



Marketing Science

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

The Fateful First Consumer Review

Sungsik Park , Woochoel Shin , Jinhong Xie

To cite this article:

Sungsik Park , Woochoel Shin , Jinhong Xie (2021) The Fateful First Consumer Review. Marketing Science 40(3):481-507.
<https://doi.org/10.1287/mksc.2020.1264>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2020, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

The Fateful First Consumer Review

Sungsik Park,^a Woochoel Shin,^b Jinhong Xie^b

^a Darla Moore School of Business, University of South Carolina, Columbia, South Carolina 29201; ^b Warrington College of Business, University of Florida, Gainesville, Florida 32611

Contact: sungsik.park@moore.sc.edu,  <https://orcid.org/0000-0002-4042-3134> (SP); wshin@ufl.edu,  <https://orcid.org/0000-0001-9751-2851> (WS); jinhong.xie@warrington.ufl.edu,  <https://orcid.org/0000-0001-6988-7664> (JX)

Received: April 3, 2020

Revised: July 15, 2020

Accepted: July 26, 2020

Published Online in Articles in Advance:
February 3, 2021

<https://doi.org/10.1287/mksc.2020.1264>

Copyright: © 2021 INFORMS

Abstract. This paper uncovers the striking power of a product's first consumer review. Our analytical model suggests that two key metrics of online consumer reviews, valence and volume, are not independent, but instead evolve interdependently. This interdependence forms a mechanism to transfer a (dis)advantage from a product's first review to both a *long-lasting* (dis)advantage in future word-of-mouth (WOM) valence and an *increasing* (dis)advantage in future WOM volume. As a result, a single consumer review can significantly influence the fate of a given product. These theoretical predictions, although seemingly unlikely, are supported by our empirical investigations. For example, more than 30% of vacuum cleaner models offered by both Amazon.com and BestBuy.com receive first reviews of opposite valence on the two platforms. Those with a negative first review subsequently suffer a loss in both valence and volume vis-à-vis their counterparts with a positive first review, even after 36 months. More strikingly, the first-review effect on WOM volume increases over time. Our findings reveal a crucial weakness in the user-generated information mechanism. As a consumption-based information source, it creates an information-availability bias such that when a product receives a negative first review, it not only suffers low initial sales, but also loses the opportunity to correct the possible negative bias via subsequent reviews. These findings have substantial implications for online sellers, e-commerce platform providers, and consumers.

History: K. Sudhir served as the senior editor and David Godes served as associate editor for this article.

Funding: W. Shin gratefully acknowledges financial support from the Brian R. Gamache Endowed Professorship at the University of Florida. J. Xie gratefully acknowledges financial support from the JCPenney Endowed Professorship at the University of Florida.

Supplemental Material: Data files and the online appendix are available at <https://doi.org/10.1287/mksc.2020.1264>.

Keywords: word of mouth • consumer review • first review • dynamics of online opinion • path dependence • online reviews




1. Introduction

In today's markets, consumers increasingly depend on online word of mouth (WOM) as a reliable source of information in their purchase decisions (Pew Research Center 2016). According to recent consumer surveys, 90% of consumers report that their buying decisions are influenced by online consumer reviews (Dimensional Research 2013), and 88% of consumers trust online reviews as much as personal recommendations (BrightLocal, 2014). Motivated by the practical significance, academic interest in online consumer reviews has surged and the literature provides extensive evidence to support the impact of online reviews on consumer purchasing behavior (e.g., Godes and Mayzlin 2004, Liu 2006, Chen and Xie 2008, Chintagunta et al. 2010, Sun 2012, Ameri et al. 2019). Recently, three meta-analyses concluded that the two key metrics of online WOM, valence and volume, are directly linked to product sales (Floyd et al. 2014, You et al. 2015, Babić Rosario et al. 2016).

Given the strong linkage between online WOM and sales, the importance of obtaining favorable consumer reviews to product success cannot be overemphasized. The literature has suggested that consumer reviews are affected by product-specific characteristics, such as functionality, brand image, and price (e.g., Li and Hitt 2008, Chen et al. 2011, Godes and Silva 2012, De Langhe et al. 2016, Luca and Zervas 2016). Thus, if the same product is available at a similar price on multiple online retail platforms, its online reviews should be similar across platforms. However, many market observations contradict this expectation, as shown in Table 1.

All three products listed in Table 1 are available at both Amazon.com and Walmart.com.¹ Although both platforms use the same five-star evaluation scale, online reviews of these products are markedly different across these two platforms. First, consider Product A, offered at the same price on both platforms. Its average review rating is 4 stars at Amazon, but only 2.2 stars at Walmart, with 303 reviews at Amazon but only four reviews at Walmart. Second, even in instances

Table 1. (Color online) Inconsistency in Online Reviews: Some Examples

		Average rating			Number of reviews			Price		
		Amazon		Walmart	Amazon		Walmart	Amazon		Walmart
A		4.0	>	2.2	303	>	4	79.00	=	79.00
B		2.5	<	4.0	8	<	81	36.68	<	39.80
C		4.2	>	3.0	328	>	2	80.99	<	89.99

Note. The products (date when the data were recorded) are as follows: A, Canon Powershot ELPH 115 (June 6, 2014); B, Toastmaster TM-103TR Stainless Steel 4 Slice Toaster Oven (April 16, 2016); C, Sealy Baby Firm Rest Crib Mattress (April 13, 2016).

where different platforms charge different prices, consumer reviews vary across the platforms, although not necessarily because of the cross-platform price differential. For example, both Products B and C are sold at a lower price at Amazon than at Walmart. However, Amazon's lower price is a disadvantage in online WOM for Product B, but an advantage in online WOM for Product C. Such cross-platform inconsistency in consumer reviews is not limited to these three products or between the two websites presented in Table 1. Rather, it exists in many product categories across various online platforms.

These observations raise several questions. Why does a product with favorable reviews on one online platform receive an unfavorable review on another? Is this a more general phenomenon than we have so far realized? What fundamental market forces drive such substantial cross-platform inconsistency? Is it possible to predict the direction of such inconsistencies in online WOM? What are the important implications to firms' online WOM management policies, especially given the significance of online WOM to a product's success? Answers to these questions are of significance both from theoretical and practical perspectives.

To answer these questions, this paper investigates the evolution of online consumer reviews, both theoretically and empirically. We begin by developing a theoretical model that characterizes the process of online WOM evolution. By explicitly modeling the influence of online reviews on consumer purchase decisions, as well as the process of updating online reviews, we discover that the two key metrics of online WOM, valence and volume, are not independent; rather, they evolve interdependently following a positive or negative feedback pattern. Specifically, our analysis reveals that at any given time, the current WOM valence positively impacts future WOM volume, and the current WOM volume positively or negatively impacts future WOM valence, depending on whether the current WOM valence has an upward or a downward bias.

We show that such interdependence between WOM valence and volume is fundamentally driven by the distinctive characteristics of consumer reviews as a unique information source. Unlike other types of product information (such as advertising and third-party product reviews), online consumer reviews are, by definition, posted by users based on their personal consumption experience, which implies that the availability of consumer reviews for a given product is *conditional* on its adoption. First, this conditional availability of consumer reviews creates the dependence of WOM volume on valence. In other words, at any given time, an increase in a product's average review rating leads to more sales, which, in turn, boosts its future review-posting volume. Second, though less intuitive, this conditional availability of consumer reviews also creates the dependence of WOM valence on volume through the updating process of WOM valence. Specifically, at any given time, a product's average review rating is updated as the weighted average of two components: (1) the previous average rating, which is based on previously posted reviews, and (2) the incremental average rating, which is based on newly posted reviews. The larger the volume of the newly posted reviews, the stronger their impact on the updated valence.

Understanding the interdependence between WOM valence and volume is vital because, as will be shown by our theoretical analysis, this interdependence endows a product's first consumer review with striking power. In today's information-rich environment, where online shoppers have a staggering amount of user-generated information detailing other shoppers' consumption experiences, it is hard to believe that a single consumer review could exert a significant influence over the fate of a given product. We find, however, that a product's first consumer review has a *persistent* impact on the product's entire WOM history, measured by both volume and valence. More surprisingly, the impact of the first review on future

Table 2. Theoretical Predictions

A1	The first-review effect on WOM valence	A product has a higher average rating after a positive first review than a negative first review.
A2	The first-review effect on WOM volume	A product has a higher number of reviews after a positive first review than a negative first review.
A3	The first-review effect on the dynamics of WOM valence	The advantage in the average rating due to the advantage from a positive first review decreases with time.
A4	The first-review effect on the dynamics of WOM volume	The advantage in the number of reviews due to the advantage from a positive first review increases with time (i.e., the snowball effect).

WOM volume does not diminish, but rather, it intensifies over time. Our theoretical analysis leads to four specific predictions regarding the first-review effect, as summarized in Table 2 (see predictions A1–A4). The first two predictions focus on the first review’s overall effect on WOM valence and volume, and the last two predictions focus on the first review’s dynamic effect on WOM valence and volume.

These predictions suggest a persistent and even increasing first-review effect, which is counterintuitive and seemingly unlikely because a single review, however persuasive it may be, is likely to be buried down in the list as subsequent reviews are posted. Thus, intuitively, other reviews should eventually cancel out the impact of the initial review. To gain some external validity, we test these theoretical predictions based on field data. Specifically, we conduct an empirical study using a sample of 177 vacuum cleaner models sold simultaneously on two different retail platforms, Amazon and BestBuy.com (referred to as “Best Buy” hereafter). We empirically examine the relationships between a product’s cross-platform difference in its first review and its cross-platform difference in future WOM valence and volume by developing an econometric model that controls other cross-platform differences (such as differences in price) and platform-specific effects. Consistent with predictions A1 and A2, our results reveal that when the same product is offered on the two platforms, both the average review rating and the number of reviews are significantly lower on the platform with a negative first review than on that with a positive first review.² Our results also support the dynamic patterns of the first-review effect predicted in A3 and A4. The first-review effect on WOM valence becomes weaker as time passes, but it does not disappear even after an extended period of time. Specifically, our estimates suggest that a negative first review decreases the average rating by 0.29 stars, even one year after the first review posting. Our results also support the counterintuitive snowball effect of the first review on WOM volume (i.e., the impact of the first review on the number of reviews increases with time). Specifically, our estimates reveal that, on average, a product with a negative first review receives 14.98 fewer reviews at the end of the 6th month and

36.49 fewer reviews at the end of the 12th month than a product with a positive first review.

We supplement our empirical analysis with various robustness checks, which generate the following results. First, the first-review effects hold for different lengths of observation window. Specifically, using the data set of one-, two-, and three-year observation windows, we can confirm that the first-review effect persists even after a significant period of time. More surprisingly, the snowball effect is even supported under an observation window as long as three years. Second, the first-review effect survives some potential alternative explanations. In particular, our analysis shows that the first-review effect is unlikely to be driven by the distinct tastes of consumers in different platforms that favor certain unobserved product characteristics. Moreover, even after accounting for an increase in volume resulting from a front-page placement, we still observe the first-review effect. Third, the first-review effect is not specific to a certain product category or platform, but is a general finding across various contexts. Our analysis reveals that the data from another category (toasters) support all of our predictions, and so do those from another pair of platforms (Amazon U.S. versus Amazon Canada). Finally, the first-review effect is robust to alternative assumptions on price variations and platform heterogeneity. More specifically, we obtain the same results even after accounting for temporal price variation. Furthermore, even when the magnitude of the first-review effect is assumed to be different across platforms, we still observe the persistence of the first-review effect. While all of the above analyses suggest the robustness of our findings, one caveat is that our analysis focuses on product categories whereby the online WOM is likely to sway consumer choices. For some other categories, however, the first-review effect may be weak or even absent. Our additional analysis shows that consumers’ reliance on nonreview information weakens the first-review effect on the WOM volume. This suggests that boundary conditions may exist for the first-review effect and that the relative strength of the influence of nonreview information is one such condition.

Taken together, our empirical studies provide evidence that supports our theoretical predictions and

demonstrates the formidable and protracted influence of a product's first review on its future online WOM valence and volume.

This paper contributes to online WOM research and practice. In recent years, scholars have made significant efforts to study the consequences of online WOM. In particular, many papers have investigated the influence of online WOM on sales (e.g., Liu 2006, Trusov et al. 2009, Chintagunta et al. 2010, Chen et al. 2011, Libai et al. 2013, Ludwig et al. 2013), as well as the boundary conditions for such influences (e.g., Forman et al. 2008, Zhu and Zhang 2010, Ho-Dac et al. 2013). Based on the results of a large number of empirical studies, recent meta-analyses concluded that two key metrics of WOM, volume and valence, significantly affect sales (Floyd et al. 2014, You et al. 2015, Babić Rosario et al. 2016). In this paper, we uncover the interdependence between these two key metrics of online WOM and derive a crucial implication of such interdependence: the powerful first-review effect. Our research findings make the following contributions.

First, whereas “big data” is one of the hottest buzzphrases in today's business world, our research brings attention to the significance of “small” data. The first review, even though it is a single data point, has the potential to sway the entire evolution path of online consumer reviews and, thus, the fate of the product. As will become clear in our theoretical model, the power of the first review is created by the interdependence between WOM valence and volume. Fundamentally, this interdependence forms a mechanism that transfers a (dis)advantage in a product's first review rating to (1) a *long-lasting* (dis)advantage in future WOM valence and (2) an *increasing* (dis)advantage in future WOM volume. Therefore, the first review, although small in size, holds significant predictive power. These findings call for proactive strategies in managing the first review. In the early stages of a product's life cycle, a special effort should be put forth to preemptively minimize the possibility of a negative first review. Moreover, if the first review turns out to be negative, immediate action should be taken to reverse its negative impact.

Second, whereas the literature has attributed various advantages to consumer reviews, such as higher credibility, up-to-date information, and nuanced communication (Chen and Xie 2008, Ranard et al. 2016), our research identifies a significant weakness of online consumer reviews. Specifically, because the availability of consumer reviews is conditional on the adoption of the product, not all products have the same chance of receiving informative consumer reviews. For example, a product with low initial sales (which can be caused simply by an unfavorable first review) would fail to generate a viable number of reviews, and such a low review volume would further harm the product's sales.

Thus, if a product receives an unfavorable initial review, consumers may lose the opportunity to learn about the product altogether, which also potentially deprives them of the chance to enjoy a high-quality product. Our data set shows that 288 out of 1,155 vacuum cleaner models received a negative first review on Amazon, which suggests that about one-fourth of all models offered by this platform may fail to provide informative review information. Uncovering this important weakness of online consumer reviews is crucial, because consumer reviews are gaining greater significance in today's market, and firms are increasingly paying attention to them compared with other communication channels. Our findings regarding the availability bias of consumer reviews indicates that blindly following this trend might lead to a suboptimal mix of marketing communication efforts. By recognizing this important limitation of online consumer reviews, firms could develop an effective communication strategy that takes into account the complementarity between consumer-generated information and firm-initiated marketing activities.

Third, our research advances the literature of review dynamics. Whereas most existing studies have focused on WOM valence dynamics (e.g., Li and Hitt 2008, Wu and Huberman 2008, Godes and Silva 2012), our research examines the dynamics of both valence and volume and, more importantly, jointly considers the evolution of both. This approach allows us to uncover the path dependence of both valence and volume from the very first review, which cannot be explained by considering the evolution of the valence independently of that of the volume. Moreover, studies of the dynamics of WOM valence have suggested opposite patterns: a downward time trend (i.e., Li and Hitt 2008, Wu and Huberman 2008) or an upward time trend (i.e., Godes and Silva 2012). Our paper, in contrast, proposes and empirically demonstrates that both a downward and an upward trend can exist simultaneously for the same product: WOM valence increases with time on a platform where the first review is negative, but decreases with time on a platform where the first review is positive. Moe and Schweidel (2012) find a positive relationship between the valence of the ratings environment and review-posting incidence. By modeling the interdependence between WOM valence and volume, our research demonstrates that (a) a (dis)advantage in the first review will transfer to a (dis)advantage in both WOM valence and volume, which can explain a positive relationship between the valence of the ratings environment and review-posting incidence, and (b) the (dis)advantage in WOM volume that results from the (dis)advantage of the first rating increases rather than decreases over time. The latter finding, fundamentally driven by the joint evolution of

WOM valence and volume, cannot be explained by consumer motivation (Moe and Schweidel 2012).

Finally, our results highlight the asymmetric influence of positive and negative first reviews on the information value of the online review process in general. A positive first review, regardless of how biased it might be, facilitates informative online WOM. Even with a significant upward bias, a positive first review leads to a high future WOM volume, which, as more reviews are posted, will effectively correct any initial bias. However, a negative first review is destructive to online WOM. An unfavorable first review reduces the number of consumers who will experience the product and potentially provide additional reviews. As a result, there is little chance to correct the initial bias, and the online WOM will fail to offer sufficient product information. This finding is in stark contrast to the work of Muchnik et al. (2013), who observe bias correction in the evaluation of news article comments after a negatively manipulated initial rating but positive herding after a positively manipulated initial rating. The crucial difference is that the valence and volume of the submitted evaluations in their context are independent of each other. In the experimental setting of Muchnik et al. (2013), participants cannot observe the ratings before clicking through the comments, suggesting that the consumption of the content (of comments) is independent of previous ratings. In the context of product reviews, however, the valence of previous reviews affects sales in the subsequent periods, because consumers tend to prefer what they have been told are better products. In addition, an evaluation can be made conditional on the purchase of the product, and, thus, the valence of previous reviews affects the volume of subsequent reviews through sales. Review volume, in turn, influences the overall valence of reviews, because bias in the first review may or may not be corrected by future review volume. Therefore, interdependence between volume and valence is key to understanding the first-review effect in the context of online product reviews.

The rest of this paper is organized as follows. In Section 2, we develop a theoretical model to study the interdependence of valence and volume of online WOM, which generates predictions about the impact of the first review. In Section 3, we empirically test each of these theoretical predictions using data collected from the field. In Section 4, we explore a boundary condition of the first-review effects both theoretically and empirically. Finally, in Section 5, we summarize our findings and discuss future research opportunities.

2. Theoretical Analysis

To develop a conceptual understanding of the first-review effect, in this section, we provide a theoretical analysis of online WOM. Our model is highly stylized

for the purposes of illustration, but includes essential components of online WOM evolution well-documented by extant empirical studies.

2.1. A Model of Online WOM

In this section, we present our model of online WOM evolution that describes how reviews are generated as well as how reviews influence consumers' purchase decisions. Consider a product of quality $\alpha \in (0,1)$. Because an individual rating for the product is determined by idiosyncratic experience of individual consumers, we model the individual rating w_i as a sampling process from a Bernoulli distribution with the success probability given as the true quality α . In other words, w_i is given as 1 (positive) with probability α but 0 (negative) with probability $1 - \alpha$. Thus, the average review rating in any given period t is $r_t = \frac{\sum_{i=1}^{n_t} w_i}{n_t}$ ($t \geq 1$), where n_t is the total number of reviews in that period.³

Next, consider a consumer who evaluates a product to purchase in period t . Her valuation of the product depends on both the perceived quality and the fit of the product with her taste. First, she infers the quality of the product by reading the reviews that previous users have written. In particular, she uses the average rating r_{t-1} as an unbiased estimator of the true quality α . Second, each consumer knows her exact preference and evaluates the fit of the product based on that preference. A consumer lowers her valuation of the product in proportion to the mismatch between her own taste and the product, which we denote by x . Because consumers are heterogeneous in their tastes, we assume that x follows $U[0,1]$. Together, the valuation of a consumer is given by $v_t = v_0 + \beta r_{t-1} - \tau \cdot x$, where v_0 is the baseline valuation (independent of the quality), β represents consumer's sensitivity to quality, and τ is the mismatch cost. Note that a consumer in this model uses the reviews to infer the quality but not the fit of the product, implying that our model is more relevant to the product category in which quality carries much more weight in the purchase decision than the fit does.

Based on her valuation v_t and the price p of the product, a consumer makes a purchase decision. Because we consider a rational consumer, her decision is not influenced by behavioral biases that have been documented in the WOM literature (e.g., Adomavicius et al. 2013). Given the uncertainty concerning product quality, we assume that consumers are risk-averse and derive the following exponential utility: $U_t = 1 - e^{-(v_t - p)}$ (for other applications of this utility formulation; see Holmström and Milgrom 1987, Holmström and Tirole 1993, Kornish and Li 2010, Feng and Xie 2012). In this specific formulation, for simplicity, we have normalized the Arrow–Pratt risk aversion parameter to be one. However, other values of the risk aversion parameter do not qualitatively change our

results as long as the parameter is greater than zero, that is, consumers are risk averse. Given this utility, the consumer buys the product if and only if her expected utility from purchase is greater than that of no purchase: $E[U_t] = 1 - e^{-(v_0 + \beta r_{t-1} - \tau x - p - \frac{\beta^2 r_{t-1}(1-r_{t-1})}{2n_{t-1}})} \geq 1 - e^0 (= 0)$ (see the appendix for a detailed derivation of the expected utility).

In each period, M consumers newly arrive in the market. Then, from the expected utility, we can derive the sales of period t as follows: $s_t = \frac{M}{\tau}(v_0 + \beta r_{t-1} - p - \frac{\beta^2 \cdot r_{t-1}(1-r_{t-1})}{2n_{t-1}})$. Note that the sales are greater with reviews than without them, which is translated into $\beta r_{t-1} - \frac{\beta^2 \cdot r_{t-1}(1-r_{t-1})}{2n_{t-1}} \geq 0$, for all r_{t-1} and n_{t-1} . To ensure that this condition holds, we assume $\beta \leq 2$. Moreover, for any realization of the review rating, the sales should not be negative. Thus, in our analysis, we focus on the parameter space where $p \leq v_0$ holds.⁴

Finally, among the consumers making a purchase, only a fraction will write a review, as noted in a recent empirical study (Anderson and Simester 2014). To capture the review-updating process in a parsimonious way, we use a constant propensity to write a review, $\delta \in (0, 1]$, implying that the total number of newly arriving reviews in period t (denoted by N_t) follows $\text{Binomial}(s_t, \delta)$.⁵ Then the total number of reviews in period t is updated as $n_t = n_{t-1} + N_t$ ($t \geq 1$).

2.2. Evolution of Online WOM

Based on the model described thus far, we investigate the evolution of online WOM. We specifically examine how the two key variables of online WOM (valence and volume) evolve from period to period. For this purpose, in this section, we fix the valence and the volume of period t (r_t and n_t) and examine how the expected valence and volume of period $t+1$ are updated. First, recall that the two metrics of online WOM affect the expected sales of the next period through $E[s_{t+1}|r_t, n_t] = \frac{M}{\tau}(v_0 + \beta r_t - p - \frac{\beta^2 \cdot r_t(1-r_t)}{2n_t})$. Because N_{t+1} follows $\text{Binomial}(s_{t+1}, \delta)$, the expected number of reviews is as follows:

$$\begin{aligned} E[n_{t+1}|r_t, n_t] &= n_t + E[N_{t+1}|r_t, n_t] = n_t + \delta \cdot E[s_{t+1}|r_t, n_t] \\ &= n_t + \frac{\delta M}{\tau} \left(v_0 + \beta r_t - p - \frac{\beta^2 \cdot r_t(1-r_t)}{2n_t} \right). \end{aligned} \quad (1)$$

Next, let h_t denote the total number of positive reviews in period t , and let H_t represent the number of newly arriving positive reviews in period t . These definitions imply $h_{t+1} = h_t + H_{t+1}$. Then by definition, the expected average rating in period $t+1$ is given as follows:

$$E[r_{t+1}|r_t, n_t] = E\left[\frac{h_{t+1}}{n_{t+1}}|r_t, n_t\right] = E\left[\frac{h_t + H_{t+1}}{n_t + N_{t+1}}|r_t, n_t\right]. \quad (2)$$

Equations (1) and (2) describe the period-to-period transition of the valence and the volume, which

determines the evolution over time of these two metrics. Interestingly, these equations show that the current state of each metric is affected by not only its own previous state but also that of the other metric. An important implication of this observation is that the valence and the volume of online WOM evolve interdependently. On systematically investigating the relationship between the two variables, we obtain the following theorem.

Theorem 1 (Interdependence Between Valence and Volume of Online Consumer Reviews). *Two key metrics of online consumer reviews, valence and volume, evolve interdependently with positive or negative feedback. Specifically, in any given period $t (\geq 1)$,*

- (a) *the current valence r_t positively influences the future volume ($E[n_{t+1}]$);*
- (b) *the volume of newly posted reviews ($N_{t+1} = n_{t+1} - n_t$) influences the valence ($E[r_{t+1}]$)*
 - i. *positively if the previous valence is biased downward (i.e., $r_t < \alpha$),*
 - ii. *negatively if the previous valence is biased upward (i.e., $r_t > \alpha$).*

For proofs, see the appendix. The theorem describes how the two metrics influence each other in their evolution. To understand part (a) of the theorem, recall from previous research that positive reviews from the previous period enhance the product's sales (You et al. 2015, Babić Rosario et al. 2016): $\frac{\partial E[s_{t+1}]}{\partial r_t} > 0$. Moreover, note that only consumers who have purchased the product post online reviews, implying that each review posting is conditional on the adoption of the product. Therefore, the volume of consumer reviews in any given period will be positively affected by the level of sales: $\frac{\partial E[n_{t+1}]}{\partial E[s_{t+1}]} > 0$.⁶ This finding implies that online consumer reviews may provide insufficient information, depending on the current level of sales. Especially, in the case of low sales, consumers could suffer from this information availability bias. Taken together, these two inequalities imply that the valence of consumer reviews increases the volume of reviews in the subsequent period: $\frac{\partial E[n_{t+1}]}{\partial r_t} = \frac{\partial E[n_{t+1}]}{\partial E[s_{t+1}]} \cdot \frac{\partial E[s_{t+1}]}{\partial r_t} > 0$.

Part (b) of the theorem discusses the impact of the volume on the future valence of the online reviews. In this case, unlike in part (a), the impact may be either positive or negative, depending on the magnitude of the current valence relative to the true quality of the product. To understand this, note that, in any period, the valence is the weighted average of the valence of all previously posted reviews and that of newly posted reviews, with the weight being the volume of each group of reviews. Thus, if the valence in the previous period is more negative than that of the newly posted reviews, a greater volume of newly posted reviews will render the valence of the current period more positive. However, if the valence in the

previous period is more positive than that of the newly posted reviews, a greater volume of newly posted reviews will turn the valence of the current period more negative. On average, the valence of the newly posted reviews is consistent with the true quality of the product. Thus, if the realized valence of already posted reviews happens to be lower than the product's true quality, a greater volume of newly posted reviews will elevate the future valence. However, if the realized valence is higher than the true quality, a greater volume of newly posted reviews will lower the future valence.

An important implication of Theorem 1 is that evolution patterns of the two key metrics of online WOM, valence and volume, can be accurately captured only when their interaction is considered. Given the evolution patterns characterized thus far, it is evident that previous reviews can influence both the valence and the volume of subsequent reviews. Such influence leads to the possibility that the very first review might have a significant impact on the evolution of both the valence and the volume of the entire set of reviews. We examine this possibility in the next subsection.

2.3. The First-Review Effect

In this section, we examine how the first review might influence both the valence and the volume of subsequent reviews. For this purpose, we start from the initial period ($t = 0$), where the very first review is posted. At this initial period, the first review is exogenously given as either positive ($r_0 = 1$) or negative ($r_0 = 0$). Then, depending on the valence of the first review, we separately trace the evolution of valence and volume using superscript $k \in \{+, -\}$, where $k = +$ if $r_0 = 1$ and $k = -$ if $r_0 = 0$. By comparing $E[n_t^+]$ and $E[n_t^-]$ as well as $E[r_t^+]$ and $E[r_t^-]$, we derive predictions regarding the impact of the first review on the valence and the volume of online WOM in every subsequent period. Note that because these predictions are made in period 0 where n_t^k and r_t^k ($t \geq 1$) have not been realized yet, we examine their expected values. However, doing so complicates the analysis and we thus resort to the numerical simulation in deriving our predictions. See Online Appendix A1 for the detailed procedure and the results of the numerical analysis.

Proposition 1 (First-Review Effect: Overall Impact). *A product's first consumer review persistently influences its future consumer reviews. Specifically, a higher rating in the first review leads to (a) a higher average rating and (b) a higher number of reviews in any given period. Formally,*

$$E[r_t^+] > E[r_t^-] \text{ and } E[n_t^+] > E[n_t^-], \text{ for all } t. \quad (3)$$

The previous proposition shows that both the valence and the volume of WOM crucially depend on the

valence of the first review. More specifically, when the first review is positive, both the average rating and the total number of reviews are higher than when the first review is negative. Below, we provide intuitions for the valence and the volume results, respectively.

In general, many idiosyncratic factors affect the valence of each individual review. Thus, one would not expect the first review to affect the average rating in subsequent periods in any significant way. Nevertheless, the first part of the proposition states that a positive first review generates a higher average rating in every subsequent period. To understand this rationale, recall that the valence in any period is the weighted average of the valence of all previously posted reviews and the average rating of newly posted reviews. Moreover, the average rating of newly posted reviews does not depend on the valence of the first review. Therefore, any advantage in the average rating of the previous period is likely to persist in the average rating of the next period, implying that the average rating in subsequent periods will be more positive if the first review is positive than if it is negative.

Proposition 1 also shows that a greater number of reviews will be generated after a positive first review than after a negative one. Two different mechanisms contribute to this result. First, recall from the first part of the proposition that a higher rating in the first review results in higher average ratings in subsequent periods. Because consumers are more likely to purchase a product when online reviews are favorable (You et al. 2015, Babić Rosario et al. 2016), sales will be greater after a positive first review than after a negative one. Then, in the subsequent period, a greater volume of WOM will follow. Second, more reviews reduce consumers' uncertainty about the quality of the product regardless of the valence of the reviews, because more reviews reflect greater sales, which in turn work as a social proof that the product quality is sufficiently high to warrant purchase (Bikhchandani et al. 1998, Zhang and Liu 2012). Thus, consumers are more attracted to a product when making a purchase decision if they see a large number of posted reviews. Hence, more reviews lead to greater sales and, in turn, to a greater volume of WOM, which establishes a positive feedback loop between sales and WOM volume. Both of these two effects (the direct impact of WOM valence on sales and the positive feedback loop between sales and volume) work together to generate a greater volume of WOM from a positive than from a negative first review.

Given the well-documented impact of WOM valence and volume on sales (You et al. 2015, Babić Rosario et al. 2016), these results have important implications. Our results suggest that a product of the same quality may generate different amounts of WOM with different average ratings, thus achieving different levels of sales, all depending on the valence of the

first review. More importantly, this impact persists over time; thus, any negative impact due to a negative first review is very hard to reverse without costly intervention. Therefore, when launching a new product, it is critical that the first review be positive.

Given that the first-review effect persists, an interesting question is how the magnitude of this effect changes over time. To examine this issue, we study the evolution of the valence and the volume of consumer reviews and obtain the following result.

Proposition 2 (First-Review Effect: Dynamics Over Time).

(a) *The first-review effect on valence diminishes over time:*

$$E[\Delta r_t] > E[\Delta r_{t+1}], \text{ for all } t, \quad (4)$$

where $\Delta r_t = r_t^+ - r_t^-$ and $\Delta r_{t+1} = r_{t+1}^+ - r_{t+1}^-$.

(b) *The first-review effect on volume intensifies over time:*

$$E[\Delta n_t] < E[\Delta n_{t+1}], \text{ for all } t, \quad (5)$$

where $\Delta n_t = n_t^+ - n_t^-$ and $\Delta n_{t+1} = n_{t+1}^+ - n_{t+1}^-$.

The previous result characterizes the evolution of the valence and the volume of WOM as a function of the valence of the first review. The first part of the proposition shows that the impact of the first review on the valence of WOM becomes weaker as time passes. This weakening occurs because any bias in the first review is corrected as more reviews arrive. Recall that the valence of an individual review is drawn from a Bernoulli distribution with a success probability α (i.e., the product's true quality). The law of large numbers suggests that the average rating will converge to this true quality over the long term. Thus, if the average rating was too high at the beginning ($r_0 = 1$), it will decrease to α ; but if it was initially too low ($r_0 = 0$), it will rise to α . Therefore, the first-review effect on the valence becomes weaker over time, although, as Proposition 1 suggests, it does not completely disappear.

In contrast, the first-review effect on WOM volume not only persists but is also reinforced as time passes, according to the second part of the proposition. This result is obtained because, in every period, the number of newly arriving reviews is greater when the first review is positive than when it is negative. Note that Proposition 1 discusses the *total* volume of WOM in each period as a function of the first review. However, the same intuition applies to the newly arriving reviews in each period. In particular, because, by part (a) of Proposition 1, the positive first review shifts the valence of overall WOM upward, consumers are more likely to purchase the product and, thus, the number of newly arriving reviews in every period is likely to be greater when the first review is positive than when it is negative. Therefore, because of the difference in the first review, the incremental number of reviews accumulates in each period, which strengthens the impact of the first review on the overall volume period by period. This result implies that a small difference at

the beginning could lead to a drastic divergence at the end. In this sense, the online WOM phenomenon is subject to a severe path dependence, where the very first review determines the evolution path, and thus the fate of a product.

Overall, our theoretical analysis implies that a negative first review is detrimental to the performance of a product in several different ways. First, with a negative first review, the valence of WOM remains negative for a significant amount of time, which reduces sales, according to recent meta-analyses (You et al. 2015, Babić Rosario et al. 2016). At the same time, the volume of WOM for the product is lower, which again leads to lower sales, as shown by the aforementioned meta-analyses. Moreover, even though a negative bias in the average rating of a product can be corrected over time, it may have little or no chance of correction because of a lower review volume. In an extreme case, a product may not even take off as a result of a negative first review, resulting in only mediocre sales at best, and thus may not obtain many additional reviews, keeping the online sentiment toward the product negative. Therefore, we should not consider the first review as just a *single* review, but rather as *the most influential* review that, as such, must be properly managed.

So far, our theoretical analysis has generated four testable predictions regarding the significance of the first review, which we present in Table 2 (A1–A4). Given their counterintuitive nature, it is necessary to subject each of these predictions to empirical tests. In the next section, we present empirical studies to test these predictions.

3. Empirical Analysis

In this section, we report the results of our empirical investigations into the first-review effect. We use vacuum cleaners as the empirical context of our analysis because, in this category, consumers' purchase decisions are likely to be heavily influenced by product reviews. Note that vacuum cleaners are privately consumed durable goods with infrequent product trials and thus are susceptible to online WOM (You et al. 2015). Furthermore, our analysis of review texts reveals that consumers of vacuum cleaners consider more vertical attributes (i.e., quality) than horizontal ones (i.e., fit) in their purchase decisions (see Online Appendix B2 for details), which indicates that the vacuum cleaner category provides an appropriate context to test our theory.

To identify the first-review effect, we compare the volume and the valence of WOM when the first review is positive with those when it is negative. Our study specifically uses a sample of products that simultaneously appear on two different retail platforms, Amazon and Best Buy, and examines how

Table 3. Summary Statistics

	Amazon		Best Buy	
	Mean	Standard deviation	Mean	Standard deviation
Proportion of negative first reviews	0.254	0.437	0.181	0.386
Average rating	3.849	0.627	4.070	0.864
Number of reviews	380.655	673.972	58.226	104.865
Price	168.014	128.128	184.850	139.899
Word count in the first review	275.712	346.265	51.345	67.934
Number of days since the first review	1,101.175	804.593	778.492	613.688

different first reviews on different platforms could make a difference in the valence and the volume of the online reviews for the same product. In what follows, we first present our data as well as initial evidence that our theoretical predictions are in line with the empirical observations. We then describe formal tests of the theory in greater detail. Later, we also discuss the robustness of our results.

3.1. Data

We collect online review information as well as product characteristics on vacuum cleaners from both Amazon and Best Buy. For completeness, we cover in our data set all the vacuum cleaners with at least one review from both Amazon and Best Buy at the time of data collection (January 6, 2015). This results in a total of 177 vacuum cleaner models in our data set. For each of these products, we collect the number of reviews, the order of each review, the date each review was posted, the rating of each review, the review texts, and the price of the product from both platforms. Table 3 provides summary statistics of this data set.

Given these data, for each product, we calculate the average rating as well as the number of reviews each month since the first review was posted, which allows us to have multiple observations of the average rating and the number of reviews for each product across time. Note, however, that different products in our data set have different numbers of observations because they were launched, and their first reviews were posted, at different points in time. Thus, for products launched closer to the data collection time, observations are truncated earlier, thus resulting in a

smaller number of observations. Although we do not expect this data structure to cause any selection bias in the overall impact of the first review, the evolution pattern of WOM might be heavily influenced by products with longer durations. To minimize such an asymmetric influence while maintaining the representativeness of the sample, we confine our observation window to one year, during which about 71% of the products in the sample cover the whole observation period, whereas the remaining 29% have observations truncated at some point before the end of the observation period. In Section 3.6.1, we consider observation windows of different lengths to demonstrate the robustness of our findings.

3.2. Initial Evidence

Our theory concerns the valence and the volume of WOM as a function of the valence of the first review. We classify the first review as positive if its rating is strictly greater than three stars and as negative otherwise, following Amazon's dichotomization scheme. In our data set, 31.07% of 177 products exhibit inconsistent first reviews across the two platforms, according to this classification. Specifically, 21 products obtain a positive first review on Amazon and a negative one on Best Buy, whereas 34 products have a negative first review on Amazon and a positive one on Best Buy. By examining these two groups of products, one can potentially identify the first-review effect. We thus calculate both the average star rating and the average number of reviews on each platform for these two groups. We report the results in Table 4. First, Table 4 shows that the first group of products (with a

Table 4. Volume and Valence of WOM by the Valence of the First Review

First Review		Average rating			Number of reviews			Sample size
Amazon	Best Buy	Amazon	Best Buy	Difference	Amazon	Best Buy	Difference	
(+)	(−)	3.974	2.835	1.138***	66.143	14.952	51.190**	21
(−)	(+)	3.046	4.264	−1.218***	41.147	18.970	22.176*	34
(+)	(+)	4.099	4.386	−0.286***	92.216	35.108	57.108***	111
(−)	(−)	3.151	2.857	0.293	55.182	17.545	37.636**	11
All samples		3.823	4.083	−0.260***	77.011	28.525	48.486***	177

Note. This table is based on the paired product-level observations in the last month of each observation window.

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (paired t -test).

Table 5. Wilcoxon Signed-Rank Test

Period	Ranks for average rating			Ranks for number of reviews			Sample size
	First review		Difference	First review		Difference	
	(+)	(−)		(+)	(−)		
One year	60.324	89.081	−28.757***	64.946	74.000	−9.054*	37
Six months	75.870	113.739	−37.870***	77.174	91.652	−14.478**	46

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

positive first review on Amazon and a negative one on Best Buy) has a higher average rating as well as a higher number of reviews on Amazon than on Best Buy. This observation is in line with the first-review effect (i.e., a positive first review leads to more reviews with a higher average rating). Second, the other group (with a negative first review on Amazon and a positive one on Best Buy) exhibits a lower average rating but a higher number of reviews on Amazon than on Best Buy. Although the average rating moves in the same direction as our theory predicts, the result on the number of reviews seems to contradict the first-review effect. However, Table 4 also suggests that this might be due to a platform-specific effect. Specifically, we find that the products sold from both platforms on average receive 0.26 fewer stars but 48.49 more reviews on Amazon, across all samples (fifth row of Table 4). Thus, we need to control for the platform-specific effect for fair comparisons. To do so, we additionally conduct a nonparametric test that does not depend on the platform-specific effect. We will also formally address this issue when we develop our empirical model in the next section.

As an alternative test, we conduct the Wilcoxon signed-rank test. In this test, we can avoid the problem of unequal baselines across the two platforms by using ranks instead of raw data. Thus, for the two groups of products with different first reviews across the two platforms, we test whether the average ranks of the products on one platform are comparable to those of the same products on the other. However, because both the average rating and the number of reviews vary over time, to determine the ranking of products, we fix the time at the end of one year and consider only products that reach this particular time period. In this sample, 37 products have inconsistent first reviews. The first row of Table 5 reports the results from the comparison of these 37 products, where lower numbers imply higher rankings. As shown in this table, products typically rank higher for both the average rating and the number of reviews across the two platforms if the first review is positive rather than negative. To see whether these results are specific to the observation timing or the size of the sample, we also run the same test for a sample of 46 products that

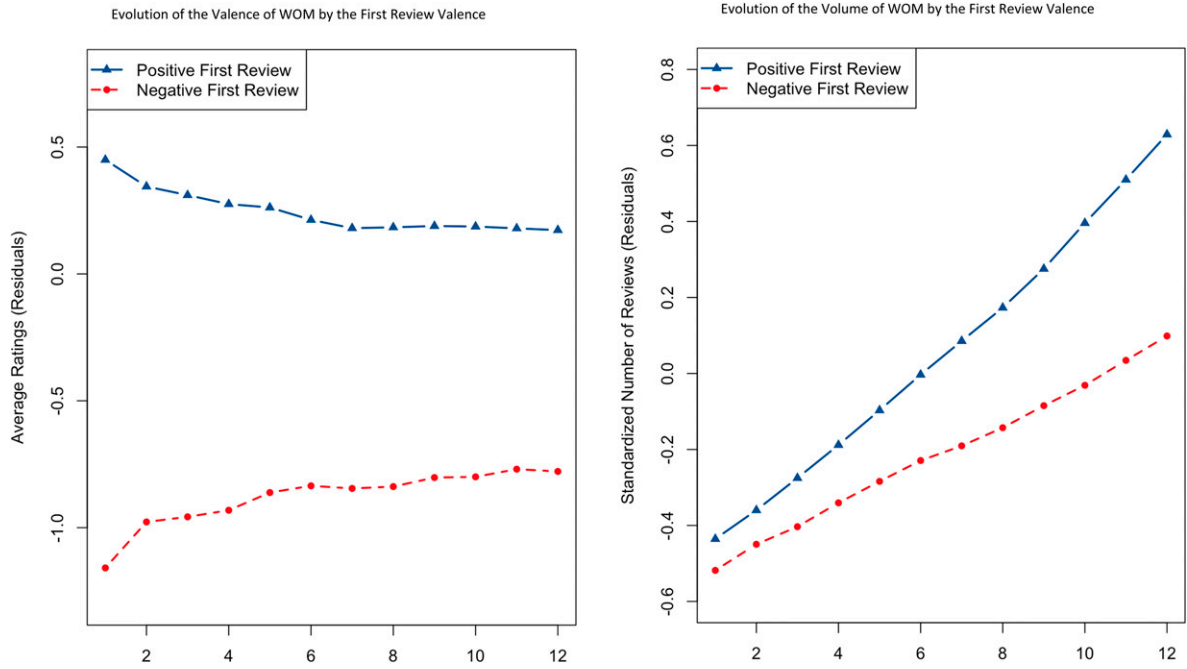
reached the end of the sixth month and report the results in the second row of Table 5. As can be seen from the table, we obtain qualitatively the same results, that is, in both metrics, the rankings are significantly higher for the same product when its first review is positive, rather than negative. Therefore, our preliminary analysis of the data supports the first-review effect: a positive first review generates more reviews with higher ratings than does a negative first review.

The above analyses have shown the static first-review effect. To examine the dynamic effect, we plot the evolution of the average ratings and the number of reviews in Figure 1. Note that in the figure, we control for both the platform-specific and the product-specific effects by plotting the residuals from the regression on these effects. The left panel of Figure 1 first confirms the static first-review effect on the WOM valence: the average rating is higher when the first review is positive than when it is negative. Moreover, it shows that the difference in the average rating decreases with time. Nonetheless, the difference does not disappear even after a year, implying that the first-review effect on the valence persists for a significant amount of time. The right panel of Figure 1 also reaffirms the static first-review effect on the WOM volume: a positive first review generates more reviews than a negative first review. More interestingly, the figure indicates that the difference in the number of reviews increases over time, thus confirming the dynamic effect of the first review on the volume. Overall, we find support from our data set for both static and dynamic first-review effects.

3.3. Identification Strategy

As noted above, our identification of the first-review effect is based on the cross-platform comparison of the review metrics of the same product. This implies that we can identify the first-review effect while controlling for unobserved heterogeneity specific to the product (for a similar approach, see also Chevalier and Mayzlin 2006). To better understand this, note that unobserved product-specific effects such as product quality may affect both the first review and later reviews. In general, such common factors for both dependent and independent variables, if not

Figure 1. (Color online) Evolution of the Valence (Left) and the Volume (Right) by the First Review Valence



Notes. To construct the figure, we respectively regress the average ratings (left panel) and the standardized number of reviews (right panel) on the product characteristics (types and bag and cord dummies) and site-specific fixed effects. Then, we compute the average of residuals for each month depending on the valence of the first review. For a fair comparison, we use a sample of products with at least 12-month-period observations. Note that the number of reviews is standardized at the platform level.

controlled, may lead to the endogeneity bias in the estimates. However, our estimation uses the within-product variation across two platforms, which, by definition, is independent of the product quality or other product-specific effects. Thus, the product-specific effects do not influence our estimation. Moreover, within-product variation in the reviews comes from each individual reviewer's unique experience of the product. Thus, within-product variation in the first review may well be exogenous to within-product variation in the other reviews (and hence, in our dependent measures, the average rating and the number of reviews). Therefore, in our identification strategy, endogeneity (due to unobserved product-specific effects) is unlikely to be an issue.

Our identification strategy also implies that the linear and separable platform-specific effects are differenced out in our estimation. However, it is possible that there might exist product-platform specific effects. Although our current approach does not directly rule out this possibility, in a later section, we show that such product-platform specific effects are unlikely to exist and that even if they do, the first-review effect still survives, using the second-differencing approach (see Section 3.6.2).

3.4. The First-Review Effect on the Valence of WOM

3.4.1. Model and Estimation. In this section, we formally test our predictions on the impact of the first

review on the valence of WOM (i.e., A1 and A3). Recall that A1 states that the average rating of a product is higher when its first review is positive than when it is negative. The same hypothesis also implies that this should hold for every period t . Moreover, A3 states that the first-review effect on the average rating diminishes over time. Note that the average rating of a product changes with time, as new reviews arrive. Thus, the average rating is specific not only to the product, but also to the time elapsed since the first review was posted. Therefore, we adopt the panel structure and consider the following model:

$$\begin{aligned} AR_{it}^j = & \beta_0^j + \beta_1 FNegative_i^j + \beta_2^j \ln(FDuration_{it}) \\ & + \beta_3 FNegative_i^j \times \ln(FDuration_{it}) \\ & + X_{it}^j \Gamma^j + \mu_i + \epsilon_{it}^j, \end{aligned} \quad (6)$$

where subscript i is the indicator for the product, subscript t is the indicator for the period, and superscript j is the indicator for the platform ($j = A, B$; A is Amazon, and B is Best Buy); AR_{it}^j is the average rating of product i at period t in platform j ; $FNegative_i^j$ is a dummy variable representing the valence of the first review of product i in platform j , following Amazon's dichotomization scheme presented above (1 = negative; 0 = positive); and $FDuration_{it}$ represents the duration of the first review of product i at period t , as measured by the number of months. Note that this is not a calendar date but the elapsed time since

the first review posting. Moreover, $FDuration_{it}$ always takes the same value as the period: $FDuration_{it} = t$. In the regression, because of a curvilinear relationship between the average rating and the duration of the first review, we use the log transformation of the duration.⁷ To capture potentially different rates of temporal change in the average ratings of the two platforms, we use platform-specific coefficients for the $FDuration$ (β_2^j).

The matrix \mathbf{X}_{it}^j contains control variables, including the log of the product price, the log of the number of words in the first review (WC), consumer search volume on Google (both at category and brand levels), product type dummies, product feature indicators, and the volume.⁸ We include these variables to control product specific and product-platform specific effects that might affect the evolution of online WOM. For example, a product's price might correlate with the valence of its reviews, the first review may be more influential if its review text is lengthy, and it is also possible that the WOM volume may influence the average ratings. Note, however, that using the volume itself will distort the estimate for the first-review effect on the valence, because the effect through the volume will be cut out. To capture the full effect of the first review on the valence while controlling for the volume, we use a volume variable that is independent of the first-review effect. In particular, we use the residual from the regression of the volume on

$FNegative$, $\ln(FDuration)$, $FNegative \times \ln(FDuration)$, and all controls in the matrix \mathbf{X}_{it}^j other than the volume variable itself (see Online Appendix B3, for more details on this first-review-independent volume variable).⁹ Last, μ_i captures the unobserved heterogeneity specific to the product, and ε_{it}^j represents an unobserved heterogeneity specific to both the product and the platform, both of which are not explained by the independent variables. We assume the errors to be normally distributed with a mean of zero.

Recall that we identify the first-review effect by the cross-website variation of the same product. To implement this identification strategy, we difference the average ratings in the two platforms for each product i and each period t . By doing this, we also eliminate the linear and separable product-level fixed effects. We thus obtain the following estimation equation:

$$\begin{aligned} \Delta AR_{it} = & \beta_0 + \beta_1 \Delta FNegative_i + \beta_2 \ln(FDuration_{it}) \\ & + \beta_3 \Delta FNegative_i \times \ln(FDuration_{it}) \\ & + \Gamma^A \mathbf{X}_{it}^A - \Gamma^B \mathbf{X}_{it}^B + \varepsilon_{it}, \end{aligned} \quad (7)$$

where $\beta_0 \equiv \beta_0^A - \beta_0^B$, $\beta_2 \equiv \beta_2^A - \beta_2^B$, $\Delta AR_{it} \equiv AR_{it}^A - AR_{it}^B$, and $\Delta FNegative_{it} \equiv FNegative_{it}^A - FNegative_{it}^B$. Note that for the control variables taking the same value across the two platforms (i.e., product type dummies and product feature indicators), Γ^A and Γ^B will not be separately identified, but only their difference will be. These coefficients capture the product-platform

Table 6. First-Review Effects on the Valence of WOM

	(1)		(2)	
	Estimate	Standard error	Estimate	Standard error
$\Delta FNegative$	-1.314***	0.228	-1.313***	0.225
$\ln(FDuration)$	-0.122***	0.034	-0.120***	0.035
$\Delta FNegative \times \ln(FDuration)$	0.412***	0.079	0.415***	0.077
$\ln(Price^A)$	-0.331	0.311	-0.340	0.310
$\ln(Price^B)$	-0.279	0.330	0.284	0.330
$\ln(WC^A)$	0.107	0.072	0.109	0.072
$\ln(WC^B)$	0.121*	0.066	-0.121*	0.066
Type handheld	-0.019	0.367	-0.007	0.370
Type robotic	0.060	0.379	0.068	0.379
Type stick	-0.097	0.321	-0.094	0.322
Type upright	-0.044	0.279	-0.046	0.280
Bagged	-0.335	0.293	-0.327	0.294
Corded	0.034	0.158	0.044	0.158
Search product ^A	-0.007	0.004	-0.013*	0.008
Search product ^B	0.007*	0.004	-0.013	0.009
Search brand ^A	-0.003	0.006	-0.004	0.007
Search brand ^B	-0.008	0.005	0.009*	0.005
First-review-independent volume ^A	0.081*	0.041	0.085**	0.042
First-review-independent volume ^B	0.015	0.030	-0.015	0.030
Month fixed effects	No		Yes	
Observations	1,488		1,488	
R ²	0.2390		0.2467	

Note. Standard errors are clustered at the product level.

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

specific effects, such as the stronger preference of Amazon consumers for a certain product type. For other controls, both Γ^A and Γ^B can be separately identified. The error term ($\varepsilon_{it} = \varepsilon_{it}^A - \varepsilon_{it}^B$) is the difference in the error terms and, thus, is also normally distributed with a mean of zero. Therefore, the error term still captures unobserved heterogeneity specific to the product. This also implies that errors are likely to be correlated across periods within a product. In this case, ordinary least squares (OLS) still produces consistent estimates of the model parameters, but the standard errors will be underestimated (Bertrand et al. 2004). Hence, we estimate our main parameters using OLS, but we also cluster standard errors by product to account for potential serial correlation (Liang and Zeger 1986). In our estimation, we only use 138 products from our data set that received at least two reviews from both platforms, because with only one review, the average rating and *FNegative* are almost perfectly correlated. Note that these 138 products include those with inconsistent first reviews across the two platforms, and also those with consistent first reviews, with the former identifying the first-review effect and the latter capturing the baseline difference across the two platforms.

3.4.2. Results. We report our estimation results in column (1) of Table 6. The table shows that the data support both of our predictions regarding the valence of WOM (i.e., A1 and A3). To see this, note that the coefficient for $\Delta FNegative$ (i.e., β_1) represents the first-review effect on the valence in the first period ($t = 1$), whereas that for $\Delta FNegative \cdot \ln(FDuration)$ (i.e., β_3) captures the dynamic effect of the first review. According to column (1) of Table 6, $\hat{\beta}_1$ is negative and significant ($\hat{\beta}_1 = -1.314$, $p < 0.01$), which implies that the average rating of the first month is lower when the first review is negative than when it is positive.

Column (1) of Table 6 also shows that $\hat{\beta}_3$ is positive and significant ($\hat{\beta}_3 = 0.412$, $p < 0.01$), suggesting that since the first month, the effect of the first review weakens over time. Based on these estimates, we examine specific predictions in A1 and A3.

First, we examine whether the first review indeed influences the average rating in every period (A1). In particular, we test whether the average rating is lower when the first review is negative than when it is positive for the entire sample period ($t = 1, 2, \dots, 12$). As reported in Table 7, for every t , we find that $\hat{\beta}_1 + \ln(t)\hat{\beta}_3$ is negative and significant ($p < 0.01$ for $t = 1, \dots, 9$; $p < 0.05$ for $t = 10, 11$; $p < 0.1$ for $t = 12$). As an alternative test, we separately estimate each of the period-specific first-review effects by interacting $\Delta FNegative$ with each period dummy and examine their signs. The results in Table 8 show that the estimates for all these interaction terms are significant ($p < 0.05$). These results confirm that the first-review effect on the valence is consistent with the prediction A1.

Second, we examine the dynamic effect of the first review (A3). As noted above, the estimate of $\Delta FNegative \cdot \ln(FDuration)$ ($\hat{\beta}_3$) shows the sign opposite to that of the estimate of $\Delta FNegative$ ($\hat{\beta}_1$). This result indicates that the difference in the average ratings caused by the first review decreases over time. Yet, the effect does not disappear for an extended period of time, as A1 suggests. For example, our estimation results predict that receiving a negative (rather than positive) first review for a product reduces the average rating by 0.58 stars by the end of the sixth month ($\hat{\beta}_1 + \ln(6)\hat{\beta}_3 = -0.58$), and by 0.29 stars by the end of the first year ($\hat{\beta}_1 + \ln(12)\hat{\beta}_3 = -0.29$). According to our estimates, even after 12 months, the average rating is expected to be lower when the first review is negative than when it is positive. More surprisingly, as Section 3.6.1 will later show, our estimation with longer observation windows

Table 7. Period-Specific First-Review Effects Based on the Estimates in Tables 6 and 9

<i>FDuration</i> (t)	Dependent variable: Valence		Dependent variable: Volume	
	$\hat{\beta}_1 + \ln(t)\hat{\beta}_3$	p -value for H_0 : $\beta_1 + \ln(t)\beta_3 = 0$	$\hat{\beta}_1 + (t-2) \cdot \hat{\beta}_3$	p -value for H_0 : $\beta_1 + (t-2) \cdot \beta_3 = 0$
1	-1.314	0.000		
2	-1.029	0.000	-0.019	0.181
3	-0.862	0.000	-0.073	0.000
4	-0.743	0.000	-0.127	0.000
5	-0.651	0.000	-0.180	0.000
6	-0.576	0.000	-0.234	0.000
7	-0.513	0.001	-0.287	0.000
8	-0.458	0.002	-0.341	0.000
9	-0.409	0.006	-0.395	0.000
10	-0.366	0.014	-0.448	0.000
11	-0.326	0.029	-0.502	0.000
12	-0.290	0.053	-0.555	0.000

Table 8. Period-Specific First-Review Effects

	Dependent variable: Valence		Dependent variable: Volume	
	Estimate	SE	Estimate	SE
$\Delta FNegative \times I(FDuration = 1)$	-1.259***	0.222		
$\Delta FNegative \times I(FDuration = 2)$	-1.042***	0.194	-0.014	0.023
$\Delta FNegative \times I(FDuration = 3)$	-0.917***	0.192	-0.072***	0.021
$\Delta FNegative \times I(FDuration = 4)$	-0.840***	0.203	-0.127***	0.016
$\Delta FNegative \times I(FDuration = 5)$	-0.638***	0.178	-0.190***	0.019
$\Delta FNegative \times I(FDuration = 6)$	-0.535***	0.184	-0.235***	0.017
$\Delta FNegative \times I(FDuration = 7)$	-0.500***	0.177	-0.294***	0.019
$\Delta FNegative \times I(FDuration = 8)$	-0.448**	0.176	-0.341***	0.024
$\Delta FNegative \times I(FDuration = 9)$	-0.355**	0.152	-0.370***	0.021
$\Delta FNegative \times I(FDuration = 10)$	-0.338**	0.149	-0.456***	0.027
$\Delta FNegative \times I(FDuration = 11)$	-0.329**	0.128	-0.509***	0.028
$\Delta FNegative \times I(FDuration = 12)$	-0.320**	0.132	-0.552***	0.031

Notes. Standard errors (SEs) are clustered at the product level. The term $I(FDuration = t)$ is an indicator that takes a value of one if $FDuration$ is equal to t and zero otherwise.

** $p < 0.05$; *** $p < 0.01$.

suggests that a negative first review significantly decreases the average ratings even after 36 months. Thus, the first-review effect persists over time. Even though the average rating should theoretically converge to the true quality in the long run, it would take a significant amount of time to do so, because a negative first review induces a lower volume of subsequent reviews, as shown in Proposition 1(b).¹⁰ Therefore, the valence of the first review has a nontrivial and persistent impact on the valence of the overall WOM.

Finally, we estimate Equation (7) with month fixed effects included in order to control for any potential seasonality. As shown in column (2) of Table 6, we find very similar estimates for both β_1 and β_3 . This implies that either seasonality or some time effects cannot explain our results.

3.5. The First-Review Effect on the Volume of WOM

3.5.1. Model and Estimation. We now turn to the impact of the first review on the volume of WOM. In particular, we test predictions A2 and A4: in any period, the number of reviews is greater when the first review is positive than when it is negative; the gap in the number of reviews not only persists, but also intensifies as time passes. To test these predictions, we consider the following model: for $t \geq 2$,

$$\begin{aligned} \widetilde{NR}_{it}^j = & \beta_0^j + \beta_1^j FNegative_i^j + \beta_2^j FDuration_{it-2} \\ & + \beta_3^j FNegative_i^j \times FDuration_{it-2} \\ & + X_{it}^j \Gamma^j + \mu_i^j + \varepsilon_{it}^j, \end{aligned} \quad (8)$$

where \widetilde{NR}_{it}^j represents product i 's standardized number of reviews received up to period t in platform j , and all independent variables are defined as in the previous section, except that we include both lagged first-review-independent valence and lagged first-review-independent volume in X_{it}^j (see Online Appendix B3 for details).¹¹ Because of the inclusion of lagged variables, we estimate the model with the data only from the second period and on. Accordingly, we use $FDuration_{it-2}$ to capture the dynamic effect after the second period. Finally, in our model, we use the dependent variable after standardizing it at the platform level: $\widetilde{NR}_{it}^j = (NR_{it}^j - \overline{NR}^j) / \sigma^j$, where \overline{NR}^j and σ^j respectively refer to the mean and the standard deviation of the number of reviews (NR_{it}^j). This is to make the two platforms comparable by removing the substantial difference in the variation of their number of reviews.¹²

As in the previous study, we identify the first-review effect by differencing Equation (8) for the two platforms for each product in each time period. We thus estimate the following equation:

$$\begin{aligned} \Delta \widetilde{NR}_{it} = & \beta_0 + \beta_1 \Delta FNegative_i + \beta_2 FDuration_{it-2} \\ & + \beta_3 \Delta FNegative_i \times FDuration_{it-2} \\ & + \Gamma^A X_{it}^A - \Gamma^B X_{it}^B + \varepsilon_{it}, \end{aligned} \quad (9)$$

where $\Delta \widetilde{NR}_{it} \equiv \widetilde{NR}_{it}^A - \widetilde{NR}_{it}^B$, and β_0, β_2 , and $\Delta FNegative_i$ are defined as before. We estimate Equation (9) using OLS with standard errors clustered at the product level.

3.5.2. Results. We report the estimation results in column (1) of Table 9. The estimates suggest that our data support both A2 and A4. Among the coefficients of our interests, the coefficient of $\Delta FNegative_i$ (β_1) represents the first-review effect on the volume in the second period ($t = 2$), and that of $\Delta FNegative_i \times FDuration_{it-2}$ (β_3) captures the dynamic effect of the first review. Table 9 shows that $\hat{\beta}_1$ is negative yet statistically insignificant ($\hat{\beta}_1 = -0.019$, $p = 0.181$), which indicates that in the second period, the first review has only an insignificant effect on the number of reviews, although its direction is consistent with the prediction in A2. This happens probably because in initial periods, the volume of reviews is too small to exhibit any significant difference (note that, e.g., in the second period, each product, on average, receives only 10.24 reviews). However, the same table also shows that $\hat{\beta}_3$ is negative and significant ($\hat{\beta}_3 = -0.054$, $p < 0.01$), which, together with the β_1 estimate, suggests that the

negative first review (compared with the positive one) decreases the number of reviews more period by period. Based on these estimates, the predictions A2 and A4 can be confirmed as follows.

First, to examine whether the first review positively influences the number of reviews in all subsequent periods, we compute $\hat{\beta}_1 + \hat{\beta}_3 \cdot FDuration_{it-2}$ for each value of t ($t = 2, 3, \dots, 12$). As shown in Table 7, we obtain negative and significant values ($p < 0.01$) in all periods except the second period ($t = 2$). Moreover, we estimate period-specific first-review effects on the volume by interacting $\Delta FNegative_i$ with period dummies and find that the estimated first-review effects are all negative and significant ($p < 0.01$) except for the second period (see Table 8). Therefore, our empirical analysis suggests that the negative first review decreases the number of reviews in the subsequent periods, which confirms A2.

Next, we examine the snowball effect (A4). Given that the static effect of the negative first review on the volume is negative, the negative (and significant) interaction between $\Delta FNegative_i$ and $FDuration_{it-2}$ (β_3) implies that such a static effect is intensified over time, thus confirming the snowball effect (A4). To illustrate this snowball effect, we calculate the estimated difference in the number of reviews (due to the first

review) in the middle and at the end of the data period (i.e., period 6 and period 12). We find that a positive first review (compared with a negative first review) is expected to increase the number of reviews by 0.23 standard deviation ($= \hat{\beta}_1 + (6 - 2) \cdot \hat{\beta}_3$) in period 6 and by 0.56 standard deviation ($= \hat{\beta}_1 + (12 - 2) \cdot \hat{\beta}_3$) in period 12. To put this in perspective, on Amazon, products with a negative first review generate 14.98 and 36.49 fewer reviews compared with those with a positive first review in period 6 and in period 12, respectively. This suggests that the difference in the number of reviews increases over time. Therefore, our analysis confirms that the first review has a significant and even increasing impact on the WOM volume. Finally, as in the valence model, we reestimate Equation (9) with month-level fixed effects included and find the qualitatively same results (see column (2) of Table 9). Thus, our results are unlikely to be explained by seasonality or time fixed effects.

We conclude this section by briefly discussing an empirical test of the mechanism behind our results. Recall from Section 2 that the first-review effect comes from the interdependence of the valence and the volume (as summarized in Theorem 1). Using the same data set, we find empirical support for both claims of Theorem 1, thus confirming the suggested

Table 9. First-Review Effects on the Volume of WOM

	(1)		(2)	
	Estimate	Standard error	Estimate	Standard error
$\Delta FNegative$	−0.019	0.015	−0.018	0.014
$FDuration$	0.065***	0.002	0.065***	0.002
$\Delta FNegative \times FDuration$	−0.054***	0.003	−0.054***	0.003
$\ln(Price^A)$	−0.128***	0.027	−0.129***	0.027
$\ln(Price^B)$	0.206***	0.025	−0.206***	0.025
$\ln(WC^A)$	0.133***	0.008	0.133***	0.008
$\ln(WC^B)$	0.063***	0.010	−0.063***	0.010
Type handheld	−0.252***	0.032	−0.247***	0.031
Type robotic	0.702***	0.047	0.704***	0.048
Type stick	0.243***	0.032	0.242***	0.033
Type upright	−0.126***	0.021	−0.125***	0.021
Bagged	−0.116***	0.023	−0.116***	0.023
Corded	−0.268***	0.027	−0.265***	0.027
Search product ^A	−1.62E-03	1.24E-03	−0.003*	0.002
Search product ^B	−2.48E-03**	1.06E-03	0.003*	0.002
Search brand ^A	0.005***	0.002	0.005***	0.002
Search brand ^B	0.013***	0.002	−0.013***	0.002
Lagged first-review-independent valence ^A	−0.004	0.009	−0.005	0.009
Lagged first-review-independent valence ^B	0.011	0.010	−0.010	0.010
Lagged first-review-independent volume ^A	1.169***	0.014	1.170***	0.014
Lagged first-review-independent volume ^B	1.097***	0.010	−1.09***	0.010
Month fixed effects	No		Yes	
Observations	1,648		1,648	
R ²	0.9776		0.9779	

Note. Standard errors are clustered at the product level.

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

theoretical mechanism for both types of first-review effects. The details of this analysis can be found in Online Appendix B4.

3.6. Robustness Checks

So far, we have identified the strong influence that the first review exerts on the valence and volume of a product's entire WOM. In this section, we examine the robustness of these findings by considering different observation windows, alternative explanations (unobserved product characteristics and the display effect), an alternative empirical model (simultaneous estimation), other contexts (a different category and a different pair of platforms), and alternative assumptions (temporal price variation and a platform-specific first-review effect). Our results show that the first-review effects remain robust to all these considerations and, thus, are truly robust findings. In what follows, we discuss the motivation, the method, and results for each of these analyses.

3.6.1. Observation Windows. In our main analysis, we use an unbalanced panel data set with a one-year observation window. We choose one year to balance between two objectives: minimizing heterogeneity in duration and minimizing the loss of information. However, it is important to determine how sensitive our results are to this specific choice of observation window. We thus consider two other observation windows (two years and three years) and estimate Equations (7) and (9). Note that as the observation period becomes longer, the percentage of products reaching

the end of the observation period decreases, whereas the total number of observations used in the estimation increases. We report the estimation results in column (1) (two-year window) and column (2) (three-year window) of Table 10 (valence study) and Table 11 (volume study). The full results are reported in Tables C1 (valence study) and C2 (volume study) in Online Appendix C. In the valence study, we find that the estimates for the main effect of $\Delta FNegative$ and its interaction with $\ln(FDuration)$ are consistent in both direction and significance with those of our main analysis in both models. Moreover, the period-specific first-review effect on the valence is negative and significant for all periods. In the volume study, the coefficients for both $\Delta FNegative$ and $\Delta FNegative \cdot FDuration$ are negative and significant, providing even stronger evidence than the main analysis.

Therefore, our results are robust to the length of the observation window. More importantly, by examining the first-review effects with longer observation windows, our analyses provide even stronger evidence for persistence as well as for the snowball effect.

3.6.2. Unobserved Product-Platform Specific Effects.

The first-review effect characterizes the relationship between the valence of the first review and the volume/valence of the entire review. However, it is possible that this relationship could be partially driven by differences in taste concerning certain unobserved product characteristics between two distinct consumer populations on the two platforms. If such unobserved product-platform specific effects exist

Table 10. Robustness Check: Valence of WOM

	(1)	(2)	(3)	(4)	(5)
	Observation window: 2 years	Observation window: 3 years	No control specification	Display effect	Platform-Specific first-review effect
$\Delta FNegative$	-1.418*** (0.242)	-1.475*** (0.259)	-1.309*** (0.226)	-1.325*** (0.226)	
$\ln(FDuration)$	-0.092*** (0.034)	-0.072** (0.036)	-0.109*** (0.031)	-0.128*** (0.035)	-0.083** (0.033)
$\Delta FNegative \times \ln(FDuration)$	0.361*** (0.080)	0.389*** (0.086)	0.408*** (0.077)	0.416*** (0.078)	
$FNegative^A$					-0.979*** (0.246)
$FNegative^B$					-1.787*** (0.273)
$FNegative^A \times \ln(FDuration)$					0.330*** (0.078)
$FNegative^B \times \ln(FDuration)$					0.534*** (0.114)
Observations	2,528	3,105	1,488	1,488	1,488
R^2	0.2670	0.2463	0.1636	0.2478	0.2634

Note. Standard errors are clustered at the product level and are reported in the parentheses.

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 11. Robustness Check: Volume of WOM

	(1)	(2)	(3)	(4)	(5)
	Observation window: 2 years	Observation window: 3 years	No control specification	Display effect	Platform-Specific first-review effect
$\Delta FNegative$	−0.097*** (0.014)	−0.065*** (0.014)	0.009 (0.043)	−0.018 (0.014)	
$FDuration$	0.034*** (0.001)	0.023*** (0.001)	0.059*** (0.015)	0.065*** (0.002)	0.074*** (0.003)
$\Delta FNegative \times FDuration$	−0.020*** (0.003)	−0.018*** (0.002)	−0.028* (0.016)	−0.054*** (0.003)	
$FNegative^A$					−0.049** (0.021)
$FNegative^B$					0.007 (0.023)
$FNegative^A \times FDuration$					−0.071*** (0.003)
$FNegative^B \times FDuration$					−0.029*** (0.004)
Observations	2,709	3,307	1,648	1,648	1,648
R^2	0.9808	0.9807	0.0279	0.9783	0.9826

Note. Standard errors are clustered at the product level and are reported in the parentheses.

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

but are omitted from the estimation, our estimate for $\Delta FNegative$ will be biased, and in extreme cases, the estimated first-review effect may merely reflect a spurious correlation. We investigate this issue through three different analyses. These analyses show that product-platform specific effects are unlikely to exist, and that even if they did, the first-review effect would survive such effects.

First, from our data, we do not find evidence that consumer tastes are different across the two platforms. In particular, we examine the average rating as one indication of the underlying preference and the number of reviews as a proxy of the popularity. In this analysis, we obtain these two preference measures across product groups defined by distinct characteristics (e.g., product types and product features) in two platforms and calculate the correlation of these measures across the two platforms. We find that the correlation between the average ratings of the two platforms is high (0.982 across types and 0.568 across features), and so is the correlation between the numbers of reviews (0.855 across types and 0.773 across features). This suggests that there is hardly any systematic difference in the preference for certain product characteristics between the two platforms' consumers.

Second, recall that unobserved product-platform specific effects would bias our results only to the extent that $\Delta FNegative$ correlates with these unobserved effects. In other words, to the extent we believe that $\Delta FNegative$ had been selected on the unobserved product-

platform specific effects, our estimate of $\Delta FNegative$ will be biased. We thus examine whether such a belief is well grounded. Specifically, we build on the result of Altonji et al. (2005), who have proved that under mild conditions, "selection on the unobservables is the same as selection on the observables."¹³ Thus, if $\Delta FNegative$ is not selected on observed controls, $\Delta FNegative$ will not be selected on the unobserved product-platform specific effects. To examine the potential selection of $\Delta FNegative$ on observed control variables, we estimate Equations (7) and (9), excluding all control variables, and then compare the estimates with those from the model with controls, that is, our original estimates in Tables 6 and 9 (for other applications of the no-control specification for concerns regarding selection on unobservables, see Cameron and Taber 2004, Altonji et al. 2005, Mayzlin et al. 2014). We report the estimation results from the no-control specification in column (3) of Tables 10 and 11. When we compare these results with those in Tables 6 and 9, we observe that the coefficients for the first review remain fairly stable with or without the inclusion of covariates. This result implies that $\Delta FNegative$ is less likely to be selected on observables and, thus, on unobservables as well. Hence, our findings on the first-review effect are unlikely to be driven by the unobserved product-platform specific effects.

Finally, we find consistent results regarding the first-review effect on the WOM volume, even after eliminating any potential unobserved product-platform

specific effects. In the presence of the product-platform fixed effects (γ_i^j), the Equation (8) can be rewritten as follows:

$$\begin{aligned}\widetilde{NR}_{it}^j &= \beta_0^j + \beta_1 FNegative_i^j + \beta_2 FDuration_{it-2} \\ &\quad + \beta_3 FNegative_i^j \times FDuration_{it-2} \\ &\quad + \mathbf{X}_{it}^j \boldsymbol{\Gamma}^j + \boldsymbol{\mu}_i + \gamma_i^j + \varepsilon_{it}^j.\end{aligned}\quad (10)$$

The first-differencing of Equation (10) across the two platforms yields the following:

$$\begin{aligned}\Delta \widetilde{NR}_{it} &= \beta_0 + \beta_1 \Delta FNegative_i + \beta_2 \Delta FDuration_{it-2} \\ &\quad + \beta_3 \Delta FNegative_i \times \Delta FDuration_{it-2} \\ &\quad + \boldsymbol{\Gamma}^A \mathbf{X}_{it}^A - \boldsymbol{\Gamma}^B \mathbf{X}_{it}^B + \gamma_i^A - \gamma_i^B + \varepsilon_{it}.\end{aligned}\quad (11)$$

In Equation (11), the product-level fixed effects ($\boldsymbol{\mu}_i$) is differenced out, yet the product-platform fixed effects ($\gamma_i^A - \gamma_i^B$) is not. Furthermore, we can remove the latter by using the second-differencing approach in which we take the difference of Equation (11) across time (for a similar approach, see Chevalier and Mayzlin 2006). In this case, even if the product-platform specific effects indeed exist, they will be eliminated, and the estimation will be free of the bias. The resulting estimation equation is given as

$$\begin{aligned}\Delta \widetilde{NR}_{it} - \Delta \widetilde{NR}_{it-1} &= \beta_2 + \beta_3 \Delta FNegative_i + \boldsymbol{\Gamma}^A \Delta \mathbf{X}_{it}^A \\ &\quad - \boldsymbol{\Gamma}^B \Delta \mathbf{X}_{it}^B + u_{it},\end{aligned}\quad (12)$$

where $\Delta \mathbf{X}_{it}^j \equiv \mathbf{X}_{it}^j - \mathbf{X}_{it-1}^j$ and $u_{it} = \varepsilon_{it} - \varepsilon_{it-1}$. Recall that the coefficient of our primary interest is β_3 , which represents an additional disadvantage from the negative first review in each period. The estimation results are reported in Table 12, which suggests that $\hat{\beta}_3$ is negative and highly significant ($p < 0.01$). This implies that, consistent with the main analysis, the disadvantage from a negative first review increases over time, thus

supporting the snowball effect. Therefore, the first-review effect on the volume survives any potential unobserved product-platform specific effects.

3.6.3. Display Effect. Recall that our theoretical model suggests that the first review makes a difference in WOM volume because a positive first review induces a higher average rating, which in turn leads to increased sales. However, it is possible that higher sales and more WOM in the case of a positive first review might stem from a greater likelihood that the product receives a first-page listing in search results on the retail platform. To examine whether the first-review effect survives this display effect, we estimate the model of volume with an additional control variable, the first-page dummy, to which we assign a value of one if the product appears on the first page at the time of our data collection and zero otherwise.¹⁴ Clearly, our first-page dummy variable is a static measure and may not be ideal for representing a potentially dynamic display effect. However, our additional analysis confirms that the first-page appearance typically remains stable (see Online Appendix B5). We thus use the first-page dummy variable to account for the display effect in our estimation. We report the results in column (4) of Table 11. The full results are reported in column (4) of Table C2 in Online Appendix C. From this analysis, we first find that the first-page display effect indeed has a significant impact on the number of reviews both on Amazon and Best Buy. More importantly, however, even after controlling for the display effect, we still find a significant and increasing first-review effect on WOM volume. Therefore, the first-review effect on WOM volume is robust to the presence of the display effect.¹⁵

3.6.4. Simultaneity Between Valence and Volume. Our theoretical analysis suggests that the valence and the volume of online WOM influence each other in their

Table 12. Second-Differencing Model: Volume of WOM

	Estimate	Standard error
$\Delta FNegative$	-0.046***	0.004
$\Delta Search\ product^A$	1.22E-03**	6.02E-04
$\Delta Search\ product^B$	1.85E-03***	5.04E-04
$\Delta Search\ brand^A$	-2.34E-03	1.63E-03
$\Delta Search\ brand^B$	-2.42E-03	1.74E-03
$\Delta Lagged\ first-review-independent\ valence^A$	0.030***	0.010
$\Delta Lagged\ first-review-independent\ valence^B$	-0.015	0.016
$\Delta Lagged\ first-review-independent\ volume^A$	0.807***	0.049
$\Delta Lagged\ first-review-independent\ volume^B$	0.642***	0.087
Observations	1,478	
R^2	0.6078	

Notes. Standard errors are clustered at the product level. First- and second-period observations are dropped in the regression because of the use of lagged proxies and second differencing.

** $p < 0.05$; *** $p < 0.01$.

Table 13. Seemingly Unrelated Regression

	Valence equation	Volume equation
$\Delta FNegative$	−1.380*** (0.140)	−0.005 (0.019)
$\ln(FDuration)$	−0.139*** (0.039)	
$FDuration$		0.074*** (0.002)
$\Delta FNegative \times \ln(FDuration)$	0.444*** (0.076)	
$\Delta FNegative \times FDuration$		−0.065*** (0.003)
Observations	1,347	1,347
R^2	0.2261	0.9764

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

evolution. Thus, our empirical study controls for the effect of the two variables on each other by including first-review-independent volume and valence respectively. However, the question remains whether the potential simultaneity between the two variables might bias the estimated first-review effects. Although we do not find evidence that such simultaneity bias exists in our data (see Online Appendices B6 and B7 for the Hausman test and the robustness checks with alternative sets of volume/valence variables), we also show that the first-review effects remain stable even after explicitly accounting for the interdependency in the coevolving process of the two variables. In particular, we jointly estimate the valence equation and the volume equation on a seemingly unrelated regression framework using feasible generalized least squares (Zellner 1962; for a similar application, see also Godes and Mayzlin 2004). The estimation results are presented in Table 13. The full results are reported in Table C3 in Online Appendix C. The table shows that coefficients of our interests are almost identical and significant, suggesting that the first-review effects are robust to the possibility of simultaneity bias.

3.6.5. Different Product Category. The empirical context for our main analysis is the market for vacuum cleaners. To strengthen the external validity of our results, we also examine the first-review effect using data from a different product category. We specifically choose toasters, because toasters and vacuum cleaners share characteristics typical of a product category that is sensitive to online WOM, such as private consumption, durability, and infrequent trials (You et al. 2015). As in our main study, we collect reviews and product information for all available toasters with at least one review on both Amazon and Best Buy, posted as of December 22, 2016. This results in a total of 113 products, among which 40.7% received inconsistent first reviews across the two platforms. We report summary statistics of this sample in Online Appendix B8.

By using the toaster data, we estimate Equations (7) and (9) with observation windows of one, two, and three years. We use the same set of control variables as in the vacuum cleaner study, except for the product feature and the brand search volume on Google (which is either unavailable or too sparse to use). For the product type, we use the toaster dummy, which

Table 14. Toasters: Valence of WOM

	(1)	(2)	(3)
	Observation window: 1 year	Observation window: 2 years	Observation window: 3 years
$\Delta FNegative$	−2.049*** (0.239)	−2.009*** (0.248)	−2.010*** (0.265)
$\ln(FDuration)$	−0.141*** (0.051)	−0.140*** (0.050)	−0.141*** (0.052)
$\Delta FNegative \times \ln(FDuration)$	0.551*** (0.096)	0.398*** (0.095)	0.396*** (0.097)
Observations	872	1,767	2,378
R^2	0.4180	0.4154	0.3823

Note. Standard errors are clustered at the product level and are reported in the parentheses.

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

separates out the toaster type (1) from the oven type (0). We report the estimation results in Table 14 (valence study) and Table 15 (volume study). The full results are reported in Tables C4 (valence study) and C5 (volume study) in Online Appendix C. These tables show that across all three observation windows, in the valence model, $\hat{\beta}_1$ is negative and significant, $\hat{\beta}_3$ is positive and significant, but the predicted first-review effect (i.e., $\hat{\beta}_1 + \ln(t) \cdot \hat{\beta}_3$) is negative and significant for all $t = 1, \dots, 12$; in the volume model, both $\hat{\beta}_1$ and $\hat{\beta}_3$ are negative and significant. These results suggest that all of our predictions, A1–A4, are supported even in the toaster category. Therefore, our findings concerning the first-review effect are robust to the choice of product category.

3.6.6. Different Platforms. In our main analysis, we compare reviews of the same products sold on both Amazon and Best Buy. To examine whether our findings are specific to this choice of platforms or can be generalized, we repeat our analysis on a different pair of platforms: Amazon U.S. and Amazon Canada.¹⁶ By using this pair, we can also reduce the concern for unobserved product-platform specific effects (as discussed in Section 3.6.2). This is because the taste difference between their consumer populations, if such exists, is of a different type than that of the Amazon–Best Buy pairing, and thus, if the first-review effect is also confirmed in this case, it is unlikely to be driven by unobserved product-platform specific effects. In this study, we also use the vacuum cleaner category and collect both reviews and product information that were available as of February 8, 2017. As before, we include in our sample all products with at least one review on both platforms, resulting in a total of 174 products. Of these, 29.9% received a first review of differing valence on each of

the two platforms. We report summary statistics in Online Appendix B8.

Based on these data, we estimate Equations (7) and (9) using the same set of controls as in the Amazon–Best Buy study. Note that we convert the prices in Amazon Canada into U.S. dollars based on contemporaneous exchange rates. We also use one year as our observation window, because only a limited number of products reach the end of the second year. We report the estimation results in the first columns of Table 16 (valence study), and Table 17 (volume study). The full results are reported in column (1) of Tables C6 (valence study) and C7 (volume study) in Online Appendix C. We obtain qualitatively the same results in this study: in the valence study, $\hat{\beta}_1 = -1.825$ ($p < 0.01$), $\hat{\beta}_3 = 0.434$ ($p < 0.01$), and $\hat{\beta}_1 + \ln(t) \cdot \hat{\beta}_3 < 0$, $\forall t = 1, \dots, 12$, ($p < 0.01$); in the volume study, $\hat{\beta}_1 = 0.049$ ($p = 0.054$) and $\hat{\beta}_3 = -0.050$ ($p < 0.01$), but the predicted first-review effects (i.e., $\hat{\beta}_1 + \hat{\beta}_3 \cdot \text{FDuration}_{it-2}$) are negative and significant in all periods other than the second and third periods, probably because of low volume in early periods.¹⁷ Therefore, all of our predictions, A1–A4, are supported on this alternative pair of platforms, and, thus, our findings regarding the first-review effect are robust to platform choice.

3.6.7. Temporal Price Variation. In our main study, we use static price data as one of the control variables. Although previous literature has taken the same approach (e.g., Li and Hitt 2008, Godes and Silva 2012), it is based on an implicit assumption that prices do not change over time, or that at least temporal price variations are not contemporaneously correlated with both independent and dependent variables. Although we do not believe that this assumption distorts our results,¹⁸ it is still useful to relax the assumption and confirm the first-review effect while using dynamic prices as a control. To conduct this analysis,

Table 15. Toasters: Volume of WOM

	(1)	(2)	(3)
	Observation window: 1 year	Observation window: 2 years	Observation window: 3 years
$\Delta \text{FNegative}$	−0.074*** (0.023)	−0.112*** (0.017)	−0.071*** (0.011)
FDuration	0.044*** (0.004)	0.021*** (0.001)	0.007*** (0.001)
$\Delta \text{FNegative} \times \text{FDuration}$	−0.050*** (0.003)	−0.022*** (0.001)	−0.017*** (0.001)
Observations	1,110	1,994	2,597
R^2	0.9714	0.9889	0.9933

Note. Standard errors are clustered at the product level and are reported in the parentheses.

** $p < 0.05$; *** $p < 0.01$.

Table 16. Amazon U.S. vs. Amazon Canada: Valence of WOM

	(1)	(2)
	Static prices	Dynamic prices
$\Delta FNegative$	-1.825*** (0.197)	-1.206*** (0.319)
$\ln(FDuration)$	-0.061 (0.045)	-0.029 (0.061)
$\Delta FNegative \times \ln(FDuration)$	0.434*** (0.083)	0.436*** (0.144)
Observations	1,249	436
R^2	0.4046	0.2696

Note. Standard errors are clustered at the product level and are reported in the parentheses.

* $p < 0.10$; *** $p < 0.01$.

we collect the historical price data for both Amazon U.S. and Amazon Canada from the website Keepa.com and merge them with the review data collected for the analysis in Section 3.6.6. However, Keepa.com has only been in service since 2014, and, thus, if a product's first review was posted prior to that date, price data are not available. Hence, in our analysis, we include only 56 products for which we could match the price history data with our review data from both Amazon U.S. and Amazon Canada. In this sample, 32.1% of the products received inconsistent first reviews across the two platforms.

Using this data set, we estimate Equations (7) and (9) with the same set of control variables as used in the previous section (other than prices), based on the one-year observation window. We report the estimation results in the second columns of Table 16 (valence study) and Table 17 (volume study). The full results are reported in column (2) of Tables C6 (valence study) and C7 (volume study) in Online Appendix C. As can be easily seen, the signs of the coefficients of our interests are consistent with our theoretical predictions, and we find supports for all the predictions, A1–A4. Therefore, the

first-review effect holds regardless of the assumption on temporal price variation.

3.6.8. Platform-Specific First-Review Effects. Recall that we base our identification of the first-review effect on the comparison of reviews for the same products with inconsistent first reviews across the two platforms. To implement this identification strategy, we have considered common coefficients for $\Delta FNegative$ and its interaction with $FDuration$ on the two platforms. In doing so, we implicitly assume that the magnitude of the first-review effect is identical across the two platforms. However, it is possible that the first-review effect could be different across platforms. Hence, we examine the sensitivity of our results to the assumption of identical versus platform-specific first-review effects.

For this purpose, we replace β_1 and β_3 in Equations (6) and (8) with platform-specific parameters β_1^j and β_3^j and estimate the following equations:

$$\Delta AR_{it} = \beta_0 + \beta_1^A FNegative_i^A - \beta_1^B FNegative_i^B + \beta_2 \ln(FDuration_{it}) + \beta_3^A FNegative_i^A \times \ln(FDuration_{it}) - \beta_3^B FNegative_i^B \times \ln(FDuration_{it}) + \mathbf{X}_{it}^A \Gamma^A - \mathbf{X}_{it}^B \Gamma^B + \varepsilon_{it}, \quad (13)$$

$$\Delta \widetilde{NR}_{it} = \beta_0 + \beta_1^A FNegative_i^A - \beta_1^B FNegative_i^B + \beta_2 FDuration_{it} + \beta_3^A FNegative_i^A \times FDuration_{it} - \beta_3^B FNegative_i^B \times FDuration_{it} + \mathbf{X}_{it}^A \Gamma^A - \mathbf{X}_{it}^B \Gamma^B + \varepsilon_{it}. \quad (14)$$

This model uses a different identification strategy: identification of β_1^j and β_3^j now derives from cross-product variation within a platform, rather than from cross-platform variation of given products. Note that the product-specific effects are still controlled for, but do not appear in the above estimation equations because they are differenced out. We report the estimation results in column (5) of Tables 10 and 11.

Table 17. Amazon U.S. vs. Amazon Canada: Volume of WOM

	(1)	(2)
	Static prices	Dynamic prices
$\Delta FNegative$	0.049* (0.025)	-0.060 (0.053)
$FDuration$	0.028*** (0.003)	0.029*** (0.006)
$\Delta FNegative \times FDuration$	-0.050*** (0.005)	-0.053*** (0.012)
Observations	1,612	491
R^2	0.9655	0.9635

Note. Standard errors are clustered at the product level and are reported in the parentheses.

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

The full results are reported in column (5) of Tables C1 and C2 in Online Appendix C. As shown in these tables, we can replicate our results with platform-specific first-review parameters. In particular, in the valence study, the signs of all the relevant first-review effect parameters for both platforms are consistent with our predictions and significant. Moreover, the predicted first-review effect based on these estimates is negative and significant in all periods and on both platforms. In the volume study, all of the first-review effect coefficients are negative and statistically significant except the estimated coefficient of $FNegative_i^B(\hat{\beta}_1^B)$, which is insignificant. Here, the predicted first-review effect is negative and significant in all periods on Amazon and after the second period on Best Buy. Therefore, we can confirm across all studies that the first review has a significant influence on both the valence and volume of online WOM, regardless of whether it is common or distinct across different platforms.

Finally, one may wonder whether the first-review effects are indeed different across platforms. To answer this question, we compare the estimates of the first-review effect parameters for each pair of platforms used in all of our analysis. First, we find that the first-review effect on the valence is greater on Best Buy than on Amazon. This is probably because any bias in the first review is more easily corrected on Amazon because of its large volume of reviews. Second, the first-review effect on WOM volume is larger on Amazon than on Best Buy. This might be due to a larger customer base on Amazon, which amplifies the sales increase more after a positive first review (versus a negative first review).

4. Extension

So far, we have theoretically uncovered the interdependence of online review valence and volume in their evolution, from which we derived four testable predictions on how an advantage in the first review translates to long-lasting advantages in the valence and even increasing advantages in the volume of later periods. All of these predictions were strongly supported by the data collected on the vacuum cleaner category from Amazon and Best Buy. We also have shown that our empirical findings are robust to a different category or a different platform. This strong empirical support may motivate one to wonder whether the magnitude of the first-review effect would differ across different product categories and, if so, what could moderate the magnitude.

In answering these questions, we note that, in reality, consumers make purchase decisions based not only on online consumer reviews but also other nonreview information such as offline WOM, firm-initiated communications, or observation of other

Table 18. Camera: Valence of WOM

	Estimate	Standard error
$\Delta FNegative$	−1.723***	0.210
$\ln(FDuration)$	−0.063	0.039
$\Delta FNegative \times \ln(FDuration)$	0.341***	0.073
Observations	1,317	
R^2	0.4105	

Notes. Standard errors are clustered at the product level.

*** $p < 0.05$; ** $p < 0.01$.

consumers' usage. Although our theoretical model as well as the empirical context (i.e., the vacuum cleaner category) focus on the categories where online WOM plays a dominant role in consumers' decisions, as the influence of nonreview information increases, the first-review effect may become weaker. We thus extend our (theoretical and empirical) analyses to cases where consumers' purchase decisions are influenced by both review and nonreview information and examine the impact of nonreview information on the strength of the first-review effect.

Recall that our theoretical model assumes that consumers form quality perception from reviews with some noise (minus the disutility from the taste mismatch). Assuming that nonreview information is overall unbiased, consumers' valuation can be revised as

$$v_t = v_0 + \beta[(1 - \gamma) \cdot r_{t-1} + \gamma \cdot \alpha] - \tau \cdot x,$$

where $\gamma \in [0, 1]$ measures their relative reliance on nonreview information. Note that this formulation includes our main model as a special case (with $\gamma = 0$). On numerically analyzing this general model (with the details deferred to Online Appendix A2), we obtain the following results.

Result 1 (Impact of Nonreview Information). *As consumers rely more on the nonreview information,*

- the first-review effect on the volume becomes weaker,*
- the first-review effect on the valence can get either stronger or weaker.*

To understand these results, first note that as consumers rely more on the nonreview information, the impact of the first review on the consumers' valuation in later periods will become smaller. This, in turn, decreases the impact of the first review on the future

Table 19. Camera: Volume of WOM

	Estimate	Standard error
$\Delta FNegative$	0.017	0.017
$FDuration$	0.063***	0.003
$\Delta FNegative \times FDuration$	−0.034***	0.003
Observations	1,879	
R^2	0.9664	

Notes. Standard errors are clustered at the product level.

*** $p < 0.01$.

sales, eventually leading to a weaker first-review effect on the number of reviews. Therefore, in a category with a weaker WOM influence, the first review has much weaker impact on the future WOM volume. This explains our first finding, that is, part (a) of Result 1. However, the same does not hold for the first-review effect on the WOM valence. The finding in part (b), though less intuitive, is the consequence of part (a). To see this, note that the weaker first-review effect on the volume implies that greater dependence on nonreview information decreases the number of newly posted reviews when the first review is positive, but increases it when the first review is negative. Thus, as the influence of nonreview information increases, a positive bias in the first review is less likely to be corrected in future ratings, but a negative bias is more likely to be corrected. Overall, the persistence of the bias of the first review in the future valence could be either stronger or weaker depending on the relative size of the above two effects. Therefore, even in a category with weaker online WOM effect, the first-review effect on the valence may be either stronger or weaker.

These two distinct theoretical predictions warrant empirical validation. According to a recent meta-analysis (You et al. 2015), online WOM is likely to have a strong influence on product sales in categories of privately consumed and infrequently purchased durable goods such as vacuum cleaners. This meta-analysis result suggests that in a category where a product is publicly consumed and, hence, other consumers' adoptions are easily observed, product sales are less likely to be affected by the online WOM (You et al. 2015). In such a category, consumers are more likely to use the nonreview information. We identify digital cameras as one category that would serve as a good example of such a category.¹⁹ We thus examine the first-review effect in the digital camera category by collecting both product and review data from Amazon U.S. and Amazon Canada. We obtain 190 cameras in our data set after matching the products with at least a single review from the two platforms and report the summary statistics in Online Appendix B8. Based on this data set, we estimate Equations (7) and (9) and report the results in Tables 18 and 19. We report full results in Tables C8 and C9 in Online Appendix C.

Although our theory is agnostic about the magnitude of the first-review effect on the valence of WOM in this category, we continue to find strong evidence for it. According to Table 18, the estimate for $\Delta FNegative(\hat{\beta}_1)$ is negative and significant ($p < 0.01$), and the estimate for $\Delta FNegative \times \ln(FDuration)(\hat{\beta}_3)$ is positive and statistically significant ($p < 0.01$). The predicted first-review effect in each period ($\hat{\beta}_1 + \ln(t) \cdot \hat{\beta}_3$) is also negative and significant ($p < 0.01$).

We also find the support for the first-review effect on the WOM volume. Table 19 shows that the coefficient for $\Delta FNegative$ is not significant, whereas the estimated coefficient for $\Delta FNegative \times FDuration(\hat{\beta}_3)$ is negative and significant ($p < 0.01$). To compare the magnitude of the first-review effect on WOM volume between vacuum cleaners and digital cameras, we jointly estimate the product-category-specific first-review effects by pooling the data of the two categories. We find no evidence of significant difference in the effect between two categories from period 2 to period 6. However, the effect becomes significantly greater for vacuum cleaners than for digital cameras from period 7 to period 12 ($p < 0.01$). This result is consistent with the above theoretical predictions. In sum, our investigation shows that the relative strength of the first-review effect varies across the product category, depending on the level of consumers' reliance on nonreview information.

5. Conclusion

Two key metrics of WOM, valence and volume, have been independently found to influence the sales of a product (Floyd et al. 2014, You et al. 2015, Babić Rosario et al. 2016). The central question of this paper is whether these two metrics are interrelated and, if so, what implications this interdependence might have for the management of online consumer reviews. Our theoretical model of consumer purchase decisions and the review updating process establishes the interdependence of valence and volume of online consumer reviews, which leads us to derive the powerful impact of the first consumer review. Specific predictions of the first-review effect are empirically validated. We summarize the main findings of the paper as follows:

- The valence and volume of online consumer reviews influence each other in their evolution paths, as a consequence of the conditional availability of online reviews.
- The valence of the first review positively affects the valence of the entire set of reviews. Although this effect may decrease over the long term, it is persistently observed for an extended time period, even 36 months after the first review posting.
- The valence of the first review also positively affects the overall volume of reviews. Because of the positive feedback loop between sales and WOM volume, this effect grows stronger over time, engendering strong path dependence for online consumer reviews.

The results of our study have important theoretical and practical implications. First, our findings bring to the fore the significance of a single review. Specifically, we demonstrate both theoretically and empirically that a product's first consumer review powerfully influences the entire evolution path of online

consumer reviews for that product and, thus, potentially determines its fate. In particular, when the first review is negative, the commanding first-review effect could curb the chances of the product's market success, thus significantly hurting the seller. Our research findings increase awareness of this important first-review effect while also highlighting the need for scholars and practitioners to develop effective preemptive strategies to reduce the chances of a product receiving a negative first review, as well as management strategies that can soften the impact of a negative first review after the fact. For example, a review platform may want to set up a policy where it suppresses reviews until the product receives a certain number of reviews so that consumers may see the reviews when the overall reviews reflect the true quality of the product.

Second, our work brings attention to a potential weakness of online consumer reviews. Our results suggest that a negative first review dramatically reduces the overall number of reviews. Thus, with a negative first review, the body of consumer reviews for that product may become less informative, either because any biased information in consumer reviews remains uncorrected or because consumers may even lose the opportunity to learn about the product at all. This finding implies that consumer reviews could become less reliable as a source of information and as a communication channel, depending on the valence of the first review. In today's market, businesses are increasingly shifting their attention and investment from firm- to user-generated information. Our research points to the need to integrate these two types of information into an effective overall communication strategy.

Third, our results on the asymmetric influence of the positive and negative first reviews also have interesting implications for promotional review policies. Recent empirical studies find evidence that firms have a high incentive to post positive reviews of their own products but negative reviews for those of their competitors (Mayzlin et al. 2014, Luca and Zervas 2016). Our findings suggest that promotional reviews, if posted as the first review, are very influential. Although such reviews may be created unethically, they actually help to facilitate informative online WOM if they promote the firms' own products, because a positive first review leads to high WOM volume. In this case, any upward bias in the initial promotional review would be corrected as more genuine consumer reviews were posted. However, posting such reviews first with the objective of harming a competitor can severely damage online WOM, because an unfavorable first review reduces sales and, thus, the number of subsequent reviews. Consequently, the initial downward bias derived from a negative promotional review is

less likely to be corrected, and the online WOM may fail to offer sufficient product information, which could severely harm the viability of the product being reviewed. Hence, it is important to consider such asymmetry when an online retail platform develops its review management policy, especially its policy regarding promotional reviews.

Our paper makes unique contributions to the literature on review dynamics. Thus far, the literature has focused on the individual reviewer's motivation and its impact on the evolution of the review valence (e.g., Li and Hitt 2008, Wu and Huberman 2008, Godes and Silva 2012, Moe and Schweidel 2012). By contrast, we examine the dynamics of both valence and volume and, more importantly, consider their evolution jointly. This approach allows us to uncover the path dependence of both valence and volume from the very first review, which cannot be explained by considering the evolution of the valence independently from that of the volume. Our most novel finding, the increasing impact of the first review on WOM volume, can be explained by the conditional availability of the reviews as modeled in our paper, but not by individual reviewer motivation, as suggested in the literature.

Finally, this paper is not without limitations, which also point to a few interesting avenues for future research. First, our study focuses on the valence of the first review. However, other characteristics of the first review, such as the timing of the review or the reviewer's reputation, might significantly influence future reviews. This suggests an interesting avenue for future research. Second, our model abstracts away from firms' strategic decisions to influence reviews, given our objective of establishing a first-review effect. Future research might investigate the impact of a firm's actions concerning its own profits, consumer surplus, and the review-generation process.

Acknowledgments

This paper is partially based on the first author's doctoral dissertation at the University of Florida. The authors thank the editor, associate editor, two anonymous reviewers, Martijn de Jong, Yong Liu, Jiwoong Shin, Steve Shugan, Catherine Tucker, Ying Xie, and Juanjuan Zhang for their helpful comments and discussions.

Appendix. Proofs

Derivation of the Expected Utility of the Consumer (EU_t)

First, note that $E[U] = E[1 - e^{-(v_0 + \beta r_{t-1} - \tau x - p)}] = 1 - e^{-(v_0 - \tau x - p)} E[e^{-\beta r_{t-1}}]$, where r_{t-1} is a sample mean of random variable w_i ($i = 1, 2, \dots, n_{t-1}$), and thus asymptotically follows $N(\alpha, \frac{\alpha(1-\alpha)}{n_{t-1}})$ by the central limit theorem. However, because consumers do not observe α , in deriving the expected utility, they use r_{t-1} and $\frac{r_{t-1}(1-r_{t-1})}{n_{t-1}}$, respectively, as estimators of α and $\frac{\alpha(1-\alpha)}{n_{t-1}}$. Then, in the perspective of consumers, $e^{-\beta r_{t-1}}$ follows a lognormal distribution $LN(-\beta r_{t-1}, \frac{\beta^2 r_{t-1}(1-r_{t-1})}{n_{t-1}})$,

implying $E[e^{-\beta r_{t-1}}] = e^{-\beta r_{t-1} + \frac{\beta^2 r_{t-1}(1-r_{t-1})}{2n_{t-1}}}$. Therefore, we have $E[U_t] = 1 - e^{-(v_0 + \beta r_{t-1} - p - \frac{\beta^2 r_{t-1}(1-r_{t-1})}{2n_{t-1}})}$.

Derivation of the Sales (s_t)

A customer purchases the product if and only if $E[U_t] = 1 - e^{-(v_0 + \beta r_{t-1} - p - \frac{\beta^2 r_{t-1}(1-r_{t-1})}{2n_{t-1}})} \geq 0$, which is equivalent to $x \leq \frac{1}{\tau}(v_0 + \beta r_{t-1} - p - \frac{\beta^2 r_{t-1}(1-r_{t-1})}{2n_{t-1}})$. Because $x \sim U[0,1]$, in each period, among M consumers arriving in the market, $M \times \frac{1}{\tau}(v_0 + \beta r_{t-1} - p - \frac{\beta^2 r_{t-1}(1-r_{t-1})}{2n_{t-1}})$ consumers will buy the product. Therefore, the total sales in period t is given as $s_t = \frac{M}{\tau}(v_0 + \beta r_{t-1} - p - \frac{\beta^2 r_{t-1}(1-r_{t-1})}{2n_{t-1}})$.

Proof of Theorem 1

Part (a). In period t , we have $E[n_{t+1}|n_t, r_t] = n_t + \delta \cdot E[s_{t+1}|n_t, r_t]$, where $E[s_{t+1}|n_t, r_t] = \frac{M}{\tau}(v_0 + \beta r_t - p - \frac{\beta^2 r_t(1-r_t)}{2n_t})$. Thus, by the chain rule, $\frac{\partial E[n_{t+1}|n_t, r_t]}{\partial r_t} = \frac{\partial E[n_{t+1}|n_t, r_t]}{\partial E[s_{t+1}|n_t, r_t]} \frac{\partial E[s_{t+1}|n_t, r_t]}{\partial r_t} = \frac{\delta M}{\tau}(\beta - \frac{\beta^2(1-2r_t)}{2n_t})$. Then, $\frac{\partial E[n_{t+1}|n_t, r_t]}{\partial r_t} \geq 0$ is equivalent to $2n_t \geq \beta(1-2r_t)$, which always holds because $n_t \geq 1$, $r_t \geq 0$, and $\beta \leq 2$. Note that the equality holds when $n_t = 1$, $r_t = 0$, and $\beta = 2$, which can happen only at $t = 0$, which is the period of the very first review. Therefore, for any $t \geq 1$, we have $\frac{\partial E[n_{t+1}|n_t, r_t]}{\partial r_t} > 0$.

Part (b). Recall that N_{t+1} denotes the volume of the newly posted reviews: $N_{t+1} \equiv n_{t+1} - n_t$. By definition, we have

$$\begin{aligned} r_{t+1} &= \frac{h_t + H_{t+1}}{n_t + N_{t+1}} = \frac{h_t}{n_t + N_{t+1}} + \frac{H_{t+1}}{n_t + N_{t+1}} \\ &= \frac{r_t \cdot n_t}{n_t + N_{t+1}} + \frac{H_{t+1}}{n_t + N_{t+1}} = r_t \cdot \left(1 - \frac{N_{t+1}}{n_t + N_{t+1}}\right) + \frac{H_{t+1}}{n_t + N_{t+1}}. \end{aligned}$$

Moreover, noting that H_{t+1} follows $\text{Binomial}(N_{t+1}, \alpha)$, by the law of iterated expectations,

$$\begin{aligned} E\left[\frac{H_{t+1}}{n_t + N_{t+1}} \middle| n_t, r_t\right] &= E\left[E\left[\frac{H_{t+1}}{n_t + N_{t+1}} \middle| n_t, r_t, N_{t+1}\right] \middle| n_t, r_t\right] \\ &= E\left[\frac{\alpha N_{t+1}}{n_t + N_{t+1}} \middle| n_t, r_t\right] \\ &= \alpha \cdot E\left[\frac{N_{t+1}}{n_t + N_{t+1}} \middle| n_t, r_t\right]. \end{aligned}$$

Then we have

$$\begin{aligned} E[r_{t+1}|n_t, r_t] &= E\left[r_t \cdot \left(1 - \frac{N_{t+1}}{n_t + N_{t+1}}\right) + \frac{H_{t+1}}{n_t + N_{t+1}} \middle| n_t, r_t\right] \\ &= r_t \cdot \left(1 - E\left[\frac{N_{t+1}}{n_t + N_{t+1}} \middle| n_t, r_t\right]\right) + E\left[\frac{H_{t+1}}{n_t + N_{t+1}} \middle| n_t, r_t\right] \\ &= r_t \cdot \left(1 - E\left[\frac{N_{t+1}}{n_t + N_{t+1}} \middle| n_t, r_t\right]\right) + \alpha \cdot E\left[\frac{N_{t+1}}{n_t + N_{t+1}} \middle| n_t, r_t\right] \\ &= r_t + (\alpha - r_t) \cdot E\left[\frac{N_{t+1}}{n_t + N_{t+1}} \middle| n_t, r_t\right] \end{aligned}$$

Therefore, $\frac{\partial E[r_{t+1}|n_t, r_t]}{\partial N_{t+1}} = (\alpha - r_t) \cdot E\left[\frac{\partial}{\partial N_{t+1}} \left(\frac{N_{t+1}}{n_t + N_{t+1}}\right) \middle| n_t, r_t\right] = (\alpha - r_t) \cdot E\left[\frac{-n_t}{(n_t + N_{t+1})^2} \middle| n_t, r_t\right] > 0$ if and only if $\alpha > r_t$.

Endnotes

¹ For brevity, we will refer to these online platforms as “Amazon” and “Walmart,” respectively, from this point on.

² In the examples given in Table 1, their cross-platform (dis)advantage in WOM valence and volume are also associated with their cross-platform (dis)advantage in the first-review rating.

³ In this formulation, r_t is a sample mean of random variable w_i ($i = 1, 2, \dots, n_t$), and thus asymptotically follows $N(\alpha, \alpha(1-\alpha)/n_t)$ by the central limit theorem, which implies that products with more extreme qualities have lower variance in their average ratings.

⁴ We use an exogenous price in our model, because we do not find empirical evidence that sellers react to user reviews by adjusting their prices (see Online Appendix B1 for this empirical analysis).

⁵ Although prior research finds that the most and the least satisfied consumers are more likely to engage in WOM (Anderson 1998, West and Broniarczyk 1998), for simplicity, we use the constant propensity to write a review.

⁶ We acknowledge that some reviews might be written without purchase of the products (Anderson and Simester 2014). However, the presence of reviews without purchase does not change the impact of the current valence on the future volume.

⁷ Even without the log transformation, we could obtain qualitatively same results, although the current model with the log transformation fits better with the data.

⁸ Among these controls, the product type dummies are indicator variables for vacuum cleaner types as classified by Amazon: canister, handheld, robotic, stick, or upright. For identification, we include only four indicators, while setting the canister type as the baseline type. As the product feature indicators, we include two dummies: one for bagged vacuum cleaners and another for corded vacuum cleaners. These are coded as one if the product has the feature and as zero otherwise. Last, all the control variables in X_{it} other than consumer search volume and the volume are time invariant.

⁹ One might view the first-review effect on the valence as consisting of the direct effect and the indirect effect, with the latter being the effect through the volume. Our first-review-independent volume variable controls the influence of the volume on the valence but still allows us to capture the full effect of the first review.

¹⁰ Within our data frame, the eventual convergence of the average rating is not observed. However, our model predicts it could happen outside of our data period. Although the linear model could even predict the reversal of the first-review effect, this is simply due to extrapolation and the limitation of the model.

¹¹ This follows our theoretical mechanism presented in Section 2.2, which states that the lagged valence and the lagged volume influence the current volume. Note that Online Appendix B4 provides empirical support for this mechanism.

¹² Note that standardization is not necessary for the valence model in Equations (6) and (7) because the variation in the average ratings is fairly comparable between the two platforms (as noted in Table 3).

¹³ The conditions are that (1) the observed control variables are randomly chosen among all possible controls and (2) the number of observed and unobserved variables are large and none of the factors dominates the distribution of the main variable (i.e., $FNegative$ or the outcome variable (i.e., valence and volume)).

¹⁴ In constructing this variable, we use product type as a search criterion because this is the first filter that the retailers provide saliently on both Amazon and Best Buy. The small assortment size of product types on Best Buy makes it unnecessary to use multiple search criteria.

¹⁵ For completeness, we also estimate the valence model with the inclusion of the first-page dummy. We obtain qualitatively the same results (see column (4) of Table 10).

¹⁶ Amazon U.S. and Amazon Canada have totally independent websites with separate consumer reviews even for identical products sold on both platforms. In January 2017, however, Amazon Canada

started providing their customers with a link to Amazon U.S. product reviews. Hence, in our analysis, we only use product reviews posted before January 1, 2017.

¹⁷The positive and marginally significant β_1 is probably due to the functional form of Equation (9). Once we estimate the period-specific first-review effects, the coefficient for the second period becomes insignificant, and the coefficients for the rest of periods are negative and significant.

¹⁸The results of the no-control specifications (see Section 3.6.2) suggest that the estimates of our primary interests are robust to the inclusion or exclusion of price. Moreover, using daily price data from both Amazon and Best Buy, we find that firms do not change their prices in response to newly posted product reviews (see Online Appendix B1 for details of this analysis).

¹⁹We conduct an additional analysis to confirm that product sales are affected by the online WOM more in the vacuum cleaner category than in the digital camera category (see Online Appendix B9 for details).

References

- Adomavicius G, Bockstedt JC, Curley SP, Zhang J (2013) Do recommender systems manipulate consumer preferences? A study of anchoring effects. *Inform. Systems Res.* 24(4):956–975.
- Altonji JG, Elder TE, Taber CR (2005) Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *J. Political Econom.* 113(1):151–184.
- Ameri M, Honka E, Xie Y (2019) Word of mouth, observed adoptions, and anime-watching decisions: The role of the personal vs. the community network. *Marketing Sci.* 38(4):567–583.
- Anderson ET, Simester DI (2014) Reviews without a purchase: Low ratings, loyal customers, and deception. *J. Marketing Res.* 51(3):249–269.
- Anderson EW (1998) Customer satisfaction and word of mouth. *J. Service Res.* 1(1):5–17.
- Babić Rosario A, Sotgiu F, De Valck K, Bijmolt TH (2016) The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *J. Marketing Res.* 53(3):297–318.
- Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? *Quart. J. Econom.* 119(1):249–275.
- Bikhchandani S, Hirshleifer D, Welch I (1998) Learning from the behavior of others: Conformity, fads, and informational cascades. *J. Econom. Perspect.* 12(3):151–170.
- BrightLocal (2014) Local consumer review survey. Accessed April 21, 2016, <https://www.brightlocal.com/wp-content/uploads/2014/07/Local-Consumer-Review-Survey-20141.pdf>
- Cameron SV, Taber C (2004) Estimation of educational borrowing constraints using returns to schooling. *J. Political Econom.* 112(1):132–182.
- Chen Y, Xie J (2008) Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Sci.* 54(3):477–491.
- Chen Y, Wang Q, Xie J (2011) Online social interactions: A natural experiment on word of mouth vs. observational learning. *J. Marketing Res.* 48(2):238–254.
- 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, Gopinath S, Venkataraman S (2010) The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Sci.* 29(5):944–957.
- De Langhe B, Fernbach PM, Lichtenstein DR (2016) Navigating by the stars: Investigating the actual and perceived validity of online user ratings. *J. Consumer Res.* 42(6):817–833.
- Dimensional Research (2013) Customer service and business results: A survey of customer service from mid-size companies. Accessed April 23, 2016, https://d16cvnquvjw7pr.cloudfront.net/resources/whitepapers/Zendesk_WP_Customer_Service_and_Business_Results.pdf.
- Feng J, Xie J (2012) Research note-performance-based advertising: Advertising as signals of product quality. *Inform. Systems Res.* 23(3-part-2):1030–1041.
- Floyd K, Freling R, Alhoqail S, Cho HY, Freling T (2014) How online product reviews affect retail sales: A meta-analysis. *J. Retailing* 90(2):217–232.
- Forman C, Ghose A, Wiesenfeld B (2008) Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Inform. Systems Res.* 19(3):291–313.
- Godes D, Mayzlin D (2004) Using online conversations to study word-of-mouth communication. *Marketing Sci.* 23(4):545–560.
- Godes D, Silva JC (2012) Sequential and temporal dynamics of online opinion. *Marketing Sci.* 31(3):448–473.
- Holmström B, Milgrom P (1987) Aggregation and linearity in the provision of intertemporal incentives. *Econometrica* 55(2):303–328.
- Holmström B, Tirole J (1993) Market liquidity and performance monitoring. *J. Political Econom.* 101(4):678–709.
- Ho-Dac NN, Carson SJ, Moore WL (2013) The effects of positive and negative online customer reviews: Do brand strength and category maturity matter? *J. Marketing* 77(6):37–53.
- Kornish LJ, Li Q (2010) Optimal referral bonuses with asymmetric information: Firm-offered and interpersonal incentives. *Marketing Sci.* 29(1):108–121.
- Li X, Hitt LM (2008) Self-selection and information role of online product reviews. *Inform. Systems Res.* 19(4):456–474.
- Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73(1):13–22.
- Libai B, Muller E, Peres R (2013) Decomposing the value of word-of-mouth seeding programs: Acceleration vs. expansion. *J. Marketing Res.* 50(2):161–176.
- Liu Y (2006) Word of mouth for movies: Its dynamics and impact on box office revenue. *J. Marketing* 70(3):74–89.
- Luca M, Zervas G (2016) Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Sci.* 62(12):3412–3427.
- Ludwig S, De Ruyter K, Friedman M, Brügggen EC, Wetzels M, Pfann G (2013) More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *J. Marketing* 77(1):87–103.
- Mayzlin D, Dover Y, Chevalier J (2014) Promotional reviews: An empirical investigation of online review manipulation. *Amer. Econom. Rev.* 104(8):2421–2455.
- Moe WW, Schweidel DA (2012) Online product opinions: Incidence, evaluation, and evolution. *Marketing Sci.* 31(3):372–386.
- Muchnik L, Aral S, Taylor SJ (2013) Social influence bias: A randomized experiment. *Science* 341(6146):647–651.
- Pew Research Center (2016) Online shopping and e-commerce. Accessed June 24, 2020, <https://www.pewresearch.org/internet/2016/12/19/online-shopping-and-e-commerce/>.
- Ranard BL, Werner RM, Antanavicius T, Schwartz HA, Smith RJ, Meisel ZF, Asch DA, Ungar LH, Merchant RM (2016) Yelp reviews of hospital care can supplement and inform traditional surveys of the patient experience of care. *Health Affairs* 35(4):697–705.
- Sun M (2012) How does the variance of product ratings matter? *Management Sci.* 58(4):696–707.

- Trusov M, Bucklin RE, Pauwels K (2009) Effects of word-of-mouth vs. traditional marketing: Findings from an internet social networking site. *J. Marketing* 73(5):90–102.
- West PM, Broniarczyk SM (1998) Integrating multiple opinions: The role of aspiration level on consumer response to critic consensus. *J. Consumer Res.* 25(1):38–51.
- Wu F, Huberman BA (2008) How public opinion forms. Papadimitrou C, Zhang S, eds. *Internet and Network Economics*. Lecture Notes in Computer Science, vol. 5385 (Springer, Berlin), 334–341.
- You Y, Vadakkepatt GG, Joshi AM (2015) A meta-analysis of electronic word-of-mouth elasticity. *J. Marketing* 79(2):19–39.
- Zellner A (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Amer. Statist. Assoc.* 57(298):348–368.
- Zhang J, Liu P (2012) Rational herding in microloan markets. *Management Sci.* 58(5):892–912.
- Zhu F, Zhang X (2010) Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *J. Marketing* 74(2):133–148.