How Do Product Attributes Moderate the Impact of Recommender Systems?

Dokyun Lee
Carnegie Mellon University

Kartik Hosanagar
The Wharton School *

Abstract

We investigate the moderating effect of product attributes and review ratings on two different impacts of a purchase-based collaborative filtering recommender system on an e-commerce site: number of product views and conversion to purchase conditional on a product view. We run a randomized field experiment with a top North American retailer's website with 184,375 users split into a recommender-treated and a control group. We tag the attributes of 37,125 unique products via Amazon Mechanical Turk and augment the data with other sources of product information on the site (e.g., review data, product descriptions, etc.).

Our study confirms that the use of a recommender increases both views and conversion rate among treated users for all products, but this increase is moderated by product attributes and review ratings. We find that a recommender's positive impact on product views is greater for utilitarian products compared to hedonic products and for experience products compared to search products. In contrast, a recommender's positive impact on conversion rate is greater for hedonic product compared to utilitarian product. Furthermore, we find that recommenders' positive impact on conversion rate is greater for products with lower average review ratings, suggesting that a recommender acts as a substitute to high review ratings. While the opposite is true for product views – recommender and high review ratings are complements in increasing views. We discuss the potential mechanisms behind our results as well as their managerial implications.

Keywords: E-commerce, Personalization, Recommender Systems, Consumer Reviews, Product Attributes, Search Cost, Electronic Markets, Awareness, Salience.

*The authors gratefully acknowledge the financial support from the Berkman Faculty Fund, the Jay H. Baker Retailing Center, the Mack Institute for Innovation Management, and the Fishman-Davidson Center for Service and Operations Management. All errors are our own.

1 Introduction

Recommender systems (RS) are ubiquitous on the web. E-commerce sites regularly use such systems to guide consumers with messages like "People who purchased this item also purchased..." so as to increase up-selling and cross-selling opportunities. Recommenders aid online shopping by reducing search cost (Anderson, 2008) and product uncertainty for consumers (Xiao and Benbasat, 2007). As such, many existing studies have shown that RS impact consumer choice and increase revenues for firms (Anderson, 2008; Bodapati, 2008; Das et al., 2007; De et al., 2010; Dias et al., 2008; Fleder and Hosanagar, 2009; Hosanagar et al., 2014; Jannach and Hegelich, 2009; Monetate, 2013; Oestreicher-Singer and Sundararajan, 2012; Thompson, 2008).

While it is well known that RS impact consumer choice and generally lead to an increase in sales volume, there is little discussion of how product-specific attributes or review ratings influence the effectiveness of recommenders. This is surprising since recommenders impact consumer search and learning, and help resolve product uncertainty, much like other e-commerce features such as reviews and product descriptions. Given that different products have different search costs (Hann and Terwiesch, 2003) and product uncertainty[1] (Dimoka et al., 2012), recommenders' impact should vary from product to product. Furthermore, user-generated reviews and other sources of information that reduce product uncertainty are likely to interact with recommenders. Therefore, the effective implementation of recommenders must account for interactions with other factors.

However, the literature does not clarify how the impact of recommenders is moderated by factors such as product type, product attributes, and other sources of product information on retailer websites, like consumer-generated reviews and item descriptions. In this study, through a large-scale randomized field experiment, we investigate how factors that influence product search cost, uncertainty, and risk interact with the positive impact of recommender system on both the view and purchase decision. In particular, product attributes such as hedonic vs. utilitarian products and search vs. experience are examined. Theoretical work in RS distinguish the impact of recommenders into *awareness* (when a recommender introduces products and changes awareness or consideration set and in the process reduces search cost) and *saliency* (the very act of recommendation acts as signal and changes purchase intention or consumers' sensitivity to uncertainty). We measure two impacts empirically by examining visit to a product page which we call product views (awareness) and product conversion conditional on views (saliency). By doing so, we provide nuanced insight

[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).

on how recommenders interact with product attributes at different stages of a product's purchase journey. We also present empirical evidence for whether recommenders serve as substitutes or complements for high review ratings, as well as how recommenders interact with other product-specific, factors such as product price and product descriptions.

Answers to the questions above can guide recommender implementation in e-commerce and also provide insight into consumer purchase behavior. According to a study by Econsultancy and Monetate (2013), while 94% of e-commerce sites consider recommendation systems to be a critical competitive advantage, only 15% of these companies were getting good returns from their investments. Seventy-two percent of those surveyed attributed failure to a lack of knowledge as to when, what, and how to deploy such a system. This is because recommenders almost always coexist with other features on the web that influence purchase decisions. Understanding whether and how the effectiveness of recommender systems varies across product categories and by review volume and valence can help managers better understand how best to deploy recommender systems.

Despite their managerial significance and despite relevant theoretical studies, the research questions posed are empirically untested or are only partially answered. This is largely because the literature lacks data that spans multiple product categories to provide the contrast needed to identify the impact of recommenders by product type. Our study attempts to fill this gap by running a randomized field experiment on an e-commerce site of a top retailer in North America.[2] In the experiment, a treated group received recommendations from a purchase-based collaborative filtering algorithm ("people who bought this also bought ...") whereas a control group received no recommendations. We utilize Amazon Mechanical Turk to efficiently content-code item attributes for more than 37,000 unique products. We analyze the resulting dataset to study how product attributes moderate the impact of recommenders on product views and conversion rate (conditional on view).

Our main results are as follows. We first confirm that the use of a recommender increases both views and conversion rate (conditional on view) among treated users for all products, but this increase is moderated by product attributes and review ratings. We find that a recommender's positive impact on product views is greater for utilitarian products compared to hedonic products and for experience products compared to search products. In contrast, a recommender's positive impact on conversion rate is greater for hedonic product compared to utilitarian product. Contrary to conjectures found in the literature, the search-experience attribute did not influence recommenders'

[2]We are not allowed to disclose the identity of the company. It is one of the top five retailers both offline and online.

effect on conversion rates. Together, these results suggest that the exact mechanism through which recommenders help increase product sales is different for hedonic versus utilitarian products and search versus experience products. Furthermore, we find that recommenders' positive impact on conversion rate is greater for products with lower average review ratings, suggesting that a recommender acts as a substitute to high review ratings. While the opposite is true for product views – recommender and high review ratings are complements in increasing views.

Our results make several contributions. From an academic standpoint, while there are many papers discussing the moderating impact of product attributes on recommender performance, they are either theoretical or based on small-scale lab experiments. This is one of the first large-scale field experiment studies to examine the moderating effects of product attributes like hedonic-utilitarian considerations, search-experience aspects, and review ratings on both the awareness and saliency impact of a recommender. We increase external validity by 1) working with a retailer that ranks among the top five in the world in e-commerce revenue and sells across a wide range of product categories, and 2) deploying a commonly used open-source implementation of a collaborative filtering algorithm. In practice, our study has several managerial implications. First, managers can determine which specific products would be better served by recommenders at different phases of the purchase journey. Second, managers will have insight into how other e-commerce features—such as product descriptions and user-generated review ratings—interact with recommenders. Managers can then optimize e-commerce sites appropriately and decide how best to combine features (e.g., reviews, more detailed descriptions, recommenders, etc.). Third, by identifying product types for which traditional recommender designs have a relatively subdued positive impact, managers can focus on developing novel recommender algorithms or other e-commerce features that are better suited for those product categories.

2 Literature on Factors Interacting with Recommender Systems

Recommender systems are known to influence consumer purchases via two main effects: *awareness* and *salience* (Fleder and Hosanagar, 2009; Xiao and Benbasat, 2007). If a consumer is not aware of a product, a recommender can help by alerting the consumer to a relevant item (Anderson, 2008; Moore and Punj, 2001; Häubl and Murray, 2003). In other words, the first effect of recommenders is a reduction of search cost for the consumers. Extant work tied to consumer search theory (Weitzman, 1979), bounded rationality (Gigerenzer and Selten, 2002), and cognitive costs (Benbasat and Todd,

1996) all suggest that consumers have a limited attention budget, incur a non-zero search cost, and ultimately make a trade-off between decision quality and search cost. Häubl and Trifts (2000) show that the use of the recommender agent reduces consumers' search effort for product information and improves purchase decisions. Häubl and Murray (2006) reproduce a similar finding.

Second, even if a consumer is aware of the product, the recommendation may boost the salience of the recommended product and increase its purchase probability (i.e., conditional on awareness). For example, the mere act of recommendation may serve as a signal and reduce product fit uncertainty as defined by Hong and Pavlou (2014), persuade the consumer that the product is better than the alternatives via an anchoring effect (Adomavicius et al., 2011), or simply lower the barrier to purchase. This salience impact has often been examined through the lens of consumer trust and perceived risk. Existing IS literature has discussed that the product-related perceived risk (Pavlou, 2003) of making a purchase on an e-commerce site may manifest through product fit uncertainty and product quality uncertainty (Hong and Pavlou, 2014). Recommenders have been shown to increase trust and reduce perceived risk in e-commerce settings (Komiak and Benbasat, 2006; Xiao and Benbasat, 2007). Further, Spiekermann and Paraschiv (2002) argue that customers have different expectations and acceptance of recommender agents based on varying degrees of perceived product risk. While the salience effect of recommenders can be attributed to reduction in perceived risk, there is evidence that it goes beyond this. Adomavicius et al. (2014) provide evidence that the rating presented by a recommender system serves as an anchor for the consumer's constructed preference. Viewers' preference ratings are malleable and can be significantly influenced by the recommendation received.

The awareness and salience effects of recommender systems are similar to the informative (shifting beliefs about product existence or prices) and persuasive (shifting preferences directly) effects of advertising (Bagwell, 2007; Butters, 1977), which are also related to search cost and product uncertainty respectively. Given different products have different search costs (Hann and Terwiesch, 2003) and fit uncertainty (Hong and Pavlou, 2014), research in marketing shows that informative and persuasive effects of advertising vary across product types (Bagwell, 2007; Nelson, 1974). Taking into account the role of search costs and product risks in driving awareness and salience effects of recommenders, we expect that the impact of recommenders is likely to also be moderated by product type. In fact, Swaminathan (2003) shows that the recommender's effect on decision quality and search behavior is moderated by product category risk and complexity. Senecal and Nantel (2004) and Aggarwal and Vaidyanathan (2005) also suggest that the impact of recommenders will

vary for search versus experience products.

This paper builds on a large stream of work on how recommenders influence consumer search and purchase behavior and seeks to examine how product attributes moderate the impact of recommenders on consumer choice in e-commerce. While the literature suggests, either directly or indirectly, that the impact of recommenders will vary across products, there is very limited empirical evidence to help us understand the phenomenon. Existing studies on the subject rely on small-scale lab experiments with few products and often present conflicting findings (Xiao and Benbasat, 2007). For example, Senecal and Nantel (2004) and Aggarwal and Vaidyanathan (2005) are both based on a study in the lab of two products (one search and one experience product) and find opposing results. We employ a large-scale field experiment with more than 37,000 unique products across a variety of different categories. The scale of our dataset not only enables greater generalizability of our findings but also allows us to consider multiple product types (search/experience, utilitarian/hedonic, durable/consumable, branded/non-branded, etc) while controlling for other relevant factors such as review ratings and product descriptions. To the best of our knowledge, this paper provides one of the first large-scale field experiment aimed at answering how product attributes moderate the different impacts of recommenders.

2.1 Hypotheses

We look at how recommenders impact product discovery and purchases. In e-commerce settings, these are observable through changes in the number of views a product receives and the rate at which these views convert to purchases. Prior research has shown that recommenders positively impact both views and purchases (Lee and Hosanagar (2018); Kumar and Hosanagar (2018)). Next we discuss how recommenders might impact different products' views and conversion to purchases differently.

The literature has identified a few key constructs of products that affect consumer search and purchase behavior. One of the most studied constructs is the search-experience quality of a product. Nelson's seminal work in the 1970s on the 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 (e.g., computers, phones). Experience goods are dominated by characteristics that can only be discerned by using the product, or they have characteristics that are subjective in nature (e.g., movies, wine). Nelson's search and experience framework has been used to explain how people

react to advertising, how they search for different products online, and, ultimately, how they make purchases (Klein, 1998; Klein and Ford, 2003). We utilize this framework to label products on a search-experience scale and explore how that interacts with the impact of recommenders.

Another relevant construct of product attributes is the hedonic-utilitarian framework of classifying products. This construct measures how consumers' purchases are driven by a "need" or a "want." A hedonic product is characterized by pleasure-oriented consumption whereas a utilitarian product is characterized by goal-oriented consumption tied to a functional need (Dhar and Wertenbroch, 2000; Khan et al., 2005). The hedonic-utilitarian characteristic has been shown to influence consumer product search behavior, purchase decisions, and even consumers' valuation of products (Hirschman and Holbrook, 1982; Bart et al., 2014; Khan et al., 2005).

While recommender systems help address search cost and product uncertainty for consumers through awareness and saliency, they are by no means the only tool used by retailers for such purposes. The most basic alternative available is the level of detail in the product description, which influences search cost and uncertainty and purchase decisions (De et al., 2013, 2010). In addition, user-generated reviews further aid in reducing product uncertainty. Therefore, we posit that a recommender's effectiveness also interacts with other product attributes, such as reviews and product descriptions. By controlling for these factors in our investigation, we holistically examine the moderating effect of product attributes on the impact of recommender systems.

Lastly, we note that the list of moderating variables we consider is by no means exhaustive. Our attention is focused on identifying attributes that: 1) are shown in the literature to influence consumer search and purchase behavior, 2) are relevant and easily available to managers in most settings, and 3) have strong theoretical foundations with well-used operational definitions. Following these criteria, we also identified several control variables that may influence a recommender's effectiveness.

Next, we discuss each variable and the 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. Additional details and sources of survey instruments for measuring product attributes are discussed in the Appendix.

Hedonic vs. Utilitarian

A characteristic often used to categorize products across industries is whether the product is predominantly 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) defines 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 products where 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' valuation of products (Hirschman and Holbrook, 1982; Bart et al., 2014; Khan et al., 2005).

With respect to online shopping, utilitarian products predominantly consist of objective attributes that serve specific functions (e.g., hammers, memory cards, and ink toners) compared to hedonic products mainly consisting of subjective attributes related to senses (e.g., touch, smell, taste). Consequently, utilitarian product shopping is mostly goal-oriented (To et al., 2007; Kim et al., 2012) and associated with higher perceived financial and functional risk compared to hedonic products (Sarkar, 2011; Chiu et al., 2014). As recommenders excel in providing alternative products for consideration in short time, consumers with limited attention budget may view more products when shopping for utilitarian products under the influence of recommenders. This may be due to the increased perceived risk of utilitarian products and the consequential fear of not finding the right product. That is, while the recommenders may also increase the number of hedonic product awareness sets, the effect size may be greater for utilitarian products. Thus, our hypothesis on how hedonic-utilitarian attributes moderate the awareness impact of recommenders is as follows:

Hypothesis 1 *A recommender's positive impact on product views will be greater for utilitarian products compared to hedonic products.*

In relation to purchase intentions, the existing literature has shown that the hedonic-utilitarian attribute moderates the trust and reuse intention with recommender systems. For example, Choi et al. (2011) suggests that consumers' trust for recommender systems and reuse intention is increased when the recommender provides a "social presence," which is defined as "*the extent to which a website allows users to experience others as psychologically present.*" This increase in trust and reuse 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.

For conversion conditional on views, the awareness impact of recommenders are no longer rel-

evant. However, the saliency impact (similar to persuasive effect in advertising) may play a role by interacting with product risk and uncertainty. Studies have shown that the effectiveness of product endorsements depends on whether the product is utilitarian or hedonic (Feick and Higie, 1992; Stafford et al., 2002). In fact, 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 has been suggested that consumers prefer opinions of people who are more like them (Feick and Higie, 1992). The collaborative filtering algorithm commonly used in e-commerce provides recommendations to a consumer based on the purchase histories of other similar consumers, and they clearly signal this. Additionally, utilitarian products are also associated with higher perceived financial and functional risk compared to hedonic products (Sarkar, 2011; Chiu et al., 2014). As higher perceived risk is associated with lower acceptance of recommenders (Xiao and Benbasat, 2007), the saliency effect of a recommender may be reduced for utilitarian products compared to hedonic products. Thus, we posit that recommender's positive effect on conversion conditional on views (saliency) will be higher for hedonic products. Our hypothesis is as follows:

Hypothesis 2 *A recommender's positive impact on conversion to purchase (conditional on views) will be greater for hedonic products compared to utilitarian products.*

Search vs. Experience

Philip Nelson's seminal work on the economics of information and advertising (Nelson, 1970, 1974) classified products into search and experience goods. Search goods consist of attributes that can be easily discerned before purchase and are dominated by objective attributes with lower informational search cost, 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 the 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 product classification influences consumers' search, consideration set formation, 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; Animesh et al., 2010). 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, and apparel are increasingly sold on e-commerce sites, these sites still find it challenging to meet consumers' information needs to convert views into sales, or to prevent high rates of return (Hong and Pavlou, 2014; Dimoka et al., 2012). A few studies have suggested several remedies, including 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. (2010, 2013); Kumar and Tan (2015)).

Nelson theorized that consumers' search for experience goods will be characterized by heavier reliance on word-of-mouth and the experiences of other consumers, since the cost of information via other routes is higher (Nelson, 1970, 1974; Klein, 1998). Consequently, Nelson hypothesized that experience good sellers will focus on persuasive and brand-focused tactics such as word-of-mouth, testimonials, and celebrity endorsements. Conversely, search good sellers will prioritize their advertising with informative and easy-to-discern facts about the products. However, it is not clear how search-experience attributes will influence recommenders' performance. Existing literature on the moderating influence of search-experience attributes 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 is based on lab experiments featuring only two products: wine and calculators. In contrast, a study by Aggarwal and Vaidyanathan (2005), once again with only two products, suggests that consumers perceive recommenders to be more effective for search goods than for experienced goods. Thus, the current literature is lacking evidence based on realistic field data and a wide variety of products.

As consumers face greater product uncertainty when exploring and searching for experience products compared to search products, a recommender's awareness impact may differ. Due to the increased uncertainty (whether it is product fit or quality) associated with experience products, consumers may engage in a higher amount of exploration and searching for experience products online. Assuming that the user trusts recommender systems, since recommenders excel in providing compelling alternative products to consider, the positive awareness impact of a recommender may be higher for experience products than for search products. Thus we develop the following hypothesis:

Hypothesis 3 *A recommender's positive impact on product views will be greater for experience products compared to search products.*

For conversion conditional on views, the awareness impact is no longer relevant and saliency impact takes over. At this stage, the influence of a recommender in driving conversions for search or experience goods depends on consumers' trust in the recommender system. If consumers trust recommenders to serve as a replacement for costly reduction of product uncertainty, the recommenders should be, relatively speaking, more impactful for experience goods. Given one cannot easily resolve product uncertainty for experience goods, saliency effect might be stronger for experience goods. Recent literature in recommender systems has dubbed the recommender agents "digitized word-of-mouth" (Chen et al., 2009) where consumers adapt and trust recommender systems as "social actors" and perceive human characteristics in them (Benbasat and Wang, 2005; Xiao and Benbasat, 2007; Komiak and Benbasat, 2004). Nelson's theory suggests that consumers rely more on word-of-mouth for experience goods and recent literature has shown that recommender systems are accepted and trusted as a form of word-of-mouth. In accordance with Nelson's theory on experience goods, we develop the following hypothesis for how search-experience attributes may moderate the conversion conditional on views (saliency impact of recommenders):

Hypothesis 4 *A recommender's positive impact on conversion to purchase (conditional on views) will be greater for experience products compared to search products.*

2.2 Consumer Reviews

It is well documented in past studies that user-generated reviews influence online consumers' purchase intentions (Chen et al., 2004; Chen and Xie, 2008; Duan et al., 2008; Gu et al., 2012; Sun, 2012; Chevalier and Mayzlin, 2006; Berger, 2014). However, results are mixed in that review ratings do not always influence consumers. The literature suggests that the effect of reviews on sales is moderated by whether the product is 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, studies show that consumers tend to discount, or even ignore, review ratings when the volume of reviews is low (Li and Wu, 2013; Duan et al., 2008; Chen et al., 2004). For niche items, the impact of reviews can be greater (Zhu and Zhang, 2010). High ratings have a more positive influence on consumer purchase intent for niche items (Tucker and Zhang, 2011). Similarly, the effect of reviews is 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 argument that consumers rely more on external informational sources such as reviews when search cost and uncertainty are greater (e.g., niche items or experience items).

Just like consumer reviews, recommender systems serve as electronic word-of-mouth (Chen et al., 2009) and influence uncertainty and search cost for online consumers (Clemons, 2008). As discussed, recommenders reduce search cost via awareness impact, by suggesting products to consider. However, how recommenders influence consumer uncertainty is slightly different from how reviews influence consumer uncertainty. While review ratings, volumes, and user-generated text actually reduce product quality uncertainty through more information (Hong and Pavlou, 2014), recommenders work to increase purchase probability in light of product uncertainty through the saliency effect, given that consumers trust recommenders (Xiao and Benbasat, 2007). There are also some similarities in that recommendation itself may work as signal of quality just as high review ratings do. Thus, it is highly likely that reviews and recommenders interact with each other to influence both consumer views and conversion conditional on views. For consumer search and views, as both review ratings and volumes do signal a certain level of quality, the view rate will only be bolstered by the recommendations – that is, among recommended products, consumers are more likely to click higher rated products to view. Thus, we hypothesize that high review rating and recommendation signals are complementary.

Hypothesis 5 *A recommender's positive impact on product views will be greater for a product with higher average ratings than for a product with lower average ratings. That is, recommenders are complements for high review ratings in views.*

However, it is not clear if recommender systems act as *substitutes* or *complements* for reviews in conversion, since they serve a similar yet slightly different purpose. If the impact of recommendations is greater for products with higher average ratings, they serve as complements. On the other hand, if the impact decreases for products with higher ratings, they act more like substitutes. Reviews mainly reduce uncertainty and provide product fit and quality information while recommenders signal a personalized fit. Since recommender systems provide personalized fit information on top of consumer reviews, it is likely that recommenders may serve as a substitute for consumer reviews and provide additional information and value for consumers. For example, if a product's rating is low or mixed and it is recommended, then the consumer may be willing to discount the product's lower rating on the assumption that the product, while not suitable for the entire buying population, is a better fit for that consumer. Conversely, a product with high average ratings might enjoy high conversion rates as is and may not receive as significant a boost in its purchase probability (conditional on viewing) from being recommended. Additionally, the cost of consuming reviews vs.

recommendations is also different in that it takes a longer time to process review summaries (mean and variance) and to read the reviews compared to getting a straightforward recommendation. Thus, consumers may be more willing to ignore reviews when the product has already been recommended. Another reason consumers may discount higher-average ratings when recommenders are present is because consumers may not agree with other consumers' reviews. Indeed, Dai et al. (2014) claim that consumers rely less on other consumers' reviews when shopping for experience products because they believe that other people's reviews are not representative of their own evaluations. Based on this discussion, our hypothesis is as follows:

Hypothesis 6 *A recommender's positive impact on product views will be greater for a product with lower average ratings than for a product with higher average ratings. That is, recommenders are substitutes for high review ratings in conversion.*

3 Data

Our main dataset is derived from a field experiment run by a partner company. The company randomly assigned their customers into either a treated group to whom a recommendation panel is shown or a control group to whom the recommendation panel is not shown. The dataset records all products viewed and purchased by users in the field experiment. 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 administered to workers on Amazon Mechanical Turk, an online marketplace for data tagging and cleaning. We now describe our experimental designs, the details of the recommender algorithms, and data limitations.

3.1 Field Experiment & Data Description

Experimental Design The field experiment was run in partnership with one of the top retailers in North America over a two-week period in August 2013. The company has both an online and offline presence and is one of the top five online retailers in the North American region by size and revenue. The experiment was run on its website and the analysis that follows is based only on its e-commerce data. 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.[3] Then, the website tracks visitors' behaviors—including users' viewing

[3]The A/B/n company outsources unique user identification to a specialized firm that utilizes a variety of data—such as IP, user-agent, login data, session ID, device, etc.—to minimize duplicate and multi-device prob-

logs and purchases—throughout the duration of the field experiment. Users are randomly chosen to be in the control group or in the treatment group; they remain in their respective groups throughout the experiment. Upon clicking and visiting the product description page of a particular item, the visitors in the treated group are shown a recommender panel, as seen in Figure 1. Visitors in the control group do not see this panel.[4] Thus, our randomization is at the user level. Finally, to check if the random assignment truly holds, at the runtime of the experiment, we had assigned a separate additional set of users (twice as large as users in our data) into each groups but did not actually treat them (i.e., these users might have label "control" or "treated" but they all do not see recommenders through the experiment). If the randomized labeling holds true, these users will behave similarly. On these users, we ran a Kolmogorov-Smirnov (KS) test on variables such as total number of item views, purchases, and wallet sizes. We could not reject the null and the non-significant p-values (reported in Appendix B) suggest that our randomization holds true.

Recommender System Design Details

There are many types of recommender systems. It is infeasible to run all of them in the field experiment setting due to the amount of resources required to implement and the 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..." (Sarwar et al., 2001; Adomavicius and Tuzhilin, 2005).[5] Our recommender system implements an item-based collaborative filtering algorithm using an open-source, machine learning framework known as Apache Mahout (mahout.apache.org). The item-based CF we implement in this experiment computes item-item similarity using past consumer purchase data. Then, the top N candidate products not previously purchased by the focal consumer are recommended. The algorithm uses purchase data of the entire website 60 days prior to the start of the

lems. While no method for unique user identification is perfect, this is the state-of-the-art approach in both research and practice.

[4]When we study the impact of recommenders, it is worth asking "relative to what?". One could study the impact of recommender systems relative to a random recommender or to a system that showcases the most popular items. These alternatives are not appealing for multiple reasons. From a practical perspective, when a consumer is on a specific product page, showing randomly generated recommendations or globally popular items will effectively show irrelevant items. Such a user experience is unacceptable to a retailer of the scale of our partner. Furthermore, it is not a format that is used by any retailer we know. The most common format used by retailers is to show other items that are either co-viewed or co-purchased with a focal item.

[5]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 a consumer bought or liked in the past. Collaborative filtering recommenders are unaware of product attributes and recommend products either purchased or liked by similar consumers, where similarity is measured by historical purchase or ratings.

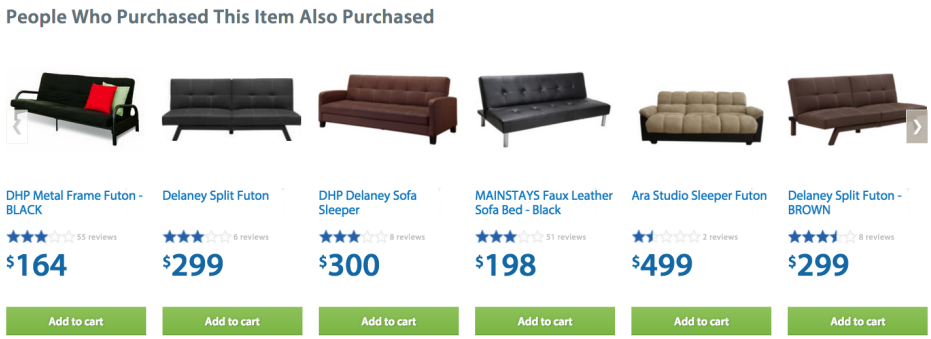

Figure 1: **Recommendation Panel** : Example of a recommender shown to a consumer. We implement the most commonly used recommender algorithm, "People who purchased this item also purchased...", or item-based collaborative filtering via the Apache Mahout framework.

experiment and the item-item similarity matrix is recomputed every three days. The recommender panel shows up to six predicted items in the recommendation panel. The exact number of products displayed is a function of the user's screen width.

Formally, following notations of Sarwar et al. (2001) we define a set of users $\mathcal{U} = \{u_1, u_2, \ldots, u_m\}$ and a set of items $\mathcal{I} = \{i_1, i_2, \ldots, i_n\}$. The users can give ratings for each item, which are stored in a $m \times n$ matrix $\mathbf{R}$ of $user \times item$, or this matrix can be values simply record views or purchases. In our case, they are based on items purchased by users. For an active user $u_a \in \mathcal{U}$, the task of the collaborative filter is to 1) compute item-item similarity matrix based on matrix $\mathbf{R}$ and 2) find top-N items to recommend to the user. In our case, we use the cosine similarity, $cos(x, y) = \frac{x \cdot y}{||x|| \cdot ||y||}$, for the distance measure of two item vectors and find the top six items.

3.2 Data Limitations

While our data are relatively clean and causal effects are easier to extract compared to observational data, our data are not without limitations. We discuss them below.

- Particular Recommender System Algorithm: A main limitation of our study is that we do not implement multiple types of recommender systems. To this end, we discuss how the result is generalizable. First, we use one of the most commonly used item-based collaborative filtering algorithm, as described in a seminal paper by Sarwar et al. (2001). Then, in implementing the algorithm, we also utilized a leading open-source platform, Apache Mahout. Our conversations with a large e-business analytics firm (which implements recommenders for many clients, including Fortune 500 companies) revealed that out of several hundred firms that implement

recommenders, only two utilized content-based recommenders. This translates to less than 1% of the firms. The rest utilized collaborative filtering. A majority of companies utilize collaborative filtering algorithms simply because content-based recommender systems require expensive attribute tagging and content analysis.[6] In setting the variety of parameters of recommender systems, such as the retraining days or the pre-computation time-period, our partner used the default parameters in the system. Thus, while we are limited to one type of algorithm, our results are broadly applicable because of the widespread use of the algorithm.

- Original recommendation set unknown: Another limitation of our study design is that we do not know which products were actually recommended via the recommender. This is not commonly tracked in many recommenders and our partner decided against investing the effort and resources needed to modify the open-source algorithm and record the actual products recommended in each session. As a result, we cannot analyze whether a specific purchase resulted from a recommendation. Instead, we can compare the purchase behavior of treated vs. control users at the aggregate level and, on account of the randomization, attribute the difference to the recommender. The implication for the study is that while the direction of effect sizes we report is correct, the magnitude of our estimates are conservative (Greevy et al., 2004). To see why, let p_{rec} =probability that a user buys a product when the product view is driven by a recommendation and let p_{base}= probability that a user buys when the product view is not driven by the recommender system. If $0 \leq \alpha \leq 1$ is the proportion of product views driven by recommendations for the treated users, then $p_{treated}$ = the probability of purchase/conversion for a treated user can be written as $p_{treated} = \alpha p_{rec} + (1 - \alpha)p_{base}$. We estimate $p_{treated}$ and p_{base}in our analysis. If $p_{treated} > p_{base}$then it follows that $p_{rec} > p_{treated}$.

3.3 Product Attribute Tagging with Amazon Mechanical Turk and K-Means Clustering

Given the data from the field experiment, we still need to label product attributes of interest. With more than 37,000 unique types of items, it is challenging to identify many product attributes at this scale. In this section, we describe how we identified and labelled the product attributes discussed in Section 2 (e.g., search vs. experience; utilitarian vs. hedonic) robustly and on a large scale by combining a crowdsourcing marketplace with clustering algorithms. We first describe the

[6]One prominent exception is Pandora, (a music genome project) which managed to content-code a large library of songs.

methodology for identifying product attributes using Amazon Mechanical Turk (AMT). AMT is a crowdsourcing 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 so users on AMT can help label product attributes in our dataset. The questions are 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 the literature (e.g., we obtain tags from at least five different Turkers, all of whom are from the US, have more than 500 completed tasks, and have an approval rate greater than 98%. We also include an attention-verification question to ensure AMT users are not randomly or casually answering the questions.). Please see the Appendix for the measurement questions we used and the complete list of strategies we implemented to ensure output quality. Ultimately, the labeling was quite consistent across AMT users and we achieve values greater than 0.8 for all of the constructs in Krippendorff's Alpha, an inter-rater reliability measure in which any value above 0.8 is accepted in the literature as a satisfactory outcome.[7] We utilize several thousand unique AMT workers answering many questions for more than 37,000 unique items.

Once we have obtained raw input labels (several questions per construct) from AMT, we utilize a clustering algorithm to label the final product attributes. Because each construct in our case involves dichotomous classification (i.e., search or experience, hedonic or utilitarian), we apply a k-means clustering algorithm (Hartigan, 1975) with $k = 2$. This is a more data-driven approach to classifying products than arbitrarily implementing a median-split or mean-split, as is usually done in the literature.

Hedonic-Utilitarian

Measurement Questions - Given the above definition of hedonic and utilitarian value of a product, rate the product above with the scale below on hedonic value and utilitarian value.	Utilitarian Product Cluster Mean	Hedonic Product Cluster Mean
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 1: **Utilitarian vs. Hedonic Product Cluster Means**: The definition is given in the Appendix.

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

To measure and classify an item as a hedonic or a utilitarian product, we surveyed the extant literature and found several operating definitions and measurement questions (Dhar and Wertenbroch (2000); Strahilevitz and Myers (1998); Bart et al. (2014); Khan et al. (2005); Babin et al. (1994)). We used multiple questions all aimed at identifying whether a product is utilitarian or hedonic. One measurement question provides a definition of hedonic goods and asks the workers to rate the product on a 1-to-7 Likert scale. This is repeated for a utilitarian definition to produce two separate measurements for utilitarian and hedonic qualities. Another approach condenses this into one scale, starting from purely utilitarian and moving to purely hedonic. For every product, we asked all three questions, as seen in Table 1, to at least five different Turkers. We then recorded 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 across all products are shown in Table 1. Figure 2 selects 30 items labeled with their third-level product category name. For each item, its location indicates the mean response to the three survey questions for that product. The separation between the hedonic and utilitarian products show clear clustering behavior. Hedonic products such as "fashion," "eye accessories," "puddings & gelatins," and "interactive stuffed toys" are clustered around the upper left side of the data space. Utilitarian products such as "diapers & training pants," "computer memory," "bookcases & desks," and "baby basics" are clustered around the lower-right side of the data space. The Appendix has the full list of questions used, question sources, and the inter-rater reliability measure.

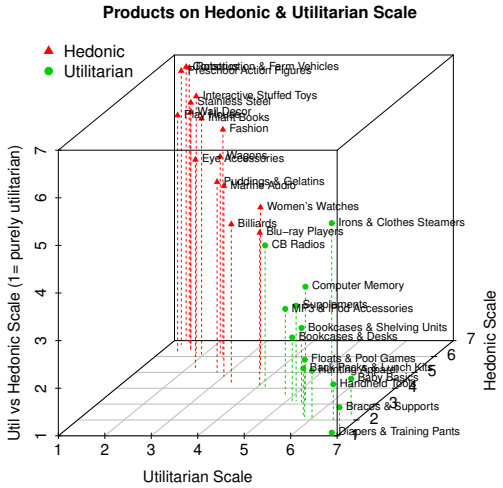

Figure 2: **Thirty Product Samples in Hedonic vs. Utilitarian Questionnaire Space:** Thirty randomly chosen items are plotted on this questionnaire space with Likert scale ranging from 1 to 7. We label each item with their third-level category name for the right balance of description, and to abstract away from the product name.

Measurement Questions [1 NOT WELL/IMPORTANT AT ALL to 7 EXTREMELY WELL/IMPORTANT]	Search Prod Cluster Mean	Experience Prod 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 the retailer and/or manufacturer about this product's attributes and features?	5.04	3.78

Table 2: **Search vs. Experience Product Cluster Means**

Search-Experience

While searching for an operational definition of search-experience, we found two sets of questions repeatedly used in the literature (Krishnan and Hartline (2001); Hsieh et al. (2005); Huang et al. (2009); Girard and Dion (2010); Klein (1998); Klein and Ford (2003)). One approach, which is used widely in marketing literature, asks the consumers to answer two questions: how well could the consumer judge the attribute or quality of the product 1) *before* they have purchased it and 2) *after* they have purchased it. If the consumers cannot judge the attributes as well before the purchase as they can after the purchase, the literature has classified those products as experience goods. For search goods, consumers can judge the quality of the product well before the purchase. Another approach asks questions related to the search cost. We combined these questions in the existing literature and asked a total of four questions on the Likert scale (all questions are listed in Table 2). Once we obtained the answers for each product from at least five different Turkers, we recorded 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 2. Figure 3 shows 30 randomly chosen items on the questionnaire Likert scale space. Items are once again labeled with their third-level product category name. The figure selects only three of the four questions for ease of visualization. Search products, for which consumers expect to judge the attributes well before purchasing item, and which do not require in-person inspection, are clustered around the bottom right side of the data space. For example, desktop computers, and printers are classified as search products while body, jewelry, and shower curtains are classified as experience products. The Appendix has the full list of questions used, question sources, and the inter-rater reliability measure.

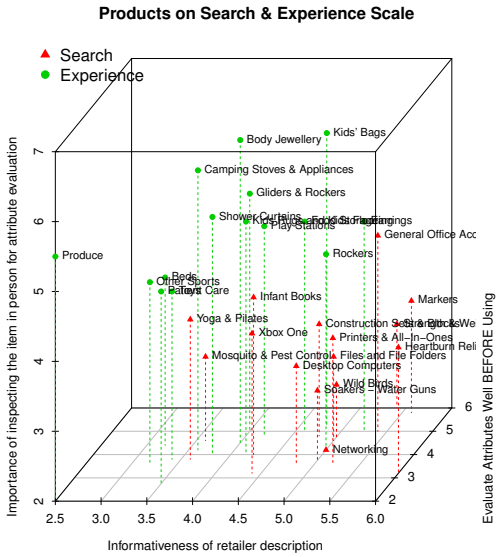

Figure 3: **Thirty Product Samples in Search vs. Experience Questionnaire Space:** Thirty randomly chosen items are plotted on this questionnaire space with Likert scale ranging from 1 to 7. Since we cannot plot four-dimensional space, we choose the three most informative questions for the three-dimensional plot. We label each item with their third-level category name for the right balance of description and to abstract away from product name. "Informativeness of Retailer Description" refers to the question "How well can you evaluate the product using only information provided by the retailer and/or manufacturer about this product's attributes and features?".

3.3.1 Control Attributes

Additionally, we include the following control attributes in the model:

1. Durability: We asked five distinct Turkers to rate on a Likert scale from 1 to 7, with 7 representing the highest product durability.
2. Description Length: The retailer provides a 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 five 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.

3.4 Data Summary

We now present a summary of our dataset along with stylized graphics. Table 3 gives the summary of the data. The dataset spans 348,489 rows of individual-item transactional records (i.e., if a user clicks and sees the product, this is a row; it also records whether the consumer ultimately purchased the product) and tracks 184,375 unique users split into 92,188 treated users and 92,187 control users. Users clicked and viewed the details of 37,215 unique items and bought 3,642 unique items

Figure 4: **Product Categories Visualized:** Starting from left to right, we draw a word cloud of category levels 1, 2, and 3, respectively. For categories 2 and 3, we limit it to 200 for readability. Some products have a 4th level category before their name, but not all of them. The size of the text is randomized.

for a total of 9,761 purchases altogether. Among users that eventually make a purchase, the total number of product views and purchases are greater for the treated group than for the control group: 234,789 (treated total product views) vs. 230,322 (control total views) and 5,068 (treated total purchases) vs. 4,693 (control total purchases). Table 3 shows that while 29,417 unique products are viewed by the treated group, the control group looks at a smaller set of 29,188 products. The difference is also shown in unique products bought at 2,082 vs. 1,982. In effect, the recommender has caused users to view and purchase more products. The products were approximately evenly split as utilitarian and hedonic goods at 18,529 vs. 18,596. There were more experience products at 21,327 than search products at 15,798.

We also collected review data for all items appearing in the dataset, a retailer's description of the item, categorization—including subcategorization—and more. There are a total of four levels of depths in the subcategories used by the retailer. The first depth has 17 subcategories, the second has 148, the third has 883, and the fourth has 492. Note that not all products have a fourth level of sub-categories but all products have a third level sub-category. At the top level, the retailer has 17 categories including house appliances, automotive, electronics, furniture, jewelry, and so on (Appendix C shows the categories table with data rows). Figure 4 shows word cloud graphics of category levels 1, 2, and 3, respectively. In the figure, we limited the number of categories shown to 200. As exhibited by the word cloud figure, the product assortment of this retailer is expansive. We carefully chose a retailer that has extensive coverage of SKUs and product categories to increase the external validity of the results.

Variable	Description	Source	Mean	SD	Min	Max
REC	Recommender system treatment condition. One means the user was randomly selected to be shown recommendations.	Treatment	0.50	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	Percentage of Amazon Mechanical Turkers who recognized the brand. Asked five Turkers per item.	AMT	0.53	0.35	0	1
DURABILITY	Durability of the item. Likert scale from 1-7 with 7 being the most durable.	AMT	4.97	1.37	1	7
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
			Counts			
UTILHEDO	Classification into utilitarian or hedonic product. One if utilitarian, 0 if hedonic.	AMT	Util	18529	Hed	18596
SEARCHEXP	Classification into search or experience product. One if search, 0 if experience.	AMT	Sea	15798	Exp	21327
			Count		**Treated**	**Control**
User ID	Unique user ID.	Site	184375		92188	92187
Unique Products Viewed	Unique products viewed by users.	Site	37125		29417	29188
Unique Products Purchased	Unique products purchased by users.	Site	3642		2082	1982
Total Views	Total number of item views.	Site	465111		234789	230322
Total Purchases	Total number of purchases.	Site	9761		5068	4693
RATINGSEXIST	The number of items with existing reviews.	Site	9631			

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

4 Empirical Strategy

As discussed in Section 2, we examine consumer product views to study the awareness impact of recommenders. Similarly, examining conversion conditional on views can shed insight on saliency impact of recommenders. Thus, the empirical specification models both views and conversion conditional on views. However, due to the data limitation of not knowing the actual recommended product (discussed in Section 3.2), we model views and conversion conditional on views separately. Following common practices in marketing and economics (Caudill, 1988; Goldfarb and Tucker, 2011), we present the conversion results with a linear probability model for several reasons. The use of a linear probability model makes the interpretation of interaction terms simple and does not require

analysis at several values, as they do in logistic regression formulation (Ai and Norton, 2003). A potential weakness of the linear probability model relative to logit model, inefficiency (Maddala, 1986), is alleviated by a large sample size in our dataset. Angrist and Pischke (2008) show that there is little difference between the limited dependent model and the linear probability model in several empirical applications. We find that our results are robust to a logistic regression specification as well.

Views For user u, item i, and session s, we model item views (binary viewed or not viewed) as shown in equation 1 with $ItemViews_{uis}$ equal to 1 if a particular item was viewed at least once for a user-item-session and 0 if otherwise. X-variables include recommender treatment indicator REC_u, item attribute vector $\overrightarrow{X_i^{view}}$, which is shown or known to the consumers before they click and view the product description page of the item (such as item price, review ratings, hedonic vs. utilitarian, etc), and recommender and item attribute interaction term ($REC_u \times \overrightarrow{X_i^{view}}$). $\overrightarrow{\theta}$ captures the coefficient of interest in our study that shows the moderating impact of item attributes on the recommender performance.

$$ItemViews_{uis} = \alpha REC_u + \overrightarrow{\beta}\overrightarrow{X_i^{view}} + \overrightarrow{\theta}(REC_u \times \overrightarrow{X_i^{view}}) + \epsilon_{uis} \tag{1}$$

Conversion Similarly, the conversion (1 if purchased or 0 if not) conditional on viewing is modeled with the following in equation 2 with mostly the same independent variables $\overrightarrow{X_i^{purchase}}$ but this vector now includes variables observable on a product description page like description length, brand prominence, etc. which are revealed when the customer views the items:

$$Conversion_{uis}|ItemViews = \alpha REC_u + \overrightarrow{\beta}\overrightarrow{X_i^{purchase}} + \overrightarrow{\theta}(REC_u \times \overrightarrow{X_i^{purchase}}) + \epsilon_{uis} \tag{2}$$

Conversion With Control for Items Recommended In the conversion specification in Equation 2, a concern may arise that the results are driven not by consumers purchasing (conditional on viewing) due to the recommendation signal (e.g. because of recommendation salience), but by the types of products recommenders are systematically recommending. That is, the observed increase in conversion rate may be because of differences in the types of products consumers become aware of due to the recommender. If we had data on what was actually recommended, we could jointly model views along with purchase or control for it. Since this data is missing, we simulate this data at

the user-item-session level with the same recommender algorithm deployed in the field experiment. We replicated the recommender system to the exact specification used in the field experiment as described in Section 3.1. To calculate the item-item similarity matrix, we obtained the purchase dataset from the same retailer (beyond the data for the two-week period in which the experiment was conducted). First, we simulate six recommendations as done in the field experiment using user session history data while excluding the current focal item. Second, we average the recommended items' product attributes denoted by $\overrightarrow{R_{uis}^{avg}}$ and control for it in this augmented specification. $\overrightarrow{R_{uis}^{avg}}$ contains the average of the following recommended item attributes: review rating, review number, brand prominence, price, utilitarian/hedonic, and search/experience.

$$Conversion_{uis}|ItemViews = \alpha REC_u + \overrightarrow{\beta}\,\overrightarrow{X_i^{purchase}} + \overrightarrow{\theta}\,(REC_u \times \overrightarrow{X_i^{purchase}}) + \overrightarrow{\gamma}\,\overrightarrow{R_{uis}^{avg}} + \epsilon_{uis} \quad (3)$$

5 Results

Results for views are presented first, followed by results for conversion conditional on views. Next, we collect secondary results and discuss alternative explanations as well as robustness.

5.1 View Results - Moderating Effect on Awareness

Table 4 provides results from the regression of equation 1 and Figure 5 graphically presents the coefficients for easier consumption. We discuss the results of having main effects only (V1), main effects + controls (V2), and then the main effects + controls + interactions (V3), respectively. The baseline result (V1) confirms the existing result that a recommender treatment (0.107) increases product views – consumers view more product under the influence of recommenders.

Next, V2 presents the linear regression results with control variables. The effect size of recommender treatment (0.118) stays stable and statistically significant. Review numbers and review ratings, both of which signal product quality, contribute to a greater number of views. Lastly, people tend to view more utilitarian products and experiential products, which could be explained by extant theories on product search and learning behavior influenced by factors such as 1) difference in perceived risk (Pavlou, 2003), 2) heterogenous taste difference, 3) hedonic browsing behavior, etc. This is not the focus of this study, however.

Finally, investigating the main specification in interaction terms, V3, we discover the main results

on how product attributes moderate recommenders' impact on consumer view behavior through awareness. Again, the main effect of the recommender stays positive and significant (0.047), not accounting for the interaction effect. Shifting our attention to interaction coefficients, $\overrightarrow{\theta_{views}}$, we note the following: The interaction term, REC × price is positive and significant and much greater than the main effect of price (-0.000002), showing net positive effect. This suggests that with recommendations, consumers are even more likely to click and view products with higher prices. This could be explained by the classic economic theory of price as a signal of quality (Wolinsky, 1983), which leads consumers to view the product more and search for more info about the product or product categories in general, with the effect now further bolstered by the recommendation. The same results hold for REC × review numbers (0.00007) and REC × review ratings (0.048), which are both positive and significant. Recommenders bolster the positive impact of higher ratings (0.0035) and review volumes (0.000014), enticing consumers to view more products.

	Main Effect Only (V1)	**Main + Controls (V2)**	**Interactions (V3)**
Constant	0.026740*** (0.000079)	-0.017626*** (0.000312)	-0.003229*** (0.000309)
Recommender Treatment	0.107511*** (0.000193)	0.118401*** (0.000194)	0.047947*** (0.000344)
Price		0.000050*** (0.000001)	-0.000002** (0.000001)
Review Numbers		0.000022*** (0.000001)	0.000014*** (0.000001)
Review Ratings		0.010030*** (0.000035)	0.003581*** (0.000036)
Utilitarian/Hedonic		0.004833*** (0.000153)	0.000855*** (0.000162)
Search/Experience		-0.010677*** (0.000149)	-0.006829*** (0.000159)
Rec X Price			0.000547*** (0.000002)
Rec X Review Numbers			0.000071*** (0.000003)
Rec X Review Ratings			0.048532*** (0.000100)
Rec X Utilitarian			0.014804*** (0.000408)
Rec X Search			-0.044335*** (0.000404)
R-squared	0.0378	0.0536	0.095
Log-likelihood	1392711.517	1458580.278	1635617.116
Deviance	326161.8293	320780.7651	306753.8574
AIC	-2785417.034	-2917138.556	-3271200.232
N	7918940	7918940	7918940

Table 4: **Linear Probability Model Results on Views : '*'= p-value <0.05, '**'= p-value <0.01, '***'= p-value <0.001**

For product attributes, V3 main specification supports hypotheses 1 and 3. While the use of recommender increases the view rate for all types of products, examining the interaction terms show that the awareness impact increases more for utilitarian products compared to hedonic products (0.014), as well as for experience products compared to search products (-0.044). As discussed in Section 2, our empirical result supports the notion that due to the increased perceived risk (functional and financial) associated with utilitarian products, recommenders' awareness impact

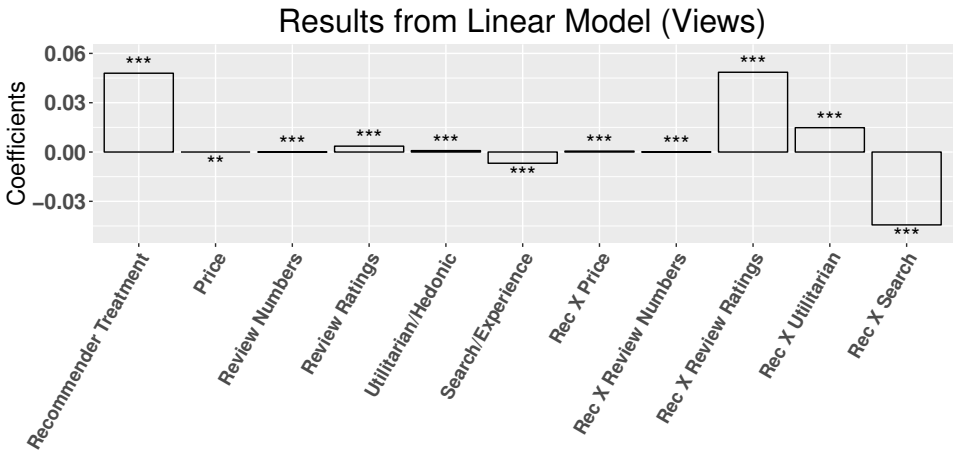

Figure 5: **View Results Linear Coefficients Compared: '*'= p-value <0.05, '**'= p-value <0.01, '***'= p-value <0.001**

is increased for utilitarian products as compared to for hedonic products. Similarly, given that recommenders are trusted (as evidenced by the fact that recommender treatment effect is positive and significant in V1), our results also support the hypothesis that recommenders have a greater impact on products views for experience products as compared to search products. The main line of reasoning is that, aided by trustworthy recommenders, consumers take part in increased exploration and search, due to the increased product uncertainty of experience products online.

Lastly, the interaction term with recommender treatment and average review ratings (0.0485) suggests that the positive impact on views from a recommender system will be strengthened for products with high review ratings. This supports our hypothesis 5 that recommenders act as complements to review ratings in increasing views.

An alternative explanation for our results may be that recommenders may work qualitatively better for specific types of products (e.g. utilitarian and hedonic) than for others. We discuss and alleviate this concern in Section 5.4.

5.2 Conversion Results - Moderating Effect on Saliency

Table 5 provides results from the regression of equations 2 and 3 and Figure 6 graphically presents the coefficients for easier consumption. Similar to the analyses on views, we present 4 specifications: C1-C4.

First we confirm that the impact of recommenders on conversion conditional on views is positive and significant in all specifications. The result corroborates the existing literature in that using a

recommender system indeed increases conversion rates and thus sales (Hosanagar et al., 2014; Lee and Hosanagar, 2014). In this experiment, the use of recommenders increased the conversion rate by approximately 6%. The impact of price on conversion is also stably negative and significant across specifications. We note that the results are highly stable not only in directions and significance, but also in magnitude across all four specifications. Thus we discuss the results with the C4 specification going forward.

	Main Effect Only (C1)	Main + Controls (C2)	Interactions (C3)	Interactions & Rec Controls (C4)
Constant	0.018318*** (0.000327)	0.035362*** (0.001080)	0.033834*** (0.001183)	0.037036*** (0.005230)
Recommender Treatment	0.001101* (0.000461)	0.000927* (0.000460)	0.004005*** (0.001067)	0.003888*** (0.001067)
Price		-0.000024*** (0.000002)	-0.000017*** (0.000003)	-0.000016*** (0.000003)
Description Length		0.000001 (0.000001)	-0.000001 (0.000001)	-0.000001 (0.000001)
Brand Prominence		0.001830** (0.000681)	0.001800** (0.000681)	0.000689 (0.000685)
Durability		-0.004635*** (0.000180)	-0.004630*** (0.000180)	-0.003861*** (0.000193)
Review Numbers		0.000002 (0.000002)	-0.000002 (0.000003)	-0.000002 (0.000003)
Review Ratings		0.001515*** (0.000109)	0.001868*** (0.000154)	0.001777*** (0.000157)
Utilitarian/Hedonic		0.003030*** (0.000485)	0.004381*** (0.000679)	0.005011*** (0.000759)
Search/Experience		0.003086*** (0.000480)	0.003056*** (0.000677)	0.003575*** (0.000678)
Rec X Price			-0.000013** (0.000004)	-0.000013** (0.000004)
Rec X Description Length			0.000005* (0.000002)	0.000005* (0.000002)
Rec X Review Numbers			0.000012** (0.000004)	0.000012** (0.000004)
Rec X Review Ratings			-0.000715*** (0.000215)	-0.000705** (0.000215)
Rec X Utilitarian			-0.002706** (0.000947)	-0.002691** (0.000947)
Rec X Search			0.000079 (0.000944)	0.000234 (0.000944)
Recommended Item Controls	No	No	No	Yes
R-squared	0.00002	0.00356	0.00368	0.00439
Log-likelihood	200593.8433	201212.3372	201233.0306	201357.4018
Deviance	6452.76733	6429.90336	6429.13978	6424.55247
AIC	-401181.6866	-402402.6744	-402432.0612	-402668.8036
N	348489	348489	348489	348489

Table 5: **Linear Probability Model Results on Purchase : '*'= p-value <0.05, '**'= p-value <0.01, '***'= p-value <0.001**

The length of product descriptions provided by the retailer had no significant influence on base conversion rate. Keeping everything else the same, conversion rate was greater for products with higher average ratings (0.0017), as indicated previously by Chevalier and Mayzlin (2006) and Sun (2012). Higher durability was associated with lower conversion rate (-0.0038). This result may be because high durability is correlated with lower purchase frequency, and thus higher perceived risk

(Jacoby et al., 1971; Pavlou, 2003). Consumers may be less willing to buy the product right after viewing it. Next, we discuss our main hypotheses and results regarding the interaction between product attributes (and review ratings) and recommender systems.

The interaction term of recommenders and utilitarian variables is negative and statistically significant (-0.0026), suggesting that the positive effect of recommenders on conversion rate is greater for hedonic products compared than for utilitarian products, thus supporting hypothesis 2. As discussed in Section 2, consumers may indeed trust recommenders to signal "People who purchased X, also purchased Y" to be reflective of opinions of people who are more like them. In this case, the effect should be greater for hedonic products than for utilitarian products (Feick and Higie, 1992). Additionally, utilitarian products serve specific purposes and are more prone to functional risks; consumers are, in general, less trustful of accepting recommenders' suggestions or are less willing to take leap of faith with utilitarian products compared to hedonic products.

With respect to search-experience attributes, the results were not statistically significant. Hence, the results do not support hypothesis 4, which stated that the the positive impact of recommenders on conversion rate will be greater for experience products (vs. search products). Furthermore, 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—does not 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 to be effective for experience products. Since our dataset spans extensive categories, we sought to replicate the results of Senecal and Nantel (2004) (that recommendations for experience products like wine were more influential than recommendations for search products like calculators) or 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 the dataset is considered in its entirety, search-experience attributes do not seem to moderate the saliency effectiveness of this widely used recommender.

The interaction term with recommender treatment and average review ratings (-0.0007) suggests that the positive impact on conversion from a recommender system will be lower for products with high review ratings. This supports our hypothesis 6 that recommenders act as substitutes to review ratings in increasing conversion conditional on views. As discussed in Section 2, the positive impact of recommenders on conversion rate is greater for products with lower review ratings. When

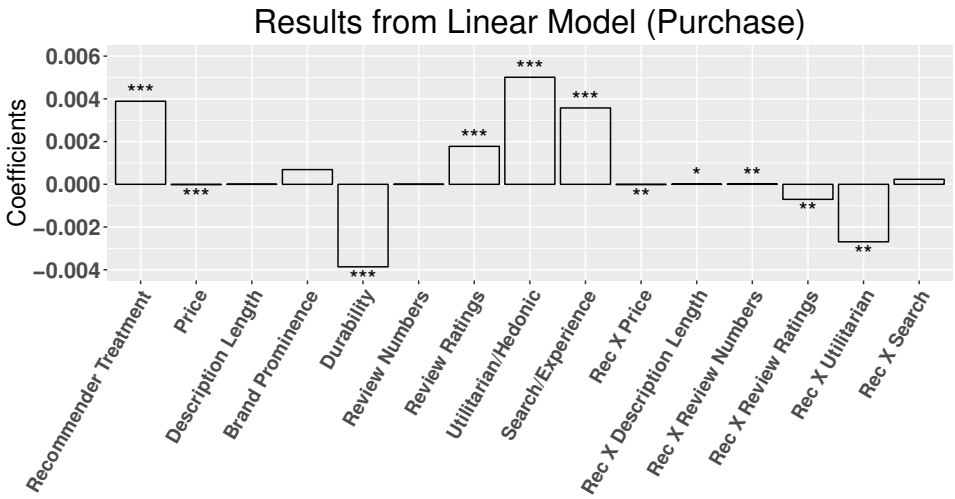

Figure 6: **Purchase Results Linear Coefficients Compared: '*'= p-value <0.05, '**'= p-value <0.01, '***'= p-value <0.001**

recommenders are present, the total impact of higher review volume is positive and significant (-0.000002 main + 0.000012 interaction= 0.00001), suggesting that the higher the review number, the higher the conversion under the recommenders.

We summarize our findings and hypotheses supported in Table 6.

Attribute Construct	**Hypotheses**	**Supported**
H1: Hedo-Util × Rec on Awareness (View)	The positive effect of recommenders to increase product views (awareness impact) will be **higher for utilitarian products**, compared to hedonic products.	Yes
H2: Hedo-Util × Rec on Saliency (Conv)	The positive effect of recommenders to increase product conversion conditional on views (saliency impact) will be **higher for hedonic products**, compared to utilitarian products.	Yes
H3: Sea-Exp × Rec on Awareness (View)	The positive effect of recommenders to increase product views (awareness impact) will be **higher for experience products**, compared to search products.	Yes
H4: Sea-Exp × Rec on Saliency (Conv)	The positive effect of recommenders to increase product conversion conditional on views (saliency impact) will be **higher for experience products**, compared to search products.	No
H5: Avg Review Rating × Rec (View)	The increase in views under the use of a recommender will be **greater for a product with higher average ratings** than for a product with lower average ratings.	Yes
H6: Avg Review Rating × Rec (Conv)	The increase in conversion rate conditional on views under the use of a recommender will be **greater for a product with lower average ratings** than for a product with higher average ratings.	Yes

Table 6: **Hypotheses and Results**

5.3 Other Results

This section collects secondary, yet non-trivial, results from our empirical investigation for documentation purposes. For brevity, we simply state the results supported by our data and the main empirical specifications. While these results are not novel enough to be hypotheses on their own, we believe them to be informative and have valuable managerial implications.

Attribute Construct	**Result Takeaways**
Recommender	Recommender increases both views and conversion rate.
Price × Rec (View)	Higher price increased the positive impact of recommenders on views.
Price × Rec (Conv)	Higher price decreased the positive impact of recommenders on conversion.
Review Volume × Rec (View, Conv)	The higher the review volume, the higher the view and conversion impact of recommender.
Desc Length × Rec	Longer description increased the positive conversion impact of recommender.

Table 7: **Other Takeaways Supported by Main Empirical Specification**

5.4 Alternative Explanation & Measurement Robustness

Alternative Explanation Due to Better Recommendation for Particular Product Types

One alternative explanation may be that in combination with historical purchase data and the nature of collaborative filters, recommenders may work qualitatively better for specific type of product over the other and makes unbalanced recommendation. For example, it is possible that recommender provides qualitatively better recommendation for utilitarian products over hedonic products, which may be driving the results for awareness part (views). This concern is reduced for conversion conditional on views due to 1) the fact that it is conditional on views and 2) the fact that we control for recommendation via simulation, as specified by Equation 3.

To this end, for products with a particular attribute (e.g., hedonic or utilitarian), we simulated user-item-session level recommendations and examined the recommender quality within each item attribute. That is, we have two sets of recommended items, when a utilitarian product is being viewed vs. when a hedonic product is being viewed. Given two sets of recommended products for each attribute, we can compare the recommendation quality. We define the "quality" by examining the recommended product's average star rating and number of reviews——essentially, proxies for product quality. Furthermore, before simulating the recommendations, we match viewed products

within each attribute to share similar distribution in other observables using a caliper approach (of 0.25 standard deviation (Lunt, 2013)) to compare recommendation for similar quality products. We then run Kolmogorov-Smirnov (KS) test to see if our recommender recommends products with different quality. We run this for hedonic vs. utilitarian and search vs. experience. For both, we could not reject the null (p-values between 0.09-0.80), which suggests that recommenders do not perform differently for one type of product over the other when considering product star ratings or the number of reviews. This is likely because the retailer we worked with is one of the largest and may have a critical mass of high quality products in a wide assortment for all types of products.

Alternative Explanation on Awareness Impact Due to Recommender Algorithm

For our results on the awareness impact of recommenders influencing product views, one concern may be that the effects observed are not the result of consumers reacting but driven by the recommender algorithm. This concern is alleviated for the results on conversion because empirical specification is both conditional on views and controls for simulated recommended items, showing that the results hold above and beyond the impact of recommender's awareness impact and driven by consumers. While the results documented does not change in either case, the theoretical mechanism explored and hypotheses proposed are under the assumption that the results are not purely driven by the recommender algorithms. We alleviate this concern by running a robustness test with modified views specification (modified from Eq 1) with session-historical simulated recommended items.

For each row in data, we simulate recommended items based on historically viewed products for that user-session. Then we obtain the average characteristics of the recommended items, which are then used as controls in a robustness test with a modified specification for views in equation 1. The results, shown in appendix D, persist in directions and the magnitudes are mostly similar, suggesting that the impact we recover goes beyond the simple impact of the recommenders. i.e., the consumers are also driving the awareness results.

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 into a search or experience product. The decision to use dichotomous classifications was for practical convenience and so that we could use existing measurement strategies. While the literature has acknowledged the shortcomings of dichotomous classification schemes, they are still commonly used, based on domi-

nant attributes (e.g., Huang et al. (2009), Senecal and Nantel (2004)). However, since these product attributes could be continuous, we repeated the analyses with a model in which the search-experience and hedonic-utilitarian attributes are denoted by a scale from 1 to 7. For the hedonic-utilitarian attribute, we used the answer from one of the questions ("how much comparative utilitarian vs. hedonic value the product offers. One purely utilitarian to 7 purely hedonic") as done in Bart et al. (2014). We invert the scale to match our dichotomous classification. For search-experience attributes, we take the average of survey instrument questions after appropriate scale inversion, as done in Hong and Pavlou (2014). In addition, for robustness we varied the questions used to obtain the classification for search and experience by doing the following[8]:

1. Using the difference between Q1, Q2 (Q2-Q1), Q3, and Q4: In total, three values. Refer to appendix for questions.
2. Using the difference between Q1, Q2 (Q2-Q1), Q3: In total, two values
3. Using the difference between Q1, Q2 (Q2-Q1), Q4: In total, two values

We obtain qualitatively similar results for all of our hypotheses and variables.

6 Conclusion and Discussion

While recommenders are prevalent in e-commerce and have been shown to increase sales volume in multiple studies, their effective use and implementation continues to elude many managers and retailers (Econsultancy and Monetate (2013)). This is partly due to the lack of knowledge regarding the factors that moderate the impact of recommenders. This study addresses this gap in the literature.

We first show that the recommenders increase both views and conversion for all products. But this positive impact is indeed moderated by product attributes, as well the other sources of product information, such as reviews and product descriptions. We find that a recommender's positive impact on product views is greater for utilitarian products compared to hedonic products. In contrast, a recommender's positive impact on conversion rate is greater for hedonic product compared to utilitarian product. This shows that the mechanism by which recommenders aid product purchase is different for different types of products. Our results support the thesis that when consumers are shopping for utilitarian products, the increased functional risk combined with the decreased search cost (due to recommender's awareness impact), make consumers view more products. But when

[8]We thank the anonymous reviewer for suggesting this.

they finally make the decision to convert, recommender's impact is greater for hedonic product due to the nature of personalized signal it carries. This suggest that managers might want to provide expert signal of quality or information geared towards reducing functional risk to maximize the recommender's conversion benefit for utilitarian product.

We also show that the positive impact of recommenders on conversion is increased for products with lower average ratings. Products with high ratings are more likely to be purchased upon being viewed and thus these products may not enjoy the same level of benefit in conversion rates. In contrast, when a product has mixed reviews, a recommendation might be more effective in convincing consumers to purchase. This suggests that recommenders act as substitutes for review ratings and offer ideas to retailers selling highly heterogeneous products such as niche products, which may attract mixed reviews. Furthermore, more detailed product descriptions increase recommender effectiveness; thus, sites that implement recommender algorithms should simultaneously invest in providing lengthier and more detailed descriptions.

We have examined the view and conversion aspect of the purchase funnel. Future studies with novel data on return rates and after-purchase customer satisfaction may offer more insight on the impact of recommender systems in all stages of the consumer purchase funnel.

References

Adomavicius, G., J. Bockstedt, S. Curley, and J. Zhang: 2011, 'Recommender systems, consumer preferences, and anchoring effects'. In: *RecSys 2011 Workshop on Human Decision Making in Recommender Systems*. pp. 35–42.

Adomavicius, G., J. C. Bockstedt, S. P. Curley, and J. Zhang: 2014, 'Do recommender systems manipulate consumer preferences? A study of anchoring effects'. *Information Systems Research*.

Adomavicius, G. and A. Tuzhilin: 2005, 'Towards the next generation of recommender systems: A survey of the state-of-the-art and possible extensions'. *IEEE Transactions on Knowledge and Data Engineering* **17**.

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).

Ai, C. and E. C. Norton: 2003, 'Interaction terms in logit and probit models'. *Economics letters* **80**(1), 123–129.

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

Angrist, J. D. and J.-S. Pischke: 2008, *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Animesh, A., V. Ramachandran, and S. Viswanathan: 2010, 'Quality Uncertainty and the Performance of Online Sponsored Search Markets: An Empirical Investigation'. *Information Systems Research* **21**(1), 190–201.

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.

Bagwell, K.: 2007, 'The economic analysis of advertising'. *Handbook of Iindustrial Organization* **3**, 1701–1844.

Bart, Y., A. T. Stephen, and M. Sarvary: 2014, 'Which products are best suited to mobile adver-

tising? A field study of mobile display advertising effects on consumer attitudes and intentions'. *Journal of Marketing Research* **51**(3), 270–285.

Benbasat, I. and P. Todd: 1996, 'The effects of decision support and task contingencies on model formulation: A cognitive perspective'. *Decision support systems* **17**(4), 241–252.

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

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).

Butters, G.: 1977, 'Equilibrium Distributions of Sales and Advertising Prices'. *Review of Economic Studies* **44**(3), 465–91.

Caudill, S. B.: 1988, 'Practitioners corner: An advantage of the linear probability model over probit or logit'. *Oxford Bulletin of Economics and Statistics* **50**(4), 425–427.

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.

Chiu, C.-M., E. T. Wang, Y.-H. Fang, and H.-Y. Huang: 2014, 'Understanding customers' repeat purchase intentions in B2C e-commerce: the roles of utilitarian value, hedonic value and perceived risk'. *Information Systems Journal* **24**(1), 85–114.

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.

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: 2008, '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.

Gigerenzer, G. and R. Selten: 2002, *Bounded rationality: The adaptive toolbox*. MIT press.

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.

Goldfarb, A. and C. Tucker: 2011, 'Online display advertising: Targeting and obtrusiveness'. *Marketing Science* **30**(3), 389–404.

Greevy, R., J. H. Silber, A. Cnaan, and P. R. Rosenbaum: 2004, 'Randomization inference with imperfect compliance in the ACE-inhibitor after anthracycline randomized trial'. *Journal of the American Statistical Association* **99**(465), 7–15.

Gu, B., J. Park, and P. Konana: 2012, 'Research note: the impact of external word-of-mouth sources on retailer sales of high-involvement products'. *Information Systems Research* **23**(1), 182–196.

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'.

Häubl, G. and K. B. Murray: 2003, 'Preference construction and persistence in digital marketplaces: The role of electronic recommendation agents'. *Journal of Consumer Psychology* **13**(1–2), 75–91.

Häubl, G. and K. B. Murray: 2006, 'Double agents: assessing the role of electronic product recommendation systems'. *Sloan Management Review* **47**(3), 8–12.

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

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.

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, C., R. D. Galliers, N. Shin, J.-H. Ryoo, and J. Kim: 2012, 'Factors influencing Internet shopping value and customer repurchase intention'. *Electronic Commerce Research and Applications* **11**(4), 374–387.

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.

Komiak, S. Y. and I. Benbasat: 2006, 'The effects of personalization and familiarity on trust and adoption of recommendation agents'. *MIS quarterly* pp. 941–960.

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

Kumar, A. and K. Hosanagar: 2018, 'Measuring the Value of Recommendation Links on Product Demand'. *Accepted at Information Systems Research.*

Kumar, A. and Y. Tan: 2015, 'The Demand Effects of Joint Product Advertising in Online Videos'. *Management Science* **61**(8), 1921–1937.

Lee, D. and K. Hosanagar: 2014, 'Impact of recommender systems on sales volume and diversity'. *Proceedings of the 2014 International Conference on Information Systems.*

Lee, D. and K. Hosanagar: 2018, 'How Do Recommender Systems Affect Sales Diversity? A Cross-Category Investigation via Randomized Field Experiment'. *Information Systems Research* **Forthcoming**.

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.

Lunt, M.: 2013, 'Selecting an appropriate caliper can be essential for achieving good balance with propensity score matching'. *American journal of epidemiology* **179**(2), 226–235.

Maddala, G. S.: 1986, *Limited-dependent and qualitative variables in econometrics*, No. 3. Cambridge University Press.

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

Monetate: 2013, 'Maximize online sales with product recommendations'. Technical report, Monetate.

Moore, R. and G. Punj: 2001, 'An Investigation of Agent Assisted Consumer Information Search: Are Consumers Better Off?'. *NA-Advances in Consumer Research Volume 28.*

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.

Sarkar, A.: 2011, 'Impact of utilitarian and hedonic shopping values on individual's perceived benefits and risks in online shopping'. *International management review* **7**(1), 58–65.

Sarwar, B., G. Karypis, J. Konstan, and J. Riedl: 2001, 'Item-based collaborative filtering recommendation algorithms'. In: *Proceedings of the 10th international conference on World Wide Web*. pp. 285–295.

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.*

Spiekermann, S. and C. Paraschiv: 2002, 'Motivating human–agent interaction: Transferring insights from behavioral marketing to interface design'. *Electronic Commerce Research* **2**(3), 255–285.

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.

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.

Swaminathan, V.: 2003, 'The impact of recommendation agents on consumer evaluation and choice: the moderating role of category risk, product complexity, and consumer knowledge'. *Journal of Consumer Psychology* **13**(1-2), 93–101.

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.

Weitzman, M. L.: 1979, 'Optimal search for the best alternative'. *Econometrica: Journal of the Econometric Society* pp. 641–654.

Wolinsky, A.: 1983, 'Prices as signals of product quality'. *The review of economic studies* **50**(4), 647–658.

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 who are potentially not proficient in English.
5. We refined our survey instrument through an iterative series of 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 to list the reasons they found them confusing. We refined the survey in this manner until nearly all queried Turkers stated that all questions were clear.
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 were dropped from the data.
7. On average, we found that the survey took a little over four minutes and it typically took at least one minute or more to completely read the questions. We considered less than 30 seconds to be too short, and discarded any message tags with completion times shorter than this 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 were "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 improve the quality of the data generated (Mason and Suri (2012); Ipeirotis et al. (2010); Paolacci et al. (2010)). Snow et al. (2008) show 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 the type of repeated labeling we implement, wherein each message is tagged by multiple Turkers, is preferable to single labeling, in which one person tags one sentence. Finally, in 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, which are commonly used inter-rater reliability measures.

The following table provides the construct we have used, the literature sources we've adapted, the measurement survey instrument and operating definitions, and the 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 motivations 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 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 8: **Survey Instrument:** We use existing and commonly used operational definitions and measurement questions to tag the items in our dataset. The Median Krippendorff's Alpha, a standard measure of inter-rater reliability, is provided and is well above the acceptable measure of 0.8.

Appendix B: Kolmogorov-Smirnov Test for Randomization

Description	**Treated VS Control p-value**
Number of Item Views	0.89
Number of Item Purchases	0.66
Wallet Size	0.37

Table 9: **User-Level Randomization Check using Kolmogorov-Smirnov Test. P-values reported.** We could not reject the null. This suggests that our user-level randomization holds true.

Appendix C: Product Categories Occurring in The Dataset

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	Health & Beauty	Holiday Gift Center	Home & Pets	Jewelry & Watches	Movies Music & Books
8422	28719	7621	50859	4015	26000
Office & Stationery	Outdoor Living	Sports & Rec	Toys	Video Games	
12352	6297	27681	20657	12032	

Table 10: **Product Categories Occurring In the Dataset:** The first level product categorization as classified by the retailer online. There are in total four levels of depths in subcategories. First depth has 18 categories, 2nd -> 149, 3rd ->884, and 4th -> 492.

Appendix D: Robustness Test for Awareness Impact: Showing the Influence Beyond Recommender Systems

We provide a robustness test for awareness results to show that the impact estimated is driven by both recommender system and consumers reacting to the recommendations. Please see Section 5.4 (Alternative Explanation on Awareness Impact Due to Recommender Algorithm) for more details.

	Interactions (V3)	**Robustness Test**
Constant	-0.003229*** (0.000309)	-0.004985*** (0.000326)
Recommender Treatment	0.047947*** (0.000344)	0.067215*** (0.000391)
Price	-0.000002** (0.000001)	-0.000005*** (0.000001)
Review Numbers	0.000014*** (0.000001)	0.000014*** (0.000001)
Review Ratings	0.003581*** (0.000036)	0.003429*** (0.000038)
Utilitarian/Hedonic	0.000855*** (0.000162)	0.000433** (0.000167)
Search/Experience	-0.006829*** (0.000159)	-0.007336*** (0.000163)
Rec X Price	0.000547*** (0.000002)	0.000580*** (0.000002)
Rec X Review Numbers	0.000071*** (0.000003)	0.000059*** (0.000003)
Rec X Review Ratings	0.048532*** (0.000100)	0.049812*** (0.000104)
Rec X Utilitarian	0.014804*** (0.000408)	0.005563*** (0.000433)
Rec X Search	-0.044335*** (0.000404)	-0.055437*** (0.000425)
Session-Historical Controls	No	Yes
R-squared	0.095	0.103

Table 11: **Linear Probability Model Results on Views with Robustness Specification: '*'= p-value <0.05, '**'= p-value <0.01, '***'= p-value <0.001**