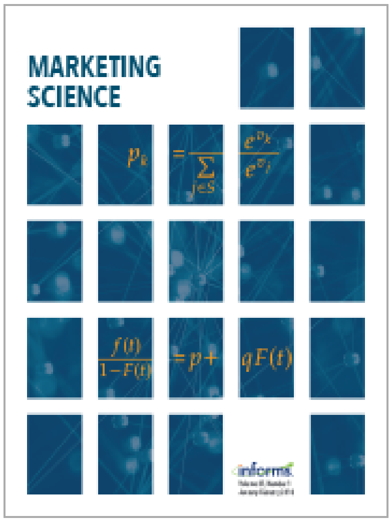


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Modeling Consumer Learning from Online Product Reviews

Yi Zhao, Sha Yang, Vishal Narayan, Ying Zhao,

MARKETING SCIENCE
Vol. 32, No. 1, January–February 2013, pp. 153–169
ISSN 0732-2399 (print) | ISSN 1526-548X (online)

http://dx.doi.org/10.1287/mksc.1120.0755

Modeling Consumer Learning from Online Product Reviews

Yi Zhao
J. Mack Robinson College of Business, Georgia State University, Atlanta, Georgia 30303,
yizhao@gsu.edu

Sha Yang
Marshall School of Business, University of Southern California, Los Angeles, California 90089,
shayang@marshall.usc.edu

Vishal Narayan
Samuel Curtis Johnson Graduate School of Management, Cornell University, Ithaca, New York 14853,
vn28@cornell.edu

Ying Zhao
Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong,
mkyzhao@ust.hk

We propose a structural model to study the effect of online product reviews on consumer purchases of experiential products. Such purchases are characterized by limited repeat purchase behavior of the same product item (such as a book title) but significant past usage experience with other products of the same type (such as books of the same genre). To cope with the uncertainty in quality of the product item, we posit that consumers may learn from their experience with the same type of product and others' experiences with the product item. We model the review credibility as the precision with which product reviews reflect the consumer's own product evaluation. The higher the precision, the more credible the information obtained from product reviews for the consumer, and the larger the effect of reviews on the consumer's choice probabilities. We extend the Bayesian learning framework to model consumer learning on both product quality and review credibility. We apply the model to a panel data set of 1,919 book purchases by 243 consumers. We find that consumers learn more from online reviews of book titles than from their own experience with other books of the same genre. In the counterfactual analysis, we illustrate the profit impact of product reviews and how it varies with the number of reviews. We also study the phenomenon of fake reviews. We find that fake reviews increase consumer uncertainty. The effects of more positive reviews and more numerous reviews on consumer choice are smaller on online retailing platforms that have fake product reviews.

Key words: learning models; choice models; product reviews
History: Received: May 6, 2011; accepted: September 18, 2012; Peter Fader served as the user-generated content special issue editor and then Preyas Desai served as the editor-in-chief and Scott Neslin served as associate editor for this article. Published online in *Articles in Advance* December 17, 2012.

1. Introduction

The Internet has provided an opportunity for consumers to share product evaluations online, facilitating a new channel for the communication of product information and word of mouth. Online consumer reviews are now widely available on websites of manufacturers, retailers, and market makers, as well as for product categories including both search goods and experiential goods. Before deciding to purchase a specific item, consumers now have online access to numerous product reviews posted by several users.

Recent research has provided considerable evidence that online product reviews have an impact on product sales. Chevalier and Mayzlin (2006) find a positive relationship between consumer book ratings and book sales. Liu (2006) finds a positive relationship between the volume of reviews of a movie and its box office revenue. Whereas prior research has addressed the link between consumer product reviews and product sales based on aggregate-level analysis, little work has examined the underpinning of such impact in affecting individual consumer purchase decisions. Such microlevel analysis can help marketers gain a deeper understanding of the impact of user-generated product reviews and more effectively respond to this new market phenomenon.

In this paper, we propose a structural model to study how online reviews affect individual consumer purchases of experiential products. Our basic premise is that consumers are uncertain about the true product quality or the extent to which a product matches their preference or usage condition. This gives rise to consumers' need for reading product reviews and learning from other consumers' usage experiences to reduce such uncertainty. This is especially true for experiential products such as books, movies, and music because, unlike other products, these are consumed solely for the pleasure and experience they provide. Consumers have been known to rely on recommendations for experiential products significantly more than other types of products (Sénécal and Nantel 2004). For this reason, online reviews are very important when consumers are choosing products they do not have first-hand experience with. By reading the post-consumption evaluations of a product by others, consumers can make a more informed decision about which product(s) to purchase.

The proposed model is built on the framework of consumer learning of product quality based on past usage experience (Erdem 1998, Narayanan et al. 2005, Zhang 2010). The basic idea is that consumers are imperfectly informed and therefore uncertain about the true quality of a product. This uncertainty is likely to be greater for experiential goods than for search goods. This is because repeated purchase of the same product item (e.g., the same movie, CD title, or book title) is quite uncommon for experiential goods. This is unlike the purchase behavior of other types of products such as groceries (where repeated purchase of the same product item is the norm) or physician adoption of new drugs (where physicians learn from both sampling and repeated prescription). So even though the consumer might have consumed similar products (say, products belonging to the same genre) in the past, there is considerable uncertainty about the quality of the specific product item (henceforth referred to as "item"). This uncertainty is associated with the likelihood that the specific item (say, the movie *Couples Retreat*) matches a consumer's preference. To cope with such uncertainty, consumers may learn based on their past usage experience with this type of product (i.e., experience with comedy movies they have viewed in the past) and other consumers' usage experiences with the focal item through consumer reviews. They then update their belief of the subjective product quality (likelihood of matching) in a Bayesian fashion. When determining whether to purchase a book title, the consumer may integrate multiple sources of information. The consumer might learn from (a) his own past experience with other titles belonging to the same genre of books as the title under consideration and (b) information about other consumers' postconsumption evaluations of that title. A novel feature of this research is that our model captures both types of learning: from one's own experience with the category and from information about others' experiences with the specific product.

Although the effect of product reviews on aggregate sales is well studied, it remains unclear if all reviews of a product have the same effect on the choice of a consumer. Conversely, there is little understanding if all consumers are affected by a set of reviews to the same extent. In addition to modeling consumer learning from one's own experience and information about others' experiences, we also model the credibility of information obtained from others. The credibility of the source of communication is directly related to the persuasiveness of communication (Sternthal et al. 1978). As such, more credible reviews for a product are likely to have a greater effect on the consumer's propensity to buy that product. It is well established that the credibility of a source of communication is greater if he or she is more similar to the recipient (Brock 1965). Accordingly, we model the review credibility as the precision with which product reviews reflect the consumer's own product evaluation. The higher the precision, the more credible the information obtained from product reviews for the consumer, and the larger the effect of reviews on the consumer's choice probabilities. We extend the Bayesian learning framework to model consumer learning on both product quality and review credibility. Our model is general enough to allow for the possibility that some consumers might find the same set of reviews to be more credible than other consumers. Finally, we allow the credibility of all reviews of a product to vary with time.

We apply the proposed model to a panel data set comprising 1,919 book purchases by 243 consumers at a large product reviews website. Our empirical analysis leads to three unique findings. First, we find evidence of stronger learning from product reviews than learning from own experience. Consumers learn less from their own experience with books similar to the one they are considering for purchase (i.e., books of the same genre) than they do from others' experiences with the focal product. Second, we find evidence that, based on the similarity between reviews for a product and their own evaluation of that product, consumers update their beliefs about the credibility of product reviews over time. Third, ignoring the learning process from others lowers the model performance and leads to biases in model estimates. Such insights cannot be obtained from reduced-form aggregate models.

Next we discuss the managerial implications of this research. Based on the parameter estimates of our structural model, and a set of general assumptions, we estimate the profit impact of online product reviews and how this profit varies with the number of reviews posted for a product. Consider a firm that manages word-of-mouth activity (Godes and Mayzlin 2009) by incentivizing consumers to post online product reviews. Such a firm might be interested in understanding the profit impact of providing incentives for more reviews. In this context, our policy simulations provide three unique insights. First, there are diminishing returns to increasing the number of reviews. So the first review of a product has a greater profit impact than the tenth review. Second, although increasing reviews always leads to greater market share, they might lead to lower profits. This happens when the marketing costs associated with eliciting more reviews are not commensurate with gross margins from product sales. Third, there is an optimum number of product reviews that the firm should spend on to maximize profits. We estimate this optimum number based on our model estimates.

In a second counterfactual experiment, we study the issue of fake product reviews. Although consumers have been exposed to product reviews for several years now, practitioners are increasingly concerned about the phenomenon wherein firms incentivize people to post fake positive reviews about the products they market. Though academics have studied this phenomenon (Mayzlin 2006, Dellarocas 2006), a deep understanding of this important and topical issue is not possible with a reduced-form aggregate-level approach. Our model provides a tool to better understand this phenomenon. We empirically demonstrate that fake reviews increase consumer uncertainty. The effects of more positive reviews and more numerous reviews on consumer choice are lower on online retailing platforms that have fake product reviews.

The remainder of this paper is organized as follows. In §2, we briefly review the relevant literature and discuss how our study extends this work. In §3, we present the data. In §4, we develop a structural model that captures consumer learning about the subjective product quality, based on both one's own and other consumers' usage experiences. We also describe the process of how the credibility of reviews evolves over time. In §5, we apply the model to a consumer choice decision on book purchases, and we discuss the empirical results and implications of this application. Section 6 concludes the paper with implications for future research.

2. Literature Review

We first briefly review two streams of literature, and then we discuss how this research contributes to those literatures.

The first stream of literature is related to the effect of online product reviews on sales of experiential products. There are mixed findings on the relationship between how positively a product is reviewed and the sales of that product. Several studies have empirically shown that positive reviews are associated with higher sales, whereas negative reviews tend to hurt sales of experiential products such as books and movies (Dellarocas et al. 2007, Chevalier and Mayzlin 2006). Several other studies did not find any statistically significant relationship (Duan et al. 2008, Liu 2006). Another finding is that buyers seem to find movies and books that have generated numerous reviews more interesting, which in turn drives greater demand, than those movies and books that have not received as many reviews (Liu 2006).[1]

The second stream of literature is related to Bayesian learning (Meyer and Sathi 1985, Roberts and Urban 1988, Luan and Neslin 2009). This process assumes that consumers are uncertain about product quality and update their quality expectation based on past experience and other factors such as marketing communication. In marketing, Erdem and Keane (1996) is a pioneering paper that studies the effect of learning from advertising. Erdem (1998) models consumer cross-category learning; i.e., consumer usage experience of one product can influence consumer quality expectation of another product made by the same company under the same brand. Mehta et al. (2003) apply the notion of consumer learning in studying consideration set formation under price uncertainty and consumer search. Erdem et al. (2004) study consumer learning of store brand quality across countries. Narayanan et al. (2005) and Narayanan and Manchanda (2009) model physician learning of the quality of new drugs from marketing instruments. Iyengar et al. (2007) model consumer learning of both service quality and usage. Zhao et al. (2011) model consumer learning in a product crisis situation.

We propose a novel modeling approach to study the impact of online product reviews on consumer choice. Our approach makes two contributions. Unlike previous studies that have examined the relationship between product reviews and consumer choice at an aggregate reduced-form level, we propose a structural individual-level choice model. We extend the Bayesian learning framework to incorporate product reviews as a source of information that influences individual-level consumer choice. Although aggregate-level descriptive analysis is convenient for establishing a general phenomenon (i.e., a product review has a significant impact on product sales), it is perhaps not appropriate for studying the microlevel mechanism under which such a phenomenon occurs. This is where the disaggregate-level structural analysis offers several advantages. From a theoretical standpoint, aggregate models are

[1] Recent work has also studied dynamics in online product reviews (Moe and Trusov 2011, Moe and Schweidel 2011). Li and Hitt (2008) and Godes and Silva (2011) find evidence of a declining trend in posted reviews as the volume of reviews increases with time. However, these papers do not study the effect of product reviews on choice, which is our main focus.

inadequate to understand the relative importance of a consumer's learning from product reviews (i.e., learning from indirect experience) vis-à-vis the consumer's learning from his or her own experience with the product (learning from direct experience). Our modeling approach yields the novel insight that consumers learn more from the indirect experience of product reviews about a book than from their own direct experience of a product category. From a methodological perspective, our model predicts consumer choice better than reduced-form models estimated with the same data. From a managerial perspective, we demonstrate how the model can be used to assess the economic effect of the much-discussed phenomenon of fake product reviews and how they might affect consumer choice. Such analysis is not possible with aggregate-level reduced-form models.

Our second contribution pertains to the study of the credibility of product-related communication. Although credibility has been extensively studied in the consumer behavior literature, and its importance in determining the persuasiveness of communication is well established (Kelman 1961, Chaiken 1980), it remains unstudied in the choice modeling literature. Modeling credibility is challenging because although product reviews and consumer choices are observed to the researcher, the credibility of reviews is unobserved. As mentioned earlier, we model credibility as the precision with which product reviews reflect the consumer's own product evaluation. The Bayesian learning literature assumes that the precision level of any source of product information (such as advertising) is constant. However, the perceived precision of information from product reviews is likely to vary over time, as she accumulates more information from her own experience to validate product reviews. We contribute to the Bayesian learning literature not just by incorporating product reviews as an additional information source but also by modeling the credibility of product reviews, how it varies over time, and how it affects individual-level consumer choice. With the exception of Zhang (2010), the Bayesian learning literature has focused on learning from consumers' own experiences and marketing instruments, ignoring the process of learning from others.

Zhang (2010) empirically models observational learning (learning based on actions taken by others) using a Bayesian updating framework. She studies the U.S. kidney market, where patients on a waiting list sequentially decide whether to accept a kidney offer. Patients draw negative quality inferences based on earlier refusals in the queue. Our study is different from this work as follows. We model learning from information sharing (learning from information shared by others). Indeed, observational learning and information sharing are distinct forms of social learning. In addition, we simultaneously model and compare learning from one's own and others' experience. Own experience with a product is not relevant in the specific institutional setting of Zhang (2010), but it is likely to occur for most choice decisions because repeat purchase behavior is quite common in most product categories, including durables. Finally, we model the consumer's perception of the credibility of product reviews and how this credibility evolves over time.

3. The Data

The data for this study were collected from a U.S. company. The full sample includes consumers who participate in a marketing program to evaluate each of their purchases in the book category. In other words, the data provide information on which book was bought by a consumer at a purchase occasion and the consumer's reviews on each of her purchases. The company regularly emails participants, urging them to post reviews, and it runs other marketing programs to ensure that all purchases are reviewed. However, like other marketing data (such as scanner panel data), there could be some missing observations, which we are unable to detect or trace.

A product review typically consists of the following: a review title, a review body, pros and cons of the product, and a rating of the product on a five-point scale. The five-point scale reads as follows. 1 = avoid it, 2 = below average, 3 = average, 4 = above average, and 5 = excellent. In this research, we focus our attention to studying learning behavior in the purchases of books. Books are experience goods and are frequently reviewed by consumers. Also, the prevalence of advertising is expected to be much less in books than in other frequently reviewed experience goods such as movies, so it is quite unlikely that consumers in our data learn about product quality from advertising. Finally, it has been established that at an aggregate level, book reviews affect book sales (Chevalier and Mayzlin 2006).

Next we discuss our sampling plan. Out of *all* books purchased at least once in the 30-month period starting July 1999, we randomly selected 1,000 books. We then classified each book into one of the following categories: romance ("romance"), science fiction/fantasy ("science fiction"), mystery and crime ("mystery"), and horror and thriller ("horror"). For the purpose of model calibration, we followed the practice in the learning literature of sampling heavy buyers (Erdem and Keane 1996) and restricted ourselves to those consumers who made at least four purchases from these 1,000 books in the data period. For consumers with fewer purchases, it is rather difficult to identify the key pattern from noise. We model the purchase of the top 150 books (by the number of purchase observations) and classify the remaining 850 books into four category-specific "other" goods. This leads to a data set of 1,919 purchases made by 243 consumers, including 798 purchases of the "other" goods. For the 1,919 purchase observations, the distribution of consumer ratings is as follows: 969 purchases were rated 5, 524 were rated 4, 257 were rated 3, 111 were rated 2, and 58 were rated 1. As is typical of online product reviews, a majority of product evaluations are quite positive. Next, we sampled all the product reviews posted on the company's website in the data period for all these books and the date when each review was posted. We finally collected data on the month and year in which each book was released, as well as its price. Table 1 reports the summary statistics.

Among the four genres of books, science fiction has the largest share, followed by the genres mystery, romance, and horror. The science fiction genre has the highest average consumer rating, followed by the genres romance, mystery, and horror. Book rating data are important in our study for both substantive and methodological reasons. First, book ratings measure consumer satisfaction or experience of a specific book. Following Chintagunta et al. (2009), we employ these data in our learning model as experience signals that are used to update consumers' quality beliefs. Because experience signals are observed in our case, they do not need to be simulated as in typical empirical applications where experience signals are often unobserved. This significantly reduces the computational burden of model estimation. Second, the review information allows us to *observe* the similarity between the consumer's rating of a product and reviewers' ratings. Such similarity enables us to empirically study consumer learning of the credibility of reviews. In our data, the mean of the absolute difference between the consumers' ratings and the mean of reviewers' ratings on commonly reviewed books is 0.09 (statistically insignificant), and the variance is 1.21.

Table 1 Summary Statistics of the Data

Variables	Romance	Science fiction	Mystery	Horror
Choice share	0.13	0.53	0.23	0.11
Buyers' ratings of purchased book				
Mean	4.11	4.30	4.02	3.98
SD	1.08	0.95	1.12	1.16
Mean of reviewers' ratings of purchased books				
Mean	4.18	4.40	4.07	4.07
SD	0.33	0.42	0.60	0.49
Number of reviews				
Mean	16.41	18.07	17.28	17.35
SD	21.56	21.65	24.91	17.31
Time since publication (in days)				
Mean	732.99	1,248.41	514.52	758.00
SD	686.97	1,154.45	611.64	612.43
Price (in $)				
Mean	12.33	20.85	17.03	15.46
SD	4.30	15.47	8.33	7.82

4. The Model

In this section, we develop a structural model that captures consumer learning about the subjective product quality and reviewer credibility, based on both their own and others' usage experiences. The general modeling context is that a consumer makes choice decisions about which item (such as a book title) to purchase. These items can be grouped into different categories (such as genres). The consumer is facing uncertainty on true product quality and the review credibility. She updates her belief on review credibility using her own experience and the reviews. She then integrates her own experience with reviews to form an expectation on the true product quality. Such expectation will then drive her choice. We describe the model in the following order: (i) modeling consumer updating of the expected quality of a category, (ii) modeling consumer updating of the perceived credibility of reviews, (iii) modeling consumer integration of usage information from the self and others, and finally, (iv) modeling consumer choice decisions.

4.1. Modeling Consumer Updating of Expected Quality of a Category

First, we model how consumer i updates her belief of the quality of category j, based on items in this category that she buys and consumes. Quality in learning models refers to only that product attribute about which consumers are uncertain (Erdem 1998) and which is not perfectly observable. We assume that consumer i holds prior beliefs about the quality of each category at each time period, which are updated when the consumer consumes an item from that category, to form posterior quality beliefs. Consider a consumer at time $t-1$ who has a prior belief of the true category quality A_{ij} given her information set $I_{i,\,t-2}$, which is distributed as follows:

$$A_{ij} \mid I_{i,\,t-2} \sim \mathrm{N}(E_{i,\,t-2}(A_{ij}),\, \sigma^2_{vij,\,t-2}), \tag{1}$$

where $E_{i,\,t-2}(A_{ij})$ is the mean and $\sigma^2_{vij,\,t-2}$ captures the uncertainty of consumer i's belief about category j at time $t-2$. The subscript ν is a notation to differentiate the perception variance from other variance parameters in the model. For example, a consumer might have a generally positive quality perception of horror books but might be highly uncertain of this perception. As we discuss later, this uncertainty reduces over time as the consumer consumes items in the category j. So the more horror books the consumer reads, the lesser this uncertainty becomes for the horror category. Next we assume that the consumer purchases item k_j at time $t-1$. Her evaluation (or experience signal) of this item after consumption is $A_{Eik_j,\,t-1}$. Consistent with the literature, we assume that this evaluation is normally distributed around the consumer's belief of the quality of category j (A_{ij}) as follows (Erdem and Keane 1996):

$$A_{Eik_j,\,t-1} \sim \mathrm{N}(A_{ij}, \sigma^2). \tag{2a}$$

This is reasonable, because the consumer's utility for an item (and hence her choice of that item) is affected not just by her quality beliefs but also by her preferences for attributes for which there is no uncertainty (such as price). So consumers could purchase books of lower quality than the category mean. The variance σ^2 is referred to as experience variability. Unlike typical learning models where consumer experience signals are unobserved, we observe evaluations of books purchased by the consumer (measured by a five-point discrete scale rating). We assume the following relationship between the latent experience signal $A_{Eik_j,\,t-1}$ and the observed evaluation/rating of the purchased book $R_{Eik_j,\,t-1}$:

$$\begin{aligned}
&A_{Eik_j,\,t-1} \in (-\infty, 1.5] \quad \text{if } R_{Eik_j,\,t-1} = 1,\\
&A_{Eik_j,\,t-1} \in (1.5, 2.5] \quad \text{if } R_{Eik_j,\,t-1} = 2,\\
&A_{Eik_j,\,t-1} \in (2.5, 3.5] \quad \text{if } R_{Eik_j,\,t-1} = 3,\\
&A_{Eik_j,\,t-1} \in (3.5, 4.5] \quad \text{if } R_{Eik_j,\,t-1} = 4,\\
&A_{Eik_j,\,t-1} \in (4.5, +\infty] \quad \text{if } R_{Eik_j,\,t-1} = 5.
\end{aligned} \tag{2b}$$

The consumer uses her experience with item k_j to update her quality belief about category j. Specifically, following Bayes' rule, her posterior belief about the category quality after consuming product item k_j in this category at time $t-1$ is as follows:

$$A_{ij} \mid I_{i,\,t-1} \sim \mathrm{N}(E_{i,\,t-1}(A_{ij}), \sigma^2_{vij,\,t-1}), \tag{3}$$

where

$$E_{i,\,t-1}(A_{ij}) = E_{i,\,t-2}(A_{ij}) + D_{ik_j,\,t-1} B_{ij,\,t-1} \cdot (A_{Eik_j,\,t-1} - E_{i,\,t-2}(A_{ij})), \tag{4a}$$

$$\frac{1}{\sigma^2_{vij,\,t-1}} = \frac{D_{ik_j,\,t-1}}{\sigma^2} + \frac{1}{\sigma^2_{vij,\,t-2}}, \tag{4b}$$

$$B_{ij,\,t-1} = \frac{\sigma^2_{vij,\,t-2}}{\sigma^2_{vij,\,t-2} + \sigma^2}. \tag{4c}$$

$D_{ik_j,\,t-1}$ is a dummy variable indicating whether the item k_j is consumed by consumer i at time $t-1$. Note that if the experience signal $A_{Eik_j,\,t-1}$ resulting from the consumption of item k_j is the same as the expected value of category quality in time $t-2$, then this expected value remains unchanged in time $t-1$. Also, $B_{ij,\,t-1}$, which denotes the relative weight of the experience signal, is greater for greater values of $\sigma^2_{vij,\,t-2}$. In other words, the more uncertain a consumer is about her prior quality belief about the category, the more she learns from consuming an item in that category.

We allow the true category qualities to be different across consumers. Specifically, we allow $A_{ij} \sim \mathrm{N}(A_j, \sigma_A^2)$. We assume the initial prior belief of the true category quality to be normally distributed; i.e., $A_{ij} \mid I_{i1} \sim \mathrm{N}(E_{i0}(A_{ij}), \sigma^2_{vij0})$. By using the assumption of rational expectation, we have $E_{i0}(A_{ij}) = A_j$ and $\sigma^2_{vij0} = \sigma^2_A$. In other words, we assume that whereas consumer i does not know the value of her true quality A_{ij} in the first period, she knows that it is normally distributed across the consumer population as $A_{ij} \sim \mathrm{N}(A_j, \sigma_A^2)$. So she rationally forms her initial prior beliefs to be the same as the population-level distribution.

4.2. Modeling Updating of Perceived Credibility of Average Review by Consumers

In this section, we present a model for the credibility of information that the consumer receives from product reviews. For any product, the consumer is exposed to product reviews that provide information about the quality of the product. The credibility of this information for the consumer is unobserved to the researcher. We model the review credibility as the precision with which product reviews reflect the consumer's own product evaluation. The higher the precision, the more credible is the information obtained from product reviews for the consumer.

Specifically, let MR_{k_j} be the population mean of the evaluations of all reviewers for the item k_j. MR_{k_j} nonsystemically deviates from consumer i's evaluation of item k_j, with the deviation following a normal distribution:

$$MR_{k_j} - A_{Eik_j,\,t-1} \sim \mathrm{N}(0, \sigma_R^2). \tag{5}$$

The consumer cannot observe the product evaluations of all reviewers but is exposed to reviews posted at or before time $t-1$. Let $R_{i'k_j,\,t-1}$ be the numerical evaluation of a specific review of product k_j posted by reviewer i' at or prior to time $t-1$. Furthermore, $\bar{R}_{k_j,\,t-1}$ and $n_{k_j,\,t-1}$ respectively denote the sample mean and number of reviews of product k_j at or prior to time $t-1$. We assume each review to be an unbiased signal of the population distribution of evaluations:

$$R_{i'k_j,\,t-1} - MR_{k_j} \sim \mathrm{N}(0, \tilde{\sigma}_R^2), \tag{6}$$

where $\tilde{\sigma}_R^2$ is the variance of the deviation between the evaluation of a specific review that consumer i is exposed to and the population mean. When this variance is large, it suggests that the review is not a good representation of the population mean. Because an individual can be both a consumer and a reviewer, we assume $\tilde{\sigma}_R^2 = \sigma_R^2$ to be consistent.[2]

Based on Equations (5) and (6), the distribution of the difference between the evaluation of consumer i and the average of the evaluations of all reviews that a consumer is exposed to is the following:

$$\bar{R}_{k_j,\,t-1} - A_{Eik_j,\,t-1} \sim \mathrm{N}\left(0, \left(1 + \frac{1}{n_{k_j,t-1}}\right)\sigma_R^2\right). \quad (7)$$

We earlier defined the credibility of reviews as the precision with which product reviews reflect the consumer's own product evaluation. It follows from Equation (7) that the variance parameter σ_R^2 measures the degree to which reviews are deviated from a consumer's own experience with the same book. Consequently, $1/\sigma_R^2$ measures the precision with which the reviewers' mean evaluation captures consumer i's own product evaluation, reflecting the similarity in taste between the reviewers and the consumer. The more deviated reviewers' mean evaluation is from consumer i's evaluation, the less credible reviews are to consumer i. Our modeling framework is general enough to allow for the perceived review credibility to be different across consumers and over time.

Next we describe how we model the evolution of the credibility of reviews for consumer i over time. We assume that consumer i has a prior belief of σ_R^2 at time $t-1$, which is gamma distributed:

$$\frac{1}{\sigma_R^2}\,\bigg|\, I_{i,\,t-2} \sim \Gamma(\alpha_{i,\,t-2}, \beta_{i,\,t-2}). \quad (8)$$

$\alpha_{i,\,t-2}$ is the shape parameter, and $\beta_{i,\,t-2}$ is the inverse scale parameter. Both take positive values only. After the consumer receives $n_{k_j,\,t-1}$ evaluations for item k_j at time $t-1$ (with the mean evaluation signal being $\bar{R}_{k_j,\,t-1}$) and consumes the item k_j to form her own evaluation $A_{Eik_j,\,t-1}$, her posterior distribution of the credibility of reviews (based on Bayes' rule) is

$$\frac{1}{\sigma_R^2}\,\bigg|\, I_{i,\,t-1} \sim \Gamma(\alpha_{i,\,t-1}, \beta_{i,\,t-1}), \quad (9)$$

where

$$\alpha_{i,\,t-1} = \alpha_{i,\,t-2} + \frac{D_{ik_j,\,t-1}}{2}, \quad \text{and} \quad (10a)$$

$$\beta_{i,\,t-1} = \beta_{i,\,t-2} + \frac{D_{ik_j,\,t-1}(\bar{R}_{k_j,\,t-1} - A_{Eik_j,\,t-1})^2}{2(1 + 1/n_{k_j,\,t-1})}. \quad (10b)$$

[2] We thank an anonymous reviewer for suggesting this specification.

The extent of updating of the scale parameter of consumer i is proportional to the deviation between the reviewers' mean evaluation and consumer i's own evaluation of the item. Also, the greater the number of reviews that consumer i is exposed to, the greater the level of updating will be.

4.3. Modeling Integration of the Consumer's Own Experience and Reviews

So far, we have discussed how consumers update their perception of the quality of a category when they consume a product from that category. We have also presented a general model of the credibility of product reviews for the consumer. Next, we describe how the consumer uses the two sources of information (own experience and product reviews) to choose an item to buy. Suppose the consumer is making a purchase decision on an item k_j^* at time t. One source of relevant information is her belief of the quality of category j at time $t-1$. From Equation (2a), we know that the evaluation of item k_j^* for consumer i is normally distributed around the category quality belief A_{ij} as follows:

$$A_{Eik_j^*t} \sim \mathrm{N}(A_{ij}, \sigma^2). \quad (11a)$$

We also know from Equation (1) that the prior category quality belief at time $t-1$ is distributed as follows:

$$A_{ij} \mid I_{i,\,t-1} \sim \mathrm{N}(E_{i,\,t-1}(A_{ij}), \sigma^2_{vij,\,t-1}). \quad (11b)$$

Equations (11a) and (11b) together imply that

$$A_{Eik_j^*t} \sim \mathrm{N}(E_{i,\,t-1}(A_{ij}), \sigma^2 + \sigma_{vij,\,t-1}). \quad (11c)$$

The other source of information is the evaluation of item k_j^* by reviewers. We assume that consumer i forms an expectation of the credibility of reviews (captured by the variance parameter σ_R^2). We also know from Equation (8) that $1/\sigma_R^2 \sim \Gamma(\alpha_{i,\,t-1}, \beta_{i,\,t-1})$. Based on this, consumer i forms the following expectation of credibility at time $t-1$:

$$E_{i,\,t-1}(\sigma_R^2) = \frac{\beta_{i,\,t-1}}{\alpha_{i,\,t-1} - 1}. \quad (12a)$$

This expectation combined with Equation (7) clearly implies that the mean evaluation of item k_j^* by reviewers is distributed as follows:

$$\bar{R}_{k_j^*t} \sim \mathrm{N}\left(A_{Eik_j^*t}, \left(1 + \frac{1}{n_{k_j^*,\,t-1}}\right)\frac{\beta_{i,\,t-1}}{\alpha_{i,\,t-1} - 1}\right). \quad (12b)$$

Then based on the Bayes rule, Equations (11c) and (12b), the expected quality of item k_j^*t for consumer i at time t, is normally distributed with the following mean and variance:

$$E_{it}(A_{Eik_j^*t}) = W_{ik_j^*,\,t-1}E_{i,\,t-1}(A_{ij}) + (1 - W_{ik_j^*,\,t-1})\bar{R}_{k_j^*t}, \quad (13a)$$

$$\mathrm{Var}_{it}(A_{Eik_j^*t}), = \frac{1}{1/(\sigma^2+\sigma^2_{vij,t-1})+((\alpha_{i,t-1}-1)/\beta_{i,t-1})/(1+1/n_{k_j^*,t-1})}, \tag{13b}$$

$$W_{Eik_j^*,t-1}, = \frac{1/(\sigma^2+\sigma^2_{vij,t-1})}{1/(\sigma^2+\sigma^2_{vij,t-1})+((\alpha_{i,t-1}-1)/\beta_{i,t-1})/(1+1/n_{k_j^*,t-1})}. \tag{13c}$$

The expected quality of the item depends on the consumer's perception of the expected quality of the category to which the item belongs and the reviews posted for that item. The mean and variance of the distribution of expected item quality both affect the consumer's expected utility for the item. She then decides on which item to buy based on her expected utilities across items in various categories.

We now present the specification for initial parameters $(\alpha_{i0}, \beta_{i0})$ related to consumer learning from reviews. Following Equation (12a), we have $\beta_{i0} = (\alpha_{i0}-1)E_{i0}(\sigma_R^2)$; α_{i0} indicates the magnitude of richness of consumer i's initial experience. Intuitively, an inexperienced consumer should have lower α_{i0} compared with an experienced customer. $E_{i0}(\sigma_R^2)$ represents consumer i's initial uncertainty of reviews. To obtain stable and theoretically meaningful estimates, we estimate $(\alpha_{i0}, E_{i0}(\sigma_R^2))$ instead of $(\alpha_{i0}, \beta_{i0})$. Moreover, since α_{i0} and $E_{i0}(\sigma_R^2)$ are both positive, we assume $\log(\alpha_{i0}) \sim \mathrm{N}(M_{\alpha 0}, V_{\alpha 0}^2)$ and $\log(E_{i0}(\sigma_R^2)) \sim \mathrm{N}(M_{R0}, V_{R0}^2)$. We allow for heterogeneity across consumers in their initial credibility of reviews. Next we present our model for the consumer's decision-making process of what item to purchase.

4.4. Modeling Consumer Choice Decisions

We discuss our utility-maximization-based approach of modeling how a consumer chooses to buy an item in the presence of information from her own experience with the category the item belongs to, and information from product reviews. Following Chintagunta et al. (2009), we assume that consumer i's utility for item k_j^* can be written as[3]

$$V_{ik_j^*t} = -\exp\left[-\hat{r}_i(\omega_i A_{Eik_j^*t} + X_{ik_j^*t}\gamma_i + \varepsilon_{ik_j^*t})\right] \tag{14}$$

where $\hat{r}_i$ is the risk aversion parameter and ω_i is the quality weight (both are greater than 0), $X_{ik_j^*t}$ is a vector of covariates such as price, and γ_i is the vector of coefficients of $X_{ik_j^*t}$; $\varepsilon_{ik_j^*t}$ is the extreme value error. Because the consumer cannot observe product quality prior to purchase, we assume that she decides based on the following expected utility:

$$E_{it}(V_{ik_j^*t}) = -E_{it}[\exp(-\hat{r}_i\omega_i A_{Eik_j^*t})] \cdot \exp[-\hat{r}_i(X_{ik_j^*t}\gamma_i + \varepsilon_{ik_j^*t})]. \tag{15}$$

Based on the theory of moment-generating function for normal distribution, this expression for the expected utility simplifies to the following:

$$\begin{aligned} E_{it}(V_{ik_j^*t}) &= -\exp\left[-\hat{r}_i\omega_i[E_{it}(A_{Eik_j^*t})] + \tfrac{1}{2}(\hat{r}_i\omega_i)^2[\mathrm{Var}_{it}(A_{Eik_j^*t})]\right] \\ &\quad \cdot \exp\left[-\hat{r}_i(X_{ik_j^*t}\gamma_i + \varepsilon_{ik_j^*t})\right] \\ &= -\exp\left\{-\hat{r}_i\{[\omega_i E_{it}(A_{Eik_j^*t})] - \tfrac{1}{2}\hat{r}_i(\omega_i)^2[\mathrm{Var}_{it}(A_{Eik_j^*t})] \right. \\ &\quad \left. + X_{ik_j^*t}\gamma_i + \varepsilon_{ik_j^*t}\}\right\}. \end{aligned} \tag{16}$$

The details of how $E_{it}(A_{Eik_j^*t})$ and $\mathrm{Var}_{it}(A_{Eik_j^*t})$ depend on the consumer's own experience with category j, and with reviews posted for the item k_j^*, were presented in Equation (13). The consumer maximizes the expected utility, which is equivalent to maximizing the following:

$$\begin{aligned} U_{ik_j^*t} = \hat{U}_{ik_j^*t} + \varepsilon_{ik_j^*t} &= \omega_i[E_{it}(A_{Eik_j^*t})] \\ &\quad - r_i[\mathrm{Var}_{it}(A_{Eik_j^*t})] + X_{ik_j^*t}\gamma_i + \varepsilon_{ik_j^*t}, \end{aligned} \tag{17}$$

where $r_i = (1/2)\hat{r}_i(\omega_i)^2$.

It is important to account for unobserved heterogeneity in consumer preferences, especially in models of learning (Shin et al. 2011). Accordingly, we model unobserved heterogeneity of the model parameters across consumers as follows: $\log(\omega_i) \sim \mathrm{N}(\bar{\omega}, \sigma_\omega^2)$, $\log(r_i) \sim \mathrm{N}(\bar{r}, \sigma_r^2)$, $A_{ij} \sim \mathrm{N}(A_j, \sigma_A^2)$, and $\gamma_i \sim \mathrm{MVN}(\bar{\gamma}, \Sigma)$, where Σ is a diagonal variance matrix. Finally, the utility of the "other" good o_j in category j is specified as

$$U_{io_jt} = O_j + \varepsilon_{io_jt}, \tag{18}$$

where O_j is the category-specific intercept for the "other" good.

It is plausible that the utilities of items within a category are correlated because of unobserved variables. To account for this possible correlation within a category, we adopt a nested logit specification. If ρ is a measurement of this correlation (within nests or categories), we can write the probability of consumer purchasing product k_j^*, conditional on the consumer purchasing category j as follows (Berry 1994, Train 2003):

$$P(D_{ik_j^*t} = 1 \mid D_{ijt} = 1) = \frac{\exp(\hat{U}_{ik_j^*t}/(1-\rho))}{\sum_{k'}\exp(\hat{U}_{ik'_jt}/(1-\rho))}. \tag{19}$$

[3] We also tested the utility function specification following Erdem and Keane (1996). However, the exponential specification leads to a better model performance, and hence we decided to adopt this specification.

Furthermore, the unconditional probability of the consumer purchasing from category j is given by

$$P(D_{ijt}=1)=\frac{[\sum_{k'}\exp(\hat{U}_{ik'_j t}/(1-\rho))]^{1-\rho}}{\sum_{j'}[\sum_{k'}\exp(\hat{U}_{ik'_{j'} t}/(1-\rho))]^{1-\rho}}. \tag{20}$$

The unconditional probability of buying item k_j^* is obtained by simply multiplying the two probabilities above. This completes the description of the model.

We include two covariates in the model: the price of the book and the time elapsed since publication (in days) for each book. The average price of a book in the science fiction genre is the greatest among the four genres, followed by mystery, horror, and romance. We predict that demand may be negatively correlated with the time since publication because publishers often invest more resources in marketing a book when it is launched. Among the four genres, mystery books are the latest releases, followed by books in the horror, romance, and science fiction genres. It is plausible that temporal variations in the reviews of a book might lead the online retailer and/or the publisher to change prices. However, we are unable to find any evidence of this on the website from which we obtained data.

4.5. Model Identification, Estimation, and Comparison

We first discuss specific features of our data that enable model identification. Our data set is different from other data used in the learning literature (e.g., scanner data and physician prescription data) in that we observe two separate pieces of information, one being the consumer's own experience signal (unbiased predictor of consumer preference), i.e., her rating of a book after consumption; and the other being reviewers' ratings of the same book. The two pieces of information allow us to separately identify consumer learning from the self and from reviewers. Also, because we observe consumers' postconsumption evaluations (i.e., book ratings), we are able to separately identify J intercepts, i.e., an intercept for each genre. In a standard choice model with J choice options, we can only identify $J-1$ intercepts without any additional information. Our approach directly follows Chintagunta et al. (2009). In their paper, the authors integrate customer satisfaction data into a learning model and use satisfaction data to infer quality perceptions.

In our empirical context, we observe the buyer's book rating, which is a good indicator of the perceived quality A_{Eijt}. It is well known that the variance of utility weight, experience variability, and variance of risk reversion cannot be jointly identified because of the "invariance scale" problem in classical learning models (Erdem 1998). The common solution for this problem is to fix the variance of utility weight to 1. However, because we have indirect information on the consumer's perceived quality in our context, we can identify these three variances without any additional constraint.

We next explain the intuition on the identification of quality weight and risk aversion coefficients from the prior quality belief. Quality weight and risk aversion coefficients do not vary over time, and therefore their effect on choice is persistent. However, the effect of prior quality belief on choice (based on the Bayesian updating theory) is decreasing over time. In other words, the effect of prior is more significant in the first few observations but minimal in the long term. These different patterns allow us to achieve the identification. Separate identification of these three parameters has been widely adopted in the learning literature. Because we can identify the quality weight parameter and the risk aversion parameter at the individual consumer level given the panel data, we can then identify the unobserved heterogeneity in these parameters across consumers. We ran some simple regressions and found some evidence on the effect of average review ratings and number of reviews on consumer choices.

The rationale for identifying the risk aversion coefficient from quality weight is as follows. The risk aversion coefficient measures the consumer's sensitivity to her uncertainty on product or quality, and this uncertainty mainly depends on number of previous purchases in the same category and number of reviews for the product or book (please refer to Equations (4b) and (13b)). However, quality weight measures the consumer's sensitivity to her evaluation on the product quality, and this evaluation mainly depends on her ratings on her previously purchased books in the same category and reviewers' ratings on the book (please refer to Equations (4a) and (13a)). Namely, the risk aversion coefficient and quality weight measure effects on two completely different variables, so we can separately identify them.

Finally, quality weights are separately identified from the true mean qualities in standard learning models, through the functional form of the distribution and by fixing the variance of the quality weight across consumers to be 1 (Erdem 1998). In addition, in our case, we observe consumer experience signals, i.e., the product ratings, which help us make inference on true mean quality. Thus, the true mean quality and the observed choices data jointly identify the quality weight.

We used the simulated maximum likelihood method to estimate the model because the discrete choice probabilities needed to construct the likelihood functions involve high-order integrals over the random variables (e.g., Keane 1993, Hajivassiliou and Ruud 1994). We used the quasi-Newton method to

Table 2 Model Simulation Results

Parameter	True value	Mean estimate (SE)
True mean quality		
$A_{j=1}$ (romance)	2.0	1.893 (0.061)
$A_{j=2}$ (science fiction)	2.5	2.476 (0.044)
$A_{j=3}$ (mystery)	3.0	2.976 (0.034)
$A_{j=4}$ (horror)	3.5	3.528 (0.029)
Intercept ("Other" goods)		
$O_{j=1}$ (romance)	3.0	3.068 (0.197)
$O_{j=2}$ (science fiction)	3.0	2.991 (0.192)
$O_{j=3}$ (mystery)	3.0	3.121 (0.196)
$O_{j=4}$ (horror)	3.0	2.980 (0.202)
Utility parameters		
$\bar{\omega}$ (utility weight)	1.0	0.980 (0.038)
$\bar{r}$ (measure of risk aversion)	−0.5	−0.624 (0.245)
$\bar{\gamma}_1$ (coefficient of price)	−0.5	−0.467 (0.030)
$\bar{\gamma}_2$ (coefficient of "time since publication")	0.5	0.575 (0.047)
Other parameters		
σ_A (SD in true category quality)	0.5	0.502 (0.018)
σ_ω (SD in true utility weights)	0.5	0.489 (0.027)
σ_r (SD in true risk aversion)	0.5	0.354 (0.101)
$\sigma_{\gamma 1}$ (SD in price coefficient)	0.5	0.493 (0.030)
$\sigma_{\gamma 2}$ (SD in coefficient of time since publication)	0.5	0.536 (0.034)
σ_j (SD of experience variability)	1.0	0.871 (0.093)
logit(ρ) (correlation among category utilities)	0.0	−0.039 (0.109)
log(α_0) (credibility parameter)	2.3	2.186 (0.221)
log(β_0) (credibility parameter)	0.7	0.567 (0.279)

maximize the log-likelihood function. The BHHH algorithm (Berndt et al. 1974) is employed to approximate the Hessian. To demonstrate that the proposed model is completely identified and that the estimation procedure is valid, we conducted a simulation study. We drew a set of covariates from a standard normal distribution and assigned "true" values for all model parameters. On estimating the model, based on the method described above, we find that we are able to statistically recover the true values for all parameters. Table 2 reports the true values and the estimates of all parameters. The 95% confidence intervals of all estimates contain the corresponding true parameter values, suggesting that our model is identified and the proposed estimation procedure is valid. We repeated the simulation exercise with several different sets of parameter values and find consistently unbiased estimates.

We estimated three models (the proposed model and two nested versions of the proposed model). Model 1 is the baseline model where we include the two covariates but do not incorporate learning. Model 2 incorporates learning from consumers' own experiences with each genre of books. Model 3 is our proposed model where we account for both learning from the self and learning from others. Table 3 reports the in-sample fit statistics of the three models. Models 2 and 3 both perform better than model 1. So incorporating learning from the self and others,

Table 3 In-Sample Model Fit Comparison

	Learn from self	Learn from others	AIC	BIC
Model 1	N	N	29,332.82	29,427.33
Model 2	Y	N	29,313.44	29,419.07
Model 3 (proposed)	Y	Y	29,139.42	29,267.29

both improve model performance. Furthermore, irrespective of the fit statistic (log likelihood, Akaike information criterion (AIC), or Bayesian information criterion (BIC)), there is a greater improvement in model performance from model 2 to model 3, than from model 1 to model 2. This suggests greater learning from product reviews than learning from own experience, which is to be expected. It may be recalled that we model consumer learning from own experience with respect to each book genre and model consumer learning from others with respect to the focal book title. Compared with the uncertainty about a genre, a consumer is likely to be much more uncertain about the match of his or her own taste with a specific book.

Next we compare the out-of-sample predictive ability of the three models (see Table 4). For this purpose, we draw a random holdout sample of 368 purchase observations (10% of our data), estimate the model on the remaining data, and compute fit statistics based on the holdout sample. We employ two measures: "hit rate (book)" is defined as the proportion of book purchases which are correctly predicted by the model, and "hit rate (category)" is defined as the proportion of purchases for which the model correctly predicts the category from which this book is purchased. Across both metrics, models 2 and 3 perform better than model 1. The improvement in hit rate is greater between models 3 and 2, than between models 2 and 1. So although modeling learning both from oneself and from others is important for the predictive performance of the model, gains in predictive performance are much greater from modeling learning from others. This provides converging evidence that in our data, there is greater learning from others than from the consumer's own past experience.

Table 4 Out-of-Sample Model Fit Comparison

	Hit rate (book) (%)	Relative increase (book) (%)	Hit rate (category) (%)	Relative increase (category) (%)
Model 1	4.40	NA	35.04	NA
Model 2	4.75	7.95	35.20	0.46
Model 3 (proposed)	4.97	12.95	37.04	5.71

Notes. Note that the hit rate at the book level based on the random choice rule is 0.65% (=1/154), and the hit rate at the category level based on the random choice rule is 25% (=1/4). Compared to these benchmarks, all of our presented models perform reasonably well.

To demonstrate the value of our structural approach, we compare our proposed model to a reduced-form model that incorporates the same set of information. This is an expanded version of model 1 with the following covariates: genre-specific intercept, average rating of the book prior to the purchase, number of reviews posted prior to the purchase, average rating of the books in the category the book belongs to prior to the purchase, genre-specific state dependence effect, price, and time since publication. We find that our proposed model provides higher hit rates (4.97% at the book level and 37.04% at the category level) than the reduced-form model (4.77% at the book level and 35.88% at the category level). This clearly demonstrates the importance of modeling the underlying learning mechanism to gain a deeper understanding about the impact of product reviews.

5. Results and Managerial Implications

5.1. Results

Next we discuss the parameter estimates and associated insights. Table 5 reports the estimates from our proposed model (model 3) and its two nested versions (models 1 and 2). The relative magnitudes of the estimated true mean quality levels are generally consistent with the market shares of the four genres of books. The mean quality of the genre with greatest market share (science fiction) is estimated to be the greatest. Overall, the mean quality estimates for the four genres across the three models are not very different. This is because the genre-specific mean quality is inferred from the consumer book review data, which are the same across three models. One difference is that the inference on genre-specific mean quality in models 2 and 3 also utilizes consumer book choice data through the initial prior specification.

The intercept estimates for the "other" goods have some noticeable changes. This is due to the different estimates of the utility weight and risk aversion coefficient across models. For example, the intercept estimates are much smaller in model 3 than in model 1. This is mainly due to the significant risk coefficient we found. In model 1, the risk coefficient is fixed to be 0 since we did not model any learning or uncertainty there, whereas in model 3, the risk coefficient estimate is significant. In model 3, the risk aversion

Table 5 Parameter Estimates (Mean and Standard Error)

Parameter	Model 1	Model 2	Model 3
True mean quality			
$A_{j=1}$ (romance)	**4.248 (0.106)**	**4.406 (0.131)**	**4.478 (0.111)**
$A_{j=2}$ (science fiction)	**4.514 (0.072)**	**4.845 (0.087)**	**4.746 (0.064)**
$A_{j=3}$ (mystery)	**4.198 (0.096)**	**4.359 (0.108)**	**4.270 (0.093)**
$A_{j=4}$ (horror)	**4.207 (0.112)**	**4.308 (0.134)**	**4.218 (0.111)**
Intercept ("other" goods)			
$O_{j=1}$ (romance)	**4.067 (0.363)**	**1.716 (0.531)**	**−13.277 (0.498)**
$O_{j=2}$ (science fiction)	**5.009 (0.352)**	**2.616 (0.526)**	**−12.372 (0.495)**
$O_{j=3}$ (mystery)	**4.099 (0.377)**	**1.837 (0.521)**	**−13.196 (0.498)**
$O_{j=4}$ (horror)	**3.458 (0.369)**	**1.150 (0.530)**	**−13.887 (0.503)**
Utility parameters			
$\bar{\omega}$ (utility weight)	**−0.591 (0.147)**	**−2.208 (0.621)**	**−2.199 (0.685)**
$\bar{r}$ (log(risk aversion))	—	**−0.198 (0.313)**	**2.350 (0.062)**
$\bar{\gamma}_1$ (coefficient of price)	0.076 (0.900)	0.042 (0.528)	0.014 (0.534)
$\bar{\gamma}_2$ (coefficient of time since publication)	**−0.665 (0.075)**	**−0.626 (0.072)**	**−0.768 (0.084)**
STD and correlation parameters			
σ_A (SD in true category quality)	**0.752 (0.062)**	**0.871 (0.064)**	**0.502 (0.026)**
σ_ω (SD in true utility weights)	**0.343 (0.081)**	**1.526 (0.532)**	**0.495 (0.229)**
σ_r (SD in true risk aversion)	—	**0.367 (0.120)**	**0.060 (0.025)**
$\sigma_{\gamma 1}$ (SD in $\bar{\gamma}_1$)	0.031 (1.808)	0.032 (0.304)	0.004 (0.367)
$\sigma_{\gamma 2}$ (SD in $\bar{\gamma}_2$)	**0.811 (0.087)**	**0.753 (0.076)**	**1.032 (0.091)**
σ (experience variability)	**1.260 (0.025)**	**1.241 (0.034)**	**1.329 (0.033)**
ρ (correlation in cross-category utilities)	0.013 (0.079)	0.005 (0.060)	0.012 (0.055)
Parameters related to learning on reviews			
$M_{\alpha 0}$ (mean of initial $\log(\alpha)$)	—	—	**6.125 (0.929)**
$V_{\alpha 0}$ (SD of initial $\log(\alpha)$)	—	—	**1.379 (0.596)**
M_{R0} (mean of initial $\log(E(\sigma_R^2))$)	—	—	**1.886 (0.118)**
V_{R0} (SD of initial $\log(E(\sigma_R^2))$)	—	—	**0.398 (0.052)**

Note. Statistically significant estimates ($p < 0.05$) appear in bold.

suggested that the expected utility of each of the four genres of the books will incur a negative component (i.e., risk coefficient times the expected variance, and there is a negative sign in front of the risk coefficient in our model specification). This negative component will drive down their expected utility, given that the expected mean quality is mainly inferred from the observed review ratings (i.e., they will not offset the negative component). For the simulated choice shares to be consistent with what is observed from the data, the intercept estimates of the four "other" goods in model 3 are smaller.

The three models yield very different estimates of the mean utility weight parameter $\bar{\omega}$. Specifically, we find that the parameter estimates from the model that ignores learning from both the self and others (model 1) are biased upwards. The risk aversion parameter, $\bar{r}$, is found to be significant in both model 2 ($=\exp(-0.198)$) and model 3 ($=\exp(2.350)$). Also, we find that irrespective of model specification, consumers are heterogeneous in their risk aversion (σ_r is significant for both models 2 and 3). Similarly, model 2 overestimates the initial consumer uncertainty of true quality ($\sigma_{vij,\,t=0}$). However, the magnitude of such uncertainty is not substantial. This is consistent with our finding of limited consumer learning from their own experience of books of the same genre. This implies that most of quality uncertainty is with respect to the subjective quality of the specific book title, which is what incentivizes an individual consumer to learn from product reviews. Furthermore, all models produce a significant estimate of experience variability (σ), suggesting that consumer experience only provides a noisy signal of the true product quality. Finally, the correlation parameter in the nested-logit specification is insignificant in all three models.

We now report the effects of the two covariates we incorporate in our application: price and time since publication. To capture the nonlinearity effect of time since publication, we include the log-transformed time since publication. Across all three models, we find that the price coefficient is insignificant, suggesting low sensitivity to prices in book choices. This could perhaps be driven by low variance in prices across books (see Table 1). As predicted, books that have been published more recently are preferred. Both nested models overestimate this coefficient, providing further evidence of the importance of accounting for learning from reviews in choice models.

A unique feature of our model is that it accounts for the credibility of reviews and allows it to vary across consumers. Credibility is measured by $1/\sigma_R^2$ (i.e., precision), where σ_R^2 is a measure of uncertainty about reviews posted by reviewers. The initial credibility coefficient is modeled as $1/E_{i0}(\sigma_R^2)$. Our estimates of mean and standard deviation for $\log(E_{i0}(\sigma_R^2))$ are 1.886 ($\mathrm{SD}=0.118$) and 0.398 ($\mathrm{SD}=0.052$), respectively. The large and significant standard deviation for $\log(E_{i0}(\sigma_R^2))$ suggests that review credibility is substantially heterogeneous. The mean of $\log(\alpha_{i0})$, magnitude of richness of experience, is 6.125 ($\mathrm{SD}=0.929$), which indicates that most consumers have experience buying books. However, high and significant standard deviation of $\log(\alpha_{i0})$ suggests our data might include some novices as well.

In summary, the following are the unique findings of this research: (a) there is a significant amount of consumer learning from online reviews of book titles, much more than that from consumers' own experience with other books in the same genre; (b) based on the similarity between reviews for a book and their own evaluation of that book, consumers update their beliefs about the credibility of product reviews over time; and (c) not accounting for consumer learning from others leads to biased inferences of the effects of some covariates on choice.

5.2. Counterfactual Simulations

A key benefit of adopting a structural modeling approach to understand the effect of product reviews is that we are able to estimate the effects of firms' marketing policies and reviewer behavior on consumer choice and market share. To illustrate the managerial relevance of this research, we present two simulations.

Simulation 1. The objective of this counterfactual experiment is to illustrate how our modeling framework can help firms evaluate the profitability of offering product review incentives. Our managerial context is that of a firm that manages word-of-mouth activity (Godes and Mayzlin 2009) by incentivizing consumers to post online product reviews. In many industries such as retail and lodging, consumers are incentivized to write a review to share their experiences. Firms such as Amazon, Apple, Macy's, and Walmart are known to regularly request their consumers to post product reviews. Consider a book retailer (or publisher) whose gross margin of selling a book is $x\%$ of the selling price (P), and the average cost of incentivizing a consumer to post one product review is $y\%$ of the selling price. As is typical of word-of-mouth marketing campaigns, we assume that the firm cannot control the content of the review posted by the reviewer but can incentivize consumers to post more reviews. If the market size is M units, it follows that the profit increase as a result of v additional reviews is given by $(\Delta M)xP - vyP$, where ΔM is the increase in demand as a result of v product reviews. So the former term in the expression for profit increase denotes increase in revenue, and the latter term denotes the marketing expenditure incurred by the publisher.

Based on this setup, we estimate the profit impact of increased customer reviews for a representative book (with all attributes set at mean levels in our data). We assume the marketing expenditure associated with an additional product review, $y = 10\%$, and a conservative market size M of 10,000 units. We then conduct policy simulations by estimating changes in market share, and consequent changes in profits, at varying levels of product margins (we vary $x = 10\%$, 12%, and 14%, based on book industry norms) and varying numbers of additional reviews (we vary $v = 10, 20, \ldots, 100$). The market share of a representative book with the number of reviews equaling the mean number of reviews per book in our data is estimated to be 1.13%.

Results of policy simulations appear in Table 6, and there are three key findings. First, as expected, there are diminishing returns to increasing the number of reviewers. Increasing the number of reviewers from 0 to 10 has a much greater profit impact than increasing the number of reviewers from 90 to 100. Second, although increasing reviewers always leads to greater market share, it might lead to lower profits. For example, the policy of spending 10% of the book price on incentivizing an additional review leads to lower profits if the number of reviews increases by 70 and if the profit margin is 10%. Third, there is an optimum number of product reviews that the firm should spend on to maximize profits. For example, if the profit margin is 14%, the optimal number of reviews is 40. These results are consistent with the mechanism of Bayesian learning, as uncertainty is greater initially and reduces over time as the consumer gains more experience with the genre and is exposed to more numerous reviews. From a managerial perspective, this suggests that although all product reviews have a positive effect on market share, reviews that are posted earlier have a greater effect than those posted later. Firms spending marketing dollars for incentivizing people to post reviews might wish to consider this dynamic effect of reviews on market share. It would be rational for firms to pay more for eliciting earlier reviews than for later reviews.

Table 6 Effect of Increasing Product Reviews on Profits

Increase in the number of reviews	Market share (in %)	Profit increase as % of baseline profit		
		Margin = 10%	Margin = 12%	Margin = 14%
00	1.13	0.00	0.00	0.00
10	1.26	7.08	7.82	8.34
20	1.36	**11.68**	13.16	14.21
30	1.41	11.50	**13.72**	15.30
40	1.45	10.62	13.57	**15.68**
50	1.45	6.19	9.88	12.52
60	1.45	1.77	6.19	9.36
70	1.46	−1.77	3.39	7.08
80	1.46	−6.19	−0.29	3.92
90	1.46	−10.62	−3.98	0.76
100	1.47	−14.16	−6.78	−1.52

Note. Baseline profit is the profit when the number of reviewers is the same as the sample mean.

Simulation 2. Fake reviews are receiving an increasing amount of attention by practitioners (Streitfeld 2011, Helft 2010). Some firms hire consumers to write fake reviews to spread more positive marketing communication. Our model provides a tool to understand the impact of this practice. Because fake reviews are more positive than authentic reviews, there should be a greater discrepancy between fake reviews of a product and the consumer's own experience of that product, which will increase consumer uncertainty. Intuitively, such increased consumer uncertainty may reduce the effect of reviews because of the lowered credibility of reviews.

Consider four online retailing platforms selling the same books. One has only authentic product reviews (termed "authentic"), and the other three have a 25%, 50%, or 75% chance of getting fake reviews (termed "fake"), respectively. For authentic reviews across the four platforms, we set the mean quality of each book as 4. Fake reviews are assumed to have a rating of 5, because they are always very positive. The experience signals of a new consumer who buys from these platforms are drawn from the distribution $N(4, \sigma^2)$. For each platform, we consider two subconditions: one involves customers with less prior experience ($\alpha_0 = 5$) and the other involves customers with more prior experience ($\alpha_0 = 50$). For each platform and under each of the two conditions, we simulate the path of the degree of uncertainty ($E_{it}(\sigma_R^2)$) perceived by the consumer as she purchases more books. Some patterns suggest the following. In the fully authentic platform (the dark lines in Figures 1–3), the uncertainty first decreases (as a result of learning from one's own experience) and then levels off. In the fake platforms, the consumer is exposed to a combination of fake and authentic reviews. We find that her uncertainty increases significantly as she purchases more books. This analysis demonstrates that fake reviews *increase* consumer uncertainty.[4] Furthermore, the effect of fake reviews on consumer uncertainty becomes larger as the likelihood of getting a fake review increases and/or the consumer prior experience decreases.

Next we illustrate how the effect of reviews differs across the four platforms. We first examine the effect of increased review ratings of a book when the consumer makes her 21st purchase on each platform (Figures 1–3). We know that this consumer perceives

[4] We set all model parameters at their mean level.

Figure 1 The Effect of Fake Reviews on Consumer Uncertainty (25% Fake Reviews)

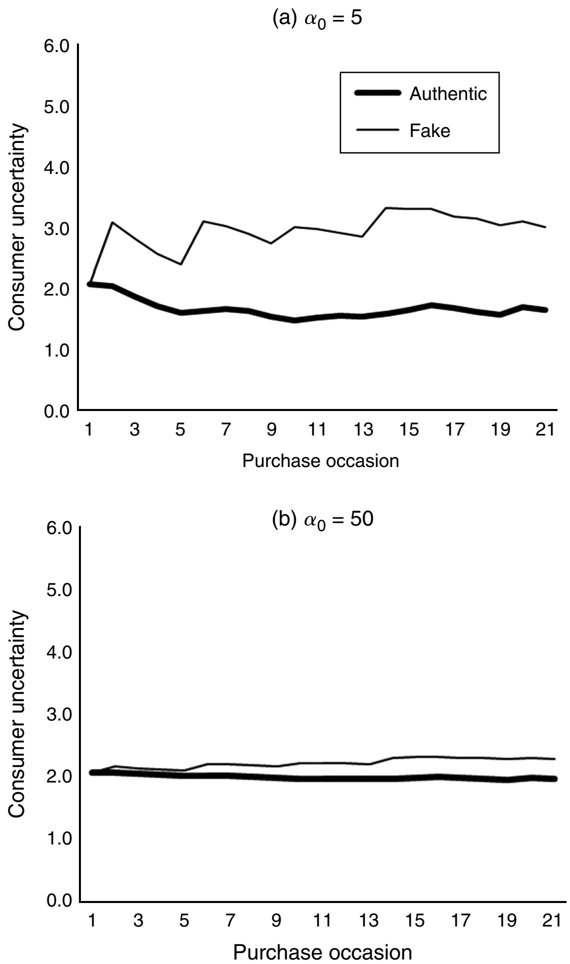
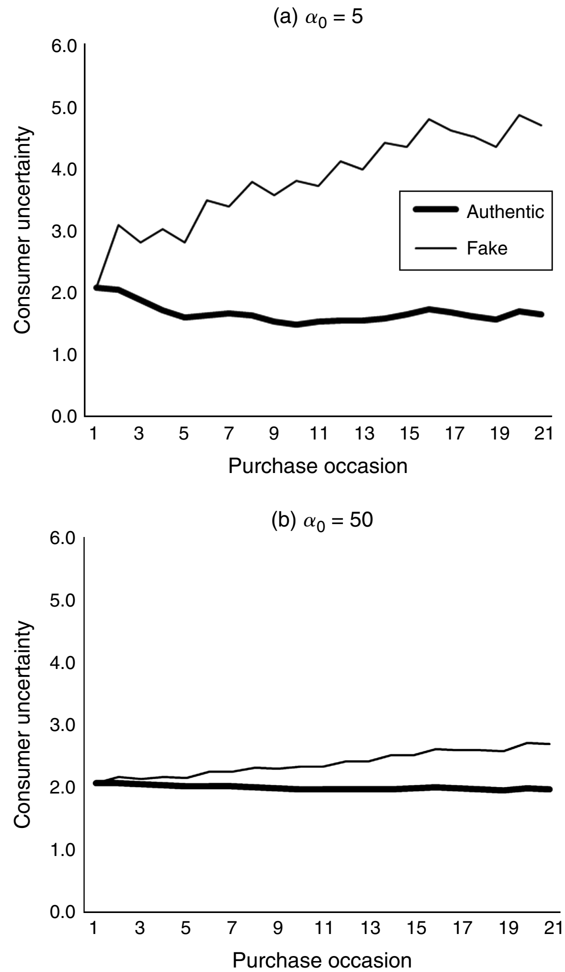


Figure 2 The Effect of Fake Reviews on Consumer Uncertainty (50% Fake Reviews)



all reviews on the fake platforms to be less credible than those on the authentic platform. We compute the relative change in the choice probability[5] of a book when its product ratings are increased by one standard deviation (but everything else is held constant). The consumer is exposed to only authentic reviews for this book on all four platforms. We repeat this exercise for each of the 150 books in our data and then calculate the average. As shown in Table 7, exposure to fake reviews reduces the effect of subsequent reviews, even when the subsequent reviews are authentic. This is because the consumer perceives all reviews on the fake platforms to be less credible, irrespective of whether they are fake or authentic.

We also examine the effect of increased number of reviews of a book on its choice probability on each platform. We perform the same analysis as before, with the exception that instead of increasing the ratings of the book, we increase the *number* of its reviews by one standard deviation. Our findings suggest that fake reviews tend to reduce the effect of more numerous reviews, even when the additional reviews are authentic.

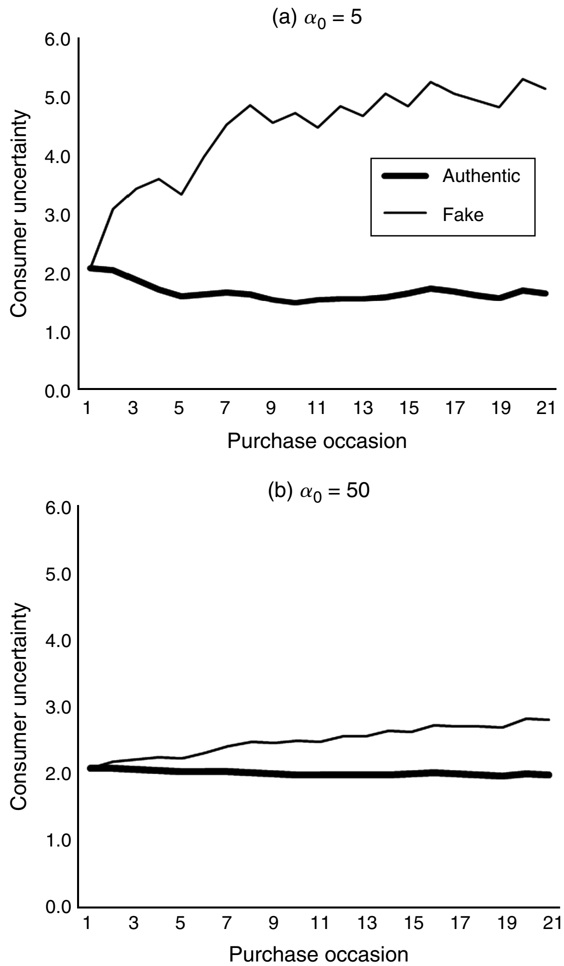
In summary, we find that the marginal effect of more positive ratings on choice probability is lower on platforms where fake reviews are common. The marginal effect of increased number of reviews is also reduced as a result of prior exposure to fake reviews. This illustrates the detrimental effect of the practice of untruthful user-generated content.

6. Conclusion

Marketing researchers have studied how consumers learn about product quality from several stimuli. Structural models of learning have been employed in the context of consumers' usage experience (Roberts and Urban 1988), advertising exposure (Erdem and Keane 1996, Ackerberg 2003, Byzalov and

[5] The choice probability change is calculated as (choice probability under the increased rating – choice probability under the current rating)/(choice probability under the current rating).

Figure 3 **The Effect of Fake Reviews on Consumer Uncertainty (75% Fake Reviews)**



Shachar 2004), umbrella branding (Erdem 1998), physician detailing in the pharmaceutical industry (Narayanan et al. 2005), and learning from observing choices made by other consumers (Zhang 2010). In this research, we propose a structural model of learning from a source of product-related information that is becoming increasingly ubiquitous—online product reviews.

The context of our research is consumer choice of experiential products such as books, music, and movies. A unique characteristic about this phenomenon is that it entails numerous sources of product information. In other words, a consumer is typically exposed to reviews posted by other consumers. It is plausible that the credibility of the reviews evolves over time as the consumer gains more direct experience and has more opportunities to evaluate how well reviews predict her own preference. We propose a very generalized specification of modeling review credibility such that the credibility of product reviews is allowed to vary over time for the same consumer (as she gains more direct experience) and across different consumers. Another feature of experiential goods purchases is limited repeated purchase of the same product item but significant experience with other products of the same type. Accordingly, we model consumer learning from their own experience with other products of the same type.

Table 7 **Average Market Share Change as a Result of Increased Ratings and Increased Numbers of Reviews (in %)**

	Fake review (0% probability)	Fake review (25% probability)	Fake review (50% probability)	Fake review (75% probability)
Increased rating of reviews	8.10	5.75	4.37	3.79
Increased number of reviews	23.11	22.33	19.94	18.46

Our analysis leads to several unique findings. Consumers learn more from reviews of a given product than they do from their own past experience of similar products. Consumers update their beliefs of the credibility of reviews based on their own experiences and ratings from reviews on the same books, so that learning from reviews varies across consumers and also over time. We demonstrate how our model can be used for decisions pertaining to word-of-mouth marketing. Specifically, we compute the profit impact of increasing the number of reviews when firms need to spend marketing resources in incentivizing consumers to post reviews. We find strong evidence of diminishing effects of product reviews on profits; a firm could even incur losses when investing in a sufficiently large number of product reviews for low-margin products. In another policy simulation, we examine the issue of fake reviews. We consider two types of online retailing platforms, one of which has only authentic product reviews (termed authentic), and the other has a possibility of getting fake reviews (termed fake). We find that fake reviews increase consumer uncertainty. The effects of more positive reviews and more numerous reviews on consumer choice are lower on online retailing platforms that have fake product reviews.

This research marks the first attempt to incorporate a novel source of product information into structural models of consumer learning. As such, our findings are neither without limitations nor comprehensive. There are several limitations in this study, suggesting future research opportunities. First, we study a single-attribute learning context (based on genre). However, our model is sufficiently general and can be readily extended to capture learning on multiple attributes (based on genre, author, etc.). Second, we model the difference between the consumer experience signal and the average review rating to be 0. This assumption can be relaxed to specifically model consumer-perceived review bias. Third, our data are not commonly available because they include consumers who evaluate all of their purchases. In the case when only a fraction of consumer purchases are evaluated, we can treat consumer experience signals on those unevaluated products as latent variables and integrate them out in model inference as in a standard learning model. Fourth, despite the associated computational burdens, it might be interesting to study whether consumers adopt forward-looking behavior in the context of the purchase of experiential products. Finally, although we model learning from consumer reviews posted on a major website that hosts such information, the presence of alternative sources of learning, such as advertising, off-line word of mouth, and product reviews from other websites, cannot be ruled out. This issue of incomprehensive data sets is perhaps generic to all research on consumer learning. We hope that this study will stimulate further interest in this challenging, interesting, and increasingly important research area.

Acknowledgments

This paper was originally intended to appear in the special issue on user-generated content, but the review process was not concluded in time. The authors thank the anonymous company that provided data for this study. They are grateful to the editor, associate editor, and two anonymous referees for extremely helpful comments. The authors also thank participants at the Marketing Science conference as well as workshops given at Duke University and the University of Texas at Dallas for their valuable comments. The second author acknowledges the financial support from the National Natural Science Foundation of China [Grants 71128002 and 71210003].

References

Ackerberg DA (2003) Advertising, learning and consumer choice in experience goods markets: An empirical examination. *Internat. Econom. Rev.* 14(3):1007–1040.

Berndt E, Hall B, Hall R, Hausman J (1974) Estimation and inference in nonlinear structural models. *Ann. Econom. Soc. Measurement* 3/4:653–665.

Berry ST (1994) Estimating discrete choice models of product differentiation. *RAND J. Econom.* 25(2):242–262.

Brock T (1965) Communicator-recipient similarity and decision change. *J. Personality Soc. Psych.* 1(6):650–654.

Byzalov D, Shachar R (2004) The risk reduction role of advertising. *Quant. Marketing Econom.* 2(4):283–320.

Chaiken S (1980) Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *J. Personality Soc. Psych.* 39(5):752–766.

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* 43(3):345–354.

Chintagunta PK, Jiang R, Jin GZ (2009) Information, learning, and drug diffusion: The case of Cox-2 inhibitors. *Quant. Marketing Econom.* 7(4):399–343.

Dellarocas C (2006) Strategic manipulation of Internet opinion forums: Implications for consumers and firms. *Management Sci.* 52(10):1577–1593.

Dellarocas C, Zhang M, Awad NF (2007) Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *J. Interactive Marketing* 21(4):23–45.

Duan W, Gu B, Whinston AB (2008) Do online reviews matter?—An empirical investigation of panel data. *Decision Support Systems* 45(4):1007–1016.

Erdem T (1998) An empirical analysis of umbrella branding. *J. Marketing Res.* 34(3):339–351.

Erdem T, Keane MP (1996) Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Sci.* 15(1):1–20.

Erdem T, Zhao Y, Valenzuela A (2004) Performance of store brands: A cross-country analysis of consumer store brand preferences, perceptions, and risk. *J. Marketing Res.* 41(1):86–115.

Godes D, Mayzlin D (2009) Firm-created word-of-mouth communication: Evidence from a field test. *Marketing Sci.* 28(4):721–739.

Godes D, Silva J (2012) Sequential and temporal dynamics of online opinion. *Marketing Sci.* 31(3):448–473.

Hajivassiliou V, Ruud P (1994) Classical estimation methods for LDV models using simulation. Arrow K, McFadden D, eds. *Handbook of Econometrics*, Vol. IV (Elsevier Science, New York), 2383–2441.

Helft M (2010) Charges settled over fake reviews on iTunes. *New York Times* (August 26) http://www.nytimes.com/2010/08/27/technology/27ftc.html.

Iyengar R, Ansari A, Gupta S (2007) A model of consumer learning for service quality and usage. *J. Marketing Res.* 44(4):529–544.

Keane M (1993) Simulation estimation for panel data models with limited dependent variables. Maddala GS, Rao CR, Vinod HD, eds. *Handbook of Statistics* (Elsevier Publishers, New York), 545–571.

Kelman HC (1961) Processes of opinion change. *Public Opinion Quart.* 25(Spring):57–78.

Li X, Hitt LM (2008) Self-selection and information role of online product reviews. *Inform. Systems Res.* 19(4):456–474.

Liu Y (2006) Word of mouth for movies: Its dynamics and impact on box office revenue. *J. Marketing* 70(3):74–89.

Luan JY, Neslin SA (2009) The development and impact of consumer word of mouth in new product diffusion. Working Paper 2009-65, Tuck School of Business at Dartmouth, Hanover, NH. http://ssrn.com/abstract=1462336.

Mayzlin D (2006) Promotional chat on the Internet. *Marketing Sci.* 25(2):155–163.

Mehta N, Rajiv S, Srinivasan K (2003) Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing Sci.* 22(1):58–84.

Meyer RJ, Sathi A (1985) A multiattribute model of consumer choice during product learning. *Marketing Sci.* 4(6):41–61.

Moe W, Schweidel D (2012) Online product opinions: Incidence, evaluation, and evolution. *Marketing Sci.* 31(3):372–386.

Moe W, Trusov M (2011) Measuring the value of social dynamics in online product forums. *J. Marketing Res.* 48(3):444–456.

Narayanan S, Manchanda P (2009) Heterogeneous learning and the targeting of marketing communication for new products. *Marketing Sci.* 28(3):424–441.

Narayanan S, Manchanda P, Chintagunta PK (2005) Temporal differences in the role of marketing communication in new product categories. *J. Marketing Res.* 42(3):278–290.

Roberts JH, Urban G (1988) Modeling multiattribute utility, risk and belief dynamics for new consumer durable brand choice. *Management Sci.* 34(2):167–185.

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

Shin S, Misra S, Horsky D (2012) Disentangling preferences and learning in brand choice models. *Marketing Sci.* 31(1):115–137.

Sternthal B, Phillips LW, Dholakia R (1978) The persuasive effect of source credibility: A situational analysis. *Public Opinion Quart.* 42(3):285–314.

Streitfeld D (2011) In a race to out-rave, 5-star Web reviews go for $5. *New York Times* (August 19) http://www.nytimes.com/2011/08/20/technology/finding-fake-reviews-online.html.

Train K (2003) *Discrete Choice Models with Simulation* (Cambridge University Press, Cambridge, UK).

Zhang J (2010) The sound of silence: Observational learning from the U.S. kidney market. *Marketing Sci.* 29(2):315–335.

Zhao Y, Zhao Y, Helsen K (2011) Consumer learning in a turbulent market environment: Modeling consumer choice dynamics after a product harm crisis. *J. Marketing Res.* 48(2):255–267.