Impact of Prior Reviews on the Subsequent Review Process in Reputation Systems

Xiao Ma
Assistant Professor
C. T. Bauer College of Business
University of Houston
xma@bauer.uh.edu

Lara Khansa *
Associate Dean, Professor
Pamplin College of Business
Virginia Tech
larak@vt.edu

Yun Deng
Ph.D. Candidate
School of Management
Huazhong University of Science and Technology
lucky.deng1122@gmail.com

Sung Kim
Peter T. Allen Professor
Wisconsin School of Business
University of Wisconsin-Madison
skim@bus.wisc.edu

ABSTRACT: Reputation systems have been recognized as particularly successful online review communities and word-of-mouth channels. Our study draws upon the Elaboration Likelihood Model (ELM) to analyze the extent that the characteristics of reviewers and their early reviews reduce or worsen the bias of subsequent online reviews. Investigating the sources of this bias and ways to mitigate it is of considerable importance given the previously established significant impact of online reviews on consumers' purchasing decisions and on businesses' profitability. Based on a panel dataset of 744 individual consumers collected from Yelp.com, we used the Markov Chain Monte Carlo (MCMC) simulation method to develop and empirically test a system of simultaneous models of consumer review behavior. Our results reveal that male reviewers or those who lack experience, geographical mobility, or social connectedness are more prone to being influenced by prior reviews. We also found that longer and more frequent reviews can reduce online reviews' biases. This paper is among the first to examine the moderating effects of reviewer and review characteristics on the relationship between prior reviews and subsequent reviews. Practically, this study offers businesses effective customer relationship management strategies to improve their reputations and expand their clientele.

KEY WORDS AND PHRASES: reputation systems, consumer review, elaboration likelihood model, hierarchical modeling, MCMC simulation, simultaneous equations model

Introduction

The ubiquity and affordability of the Internet and mobile connectivity have contributed to the transformation of the World Wide Web into an interactive medium for connecting people. This seamless connectivity has fueled an online crowd movement, referred to as crowdsourcing [24, 41], that encompasses, among other functions, individuals' willingly rating products and services. These contributions, loosely grouped under the umbrella term "word-of-mouth" (WOM), are defined as any form of "informal communications directed at other consumers about the ownership, usage, or characteristics of particular goods and services and/or their sellers" [62, p. 261]. Electronic WOM (eWOM), a particular WOM phenomenon whereby consumers share their experiences with other consumers online [40], has become "an important source of information to consumers, substituting and complementing other forms of business-to-consumer and offline WOM communication about product quality" [17, p. 345]. Reputation systems, such as Yelp, have been recognized as particularly successful eWOM channels [20, 21, 22, 23, 39, 42]. Yelp, like most online review communities, provides its reviewers with the infrastructure backbone and the technological means to share with others their experiences with a variety of products and services.

Considerable prior research has theorized about and empirically validated the significant effect that online reviews have on consumers' purchasing decisions and on businesses' profitability [11, 25, 60]. Despite their significant influence on sales, some of these online reviews have been plagued by self-selection bias [22, 28, 42, 47]. Li and Hitt [42] explained that early adopters of products are usually those who are most eager to experiment with them, and this eagerness is not truly representative of the market as a whole. This self-selection bias causes early reviews of products and services to be positively biased, thus misleading subsequent customers. As the rating environment matures, online reviews gradually become less positive [29, 46]. Dellarocas et al. [22] documented the presence of this self-selection bias in Yahoo's reviewer ratings during the first weekend of the release of a movie. Understanding the characteristics and determinants of online reviews is therefore imperative to mitigate their bias, especially considering the acclaimed importance of Web 2.0 for online purchases and the high speed with which information travels online. This understanding will also enable practitioners to develop effective strategies that leverage these determinants to boost their revenues and increase their profitability. Our thorough survey of the literature revealed that no prior study has attempted to investigate the online review process and its significant determinants. The purpose of this study is to develop and test a model of the determinants of online reviews in reputation systems and fill this gap in the literature.

In searching for a theoretical underpinning for our effort to understand the mechanics by which prior Yelp reviewers' ratings influence subsequent ones, we found support for our research in the Elaboration Likelihood Model (ELM) [52]. We selected the ELM as the underlying theory for our study for the following two reasons enumerated in [8]: (1) the ELM studies two routes of influence processes, the central route and the peripheral route, and their effects on individuals' perceptions and behaviors, and (2) the ELM captures differential outcomes of these influence processes based on individual and message characteristics. Individuals who scrutinize and critically deliberate on their experiences with products and services are said to choose a central deliberation route. The ratings of these centrally deliberating reviewers are less likely to be swayed by prior reviewers' ratings. On the other hand, individuals who simply rely on cues or on others' judgment to make decisions and who think superficially about their experiences are said to adopt a peripheral deliberation route, or, in other words, they use peripheral cues as heuristics [36, p. 51]. These less involved reviewers are more likely to be influenced by prior reviewers' ratings. Although the ELM illustrates well how personal characteristics can change the decision-making process, it does not discuss how the characteristics of a written message can also affect this process. Practitioners need to consider both individual differences and variations in the written message when characterizing online review behavior. Motivated by this practical need, this study extends the ELM by also investigating the moderating effect of online reviews themselves.

This paper makes several contributions to theory, methodology, and practice. From a theoretical perspective, to the best of our knowledge ours is the first study that has examined the moderating variables that affect the extent to which online reviews are more or less independent of previous ones. This is important, given prior findings that online reviews significantly affect sales and profitability [15, 17, 19, 25, 30, 51, 66]. To support our model and develop our hypotheses, we extended the applicability of the ELM to the context of online reviews, thus paving the way for future research along similar lines. We also extended the ELM by accounting not only for reviewer characteristics but also for the characteristics of the written reviews themselves. Accordingly, we have categorized the moderators between prior reviews and subsequent ones into those that characterize the reviewers and the others that relate to the reviews themselves. In the process, we proposed new variables gauging reviewers' social connectedness, geographical mobility, and time since their last review. Our work also complements prior studies that have investigated the direct effects of reviewers' experience, gender, and the length of their reviews.

From a methodological perspective, we tested our hypotheses using a multilevel model and the Markov Chain Monte Carlo (MCMC) simulation method. This method has several advantages over the previous models used to study consumer review behaviors. First, this is the first time that a multilevel structured model estimated by a MCMC simulation method has been used to analyze consumer review behavior in the Information Systems (IS) field. This model is robustly specified using the ordered probit model. Second, this model effectively corrects for any reviewer self-selection bias, and thus is preferred over the Ordinary Least Squares (OLS) framework because OLS generally produces inconsistent estimations. Third, our model successfully demonstrates the advantage of the hierarchical modeling framework. It predicts ratings correctly more than 50% of the time, which is significantly better than the more common 20% accuracy. Finally, we showed that our model outperformed traditional modeling frameworks by over 18%.

The paper is organized as follows: In the next section, we discuss the ELM as our theoretical framework, elaborate on the rationale for selecting reviewer and review characteristics as moderators of the relationship between the average of prior ratings and subsequent ratings, and develop our hypotheses and model. We then present our data and variables, followed by the modeling of customer rating dynamics, including the rationale for choosing and specifying the models. In the subsequent section, we report our results, model diagnostics, and ad-hoc analyses. We conclude this paper by discussing the significance of the study, its limitations, and potential avenues for further research.

Theoretical Framework and Hypotheses

Figure 1 illustrates our conceptual model that relates consumers' prior expectations of a product or a service, as measured by the average rating of prior reviews, to subsequent ratings and to two types of moderators that affect this relationship: reviewer characteristics and review characteristics. It has been shown that without any other dependable and readily available way to assess a product or a service before consumption, consumers tend to build their expectations on the average rating of prior reviews [1]. These prior expectations serve as a foundation, or level of reference, for postconsumption evaluations. Much empirical research [12, 53] supports such a positive and direct impact of prior expectations on subsequent ones. We use the ELM to explain how the characteristics of the reviewers and their reviews moderate the direct relationship between the average rating of prior reviews and subsequent ones.

------Insert Figure 1------

We propose that reviewer characteristics, namely their prior experience, geographical mobility, social connectedness, and gender are significant moderators of the relationship between prior reviews and subsequent ones. The ELM investigates how individual characteristics, in particular "motivation, ability, personality trait" [57, p. 274], determine the route a person chooses for information processing. Sussman and Siegal [56, p. 50] argued that "different people can be influenced by the same message in different ways." This view is also supported by the psychological choice model [31] in which the relationship between the influencer (average prior reviews) and the response (review rating) is moderated by personal factors (reviewer characteristics). Tam and Ho [57] used the ELM to investigate the effect of personal disposition on people's elaboration of messages and decision outcomes. Even in real life, Web retailers, such as Amazon, and online streaming systems, such as Netflix, have introduced personalized Web portals and movie rating systems to account for unique individual characteristics and predispositions. We posit that experienced and geographically mobile reviewers have the means and knowledge, and thus the ability, to deliberate by using a central deliberation route. Social connectedness is also particularly relevant in an online setting because reviewers who are connected to more friends are more motivated to think centrally within their networks; this is partly because their reputation is at stake in a large social network. We also consider gender as an agglomeration of personality traits and predispositions that are prone to affect reviewers' deliberation route.

We also posit that review characteristics, namely review length and time interval since last review, are important moderators of how prior reviews affect subsequent ones. This effect was suggested by findings in prior literature. For example, Mudambi and Schuff [48] observed a significant effect of review characteristics, including length, on a consumer's perception of the helpfulness of the review. In particular, longer review statements were perceived as more valuable by consumers [9]. Similarly, Forman et al. [26] found that reviewer characteristics, including names and geographical locations, are used by consumers to evaluate the helpfulness of online reviews in making online purchase decisions. Chen et al. [16] found that online reviews with 80% or higher helpfulness votes have a significant impact on consumers' purchasing behaviors and a positive influence on sales. In the realm of the ELM, Sussman and Siegal [56] investigated what specific aspects of a message influence the route that readers use to process it. Angst and Agarwal [4] showed that the quality of messages about the value and safety of electronic health record (EHR) systems can be used with individuals who are highly concerned about privacy to persuade them to adopt EHR systems. Similarly, Bhattacherjee and Sanford [8] showed that the quality of an argument positively affects individuals' attitudes toward IT acceptance. We propose that the length of a review indicates how thoroughly it has been written and thus, directly affects the extent to which a reviewer adopts the central deliberation route. Similarly, the interval since the last review is indicative of a reviewer's activity level and again directly affects his or her need, or lack thereof, to adopt the peripheral deliberation route.

Reviewer Characteristics

In applying the ELM within our context, consumers who have extensive experience are more likely to deliberate internally about their experiences and are less likely to rely on prior reviews when contributing their own. Experienced reviewers know more about the intricacies of the particular product or service, understand their individual expectations better, and are better equipped to think internally about every aspect of their individual experiences. This is supported by findings from the literature. Klein and Ford [38], for instance, found that consumers' online experiences negatively moderate their trust in different information sources. Cheema and Papatla [15] also showed that experienced Internet users are less likely to rely on online information sources. Similarly, Zhu and Zhang [66] found that savvy users of the Internet are less likely to trust online reviews, but novice Internet users are more readily influenced by them. Based on the ELM and on these prior findings, we expect experienced reviewers to truly reflect on their personal experiences and rely less on prior reviews in their rating of products and services. Thus,

H1: *The impact that the average rating of a product manufacturer or service provider in prior periods has on his or her subsequent rating is weaker for more experienced reviewers.*

Geographically mobile reviewers are better able to form their own benchmarks upon evaluating their experiences, and as a result have less need to rely on prior reviews. Geographical mobility could be a precursor or a result of a multitude of cognitive and demographical reviewer characteristics. For instance, prior research [7] found that geographically mobile reviewers had systematically higher incomes, and were more likely to choose expensive organizations (restaurants, hotels, etc…) and avoid traditional chain organizations. They might also have gained more knowledge about specific products and services because of their geographical mobility. Regardless of the specific characteristic that is associated with geographical mobility (higher income, education, knowledge, etc…), geographically mobile reviewers are better able to process information and think for themselves. Further, Wasko and Faraj [61, p. 39] argued that "in the absence of personal acquaintance, similarity, or the likelihood of direct reciprocity," reviewers must be motivated by the expectation of personal benefits. Geographically mobile reviewers might expect to reap the benefits of sharing their candid experiences at a later time when they need others' advice. Given that elaboration "is a function of both ability and motivation" [4, p. 357], we expect that reviewers with higher geographical mobility will tend to adopt a central route of deliberation and are less likely to rely on prior reviews when contributing their own. Therefore,

H2: *The impact that the average rating of a product manufacturer or service provider in prior periods has on his or her subsequent rating is weaker for reviewers with more geographical mobility.*

Consumers connected to a large number of friends tend to rely more on their inner networks and less on total strangers [28, 43]. Strong relationships, such as those with friends, have been shown more likely than weaker ties to be used as sources of information [66]. Kim and Prabhakar [37, p. 540] demonstrated that "if one gets positive WOM referrals on e-commerce from a person with strong personal ties, the consumer may establish higher levels of initial trust in e-commerce." This conjecture is also supported by the herding literature [6] that predicts that members of a herd are more likely to trust each other than trust outsiders. Further, the larger the reviewer's social network, the higher his or her motivation to write a truthful review to preserve his or her social image within the network [50]. Drawing support from the ELM and these prior findings, we propose that consumers with a large network of friends are more likely to adopt a central route of deliberation and rely less on others' reviews. We thus hypothesize that:

H3: *The impact that the average rating of a product manufacturer or service provider in prior periods has on his or her subsequent rating is weaker for reviewers with a large number of friends*.

It has been argued that females "exhibit greater sensitivity to the particulars of relevant information when forming judgments than are males" [45, p. 63] and "engage in more detailed elaboration of specific message content" [45, p. 64] relative to males. Meyers-Levy and Maheswaran [45, p. 65] found that "because females exhibit a greater proclivity to engage in elaborate and detailed message processing, they should be more likely than males to elaborately store message material and employ a detailed strategy at recognition, regardless of whether variations in the extremity of the incongruent cues enhance or inhibit such storage." The authors also found that females are more able to discriminate between congruent and bogus message content than did males [45, p. 68]. On the other hand, men were found to employ a "schema-based" [45, p. 65] strategy that consists of identifying an overall theme and making judgments accordingly. Based on these prior findings, we expect females to elaborate centrally to portray their own experiences as accurately as possible, regardless of how good or bad these experiences were. They are less prone to be swayed, one way or another, by others' reviews. Comparatively, we expect males to rely more on prior reviews to try to detect the overall sentiment from prior reviews before making their own judgment. Therefore,

H4: *The impact that the average rating of a product manufacturer or service provider in prior periods has on his or her subsequent rating is weaker for female reviewers*.

Review Characteristics

Longer reviews allow for information diagnosticity [48] and are both inductive and indicative of a central route of thinking. A larger amount of information and more details about "how and where the product was used in specific contexts" [48] encourage reviewers to think more about the different facets of their experiences and are more likely to induce a central route of deliberation. Longer reviews could also be indicative of a reviewer's state of mind. Disgruntled reviewers write long negative reviews to vent their frustrations [20, 63] and retaliate against the service provider or product manufacturer who disappointed them. Consumers who are not vindictive but rather altruistic also tend to write longer reviews to warn others and spare them the misfortune they have experienced [59]. On the other hand, consumers who had a good experience with a purchase also tend to write longer reviews to express their satisfaction. Regardless of the reason for wanting or needing to write a longer review, consumers think more when they write more and tend to deliberate centrally the more they think. According to the ELM, these consumers are thus expected to rely less on prior reviews when contributing their own. Thus,

> **H5**: *The impact that the average rating of a product manufacturer or service provider in prior periods has on his or her subsequent rating is weaker for longer reviews.*

The time interval between reviews is indicative of consumers' levels of activity and involvement because more active reviewers have shorter waits between reviews because of the frequency of their experiences. Active reviewers, according to the ELM, are more involved, and thus, are more likely to be independent in their ratings and less likely to be swayed by prior reviews. Less active reviewers, however, as quantified by a longer lapse between reviews, deliberate in a more peripheral or "heuristic" fashion [36, p. 51] and thus, are more prone to be affected by prior reviews. Based on the ELM, we conjecture that the effect that prior reviews have on subsequent reviews is stronger for longer time intervals, i.e. for less active reviewers. Active reviewers have more prior information about products and services and devote little time to reading reviews or obtaining product information from others. Consequently, we hypothesize that:

H6: *The impact that the average rating of a product manufacturer or service provider in prior periods has on his or her subsequent rating is stronger for longer time intervals between reviews*.

Data and Variables

Samples and Data Collection

Figure 2 below illustrates our process of sampling and data collection. Given Yelp's lack of publicized Application Programming Interfaces (API), it is not possible to obtain a purely random sample from Yelp.[1] The next best approach is to obtain a random sample from a really large snowball sampling frame of Yelp reviewers. A snowball sampling technique is best suited for collecting a large set of observations that belong to a lesser-known population [10]. To enhance our sample's randomness, we selected seed reviewers from multiple geographical locations [10] and collected as many as 193,889 unique reviewer IDs [5]. Our dataset covers a maximum of almost seven years—from 2004 until January 2011. The female-to-male ratio in the final sample is 1.33:1. Figure 3 plots the distribution of the number of reviews in our sample by a specific reviewer. This log-log distribution is highly consistent with count data that are prone to have a long tail [3].

------ Insert Figures 2 and 3 ------

For estimation purposes, we calibrated our model on the data of those reviewers who had generated at least 10 reviews.[2] After purging ineligible reviewers, we finally obtained a detailed panel-dataset of 61,029 reviews by 744 reviewers.

Variable Descriptions

[1] There is also no publicly available random sample of Yelp reviewers.
[2] Although this step could introduce a sample selection problem, we do not expect its impact on our results to be significant. The total number of reviews excluded because of this step constitutes only about 1.75% of our original review dataset. Moreover, we did additional robustness analyses and found that this step did not affect the quality of our findings.

Average Rating. To measure the average rating ($AvgRate_{jt}$), we used the rounded average rating (rounded to the nearest half-star) publicized by Yelp, instead of calculating the exact average rating of a business at a particular time. This is to ensure that our results will be consistent.[3]

Reviewer Characteristics. To test the effect of cumulative experience, we used the log-transformed cumulative number of reviews by reviewer *i* contributed before time *t* ($CumuExpr_{it}$). Geographical mobility ($GeoMobil_i$) represents the degree of a reviewer's movement in physical locations and thus was computed as the ratio of the number of reviews by reviewer *i* for those businesses located outside reviewer *i*'s home state to the cumulative number of reviews by reviewer *i*. Number of friends ($Friends_i$) is reviewer *i*'s log-transformed number of connected friends within the Yelp community. Finally, we inferred reviewer *i*'s gender ($Gender_i$) by matching his or her first name with thousands of popular names collected from mongabay.com.[4]

Review Characteristics. Text length ($Text_{ijt}$), time interval ($When_{ijt}$), and rating (Y^R or $Rating_{ijt}$) are critical review characteristics. Text length is the log-transformed number of words in a review. Time interval is the log-transformed number of days that elapsed between two consecutive reviews. Rating is simply the numeric rating score.

[3] The rounded average rating allows us to better identify the true effect of this construct, because it is consistent with what was displayed on Yelp. Suppose at t_1 the exact average rating of a business is 3.75, in which case Yelp displays a 4-star overall rating; at t_2 the exact average rating becomes 4.24 because a new 5-star is created. Even though the exact average rating changes by almost 0.5 stars, Yelp will not update the displayed version because it is still in the same range (i.e., between 3.75 and 4.25). According to our theory, under this situation the subsequent rating will not be influenced because reviewers observe the same overall rating of a business on Yelp. If instead we used the exact average rating, any change in the subsequent ratings would have been mistakenly attributed to the change in the average rating, when in fact it should have been attributed to random errors.

[4] If this was unsuccessful, we manually verified the reviewer's gender using his or her photo uploaded onto the profile pages. Besides, we devised a procedure to better detect reviewers' gender. In selecting our random sample, we chose only reviewers for whom a potentially meaningful first name was available, e.g., the first name contained only letters, contained at least three letters, and did not have numbers or unusual characteristics. We believe this procedure did not affect the randomness of our sample, because the percentage of reviewers disqualified by the procedure is negligible. If gender remained unascertainable after matching popular first names and checking profile photos, the name was replaced with the sample average instead of being treated as missing data so as not to reduce the size of our sample. Meanwhile, we conducted a robustness analysis using only those reviewers whose gender was ascertainable; the findings were qualitatively the same as with the full-size sample.

Control Variables. To estimate the influence of the average rating, we needed to control for alternative explanations, i.e., quality heterogeneity among different businesses. Business heterogeneity was controlled for by three components: (1) We used an average price per person to control for the influence of product quality that is reflected by the differentiated price levels ($Price_{jt}$)[42]; (2) We used fixed-effects dummy variables to control for any unobserved business heterogeneity that is time-invariant ($BusFixed_{jt}$) [47]; (3) We further controlled for business heterogeneity using the variance of previous ratings that controls for any other ratings effect ($RatVarnc_{jt}$); the total number of reviews for business *j*, which controls for different levels of popularity ($Reviews_{jt}$); whether business *j* is located in a metropolitan area, which controls for cultural differences ($City_{jt}$); and four dummy variables of the category of business *j*, which control for different service types ($Category_{jt}$). Besides business heterogeneity, time heterogeneity was also accounted for [29]. This was done by including six dummy variables to control for any invariant year effects ($Year_{ijt}$). Moreover, review quality may be indicative of some underlying influence on how that review was generated [16]; thus, we controlled for review quality by the number of usefulness votes a review received ($UseflVot_{ijt}$). Finally, to control for any lagged effect of reviews by the same reviewer, we used the same reviewer's rating at time *t*-1 ($Rating_{i,t-1}$) [44]. Table 1 below summarizes the descriptive statistics of the abovementioned variables.

-----Insert Table 1-----

Modeling Review Behaviors

We specified a multilevel or hierarchical structured model for the review behaviors. A multilevel structure comes natural to our panel data and can be understood as having review activities at the bottom-level unit of analysis (i.e., rating score, review timing, and review length) and reviewer characteristics (i.e., geographical mobility, friends network, and gender) at the top-level unit of analysis [33, 54]. The rationale is that the reviews of one reviewer tend to be dramatically distinct from those of another reviewer, especially in terms of the intra-reviewer correlation [53]. For example, although one reviewer tends to give mostly 4-star ratings, another reviewer may give mostly 2-star ratings. As such, unless a hierarchical structure is incorporated, estimation results are likely to be inconsistent [27], especially for those parameters that represent the effects of variables at one level of analysis on variables at another level (H2 to H4).

In accordance with previous studies, we constructed a simultaneous model [46, 65]. Specifically, a *selection* equation was used to model whether a reviewer would have continued to generate reviews (henceforth Module S) as a way to correct for any reviewer self-selection bias [17, 35]; and a *rating* equation emulated the act of generating the actual numerical ratings (henceforth Module R).[5] We included Module R because testing our hypotheses required constructing a structural model in which the rating decisions are modeled as a function of a set of independent variables. Module S is especially important here to correct for any reviewer self-selection bias [2, 32]. The details of Module R are given below, and those of Module S are included in the Appendix.

Dynamic Decision of Rating—Module R

The most prevalent mechanism for product ratings uses an integer value between 1 and 5 to assess the overall quality of a business [65]. Econometric modelers have argued that such ordinal and censored data require different model specifications from those that are appropriate for normally

---

[5] It is important to note that we do not necessarily assume a specific sequential order that the *selection* and *rating* decisions had to follow. To lower the computational burden of estimating simultaneous models, we used widely accepted simplifying procedures [55, 64]. For example, we estimated Module S and inserted a bias-correction term into Module R. This does not necessarily mean that a reviewer has to first decide whether he or she wants to contribute this review before having a numerical rating in mind; rather, it simply allows a faster and still consistent estimation procedure [64]. We made no assumption about what comes first. In fact, this process could be very complex and highly dependent on the idiosyncrasies of the reviewers, the reviewed businesses, and the review occasions. We thank an anonymous reviewer for reminding us of the risks associated with making untenable assumptions about the sequential order of review activities.

distributed data [13, 34, 64, 65]. Therefore, we followed their ordered probit specification in parameterizing the rating's decision. The unit of analysis is every review of each reviewer. We analyzed it at the review level instead of at the reviewer-firm or reviewer-day levels to maximally account for the dynamics of reviewer behavior. This is because a reviewer conceivably could write multiple reviews for the same firm or on the same day.

Let U_{ijt} denote the latent utility, αs denote the intercepts, βs denote the coefficients for the research and control variables, δs denote the three fixed effects, and k denote a realized value of a rating with $k \in [1,\mathrm{K}]$ where K is the highest rating allowed; κ_1 through κ_K are cutoffs—parameters to help identify intervals for each rank of the ratings. Module R is, therefore, specified as follows:

$$\Pr(\text{Rating}_{ijt} = k) = \Pr(\kappa_{k-1} < U_{ijt} \le \kappa_k)$$
$$\begin{aligned} U_{ijt} = {} & \text{AvgRate}_{jt}(\beta_{0,1} + \beta_{0,2}\text{CumuExpr}_{it} + \beta_{1,1}\text{GeoMobil}_i + \beta_{2,1}\text{Friends}_i + \beta_{3,1}\text{Gender}_i + \beta_{0,3}\text{Text}_{ijt} + \beta_{0,4}\text{When}_{ijt}) \\ & + \beta_{0,5}\text{CumuExpr}_{it} + \beta_{0,6}\text{Text}_{ijt} + \beta_{0,7}\text{When}_{ijt} + \beta_{0,8}\text{Price}_{jt} + \beta_{0,9}\text{RatVarnc}_{jt} + \beta_{0,10}\text{Reviews}_{jt} + \beta_{0,11}\text{City}_{jt} \\ & + \beta_{0,12}\text{Rating}_{i,t-1} + \beta_{0,13}\text{UseflVot}_{ijt} + \delta_j + \delta_{jt} + \delta_{ijt} + \alpha_0 + \alpha_1\text{GeoMobil}_i + \alpha_2\text{Friends}_i + \alpha_3\text{Gender}_i + \lambda\rho + e_{ijt} \end{aligned}$$ [6]

Unobserved Individual Heterogeneity

Controlling for unobserved reviewer heterogeneity is a necessary component of a rigorous econometric model of human and organizational behavior, especially when datasets contain characteristics such as panel-data observations and hierarchical structures. In a panel-data model like ours unobserved heterogeneity mainly takes the form of unobserved effects possibly occurring at different hierarchical levels, i.e., the time series within-reviewer level and the cross-sectional

---

[6] In this equation, $e_{ijt} \sim \mathrm{N}(0,1)$ is required for model identification purposes. To correct for self-selection bias, λ is inserted into this module as the self-selection bias-correction term, and ρ is its coefficient. Thus, a non-zero estimated $\hat{\rho}$ would indicate that the traditional type of self-selection bias exists [32, 64].

individual reviewer level [32].[7] Either a fixed- or random-effect model may be specified at the cross-sectional individual customer level [52]. However, a fixed- or random-effect model only accommodates the unobserved effects in the baseline behavior represented by the intercept term; in other words, if unobserved effects occur in the slope behavior represented by the regression coefficients, as evidenced by much research in the social science field [32, 55], a hierarchical/multilevel model (HLM) is deemed more appropriate and flexible [54]. The appendix elaborates on the relevant details.

Estimation and Results

Estimation Procedure

To consistently estimate our model, we used the Markov Chain Monte Carlo (MCMC) simulation method. We specified a hierarchical model that accommodates the multilevel structure of the consumer review behavior data. Although this hierarchical structured model enables us to incorporate a sophisticated amount of reviewer heterogeneity through various random effects, it nevertheless requires special care in estimation. Due to the high dimensionality of the model, standard estimation procedures, such as OLS, or Maximum Likelihood Estimator, are either inadequate to produce final estimates or infeasible to complete the estimation in a reasonable amount of time. Comparatively, simulation is a practical alternative proven to produce consistent estimates with a sufficiently large number of draws [58]. Key to our model of consumer reviews is the ordered probit formulation that requires estimation of a set of nonlinear cutoff points. This motivated us to use the flexible MCMC simulation that is especially suited to our empirical needs [58]. More detailed explanations of the estimation procedure are available in the appendix.

[7] We also realized the importance of controlling for product heterogeneity and time heterogeneity. Thus, besides reviewer heterogeneity, our modeling approach also included dummy variables to control for product and time heterogeneity.

Model Diagnostics: Reviewer Heterogeneity

To assess and illustrate the validity of incorporating reviewer heterogeneity in the parameter estimation, we analyzed an additional model as the traditional pooled regression. The comparison of results between this and our final model is listed in Table 2, which clearly demonstrates the superiority of the hierarchical modeling framework over the traditional method. The hierarchical modeling framework not only decreases the number of variables necessary in the time-series estimation procedure, but it is also able to yield a much larger log-likelihood (Δ =(-64,412) – (-78,800) = 14,388, a 18.3% boost) and hence a much better fit with the dataset. Moreover, both Mean Absolute Deviance (MAD) and Root Mean Square Error (RMSE) are lower, and Spearman's correlation between predicted and actual outcomes is higher in the hierarchical model. Finally, the hierarchical modeling framework significantly boosts the hit rate, which is the most important predictive performance criterion of a discrete response model.

-----Insert Table 2-----

Tests of Hypotheses

Table 3 presents the estimation results of our hierarchical model. Although the relationship between the average rating of a product or a service and a subsequent rating is not formally hypothesized, we begin by interpreting this result because it constitutes the main effect upon which all of our hypotheses are based. As previously noted, we expected that a subsequent rating relates positively with the average rating of a business. Our findings support this expectation ($\beta_{0,1}$ = 2.498, *se* = 0.044, $p < .001$). To better understand its practical meaning, we calculated the marginal effect of the average rating by using the formula in [64, p. 506]. On average, increasing the average rating of a business from 4 stars to 5 stars will lead 30.5% more reviewers to subsequently follow and give the same 5-star rating.

-----Insert Table 3-----

We used the parameter estimates shown in Table 3 to test our hypotheses. H1 predicts a weaker relationship between a previous average rating and a subsequent rating by more experienced reviewers. H1 is supported ($\beta_{0,2}$ = -1.040, *se* = 0.064, $p < .001$). The result of marginal effects implies that when cumulative experiences increase by 1 unit (after log transformation), an increase in the average rating will lead only 27.4% more reviewers to follow the average rating, a 3.1% drop compared with the above 30.5% base level. Thus, the subsequent rating is less sensitive to the average rating.

H2 postulates a weaker relationship between a previous average rating and a subsequent rating by more geographically mobile reviewers, and this is supported ($\beta_{1,1}$ = -0.172, *se* = 0.023, $p <$ 0.001). For those reviewers with a 1-unit higher level of mobility, an increase in the average rating will lead only 26.2% more reviewers to follow the average rating, a 4.3% drop compared with the 30.5% base level.

H3 predicts that the relationship between a previous average rating and a subsequent rating weakens for more connected reviewers. H3 is supported ($\beta_{2,1}$ = -0.646, *se* = 0.005, $p < 0.001$). It means that having 1 more unit of friends (after log transformation) and being able to communicate with others about local businesses significantly lessens the sensitivity of the subsequent rating to changes in the average rating: An increase in the average rating will lead only 14.7% more reviewers to follow the average rating, a dramatic 15.8% drop compared with the 30.5% base level.

H4 states that the relationship between a previous average rating and a subsequent rating weakens for female reviewers, which is supported ($\beta_{3,1}$ = -0.087, *se* = 0.017, $p < 0.001$). The result of marginal effects reveals that an increase in the average rating will lead 28.1% more female reviewers to follow the average rating, about 2.4% less than for males.

H5 posits a weaker relationship between a previous average rating and a subsequent rating if a review is longer; H5 is supported ($\beta_{0,3}$ = -0.130, *se* = 0.034, $p < 0.001$). This indicates that when reviewers write a review about 1 unit longer (after log transformation), an increase in the average rating will lead only 25.9% more reviewers to follow the average rating, a 4.6% drop compared with the 30.5% base level.

H6, the last of our hypotheses, postulates that the relationship between a previous average rating and a subsequent rating is stronger for a longer interval between reviews ($\beta_{0,4}$ = 0.360, *se* = 0.053, $p < 0.001$); H6 is also supported. This result means that if review incidents are 1 unit further apart in time (after log transformation), reviewers in their own reviews of a business are more reliant (at that moment) on the average rating created by other reviewers. An increase in the average rating will lead as many as 40.9% more reviewers to follow the average rating, a 10.4% increase compared with the 30.5% base level.

Robustness Analyses

Our findings show consistency despite our smaller sample after excluding reviewers with insufficient reviews. To test for this consistency, we analyzed both the original and reduced datasets using a simple Ordered Probit regression model. Results are presented in Table 4, where Model (1) used the original sample and Model (2) the reduced sample. The results of the hypothesized effects are highly stable across the two samples; this is because the number of excluded reviewers constituted only about 1.75% of the original dataset.

------Insert Table 4------

We also used the reduced data sample to compare the hypothesized results under both the simple regression model and the hierarchical model. Our findings were shown to be generally robust. In Table 4, Model (2) is as described above, and Model (3) shows the same final estimation results as in Table 3. The results under starkly different modeling approaches remain largely consistent. The main effect of the average rating remained highly significant. The moderation effects of $CumuExpr_{it}$, $GeoMobil_i$, $Text_{ijt}$, and $When_{ijt}$ are consistent with our predictions. Meanwhile, the moderation effects of $Friends_i$ and $Gender_i$ are not significant in the simple regression model but are significant in the hierarchical model. Therefore, the effects of both review characteristics remained significant across models, and two of the four effects of reviewer characteristics remained significant. Note that the marginal effects under the simple regression model shrank significantly in magnitude. The main reason for this reduction is that a simple regression model is incapable of accounting for the hierarchical structure of the data and incapable of precisely estimating individualized effect sizes, given that different amounts of data are available for the reviewers. This may also be the reason why the simple model could not effectively estimate a significant effect for $Friends_i$ and $Gender_i$.

We empirically checked whether there existed selective review patterns in our data that could have led to the results we found, such that the weakened relationship between previous average ratings and subsequent ratings for some segments of reviewers (i.e., more experienced, more mobile, more connected, and female reviewers) selectively reviewing products.[8] After analyzing several robustness regression models, we did not find any significant evidence that these scenarios existed in our data. We conclude that our results are robust.

Discussion

Our goal in this study was to examine the moderating role of the characteristics of online reviewers and their reviews on subsequent reviews. We proposed a theoretical framework grounded in the ELM and tested it against data collected from 744 Yelp reviewers. Our findings indicate that

[8] Those chain stores and stores that had received relatively few reviews by the time a review was generated are expected to exhibit a weaker relationship between previous average ratings and subsequent ratings because the reviews could be highly noisy. We thank an anonymous reviewer for pointing out these alternative explanations of our results.

reviewer characteristics (i.e., experience, geographical mobility, social connectedness, and gender), and review characteristics (review length and time interval since last review) significantly moderate the extent to which consumers' online ratings are biased by earlier ones. Based on the ELM, we found that reviewers who take a central deliberation route painstakingly process their personal experiences with products and services, so they are less prone to rely on others' reviews while rating their own experiences. Specifically, we found that reviewers with more experience, higher geographical mobility, and a larger number of friends, and female raters rely less on prior reviews in the course of contributing their own ratings. Similarly longer reviews were found to negatively moderate the relationship between prior ratings and subsequent ones. On the other hand, the longer the time interval since their last review, the more biased online raters are by earlier reviews and ratings.

Theoretical Contributions

Our study offers several contributions to theory. First, despite the strong evidence of the significant effect of online reviews on the sales and profitability of businesses [15, 17, 19, 25, 30, 51, 66], prior research has not adequately investigated the antecedents of online reviews. Understanding how online reviews are affected by previous ones is crucially important to gauging the extent of their independence or bias, especially given the significant effect that online reviews have on sales. This study, through its investigation of the moderating effects that reviewer and review characteristics have on the relationship between prior online reviews and subsequent ones, constitutes a step toward fulfilling this void in the literature.

Second, we drew theoretical support from the ELM to explain the observed differences in the extent to which online reviews are independent of earlier ones. By adopting the ELM as the underlying theoretical framework for our study, we have provided empirical evidence that its applicability extends to the context of online consumer reviews, and we have as such paved the way for subsequent research along those lines. Further, previous research using the ELM focused on the characteristics of the individual (equivalent to online reviewers in our context) or the message (equivalent to prior reviews in our context) [4, 8, 56]. We have extended the ELM by incorporating the message characteristics of subsequent reviews (rather than those of prior reviews) as moderators of review bias. This is especially relevant because review characteristics have been shown to significantly affect the ratings of online reviews and because it would be counterintuitive to account only for reviewer characteristics and ignore the attributes of the reviews they wrote. Reviews are written by reviewers; therefore, it would be arbitrary to only consider the characteristics of reviewers without taking into account the characteristics of their reviews.

Third, consistent with the ELM literature, we have divided the moderators between prior reviews and subsequent ones into two categories: (1) reviewer characteristics (conceptualized as number of friends, geographical mobility, cumulative experience, and gender), and (2) review characteristics (conceptualized as time interval between reviews and review length). Our study offers several important theoretical contributions within each of these categories.

This study is the first to demonstrate the significance of reviewers' social connectedness and geographical mobility in the context of online reviews and then to extend to that context the study of the moderating role of reviewers' experience and gender. We are the first to show that those reviewers in our sample with the largest number of friends rely less on the aggregate information of prior reviews and contribute more independent reviews as a result. These observations are enlightening because they contradict normative findings from prior literature [6] that were obtained through analytical modeling and that suggested that consumers fare better by following the crowd than by seeking private information from their own networks. Our findings are in line with the literature on peer influences and with the bandwagon effect, where it has been found that in both cases consumers adopt the preferences of their group affiliation [49]. Our study is also the first to examine the moderating effect of geographical mobility in the context of online reviews. We showed that in contributing their own reviews, reviewers with greater geographical mobility rely less on prior reviews. This seems contradictory to the common belief that because time is money, geographically mobile consumers will not take time to write reviews and to save time would rather check what others have written. Instead, our results show that geographically mobile reviewers contribute independent reviews. They do this, perhaps, because they appreciate online reviews and have relied on them themselves, all the more so because of their mobile life, before selecting a product or a service. We also found that the more experienced the reviewer, the less they tend to rely on prior reviews. This is in line with the results in prior literature [14, 15]. For example, Chandy et al. [14] found that consumers who are knowledgeable about an advertised product do not take its advertising into consideration. The final moderating reviewer characteristic that we studied was gender. Consistent with the results in prior literature, we found that women are more independent in their reviews because they rely less on what others have written. Men, on the other hand, tend to be influenced by prior reviews.

This study is also the first to investigate the moderating effect of time since last review on the extent that online reviews depend on previous ones; at the same time, it also complements prior studies that have only investigated the direct effect of review length on ratings. We are the first to show that all things being equal, and specifically without regard for a reviewer's cumulative experience, the shorter the time since the last review, the more the review is independent of prior reviews. This is an interesting result that reveals the same tendency toward short-term memory among online reviewers that has already been established in the marketing literature. Our findings also reveal that the longer the time elapsed since their last review, the more reviewers tend to rely on others' prior reviews and the less their reviews tend to be independent. We have also shown that lengthier reviews make reviewers rely less on prior reviews and consequently cause them to contribute more independent reviews. Besides being the first to highlight the significantly negative moderating effect of review length on online reviews, our findings provide theoretical support for prior studies that have linked review length to more extreme ratings [20, 63] and higher helpfulness ratings [48]. As reviewers think more deliberately, they delve more deeply into the details of their experiences, and they therefore tend to be either more satisfied or dissatisfied, which could explain the extreme ratings found in longer reviews. Similarly, longer reviews encourage their reviewers to contribute reviews that are more transparent and truthful and as a result are perceived as more helpful.

Methodological Contributions

We adopted a relatively underexploited modeling approach to deal with the challenge in this study and showcased its implementation in the context of online product reviews. In comparison, most of the previous studies of online product reviews have used OLS [18, 39], OLS with random effects (i.e. [23]), or OLS with fixed effects [25, 42]. Because each of these traditional methods has its intrinsic limitation, our modeling approach provides at least an alternative way of examining the problem of interest. As we have shown in Table 4, our approach produces robust estimation of key parameters at least in the current context. Thus, the hierarchical modeling that we used provides a significant addition to the literature of online product reviews, and possibly any other research questions that involve a panel data structure.

Our model innovatively solves for the simultaneous decisions involved in self-selection and in ratings. Our approach takes into account the issues of biased sampling methods and inconsistent estimation; by comparison, many other studies simply assumed their samples were unbiased [18, 23, 25, 39]. Our methodology showcases an analytical framework that can be easily extended and applied to other questions regarding online users' activities, such as in those of question-and-answer communities.

Our findings because of this methodology are more generalizable than those of previous research. The novel dataset used in this study not only includes the complete sequences of product reviews by hundreds of reviewers, but also links each of these reviews with characteristics of products, services, reviewers, and reviews that vary by time. Moreover, the dataset enables us to study product review behavior pertaining to a variety of businesses; this is in contrast to previous studies that used review data for only one type of product, such as movies [23, 25, 39, 65], books (e.g., [42]), or craft beer (e.g., [18]).

Finally, our model correctly predicted over half of the ratings, compared with a 20% chance of random predictive accuracy. It also outperforms traditional models by over 18%. Therefore, our model is highly robust, and the results of our hypotheses testing are efficient. The results convey inferences about reviewer preferences that can help business owners more effectively customize their marketing strategies.

Practical Implications

As mentioned earlier, our major discovery is the measurable conditions under which subsequent consumer reviews are influenced by a product's or service's average rating. For business owners who strive to improve their online reputation as well as enhance their customers' purchasing experiences, our findings are promising and provide guidelines to implement in their marketing, especially in customer relationship management (CRM) strategies. Some customers, or customers under some conditions, are more prone to trust other "stranger" customers. Business owners should base their marketing campaigns on the right target group of customers according to the status of their business's online reputation. If a business enjoys overall favorable acknowledgement in major online reputation systems and thus has a higher average rating, the owner should broaden its marketing scope and meanwhile, focus more on those inexperienced reviewers who tend to be in restricted geographical areas, have smaller personal networks, generally have less to say in online reviews or have recently tended to be inactive consumers. Marketers may easily identify and further target them within local neighborhoods. Businesses in this "favored" situation could benefit from trying to communicate and market to those less informed people in nearby towns and by sending them coupons. When communicating with this group of potential customers, marketers should consider their unique characteristics and provide critical information (e.g., how to access local libraries with Internet devices) to make them informed of major reputation systems they may use later to create a positive review.

In contrast, if a business suffers because earlier customers have criticized it on online review websites (a "less favored" situation), marketers in this case should narrow their focus, bringing it to bear on those who favor their products or services and who tend to be "independent" and active reviewers. Because lower cost and higher satisfaction lead to favorable reviews, providing discounts to already interested customers will enhance their satisfaction even more and thereby quickly turn the situation around. After its online reputation rating goes above average, the owner of a business in this situation might consider switching to a marketing strategy suited for the "favored" situation, as explained above. This study has also demonstrated, perhaps counterintuitively, that even when online reviewers are predisposed to being swayed by prior online reviews because of some reviewer-specific characteristics such as gender, lack of experience, inadequate geographical mobility, or mediocre social connectedness, these biases may be counterbalanced by the characteristics of the reviewers' reviews such as review length and frequency. Conversely, even when reviews are not sufficiently long or frequent, it might still be possible for reviewers to contribute independent reviews if they are females or if they are sufficiently experienced, mobile, or socially connected to friends.

From the perspective of customer welfare, however, reliance on prior reviews with average ratings worsens self-selection bias [42]. The objective of reputation systems like the one examined in this study is to accurately depict reality as a way to convey to their users their best advice, thus assisting them in achieving their objectives and meeting their demands. This implies that a reputation system should attempt to filter out or at least highlight potential biases to help others make better decisions. Currently Yelp provides snapshots of each of its reviewers that are indicative of their characteristics, including experience, geographical mobility, number of friends, and gender. Given the significant effects found in our model, it would be beneficial for reputation systems to create indices for each of the significant factors affecting online reviews and provide them to their users so that they can make more informed decisions. Reputation systems could enhance their benefits by using reliability scores to rank the reviewers themselves. Doing so would encourage reviewers to filter out their biases and put more effort into drawing a true picture of their experiences.

Limitations and Future Research

Our study has several possible limitations. First, the way we measured the cross-sectional reviewer characteristics limited our ability to rule out alternative explanations. For example, although it is reasonable to argue that having more "Yelp friends" tends to motivate a reviewer to elaborate on his or her own needs and experiences, being wealthy may possibly lead a reviewer to be more successful and hence have more friends. Those who review more out-of-state businesses travel more and thus are potentially wealthier. In other words, income, rather than geographical mobility or a network of friends, may be the ultimate unobserved driver of central deliberation. Moreover, females may have systematically different tastes for services than males, creating another consumer selection issue.[9] Based on Table 1, the correlations between geographical mobility and number of friends are negative and not statistically significant (ρ = -0.01, p = 0.2156). This result implies a lack of a strong

[9] We thank an anonymous reviewer for bringing these alternative explanations to our attention.

common alternative factor to explain the cross-sectional effects of both geographical mobility and networks of friends. Nonetheless, we acknowledge that our theoretical arguments may not be the only explanation of our findings on these cross-sectional reviewer characteristics. Second, although our model offers an integrated framework that explains many factors relevant to the information processing stage of online reviews, it does not account for many other characteristics that have been found to affect online reviews. For instance, Zhu and Zhang [66] differentiated between "experience goods" and "search goods" when studying the influence of online reviews on sales. Until all these products, reviewers, and review characteristics are accounted for and examined in future research, our findings should not be generalized beyond their original intent. Third, although we took several steps to enhance the randomness of our sample, which appears reasonably representative of the population, our sample is still not strictly random. Moreover, our data were collected from a single website, which limits the generalizability of our findings. Thus, readers need to approach our findings with some caution. Fourth, "number of friends" is a time-invariant variable because we could only track each individual in our sample for a short time (two months) and hence, could not recover the historical time-variant changes before that time window.

We foresee many avenues for future research. First, the significant impact of social connectedness suggests that social networks can influence online reviewers' decision making and possibly sway their ratings. Still to be determined is how the social status of various reviewers, i.e. their centrality with respect to their social network, affects the importance that other online reviewers attribute to their reviews. Second, we have assumed that reviewers' responses to others' reviews are independent of the technologies they use. It would be interesting to test this assumption and investigate the marketing implications of those reviewers who use handheld devices (e.g., smart phones) to read reviews and contribute their own, because these reviewers may either have more time to digest these reviews and correct for their bias or they may quickly tire of reading and scrolling between dozens of pages of reviews. Third, the implications of review texts remain underexplored, even though data- and text-mining have recently emerged as a hot topic. Future research may attempt to directly identify those reviewers who tend to rely on previous reviews or ratings instead of inferring who they are. One way to do so is to parse the lines of their reviews and pinpoint what aspect of a reviewer's review tends to be influenced by the reviews of others. In addition, future research may also reveal practical insights if researchers try to more accurately identify those reviewers most likely to influence the ratings of newcomer reviewers. Fourth, our study assumes that a review is generated immediately after a purchase. If the length of the postpurchase delay before a review is written possibly influences how reviewers recall and assess an overall purchase experience, our results could be altered for not only the effect of the time interval but also for other effects. More powerful text-mining techniques may soon be used to first infer how much time has elapsed between a purchase and the corresponding review and then assess its implications for our findings.

Conclusions

Reputation systems have metamorphosed into an essential source of online reviews of products and services. Drawing on the ELM, we dissected factors that directly determine and moderate online review decisions. Through analyzing system-generated data, we were able to identify characteristics of reviewers and their reviews that significantly moderate the biases of online reviews. Overall, our results reveal that online reviews are not only determined by economic factors (i.e. quality and price) but also by a multitude of reviewer and review characteristics often ignored in prior research. We hope that this paper will stimulate constructive ideas for improvement for reputation systems' decision makers and motivate companies to leverage this free information to better satisfy their customers.

References

1. Acemoglu, D., and Ozdaglar, A. Opinion dynamics and learning in social networks. *Dynamic Games and Applications, 1,* 1 (2010), 3–49.
2. Amemiya, T. *Advanced Econometrics*. Cambridge, MA: Harvard University Press, 1985.
3. Anderson, C. *The Long Tail: Why the Future of Business is Selling Less of More*. New York, NY: Hyperion Books, 2006.
4. Angst, C.M., and Agarwal, R. Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion. *MIS Quarterly, 33,* 2 (2009), 339–370.
5. Atkinson, R., and Flint, J. Accessing hidden and hard-to-reach populations: Snowball research strategies. *Social Research Update, 33*, 1 (2001).
6. Banerjee, A.V. A simple model of herd behavior. *Quarterly Journal of Economics, 107,* 3 (1992), 797–817.
7. Bell, J.E. Mobiles: A neglected market. *Journal of Marketing*, *33*, 2 (1969), 37–44.
8. Bhattacherjee, A., and Sanford, C. Influence processes for information technology acceptance: An elaboration likelihood model. *MIS Quarterly, 30,* 4 (2006), 805–825.
9. Schindler, R.M., and Bickart, B.A. Perceived helpfulness of online consumer reviews: The role of message content and style. *Journal of Consumer Behavior, 11*, 3 (2012), 234–243.
10. Blanken, P., Hendricks, V.M., and Adriaans, N.F.P. Snowball sampling: Methodological analysis? In V.M. Hendricks, P. Blanken, and N.F.P. Adriaans (eds.), *Snowball Sampling: A Pilot Study on Cocaine Use*. Rotterdam, the Netherlands: Instituut Voor Verslavingsonderzoek (IVO), 1992, pp. 83–100.
11. Bolton, G.E., Katok, E., and Ockenfels, A. How effective are electronic reputation mechanisms? An experimental investigation. *Management Science, 50*, 11 (2004), 1587–1602.
12. Brown, S.A., Venkatesh, V., and Goyal, S. Expectation confirmation in technology use. *Information Systems Research, 23*, 2 (2012), 474–487.
13. Cameron, A.C., and Trivedi, P.K. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press, 2005.
14. Chandy, R., Tellis, G., Macinnis, D., and Thaivanich, P. What to say when: Advertising appeals in evolving markets. *Journal of Marketing Research, 38,* 4 (2001), 399–414.
15. Cheema, A., and Papatla, P. Relative importance of online versus offline information for Internet purchases: Product category and Internet experience effects. *Journal of Business Research, 63,* 9-10 (2010), 979–985.
16. Chen, P., Dhanasobhon, S., and Smith, M. All reviews are not created equal: The disaggregate impact of reviews on sales on Amazon.com. Working paper, Carnegie Mellon University, 2008 (available at http://ssrn.com/abstract=918083).
17. Chevalier, J.A., and Mayzlin, D. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research, 43*, 3 (2006), 345–354.
18. Clemons, E.K., Gao, G., and Hitt, L.M. When online reviews meet hyperdifferentiation: A Study of the craft beer industry. *Journal of Management Information Systems, 23,* 2 (2006), 149–171.
19. Cui, G., Lui, H-K., and Guo, X. The effect of online consumer reviews on new product sales. *International Journal of Electronic Commerce, 17*, 1 (2012), 39–57.
20. de Matos, C.A., and Rossi, C.A.V. Word-of-mouth communications in marketing: A meta-analytic review of the antecedents and moderators. *Journal of the Academy of Marketing Science, 36*, 4 (2008), 578–596.
21. Dellarocas, C. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science, 49*, 10 (2003), 1407–1424.
22. Dellarocas, C., Awad, N.F., and Zhang, X. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing, 21,* 4 (2007), 23–45.
23. Dellarocas, C., Gao, G., and Narayan, R. Are consumers more likely to contribute online reviews for hit or niche products? *Journal of Management Information Systems, 27*, 2 (2010), 127–157.
24. Doan, A., Ramakrishnan, R., and Halevy, A.Y. Crowdsourcing systems on the World Wide Web. *Communications of the ACM, 54*, 4 (2011), 86.
25. Duan, W., Gu, B., and Whinston, A.B. Do online reviews matter? An empirical investigation of panel data. *Decision Support Systems 45*, 4 (2008), 1007–1016.

26. Forman, C., Ghose, A., and Wiesenfeld, B. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, *19*, 3 (2008) 291–313.
27. Gelman, A., and Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press, 2006.
28. Ghose, A.S., and Han, S. An empirical analysis of user content generation and usage behavior on the mobile Internet. *Management Science*, 57, 9 (2011), 671–1691.
29. Godes, D., and Silva, J.C. Sequential and temporal dynamics of online opinion. *Marketing Science 31,* 3 (2012), 448–473.
30. Gu, B., Park, J., and Konana, P. Research note: The impact of external word-of-mouth sources on retailer sales of high-involvement products. *Information Systems Research*, *23*, 1 (2012), 146–163.
31. Hansen, F. Psychological theories of consumer choice. *Journal of Consumer Research 3*, 3 (1976), 117–142.
32. Heckman, J.J. Sample selection bias as a specification error. *Econometrica*, *47*, 1 (1979), 153–161.
33. Hofmann, D.A. An overview of the logic and rationale of hierarchical linear models. *Journal of Management, 23*, 6 (1997), 723–744.
34. Hsiao, C. *Analysis of Panel Data*. New York: Cambridge University Press, 2003.
35. Hu, N., Zhang, J., and Pavlou, P.A. Overcoming the J-shaped distribution of product reviews. *Communications of the ACM, 52*, 10 (2009), 144–147.
36. Kim, D., and Benbasat, I. Trust-related arguments in Internet stores: A framework for evaluation. *Journal of Electronic Commerce Research, 4*, 2 (2003), 49–63.
37. Kim, K., and Prabhakar, B. Initial trust, perceived risk, and the adoption of Internet banking. In *Proceedings of the 21st International Conference on Information Systems*, Brisbane, Australia, 2000.
38. Klein, L.R., and Ford, G.T. Consumer search for information in the digital age: An empirical study of pre-purchase search for automobiles. *Journal of Interactive Marketing, 17*, 3 (2003), 29–49.
39. Koh, N.S., Hu, N., and Clemons, E.K. Do online reviews reflect a product's true perceived quality? An investigation of online movie reviews across cultures. *Electronic Commerce Research & Applications, 9*, 5 (2010), 374–385.
40. Kumar, N., and Benbasat, I. Research note: The influence of recommendations and consumer reviews on evaluations of websites. *Information Systems Research*, *17*, 4 (2006), 425–439.
41. Leimeister, J.M., Huber, M., Bretschneider, U., and Krcmar, H. Leveraging crowdsourcing: Activation-supporting components for IT-based ideas competition. *Journal of Management Information Systems, 26*, 1 (2009), 197–224.
42. Li, X., and Hitt, L.M. Self-selection and information role of online product reviews. *Information Systems Research, 19*, 4 (2008), 456–474.
43. Lim, K.H., Sia, C.L., Lee, M.K.O., and Benbasat, I. Do I trust you online, and if so, will I buy? An empirical study of two trust-building strategies. *Journal of Management Information Systems, 23,* 2 (2006), 233–266.
44. Ma, X., and Kim, S.S. Revisiting self-selection biases in e-word-of-mouth: An integrated model and Bayesian estimation of multivariate review behaviors. In *Proceedings of the 32nd International Conference on Information Systems.* Shanghai, China: AIS Electronic Library, 2011.
45. Meyers-Levy, J., and Maheswaran, D. Exploring differences in males' and females' processing strategies. *Journal of Consumer Research, 18*, 1 (1991), 63–70.
46. Moe, W.W., and Schweidel, D.A. Online product opinions: Incidence, evaluation and evolution. *Marketing Science, 31*, 3 (2012), 372–386.
47. Moe, W.W., and Trusov, M. The value of social dynamics in online product ratings forums. *Journal of Marketing Research, 48*, 3 (2011), 444–456.
48. Mudambi, S.M., and Schuff, D. What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly, 34*, 1 (2010), 185–200.
49. Nikolaeva, R., and Sriram, S. The moderating role of consumer and product characteristics on the value of customized on-line recommendations. *International Journal of Electronic Commerce, 11,* 2 (Winter 2006–7), 101–123.
50. Park, Y.-H., and Bradlow, E.T. An integrated model for bidding behavior in Internet auctions: Whether, who, when, and how much. *Journal of Marketing Research, 42*, 4 (2005), 470–482.

51. Pathak, B., Garfinkel, R., Gopal, R.D., Venkatesan, R., and Yin, F. Empirical analysis of the impact of recommender systems on sales. *Journal of Management Information Systems*, *27,* 2 (2010), 159–188.
52. Petty, R.E., Cacioppo, J.T., and Schumann, D. Central and peripheral routes to advertising effectiveness: The moderating role of involvement. *Journal of Consumer Research, 10*, 2 (1983), 135–146.
53. Phillips, D.M., and Baumgartner, H. The role of consumption emotions in the satisfaction response. *Journal of Consumer Psychology, 12*, 3 (2002), 243–252.
54. Raudenbush, S.W., and Bryk, A.S. *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, CA: Sage Publications, 2002.
55. Rossi, P.E., Allenby, G.M., and McCulloch, R. *Bayesian Statistics and Marketing*. The Atrium, Southern Gate, Chichester, UK: Wiley, 2006.
56. Sussman, S.W., and Siegal, W.S. Information influence in organizations: An integrated approach to knowledge adoption. *Information Systems Research, 14*, 1 (2003), 47–65.
57. Tam, K.Y., and Ho, S.Y. Web personalization as a persuasion strategy: An elaboration likelihood model perspective. *Information Systems Research, 16*, 3 (2005), 271–291.
58. Train, K.E. *Discrete Choice Methods with Simulation* (2nd ed.). New York, NY: Cambridge University Press, 2009.
59. Walsh, G., Gwinner, K.P., and Swanson, S.R. What makes mavens tick? Exploring the motives of market mavens' initiation of information diffusion. *Journal of Consumer Marketing, 21*, 2/3 (2004), 109–122.
60. Wangenheim, F., and Bayón, T. The chain from customer satisfaction via word-of-mouth referrals to new customer acquisition. *Journal of the Academy of Marketing Science, 35* (Summer 2007), 233–249.
61. Wasko, M.M., and Faraj, S. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly, 29,* 1 (2005), 36–67.
62. Westbrook, R.A. Product/Consumption-based affective responses and postpurchase processes. *Journal of Marketing Research,* 24, 3 (August 1987), 258–270.
63. Wetzer, I.M., Zeelenberg, M., and Pieters, R. Never eat in that restaurant, I did! Exploring why people engage in negative word-of-mouth communication. *Psychology and Marketing, 24*, 8 (2007), 661–680.
64. Wooldridge, J.M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press, 2002
65. Ying, Y., Feinberg, F., and Wedel, M. Leveraging missing ratings to improve online recommendation systems. *Journal of Marketing Research, 43*, 3 (2006), 355–365.
66. Zhu, F., and Zhang, X.M. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, 74, 2 (2010), 133–148.

Appendix

Self-selection Decision and Bias Correction—Module S

We need to correct for the consumer self-selection bias because had the consumers decided not to create any review, we could not have included them in our sample. We calibrated a binary probit model of self-selection decisions consistent with [32, 64] on our sample of *selection* and *nonselection* consumers. Let Y^S denote a reviewer's selection decision, U^S denote the latent utility, and $\mathbf{X}^S$ denote the research variables; in this example, α is the intercept, and $\boldsymbol{\beta}$ denotes a vector of coefficients corresponding to these variables. A binary probit model states that:

$$Y^S = 1 \text{ if } U^S = \alpha^S + \mathbf{X}^S \boldsymbol{\beta}^S + e^S > 0$$

Under such a specification, Module R (the Rating module) can be corrected for reviewer self-selection bias through incorporating as one of its variables a term equaling $\lambda = \lambda(\alpha^S + \mathbf{X}^S \boldsymbol{\beta}^S) \equiv \dfrac{\phi(\alpha^S + \mathbf{X}^S \boldsymbol{\beta}^S)}{\Phi(\alpha^S + \mathbf{X}^S \boldsymbol{\beta}^S)}$ where $\phi(\cdot)/\Phi(\cdot)$ is the inverse Mills ratio. The coefficient for this bias-correction term is denoted as ρ.

Unobserved Individual Heterogeneity

Adopting the specifications suggested by Ying et al. [65], we incorporated an individual reviewer's random-effect term into each of the regression coefficients in both modules as well as into the cutoff parameters in Module R. In addition, for each of these regression coefficients that incorporate a random effect, the demographic variables are used as higher level variables to facilitate a further specification of the HLM. In the equations below, Γ is a higher level coefficient; note that superscript T denotes the transpose of a matrix.

$$\begin{pmatrix} \alpha_i \\ \boldsymbol{\beta}_i \end{pmatrix} \sim MNV(\boldsymbol{\mu}_i, \Sigma) \text{ where } \boldsymbol{\mu}_i = \Gamma^{\mathrm{T}} \mathbf{d}_i$$

$$\log(\Delta \boldsymbol{\kappa}_i) \sim MNV(\boldsymbol{\mu}_i^c, \Sigma^c)$$

Details of Estimation Procedure

We adopted and customized the Markov Chain Monte Carlo (MCMC) simulation procedure for our specific empirical needs. These were adjusted to be appropriate for specific situations within our model. In summary, we first estimated Module S to infer λ_i. We used a Gibbs sampler [52]. The result is a consistent and heterogeneous estimate of the bias-correction term λ_i for each reviewer. After obtaining λ_i, we proceeded to estimate Module R.

Module S. Conditional distributions of the set of unknown parameters were calculated under some mild assumptions for model specifications. For Module S, the set of unknown parameters were $\{\alpha, \boldsymbol{\beta}\}$ with noninformative priors $\bar{\beta} = 0$ and $A^{-1} = 1 \times 10^2$. Then a chain of conditional random draws were made repetitively from the conditional distributions mentioned above. We implemented this module in R by using function *rbprobitGibbs* developed by Rossi et al. [55]. After obtaining a matrix of $\{\hat{\alpha}_r, \hat{\boldsymbol{\beta}}_r\}$ where r indexes each iteration of the simulation and $R = 100{,}000$, the first 50,000 draws were discarded as the burn-in period, and every fifth one of the last 50,000 draws was used to report results, in order to reduce autocorrelations of the MCMC chains.

Module R. We again used noninformative priors $\bar{\beta} = 0$, $A^{-1} = 1 \times 10^2$ because prior literature is not informative about the effects examined in our model, especially those being hypothesized. We used a combination of functions *rmultireg* and *rordprobitGibbs* in R and drew from multivariate posterior draws of the reviewer-specific parameter estimates to obtain Level 2 hyperparameters $\{\Gamma^R, \Sigma^R\}$.

Figure 1. A Model of the Determinants of Online Reviews in Reputation Systems

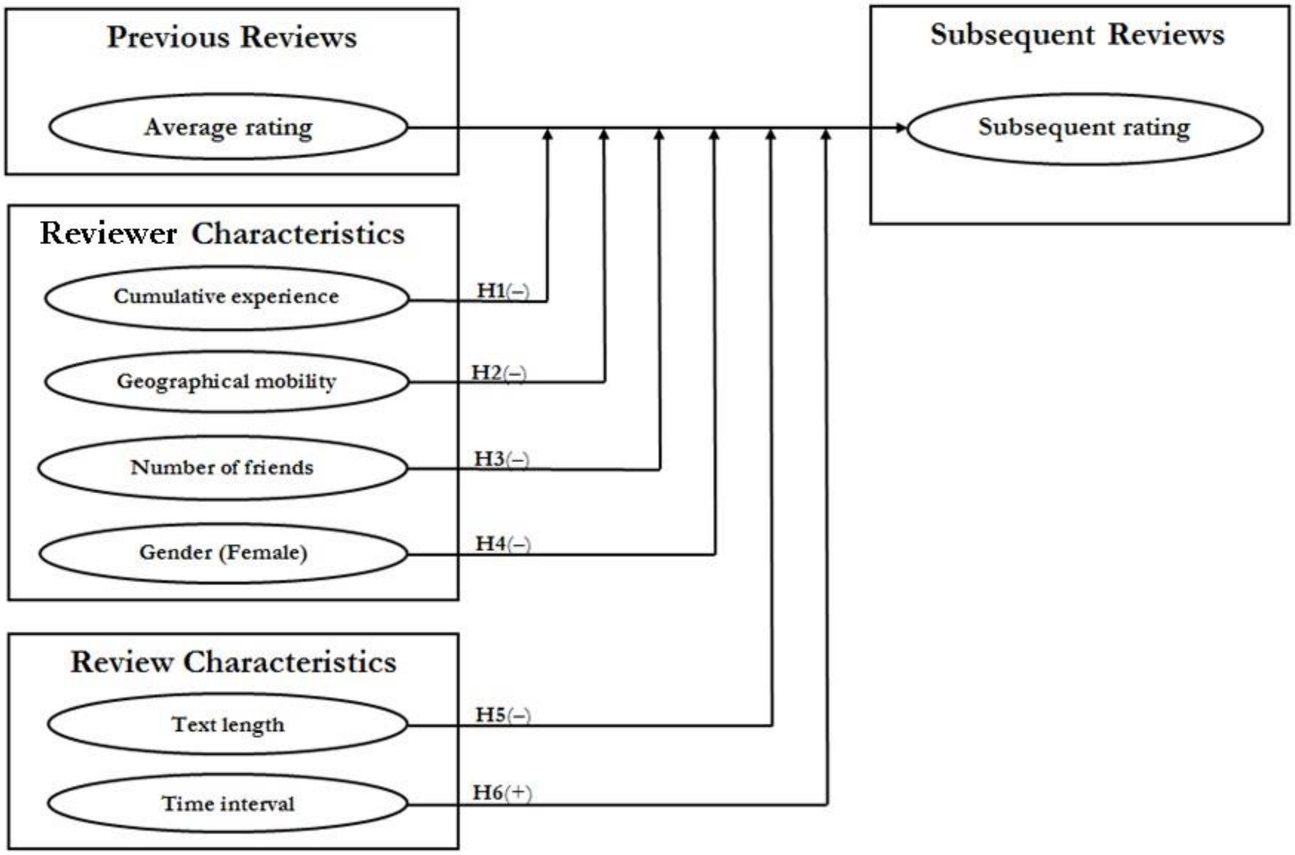


Figure 2. Snowball Sampling and Data Collection Procedures

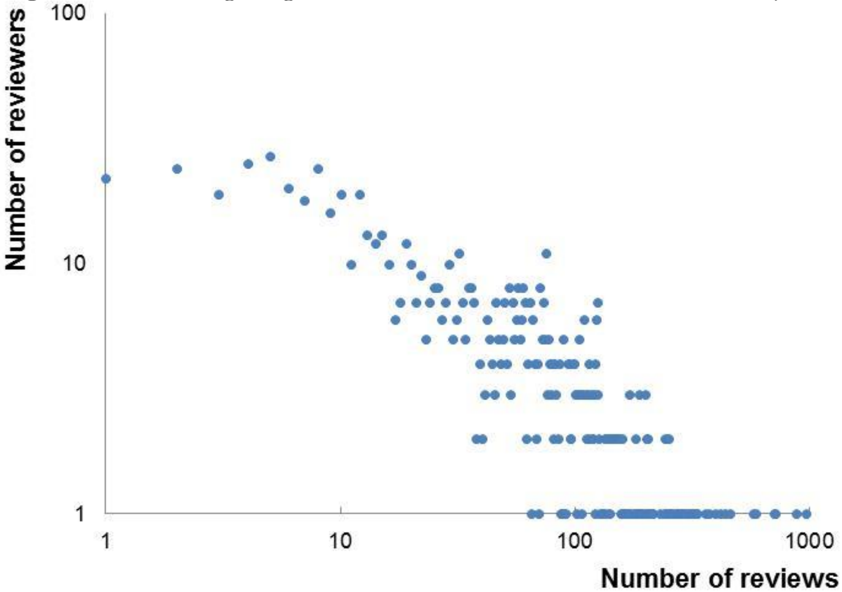
Stage1: Seeding
- Construct a list of 10 largest cities in California from Wikipedia
- Collect 10 most recent reviews on each city's Yelp Welcome page at a random time
- From these 100 reviews, obtained 78 unique reviewer IDs as seed

Stage2: Sampling
- Visit the Friends page of each seeded reviewer. For each seed, get the reviewer ID for all of her or his friends. Duplicate reviewer IDs are removed
- Obtained about 2,500 unique (friend) reviewer IDs as the Wave 1 Reviewer List
- Visit the Friends page of those in the Wave 1 Reviewer List. Likewise, get the reviewer ID for all of her or his friends. Duplicate reviewer IDs are removed
- Obtained a total of 193,889 unique reviewer IDs as the Reviewer Sampling Frame
- Randomly select 900 reviewer IDs from the Reviewer Sampling Frame as our Reviewer Sample

Stage3: Data Collection
- Visit the Reviews page of those in the Reviewer Sample
- Get all reviews by each reviewer. Also collect reviewer characteristics
- Obtained a total of 62,114 reviews by 900 reviewers

Stage4: Additional Data Collection
- Based on the businesses identified by the collected reviews, construct a complete list of 46,407 unique business IDs as the Business Sample
- Visit the homepage of businesses in the Business Sample
- Get all ratings for each business. Also collect business characteristics

Figure 3. Log-Log Distribution of Number of Reviews by Each Reviewer



Notes: This figure plots the number of reviewers ("Number of reviewers" on vertical) who contributed a total number of reviews ("Number of reviews" on horizontal) in our original sample. Both vertical and horizontal scales are base-10 log transformed to overcome the lack of visual precision caused by a highly skewed distribution. Clearly, several reviewers have contributed a large number of reviews (almost 1,000), constituting the long tail. On the original scale, this distribution has the following statistics: mean = 69.02, median = 38, standard deviation = 89.32, skewness = 4.10, Kurtosis = 27.35.

Table 1. Descriptive Statistics and Correlation Matrix of Key Variables

	Variables	N	Original Scale				Transformed Scale[1]												
			Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	1	2	3	4	5	6	7	8	9	10	11
Time-Variant Variables																			
1	$Rating_{ijt}$	61,029	3.73	1.117	1	5	3.73	1.117	1.00										
2	$Price_{jt}$	61,029	14.64	6.898	5	35	14.64	6.898	0.04	1.00									
3	$AvgRate_{jt}$	61,029	3.50	0.672	1	5	3.50	0.672	0.56	0.01	1.00								
4	$CumuExpr_{it}$	61,029	92.34	122.892	2	963	4.45	1.422	-0.03	0.03	0.00	1.00							
5	$Text_{ijt}$	61,029	171.98	128.514	1	997	4.85	0.830	-0.03	0.10	-0.01	0.10	1.00						
6	$When_{ijt}$	61,029	11.24	30.315	1	1,089	1.44	1.319	0.02	0.01	0.03	-0.23	-0.08	1.00					
7	$Rating_{i,t-1}$	61,029	3.73	1.117	1	5	3.73	1.117	0.08	0.01	0.04	-0.04	-0.01	0.03	1.00				
8	$UseflVot_{ijt}$	61,029	3.31	4.772	0	58	0.75	1.063	0.02	0.05	0.05	0.32	0.28	-0.17	0.00	1.00			
9	$City_{jt}$	61,029	0.65	0.474	0	1	0.65	0.474	0.00	-0.03	0.01	0.02	0.01	-0.01	0.01	0.05	1.00		
10	$Reviews_{jt}$	61,029	176.49	313.088	1	3,496	3.95	1.793	0.04	-0.08	0.07	-0.05	0.03	0.03	0.02	-0.02	0.27	1.00	
11	$RatVarnc_{jt}$	61,029	1.22	0.493	0.07	3.49	1.22	0.493	0.01	-0.01	0.02	0.01	-0.01	0.00	0.01	-0.01	0.06	0.09	1.00
Time-Invariant Variables																			
1	$GeoMobil_{i}$	744	0.18	0.253	0	1	0.18	0.253	1.00										
2	$Friends_{i}$	744	120.27	245.396	1	3,658	3.67	1.297	-0.01	1.00									
3	$Gender_{i}$	744	0.53	0.400	0	1	0.53	0.400	-0.01	0.00	1.00								

Note: Natural log transformation was used to mitigate the issue of skewed distribution. If transformation was applied (e.g., CumuExpr, Text, When, etc.), the transformed scale was presented; if transformation was not applied (e.g., Rating, Price, AvgRate, etc.), the original scale was presented in the "transformed scale" column.

Table 2. Comparison of Models With and Without Reviewer Heterogeneity

	1	2
Description	Traditional Pooled Regression	Hierarchical Modeling
Log-likelihood	-78,800	-64,412
AIC	182,898	154,111
BIC	296,981	268,139
MAD	0.678	0.607
RMSE	1.102	0.971
Spearman's rho	0.507	0.590
Hit rate[1]	0.472	0.527

Notes: Hit rate is calculated as follows: (1) first, a predicted rating is determined if a rating of 1 to 5 has the highest expected probability; (2) then, a prediction correctly hits the actual outcome only if the predicted rating is exactly the same as the actual observed rating; (3) hit rate is then equal to the ratio of the sum of correct hits to the total number of observations.

Table 3. Estimation Results

		Coefficient Estimates[1]			Random Effects[1]		
Variables		μ	Std. Dev.	Sig.[2]	Diag.(Σ)	Std. Dev.	Sig.[2]
Level 1 Hypothesized Effects							
Reviewer Characteristics							
$AvgRate_{jt}$ X $CumuExpr_{it}$	$\beta_{0,2}$	-1.040	0.064	***	3.953	0.017	***
$AvgRate_{jt}$ X $GeoMobil_{i}$	$\beta_{1,1}$	-0.172	0.023	***	---	---	
$AvgRate_{jt}$ X $Friends_{i}$	$\beta_{2,1}$	-0.646	0.005	***	---	---	
$AvgRate_{jt}$ X $Gender_{i}$	$\beta_{3,1}$	-0.087	0.017	***	---	---	
Review Characteristics							
$AvgRate_{jt}$ X $Text_{ijt}$	$\beta_{0,3}$	-0.130	0.034	***	2.320	0.016	***
$AvgRate_{jt}$ X $When_{ijt}$	$\beta_{0,4}$	0.360	0.053	***	3.786	0.020	***
Level 1 Nonhypothesized Effects							
Previous Reviews							
$AvgRate_{jt}$	$\beta_{0,1}$	2.498	0.044	***	3.441	0.014	***
Main Effects							
$CumuExpr_{it}$	$\beta_{0,5}$	-0.304	0.035	***	2.023	0.016	***
$Text_{ijt}$	$\beta_{0,6}$	0.109	0.039	**	2.560	0.024	***
$When_{ijt}$	$\beta_{0,7}$	0.421	0.017	***	1.093	0.008	***
Control Variables							
$Price_{jt}$	$\beta_{0,8}$	-0.264	0.008	***	0.341	0.006	***
$RatVarnc_{jt}$	$\beta_{0,9}$	0.009	0.005	*	0.052	0.007	***
$Reviews_{jt}$	$\beta_{0,10}$	0.375	0.020	***	1.146	0.007	***
$City_{jt}$	$\beta_{0,11}$	0.389	0.084	***	3.424	0.019	***
$Rating_{i,t-1}$	$\beta_{0,12}$	-0.019	0.023	***	1.561	0.013	***
$UseflVot_{ijt}$	$\beta_{0,13}$	0.636	0.055	***	2.551	0.015	***
(Fixed effects) $BusFixed_{j}$ $Category_{jt}$ $Year_{ijt}$		---	---		---	---	
Level 2							
$Intercept_{i}$	α_{0}	4.204	0.057	***	---	---	
$GeoMobil_{i}$	α_{1}	-0.141	0.026	***	---	---	
$Friends_{i}$	α_{2}	-0.607	0.007	***	---	---	
$Gender_{i}$	α_{3}	-0.109	0.017	***	---	---	

Notes:
1. For each module, estimation results of the coefficients (under column "μ") are reported as well as the square root of the variances of their random effects (under column "Diag.(Σ)") except for the Level 2 coefficients and cross-level interactions. The corresponding standard error is reported adjacent to each of these results, (under columns "Std. Dev.").
2. Significance level is reported as: † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 4. Robustness Checks: Results Comparison of the Rating Module

Variables		Pooled Regression on Original Sample (Model 1)		Pooled Regression on Final Sample (Model 2)		Hierarchical Modeling on Final Sample (Model 3)	
		Coefficient Estimate	Marginal Effect	Coefficient Estimate	Marginal Effect	Coefficient Estimate	Marginal Effect
$AvgRate_{jt}$	$\beta_{0,1}$	1.057 (0.010)	14.9%***	1.059 (0.010)	15.3%***	2.498 (0.044)	30.5%***
$AvgRate_{jt}$ X $CumuExpr_{it}$	$\beta_{0,2}$	– 0.019 (0.008)	-0.5%*	– 0.019 (0.008)	-0.5%*	– 1.040 (0.064)	-3.1%***
$AvgRate_{jt}$ X $GeoMobil_{i}$	$\beta_{1,1}$	– 0.083 (0.029)	-2.0%**	– 0.077 (0.029)	-1.9%**	– 0.172 (0.023)	-4.3%***
$AvgRate_{jt}$ X $Friends_{i}$	$\beta_{2,1}$	– 0.004 (0.006)	-0.1%	– 0.004 (0.006)	-0.1%	– 0.646 (0.005)	-15.8%***
$AvgRate_{jt}$ X $Gender_{i}$	$\beta_{3,1}$	0.017 (0.018)	+0.4%	0.013 (0.018)	+0.3%	– 0.087 (0.017)	-2.4%***
$AvgRate_{jt}$ X $Text_{ijt}$	$\beta_{0,3}$	– 0.066 (0.009)	-1.6%***	– 0.064 (0.009)	-1.6%***	– 0.130 (0.034)	-4.6%***
$AvgRate_{jt}$ X $When_{ijt}$	$\beta_{0,4}$	0.030 (0.007)	+0.8%***	0.029 (0.007)	+0.7%***	0.360 (0.053)	+10.4%***
Sample Size (# of reviewers)		900		744		744	
Sample Size (# of observations)		62,114		61,029		61,029	

Notes: This table shows a reasonably high level of consistency between the rating module's estimation results under different specifications and with different samples. Model (1) is estimated on the original sample of 900 reviewers' data with a general Ordered Probit regression model; Model (2) is estimated with the same Ordered Probit regression model but on our final sample of 744 reviewers' data; Model (3) is estimated on the same final sample but with hierarchical modeling, and hence is exactly the same as our reported results in Table 3. Both coefficient estimates and marginal effects are reported. Marginal effect indicates how *more* likely a subsequent rating is going to follow the average rating: The "increased probability" that a typical reviewer will give a 5-star rating if the average rating of a business increases from 4 to 5. Under this situation, for example, it implies that this many more reviewers (in percentage terms) will rely on the average rating and give a 5-star review. Marginal effect is calculated for a typical reviewer whose variables take their corresponding average values for continuous variables and the base level for categorical and dummy variables.

1. For the main effect (AvgRate), marginal effect is the increased likelihood of following the average rating, as described above; for all other effects (moderations), marginal effects represent the amount and direction of changes to this baseline likelihood at the main effect level.
2. For each coefficient estimate, the corresponding standard error is reported immediately below it in parentheses.
3. † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.
4. To be succinct, this table only includes results of the hypothesized effects; additional details are available upon request.