost of information via other routes are more costly (Nelson, 1970, 1974; Klein, 1998). Consequentially, Nelson hypothesized that experience goods sellers will focus on persuasive and brand-focused tactics such as word-of-mouth, testimonials, and celebrity endorsements while search goods sellers will prioritize their advertising with informative and easy to discern facts about the products. However, it is not clear how search-experience attribute will influence recommenders' performance. Extant literature on the moderating influence of search-experience attribute on the power of recommenders is limited and conflicting. Senecal and Nantel (2004) found evidence that consumers are more influenced by recommendations for experience products than for search products. However, this study has a limited external validity due to the artificial nature of lab experiment in recommender settings and the fact that it is based on only two products, wine and calculators. Contrastingly, a study by Aggarwal and Vaidyanathan (2005), with again only two products, suggest a conflicting result. Aggarwal and Vaidyanathan (2005) claim that consumers perceived recommenders to be more effective for search goods than for experienced goods. Thus, the extant literature is lacking in both results based on realistic field data and based on an expansive list of products.

Ultimately, the power of a recommender to result in conversion for search or experience goods depends on consumers' trust of the recommender system. If the consumers trust recommenders to serve as a replacement for costly search, the recommendations should be more effective when used for experience goods. Recent literature in recommender systems has dubbed the recommender agents as "digitized word-of-mouth" (Chen et al., 2009) where consumers adapt and trust recommender systems as "social actors" and perceive human characteristics (Benbasat and Wang, 2005; Xiao and Benbasat, 2007; Komiak and Benbasat, 2004). Essentially, consumers are increasingly trusting recommenders to replace searching when the search cost is high. Nelson's theory suggest that consumers rely more on word-of-mouth for experience goods and recent literature have shown that recommender systems are accepted and trusted as a form of word-of-mouth. While the baseline conversion rate for search goods online may be higher due to lowered search cost, product information uncertainty, and product fit uncertainty (Dimoka et al., 2012; Hong and Pavlou, 2014), recommenders may be better received by consumers for experience goods based on Nelson's theory. In accordance with Nelson's theory on experience goods and the role of recommender systems online, we develop the following hypothesis.

Hypothesis 3 *The base conversion rate for search goods will be higher in online settings.*

Hypothesis 4 *The increase in conversion rate under the use of a recommender will be higher for experience goods, compared to search goods.*

To measure and classify an item into a search or a experience product, we surveyed the extant literature and found several operating definitions and measurement questions. We found two sets of questions repeatedly used in the literature. One set of questions, used widely in marketing literature, asks the consumers to answer two questions: how well could you judge the attribute or quality of the product 1) *before* they have purchased it and 2) *after* they have purchased it. If the consumers can judge the attributes not so well before the purchase but well after the purchase, the literature has classified those products as experience goods while for search goods, consumers can judge the quality of the product well even before the purchase. Another set of questions asked similar questions related to the search cost. We combined these questions in the extant literature and asked in total 4 questions on the Likert scale. Once we obtained the answers for each product from at least five different Turkers, we took the mean value for each answer. Finally, we used the k-means clustering algorithm (Hartigan, 1975) to classify products into two clusters: search or experience. The cluster means for search and experience products are shown in Table 5. The Appendix has the full list of questions used, question sources, and the inter-rater reliability measure.

Consumer Reviews

It is well documented in the literature that user-generated reviews influence online consumers' purchase intentions (Chen et al., 2004; Chen and Xie, 2008; Duan et al., 2008; Sun, 2012; Chevalier and Mayzlin, 2006; Berger, 2014). However, results are mixed in that review ratings always do

Measurement Questions [1 NOT WELL/IMPORTANT AT ALL to 7 EXTREMELY WELL/IMPORTANT]	Search Good Cluster Mean	Experience Good Cluster Mean
How well could you judge the attributes or quality of this product even BEFORE you purchased or used it?	4.82	3.66
How well could you judge the attributes or quality of this product even AFTER you purchased or used it?	6.36	6.31
How important is it for you to see, touch, hear, taste, smell (whichever applies) this product IN PERSON to evaluate its attributes?	3.15	5.36
How well can you evaluate the product using only information provided by retailer and/or manufacturer about this product's attributes and features?	5.04	3.78

Table 5: **Search VS. Experience Product Cluster Means**

not influence consumers, while other studies show that the effect of reviews on sales are moderated depending on the nature of the product — which can increase search-cost — whether it's a niche or experiential item (Li and Wu, 2013; Duan et al., 2008; Dai et al., 2014; Chen and Xie, 2008; Zhu and Zhang, 2010). Specifically, consumers tend to discount or even ignore review ratings when the volume of the review is low (Li and Wu, 2013; Duan et al., 2008; Chen et al., 2004). For niche or less-popular items, the impact of reviews can be greater (Zhu and Zhang, 2010). Specifically, high ratings have a more positive influence on consumer purchase intentions for niche items (Tucker and Zhang, 2011). Similarly, the herding effect for purchase has been found to be more salient for experience goods than for search goods in online settings (Li and Wu, 2013). Ultimately, all of these results are consistent with the search-cost argument in which consumers rely more on external informational sources like reviews when the search-cost is higher (e.g., niche item or experience items). Consumers rely on reviews as a source of information and do so selectively based on the search-cost related to products.

Recommender systems are electronic word-of-mouth (Chen et al., 2009) and reduce uncertainty and search-cost for consumers online (Clemons, 2008) just as consumer reviews do. However, it is not clear if recommender systems act as substitutes or complements to reviews since they serve similar yet slightly different purpose. Reviews mainly reduce uncertainty and provide product fit information while recommenders increase awareness and provide personalized product fit information. The cost of consumption is also different in that it takes a longer time to process review ratings (mean and variance) and to read the reviews compared to getting a straight forward recommendation - given that consumers trust recommenders. In fact, the acceptance and pervasiveness of recommenders in e-commerce have grown so much that a majority of consumers now expect and prefer websites that provide personalized recommendations (Accenture, 2012). Since recommender systems provide personalized fit information on top of consumer reviews, it's likely that recommenders may serve as a substitute for consumer reviews and provide additional information and value for consumers. If the consumers do trust the recommenders as the Accenture survey suggests, it is possible that with the existence of personalized recommendations, consumers may discount other people's reviews. Another reason consumers may discount higher review ratings when recommenders are present is because consumers may not agree with other consumers' reviews. Indeed, Dai et al. (2014) claims that consumers rely less on other consumers' reviews when shopping for experience products because consumers believe that other people's reviews are not representative of their own evaluations. Since the cost of consumption is lower for recommender systems, in extreme cases and depending on products, consumers may not even bother to check the reviews. In summary, consumers may trust personalized recommendations from the website more than review data.

Based on discussed theory, we posit that while higher review ratings may increase conversion rates in the absence of a personalized recommender system, with the presence of a personalized recommender system, its positive influence may be lessened. It is likely that given the personalized recommendation by an algorithm, the high review ratings may have less impact on conversion. In other words, recommenders act as substitutes for reviews. Our hypotheses are as follows:

Hypothesis 5 *The base conversion rate will be increased for products with higher review ratings.*

Hypothesis 6 *The positive impact on conversion from high review ratings will be lessened under the presence of a recommender system.*

All hypotheses are listed in Table 8.

3.2 Control Attributes

In addition to the attributes discussed above, we include the following control attributes in the model:

1. Durability: We asked 5 distinct Turkers to rate on a Likert scale from 1 to 7, with 7 being extremely durable, on how durable the product is.

2. Description Length: The retailer provides description of all products sold on the website. We get the length to proxy for the amount of information provided.

3. Brand Awareness Proxy: We asked 5 distinct Turkers if they recognized the brand of the item. We then take the percentage of the Turkers who answered "Yes" as a proxy measure for brand prominence.

4 Model & Results

4.1 Model

The conversion rate given recommendation treatment for user u and product i's attributes are modeled as a logistic regression.

$$
\begin{aligned}
log(\frac{P(conv)_{iu}}{1-P(conv)_{iu}}) \quad = \quad & \beta_0 + \beta_1 PRICE_i + \beta_2 REC_u + \beta_3 UTILHEDO_i + \beta_4 SEARCHEXP_i + \\
& \beta_5 DURABILITY_i + \beta_6 BRAND_i + \beta_7 DESLEN_i + \beta_8 AVGRATING_i + \\
& \beta_9 RATINGNUMB_i + \beta_{10} PRICE_i \times REC_u + \beta_{10} UTILHEDO_i \times REC_u + \\
& \beta_{11} SEARCHEXP_i \times REC_u + \beta_{10} DESLEN_i \times REC_u + \\
& \beta_{10} AVGRATING_i \times REC_u + \beta_{11} RATINGNUMB_i \times REC_u + \epsilon_u
\end{aligned}
$$

We estimate the parameters β_{0-11} by maximum likelihood estimator.

4.2 Results

Table 6 provides results from running the logistic regression. We first discuss the baseline hypotheses and results before presenting the main results on interaction with the recommenders. The stand-alone main effect results confirm previous literature. The impact of price on conversion is negative and significant (-0.0016) while the effect of recommenders is positive and significant (0.17998). The result corroborates the extant literature in that using a recommender system indeed increases conversion rates, and thus the sales volume (Hosanagar et al., 2014; Lee and Hosanagar, 2014). In this experiment, the use of recommenders increased the baseline conversion rate by 5.9%. Description

Variables	Estimate	Std Error	Log odds
Constant	-3.40089***	0.05973	0.033
PRICE	-0.00163***	0.00021	0.998
REC	0.17998**	0.05886	1.197
DESLEN	-0.00004	0.00008	0.999
AVGRATING	0.10651***	0.00818	1.112
RATINGNUMB	-0.00007	0.00015	0.999
UTILHEDO (UTIL=1)	0.26382***	0.03718	1.301
SEARCHEXP (SEA=1)	0.14281***	0.03561	1.153
BRAND	0.06783	0.03636	1.070
DURABILITY	-0.19567***	0.00841	0.822
REC X PRICE	-0.00115***	0.00031	0.998
REC X DESLEN	0.00027**	0.0001	1.000
REC X AVGRATING	-0.03562**	0.0114	0.965
REC X RATINGNUMB	0.00019	0.00016	1.000
REC X UTILHEDO	-0.16565**	0.05169	0.847
REC X SEARCHEXP	0.00654	0.04989	1.006

Table 6: **Logistic Regression Results Table : '*'= p-value <0.05, '**'= p-value <0.01, '***'= p-value <0.001**

length of products provided by the retailer had no significant influence on base conversion rate. Keeping everything else the same, higher average product review ratings increase the conversion rate (0.10651) as shown previously by Chevalier and Mayzlin (2006) and Sun (2012). Higher durability was associated with lower conversion rate (-0.19567). This result is likely since high durability is correlated with higher price and lower purchase frequency, and thus higher perceived risk (Jacoby et al., 1971; Pavlou, 2003) and lower willingness to purchase, especially in online settings. Lastly, our proxy variable for brand prominence was directionally positive and weakly significant with p-value = 0.6. Next, we discuss our main hypotheses and results regarding interaction between product attributes (and reviews) and recommender systems.

Hedonic VS. Utilitarian

The main effect of hedonic-utilitarian attribute (1 if utilitarian, 0 if hedonic) show higher conversion rate for utilitarian products online at 0.26382. The effect is statistically significant and greater than any other effects including the use of recommender systems. This supports our hypothesis that the base conversion rate for utilitarian goods will be higher in online settings keeping everything else constant. As To et al. (2007) suggests, consumers are utilizing e-commerce more for utilitarian purposes. Hedonic products often have attributes related to sense and beauty that consumers need to experience beforehand and is less bought online where price, convenience, and reduced search-cost may be the primary reasons for conversion. Interaction term with recommender treatment is negative and statistically significant at -0.16565. This suggests that while consumers purchase utilitarian products more in general in e-commerce settings, recommenders increase conversion more for hedonic products. The use of recommenders for utilitarian products still increases conversion since the main effect minus the interaction term is positive ($0.17998 - 0.16565 = 0.01433$).

The result is consistent with the story that consumers buying online are mainly motivated by utilitarian reasons of price, convenience, and reduced search-cost. Given that recommenders primarily serve as another source of information to increase the awareness set and to reduce search-cost, utilitarian products, which already have lower search-cost on the internet, benefit less from the use of recommenders. Recommenders are effective for hedonic goods.

Search VS. Experience

The main effect of the search-experience attribute (1 if search, 0 if experience) shows a higher conversion rate for search products online at 0.14281. The effect is statistically significant and positive, thus supporting our hypothesis that the base conversion rate for search goods will be higher in online settings. This corroborates existing theory that (Nelson, 1970; Pavlou, 2003; Dimoka et al., 2012) search goods with less informational cost attributes have less deterrent for purchase in online settings. However, the interacted term with recommender treatment was not statistically significant while directionally positive. The results do not support our hypothesis that the conversion rate will be higher for experience goods under the use of a recommender system. The results suggest that the original conjecture by Nelson (1970), that consumers will rely more on word-of-mouth and experience of others for experience goods, doesn't seem to carry over to a recommender system. While recommenders are theorized as "digitized word-of-mouth" (Chen et al., 2009), it is possible that a simple signal such as "other consumers who've purchased this item also purchased" does not provide enough details or reduction in uncertainty to particularly work well on experience products. Another explanation may be that consumers do not believe other consumers' preferences accurately reflect their tastes as suggested by Dai et al. (2014) in cases of reviews. Since our dataset spans expansive categories of products sold on websites, we sought to replicate results of Senecal and Nantel (2004) (recommendations for experience products like wine were more influential than recommendations for search products like calculators) and Aggarwal and Vaidyanathan (2005) (that recommenders are received more favorably for search goods). Depending on the product category chosen, we were able to replicate the results that support both arguments. However, when everything in the dataset is considered, search-experience attributes do not seem to moderate the effectiveness of this particular recommender.

Consumer Reviews

The main effect of average review ratings had a positive impact on conversion at the baseline at 0.10651. This means that approximately 2 additional stars out of 5 in review ratings increases log-odds ratio as much as the use of recommender systems[7]. Contrary to a previous study (Duan et al., 2008) that showed that review volumes are associated with higher sales, our results show that once the recommenders are accounted for, the review volume does not have any impact on baseline conversion rates in e-commerce settings (RATINGNUMB coefficient is -0.00007 and statistically not significant). The interaction term with recommender treatment and average ratings suggest that the positive impact on conversion from high review ratings will be lessened under the presence of a recommender system with estimate at -0.03562. This supports our hypothesis that consumers rely less on high average ratings once the recommenders are introduced.

To further investigate the interaction between review ratings, review volume, and recommender systems, we run multiple specifications in Table 7. The model on the first column, with only review volume, corroborates the results by Duan et al. (2008), which claimed that high review volumes increase conversion. Column 2 confirms that higher average rating increases the baseline conversion rate. However, once recommenders are accounted for, the rating volume does not matter and the positive impact of high average rating is lessened. Ultimately, our results suggest that recommenders serve as substitutes for review volume and higher review ratings in causing conversion.

Lastly, we summarize our findings and hypotheses supported in Table 8. We also summarized other takeaways in Table 9.

[7] That is, the increase in log-odds ratio from using a recommender, 0.179, is approximately twice that of 0.106. However, it is likely that increase in conversion is nonlinear for average star ratings from 0 to 5.

	1	2	3
Constant	-3.93436***	-4.06221***	-4.11774***
RATINGNUMB	0.00012*		-0.00002
AVGRATING		0.05053***	0.07224***
REC			0.10843**
REC X RATINGNUMB			0.0002
REC X AVGRATING			-0.04396***

Table 7: **Multiple Specifications for Review Related Variables : '*'= p-value <0.05, '**'= p-value <0.01, '***'= p-value <0.001**

Attribute Construct	**Hypotheses**	**Supported**
Hedonic-Utilitarian	The base conversion rate for utilitarian goods will be higher in online settings	YES
Hedo-Util × Rec	**The increase in conversion rate under the use of a recommender will be higher for hedonic goods, compared to utilitarian goods**	**YES**
Search-Experience	The base conversion rate for search goods will be higher in online settings	YES
Sea-Exp × Rec	**The increase in conversion rate under the use of a recommender will be higher for experience goods, compared to search goods**	**NO**
Review Rating	The base conversion rate will be increased for products with higher review ratings	YES
Review Rating × Rec	**The positive impact on conversion from high review ratings will be lessened under the presence of a recommender system**	**YES**

Table 8: **Hypotheses and Results**

Attribute Construct	**Result Takeaways**
Durability	The higher the durability, the lower the baseline conversion rate online.
Price	The higher the price, the lower the baseline conversion rate. Additionally, the higher the price, the lower the benefit of recommender.
Description Length	Description length did not influence the baseline conversion rate. However, longer description increased the benefit of a recommender.
Brand	Brand prominence showed weak positive effect (p-value = 0.06) on baseline conversion rate.
Review Volume	The higher the review volume, the higher the conversion rate. However, once recommenders are accounted for, high review volume did not influence conversion.

Table 9: **Other Takeaways**

5 Conclusion and Discussion

While recommenders are prevalent in e-commerce and have been shown to increase sales volume in multiple studies, effective use and implementation of recommenders still elude a majority of e-commerce managers and retailers as shown in studies such as Econsultancy and Monetate (2013). We believe that this is due to the lack of holistic investigation of conversion process that influence purchase decisions other than the recommenders. This study addresses this gap and adds empirical results.

This paper examined the interaction between a recommender system and product attributes along with reviews in e-commerce setting. Several product attributes were found to influence the power of recommenders in causing consumers to ultimately buy products. Our results reproduced several baseline hypotheses regarding the impact of product attributes on e-commerce shopping and extended existing baseline hypotheses to incorporate the impact on and interaction with recommender systems. The results show rich interaction between the effectiveness of recommenders and a variety of product attributes and review. We show that recommenders act as substitutes for high average review ratings and review volumes. Additionally, we find that baseline positive impact on conversion from recommenders are reduced for utilitarian products compared to hedonic products while search-experience quality did not have any impact. We also find that the higher the price, the lower the positive impact of recommenders, while providing longer product descriptions increased the power of recommenders.

Given these findings, managers have several key takeaways for implementing effective recommender strategies. Our study suggests effective ways to utilize recommender systems. For example, since our results suggest that recommenders act as substitute for high review volume and higher average rating for conversion, e-commerce sites with low review volumes could prioritize recommender implementations. We also show that a longer product description increases recommender effectiveness. While sites selling utilitarian products may still benefit from the use of recommenders, the benefit was not as substantial as using it on hedonic products. Utilitarian product sellers may want to utilize the limited webspace for other content before a recommender system.

One shortcoming of our paper is that we used only one type of recommender system: purchase-based collaborative filtering. However, we carefully chose an algorithm (i.e., collaborative filtering over content-based) that is most widely used after researching industry reports and companies in this area [8], and utilized an open-source implementation (Apache Mahout) most widely used by e-commerce sites. We believe that our results have high external validity due to the retailer we worked with and the expansive list of products covered in the study.

References

Accenture: 2012, 'Today's Shopper Preferences: Channels, Social Media, Privacy and the Personalized Experience'. Technical report, Accenture Interactive.

Adomavicius, G. and A. Tuzhilin: 2005, 'Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions'. *Knowledge and Data Engineering, IEEE Transactions on* **17**(6), 734–749.

Aggarwal, P. and R. Vaidyanathan: 2005, 'Perceived effectiveness of recommendation agent routines: search vs. experience goods'. *Int. J. Internet Marketing and Advertising* **2**(1/2).

Anderson, C.: 2008, *The long tail: Why the future of business is selling less of more*. Hyperion Books.

Arrow, K. J.: 1963, 'Uncertainty and the welfare economics of medical care'. *The American economic review* pp. 941–973.

Babin, B. J., W. R. Darden, and M. Griffin: 1994, 'Work and/or fun: measuring hedonic and utilitarian shopping value'. *Journal of consumer research* pp. 644–656.

Bakos, J. Y.: 1997, 'Reducing buyer search costs: Implications for electronic marketplaces'. *Management science* **43**(12), 1676–1692.

Bart, Y., A. T. Stephen, and M. Sarvary: 2014, 'Which products are best suited to mobile advertising? A field study of mobile display advertising effects on consumer attitudes and intentions'. *Journal of Marketing Research* **51**(3), 270–285.

Benbasat, I. and W. Wang: 2005, 'Trust in and adoption of online recommendation agents'. *Journal of the Association for Information Systems* **6**(3), 4.

Bergemann, D. and D. Ozmen: 2006, 'Optimal pricing with recommender systems'. In: *Proceedings of the 7th ACM conference on Electronic commerce*. pp. 43–51.

Berger, J.: 2014, 'Word of mouth and interpersonal communication: A review and directions for future research'. *Journal of Consumer Psychology* **24**(4), 586–607.

Bodapati, A. V.: 2008, 'Recommendation Systems with Purchase Data'. *Journal of Marketing Research* **45**(1), 77–93.

Buhrmester, M., T. Kwang, and S. D. Gosling: 2011, 'Amazon's Mechanical Turk: a New Source of Inexpensive, Yet High-Quality, Data?'. *Psychological Science* **6**(1).

[8] One of the largest e-business and A/B/n testing company that implements recommenders reported that out of about 300 company clients, only 3 were using content-based recommenders and most companies were using purchase-based collaborative filtering recommenders.

Chen, P.-Y., Y.-C. Chou, and R. J. Kauffman: 2009, 'Community-Based Recommender Systems: Analyzing Business Models from a Systems Operator's Perspective'. In: *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*. pp. 1–10.

Chen, P.-Y., S.-y. Wu, and J. Yoon: 2004, 'The impact of online recommendations and consumer feedback on sales'. *ICIS 2004 Proceedings* p. 58.

Chen, Y. and J. Xie: 2008, 'Online consumer review: Word-of-mouth as a new element of marketing communication mix'. *Management Science* **54**(3), 477–491.

Chevalier, J. A. and D. Mayzlin: 2006, 'The Effect of Word-of-Mouth on Sales: Online Book Reviews'. *Journal of Marketing Research* **43**, 345–354.

Choi, J., H. J. Lee, and Y. C. Kim: 2011, 'The influence of social presence on customer intention to reuse online recommender systems: the roles of personalization and product type'. *International Journal of Electronic Commerce* **16**(1), 129–154.

Clemons, E. K.: 2008, 'How information changes consumer behavior and how consumer behavior determines corporate strategy'. *Journal of Management Information Systems* **25**(2), 13–40.

Dai, H., C. Chan, and C. Mogilner: 2014, ''Don't Tell Me What to Do!'People Rely Less on Consumer Reviews for Experiential than Material Purchases'. *People Rely Less on Consumer Reviews for Experiential than Material Purchases (September 8, 2014)*.

Das, A. S., M. Datar, A. Garg, and S. Rajaram: 2007, 'Google news personalization: scalable online collaborative filtering'. In: *Proceedings of the 16th international conference on World Wide Web*. pp. 271–280.

De, P., Y. Hu, and M. S. Rahman: 2010, 'Technology usage and online sales: An empirical study'. *Management Science* **56**(11), 1930–1945.

De, P., Y. Hu, and M. S. Rahman: 2013, 'Product-oriented web technologies and product returns: An exploratory study'. *Information Systems Research* **24**(4), 998–1010.

Degeratu, A. M., A. Rangaswamy, and J. Wu: 2000, 'Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes'. *International Journal of Research in Marketing* **17**(1), 55–78.

Dhar, R. and K. Wertenbroch: 2000, 'Consumer choice between hedonic and utilitarian goods'. *Journal of marketing research* **37**(1), 60–71.

Dias, M. B., D. Locher, M. Li, W. El-Deredy, and P. J. Lisboa: 2008a, 'The value of personalised recommender systems to e-business: a case study'. *Proceedings of the 2008 ACM conference on Recommender systems*.

Dias, M. B., D. Locher, M. Li, W. El-Deredy, and P. J. Lisboa: 2008b, 'The value of personalised recommender systems to e-business: a case study'. In: *Proceedings of the 2008 ACM conference on Recommender systems*. pp. 291–294.

Dimoka, A., Y. Hong, and P. A. Pavlou: 2012, 'On product uncertainty in online markets: Theory and evidence'. *MIS Quarterly* **36**.

Duan, W., B. Gu, and A. B. Whinston: 2008, 'Do online reviews matter? An empirical investigation of panel data'. *Decision support systems* **45**(4), 1007–1016.

Econsultancy and Monetate: 2013, 'The Realities of Online Personalization'. Technical report, Econsultancy and Monetate.

Feick, L. and R. A. Higie: 1992, 'The effects of preference heterogeneity and source characteristics on ad processing and judgements about endorsers'. *Journal of Advertising* **21**(2), 9–24.

Fleder, D. and K. Hosanagar: 2009, 'Blockbuster culture's next rise or fall: The effect of recommender systems on sales diversity'. *Management Science* **55**(5), 697–712.

Girard, T. and P. Dion: 2010, 'Validating the search, experience, and credence product classification framework'. *Journal of Business Research* **63**(9), 1079–1087.

Girard, T., P. Korgaonkar, and R. Silverblatt: 2003, 'Relationship of type of product, shopping orientations, and demographics with preference for shopping on the Internet'. *Journal of Business and Psychology* **18**(1), 101–120.

Hann, I.-H. and C. Terwiesch: 2003, 'Measuring the frictional costs of online transactions: The case of a name-your-own-price channel'. *Management Science* **49**(11), 1563–1579.

Hartigan, J. A.: 1975, 'Clustering algorithms'.

Hinz, J. D. O. and D.-K. J. Eckert: 2010, 'The impact of search and recommendation systems on sales in electronic commerce'. *Business & Information Systems Engineering* **2**(2), 67–77.

Hirschman, E. C. and M. B. Holbrook: 1982, 'Hedonic consumption: emerging concepts, methods and propositions'. *The Journal of Marketing* pp. 92–101.

Hong, Y. and P. A. Pavlou: 2014, 'Product Fit Uncertainty in Online Markets: Nature, Effects, and Antecedents'. *Information Systems Research* **25**(2), 328–344.

Hosanagar, K., D. Fleder, D. Lee, and A. Buja: 2014, 'Will the Global Village Fracture into Tribes: Recommender Systems and Their Effects on Consumers'. *Management Science* **60**(4), 805–823.

Hsieh, Y.-C., H.-C. Chiu, and M.-Y. Chiang: 2005, 'Maintaining a committed online customer: a study across search-experience-credence products'. *Journal of Retailing* **81**(1), 75–82.

Huang, P., N. H. Lurie, and S. Mitra: 2009, 'Searching for experience on the web: an empirical examination of consumer behavior for search and experience goods'. *Journal of Marketing* **73**(2), 55–69.

Ipeirotis, P. G., F. Provost, and J. Wang: 2010, 'Quality Mangagement on Amazon Mechanical Turk'. *HCOMP*.

Jacoby, J., J. C. Olson, and R. A. Haddock: 1971, 'Price, brand name, and product composition characteristics as determinants of perceived quality.'. *Journal of Applied Psychology* **55**(6), 570.

Jannach, D. and K. Hegelich: 2009, 'A case study on the effectiveness of recommendations in the mobile internet'. In: *Proceedings of the third ACM conference on Recommender systems*. pp. 205–208.

Jannach, D., L. Lerche, F. Gedikli, and G. Bonnin: 2013, 'What recommenders recommend – An analysis of accuracy, popularity and sales diversity effects'. *User Modeling, Adaptation, and Personalization - Lecture Notes in Computer Science* **7899**, 25–37.

Johnson, E. J., W. W. Moe, P. S. Fader, S. Bellman, and G. L. Lohse: 2004, 'On the depth and dynamics of online search behavior'. *Management Science* **50**(3), 299–308.

Khan, U., R. Dhar, and K. Wertenbroch: 2005, 'A behavioral decision theory perspective on hedonic and utilitarian choice'. *Inside consumption: Frontiers of research on consumer motives, goals, and desires* pp. 144–165.

Kim, Y. and R. Krishnan: 2013, 'On Product-level Uncertainty and Online Purchase Behaviour: An Empirical Analysis'. *Working*.

Klein, L. R.: 1998, 'Evaluating the potential of interactive media through a new lens: Search versus experience goods'. *Journal of business research* **41**(3), 195–203.

Klein, L. R. and G. T. Ford: 2003, 'Consumer search for information in the digital age: An empirical study of prepurchase search for automobiles'. *Journal of interactive Marketing* **17**(3), 29–49.

Komiak, S. X. and I. Benbasat: 2004, 'Understanding customer trust in agent-mediated electronic commerce, web-mediated electronic commerce, and traditional commerce'. *Information Technology and Management* **5**(1-2), 181–207.

Krishnan, B. C. and M. D. Hartline: 2001, 'Brand equity: is it more important in services?'. *Journal of Services Marketing* **15**(5), 328–342.

Lauraeus-Niinivaara, T., T. Saarinen, and A. Oorni: 2007, 'Knowledge and Choice Uncertainty Affect Consumer Search and Buying Behavior'. In: *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on*. pp. 82–82.

Lee, D. and K. Hosanagar: 2014, 'Impact of Recommender Systems on Sales Volume and Diversity'. *Proceedings of the 2014 International Conference on Information Systems*.

Li, X. and L. Wu: 2013, 'Measuring effects of observational learning and social-network Word-of-Mouth (WOM) on the sales of daily-deal vouchers'. In: *System Sciences (HICSS), 2013 46th Hawaii International Conference on*. pp. 2908–2917.

Liang, T.-P. and J.-S. Huang: 1998, 'An empirical study on consumer acceptance of products in electronic markets: a transaction cost model'. *Decision support systems* **24**(1), 29–43.

Mason, W. A. and S. Suri: 2012, 'Conducting behavioral research on Amazon's Mechanical Turk'. *Behavior Research Methods* **44**(1), 1–23.

Matt, C., T. Hess, and C. Weiß: 2013, 'The Differences between Recommender Technologies in their Impact on Sales Diversity'.

Monetate: 2013, 'Maximize Online Sales with Product Recommendations'. Technical report, Monetate.

Nelson, P.: 1970, 'Information and Consumer Behavior'. *Journal of Political Economy* **78**(2), 311–329.

Nelson, P.: 1974, 'Advertising as Information'. *Journal of Political Economy* **82**(4), 729–754.

Oestreicher-Singer, G. and A. Sundararajan: 2012, 'Recommendation networks and the long tail of electronic commerce'. *MIS Quarterly* **36**(1), 65–84.

Paolacci, G., J. Chandler, and P. G. Ipeirotis: 2010, 'Running Experiements on Amazon Mechcanical Turk'. *Judgement and Decision Making* **5**(5).

Pavlou, P. A.: 2003, 'Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model'. *International journal of electronic commerce* **7**(3), 101–134.

Sen, S. and D. Lerman: 2007, 'Why are you telling me this? An examination into negative consumer reviews on the web'. *Journal of interactive marketing* **21**(4), 76–94.

Senecal, S. and J. Nantel: 2004, 'The influence of online product recommendations on consumers online choices'. *Journal of retailing* **80**(2), 159–169.

Sheng, V., F. Provost, and P. G. Ipeirotis: 2007, 'Get another label? Improving data quality and data mining using multiple, noisy labelers'. *14th ACM SIGKDD* pp. 614–622.

Snow, R., B. O'Connor, D. Jurafsky, and A. Y. Ng: 2008, 'Cheap and fast - but is it good? Evaluating non-expert annotations for natural language tasks'. *Empirical Methods in Natural Language Processing*.

Stafford, M. R., T. F. Stafford, and E. Day: 2002, 'A contingency approach: The effects of spokesperson type and service type on service advertising perceptions'. *Journal of Advertising* **31**(2), 17–35.

Stigler, G. J.: 1961, 'The economics of information'. *The journal of political economy* pp. 213–225.

Strahilevitz, M. A. and J. Myers: 1998, 'Donations to charity as purchase incentives: How well they work may depend on what you are trying to sell'. *Journal of consumer research* **24**(4), 434.

Sun, M.: 2012, 'How does the variance of product ratings matter?'. *Management Science* **58**(4), 696–707.

Thompson, C.: 2008, 'If you liked this, you're sure to love that'. *The New York Times* **21**.

To, P.-L., C. Liao, and T.-H. Lin: 2007, 'Shopping motivations on Internet: A study based on utilitarian and hedonic value'. *Technovation* **27**(12), 774–787.

Tucker, C. and J. Zhang: 2011, 'How does popularity information affect choices? A field experiment'. *Management Science* **57**(5), 828–842.

Weathers, D., S. Sharma, and S. L. Wood: 2007, 'Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods'. *Journal of Retailing* **83**(4), 393–401.

Wu, D., G. Ray, X. Geng, and A. Whinston: 2004, 'Implications of reduced search cost and free riding in e-commerce'. *Marketing Science* **23**(2), 255–262.

Xiao, B. and I. Benbasat: 2007, 'E-commerce product recommendation agents: use, characteristics, and impact'. *Mis Quarterly* **31**(1), 137–209.

Zhu, F. and X. Zhang: 2010, 'Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics'. *Journal of marketing* **74**(2), 133–148.

Appendix A: Amazon Mechanical Turk Strategy & Survey Instrument for Item Attribute Tagging

Following best-practices in the literature, we employ the following strategies to improve the quality of attribute tagging by the Turkers in our study.

1. For each message, at least 5 different Turkers' inputs are recorded.
2. We restrict the quality of Turkers included in our study to comprise only those with at least 500 reported completed tasks and 98% or better reported task-approval rates.
3. We screened out the workers by giving them a simple test to see if they understood the instructions. Those who failed were banned from participating.
4. We use only Turkers from the countries where English is the primary language to filter out those potentially not proficient in English.
5. We refined our survey instrument through an iterative series of about several pilot studies, in which we asked Turkers to identify confusing or unclear questions. In each iteration, we asked 10-30 Turkers to identify confusing questions and the reasons they found those questions confusing. We refined the survey in this manner till almost all queried Turkers stated no questions were confusing.
6. To filter out participants who were not paying attention, we included an easily verifiable attention question. Responses from Turkers that failed the attention test are dropped from the data.
7. On average, we found that survey took a little over 4 minutes and it typically took at least 1 minute or more to completely read the questions. We defined less than 30 seconds to be too short, and discarded any message tags with completion times shorter than that duration to filter out inattentive Turkers and automated programs ("bots").
8. Once a Turker tags more than 100 messages, a couple of tagged samples are randomly picked and manually examined for quality and performance. This process identified several high-volume Turkers who completed all surveys in less than 15 seconds and tagged several thousands of messages (there were also Turkers who took time to complete the surveys but chose seemingly random answers). We concluded these were automated programs. These results were dropped, and the Turkers "hard blocked" from the survey, via the blocking option provided in AMT.

The existing AMT literature has documented evidence that several of the strategies implemented above improves the quality of the data generated (Mason and Suri (2012); Ipeirotis et al. (2010); Paolacci et al. (2010)). Snow et al. (2008) demonstrates that combining results from a few Turkers can produce data equivalent in quality to that of expert labelers for a variety of tagging and content-coding tasks. Similarly, Sheng et al. (2007) document that repeated labeling of the type we implement wherein each message is tagged by multiple Turkers, is preferable to single labeling in which one person tags one sentence. Finally, evaluating AMT based studies, Buhrmester et al. (2011) concludes that (1) Turkers are demographically more diverse than regular psychometric studies samples, and (2) the data obtained are at least as reliable as those obtained via traditional methods as measured by psychometric standards such as Cronbach's Alpha or Krippendorff's Alpha, commonly used inter-rater reliability measures.

The following table provides the construct we've used, literature sources we've adapted the measurement survey instrument and operating definitions, and inter-rater reliability measure achieved.

Construct & Measurement Question Sources (Krippendorff's Alpha)	Measurement Questions (Likert Scale from 1- Least 7-Most)
Hedonic VS. Utilitarian (0.9455) Adapted from Dhar and Wertenbroch (2000); Strahilevitz and Myers (1998); Bart et al. (2014); Khan et al. (2005); Babin et al. (1994)	Product consumption is driven by different motives. A couple of example motivation are based on the idea of hedonic (want) consumption vs. utilitarian (need) consumption. Hedonic, Pleasure-oriented consumption is motivated mainly by the desire for sensual pleasure, fantasy, and fun (e.g., movies, perfume, an art piece). Utilitarian, goal-oriented consumption is motivated mainly by the desire to fill a basic need or accomplish a functional task (e.g., paper clips, dishwashing agent, vacuum cleaner). **Given the above definition of hedonic and utilitarian value of a product, rate the product above in the scale below on hedonic value and utilitarian value.** • **Hedonic Value [1 NOT AT ALL HEDONIC to 7 PURELY HEDONIC]** • **Utilitarian Value [1 NOT AT ALL UTILITARIAN to 7 PURELY UTILITARIAN]** • **Please give the scale on how much comparative utilitarian VS hedonic value the product offers. [1 PURELY UTILITARIAN to 7 PURELY HEDONIC]**
Search VS. Experience (0.8433) Adapted from Krishnan and Hartline (2001); Hsieh et al. (2005); Huang et al. (2009); Girard and Dion (2010); Klein (1998); Klein and Ford (2003)	• **How well could you judge the attributes or quality of this product even BEFORE you purchased or used it? [1 NOT WELL AT ALL to 7 EXTREMELY WELL]** [For example, some products are easy to judge the attributes/quality of BEFORE you've purchased or used them (e.g., Computers, Printer Ink) while others (e.g., Movies, Food, Wine) are not.] • **How well could you judge the attributes or quality of this product even AFTER you purchased or used it? [1 NOT WELL AT ALL to 7 EXTREMELY WELL]** [For example, some products are easy to judge the attributes/quality of AFTER you've purchased or used them (e.g., Movies, Food, Wine)] • **How important is it for you to see, touch, hear, taste, smell (whichever applies) this product IN PERSON to evaluate its attributes? [1 NOT IMPORTANT AT ALL to 7 EXTREMELY IMPORTANT]** [For example, you may want to touch a piece of clothing to determine the quality of fabric, but this may not be necessary for printer toner or vitamin.] • **How well can you evaluate the product using only information provided by retailer and/or manufacturer about this product's attributes and features? [1 NOT WELL AT ALL to 7 EXTREMELY WELL]**
Durability	**Please rate how durable the product is** Some products are extremely durable and do not quickly wear out (e.g., Cars and Mobile Phones) while others are less durable and wear out quickly (e.g., Food, Gasoline, Papers, Medications). Assume average usage and no accident.
Brand Prominence Proxy	**Have you heard of the brand/company that made this product?**

Table 10: **Survey Instrument:** We use existing and widely used operational definitions and measurement questions to tag the items in our dataset. Median Krippendorff's alpha, a standard measure of inter-rater reliability measure is provided and are well above the acceptable measure of 0.8.

Appendix B: Measurement Robustness

For both hedonic-utilitarian and search-experience attributes, we utilized the clustering algorithm to classify a product dichotomously into a hedonic or utilitarian product, as well as a search or experience product. The decision to use dichotomous classifications was for practical convenience and to use existing measurement strategies. While the literature has acknowledged the shortcomings of dichotomous classification schemes, it is still commonly used in the literature based on dominant attributes (e.g., Huang et al. (2009), Senecal and Nantel (2004)). However, since these product attributes could be continuous qualities, we repeated analyses in which the search-experience and hedonic-utilitarian attributes are denoted by a scale from 1 to 7. We obtain qualitatively similar results.