



Marketing Science

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

Online Product Opinions: Incidence, Evaluation, and Evolution

Wendy W. Moe, David A. Schweidel,

To cite this article:

Wendy W. Moe, David A. Schweidel, (2012) Online Product Opinions: Incidence, Evaluation, and Evolution. *Marketing Science* 31(3):372-386. <https://doi.org/10.1287/mksc.1110.0662>

Full terms and conditions of use: <http://pubsonline.informs.org/page/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 © 2012, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

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

Online Product Opinions: Incidence, Evaluation, and Evolution

Wendy W. Moe

Robert H. Smith School of Business, University of Maryland, College Park, Maryland 20472,
wmoe@rhsmith.umd.edu

David A. Schweidel

Wisconsin School of Business, University of Wisconsin–Madison, Madison, Wisconsin 53706,
dschweidel@bus.wisc.edu

Whereas recent research has demonstrated the impact of online product ratings and reviews on product sales, we still have a limited understanding of the individual's decision to contribute these opinions. In this research, we empirically model the individual's decision to provide a product rating and investigate factors that influence this decision. Specifically, we consider how previously posted ratings may affect an individual's posting behavior in terms of *whether* to contribute (incidence) and *what* to contribute (evaluation), and we identify *selection effects* that influence the incidence decision and *adjustment effects* that influence the evaluation decision.

Across individuals, our results show that positive ratings environments increase posting incidence, whereas negative ratings environments discourage posting. Our results also indicate important differences across individuals in how they respond to previously posted ratings, with less frequent posters exhibiting bandwagon behavior and more active customers revealing differentiation behavior. These dynamics affect the evolution of online product opinions. Through simulations, we illustrate how the evolution of posted product opinions is shaped by the underlying customer base and show that customer bases with the same median opinion may evolve in substantially different ways because of the presence of a core group of “activists” posting increasingly negative opinions.

Key words: online social media; word of mouth; social dynamics; Bayesian estimation; Internet marketing; user-generated content

History: Received: December 17, 2009; accepted: May 23, 2011; Russell Winer served as the special issue editor and Sanjog Misra served as associate editor for this article. Published online in *Articles in Advance* August 25, 2011.

Introduction

The post-Internet marketplace is no longer limited to one-way communications from sellers to buyers. Instead, consumers have become much more active in influencing and altering the nature of conversations around brands and products. Facilitated by developments in online technologies, consumers can easily contribute their thoughts and opinions to the marketplace through discussion groups, product ratings and reviews, and blogs. As a result, consumers have begun to talk with each other on a scale larger than marketers have previously experienced. However, this new environment is not without risks for marketers. In particular, marketers are increasingly losing control over the dialogue taking place around their products and brands. Although this can be a positive development as consumers become more engaged and generate an increased level of “buzz” in the market, it can also have adverse consequences if the tone and content turn negative.

Of even more concern is that extant research has shown the existence of systematic biases in online consumer product ratings. Several researchers have shown empirically that posted product ratings and reviews become increasingly negative as ratings environments mature (Li and Hitt 2008, Godes and Silva 2012, Moe and Trusov 2011). Schlosser (2005) also showed in a lab environment that posters adjust their product evaluations depending on the opinions expressed by others. As these studies demonstrate, a ratings environment can take on a life of its own, sometimes to the detriment of the product or brand to which it is dedicated. In some cases, the posted content provides a fair evaluation of the product or brand. However, as illustrated in the aforementioned studies, posted content can also reflect the influence of others.

The objective of this research is to empirically examine the behavior of individuals providing product ratings in an effort to better understand how

expressed opinions (as reflected in posted product ratings) systematically evolve over time.¹ Specifically, we investigate the role that others' ratings can have in influencing posting behavior. We consider two separate effects that may influence the subsequent evolution of product opinion. First, previously posted ratings may affect the incidence with which individuals choose to contribute their own opinions, which we refer to as a *selection effect*. Second, whereas some customers may prefer to provide their comments such that they will stand out from the crowd, others may prefer to be consistent with the majority. Thus, in addition to selection effects, there may also be *adjustment effects* where those individuals who ultimately decide to post revise their evaluations upward or downward based on previously posted comments.

In contrast to much of the existing research on online ratings that examined posted ratings at the product level, we model posted ratings at the level of the individual consumer. Many individuals actively contribute their ratings for a variety of different products. Examining an individual's behavior across products allows us to identify and measure the influence that previously posted ratings (which vary across products) have on these two aspects of posting behavior.

We model an individual's posting incidence decision (i.e., *whether* to post) as a probit process subject to the fact that he or she has had experience with the product. Because we do not observe whether or not an individual has purchased and/or experienced the product, we incorporate a *latent experience* component in the incidence model. We simultaneously estimate the individual's evaluation decision (i.e., *what* to post) as an ordered probit process governing the number of stars provided on a five-star scale. In both the incidence and the evaluation models, we explicitly examine (1) the role of a consumer's postpurchase product evaluation, (2) the effect of previously contributed opinions, and (3) heterogeneity across individuals.

Across individuals, our results show that positive ratings environments increase posting incidence, whereas negative environments discourage posting. Our results also show that there are substantial differences across individuals in terms of how they respond to previously expressed opinions. Specifically, we show that individuals who post infrequently

are more positive and likelier to contribute to environments that exhibit a consensus of opinions (i.e., lower opinion variance). When these individuals post, they adjust their ratings upward in more-positive ratings environments, thereby exhibiting "bandwagon" behavior. In contrast, we find that highly active posters are more negative in their evaluations and more prone to post in dissentious environments (i.e., environments with higher opinion variance). When they post, they adjust their ratings downward in the presence of more-positive ratings, thereby differentiating themselves. A key result of our research is that these differences across individuals result in a systematic shift in the composition of the posting population, with more negative and active posters increasingly contributing over time—a dynamic that has not been posited previously as a potential explanation for the observed downward trend in online ratings.

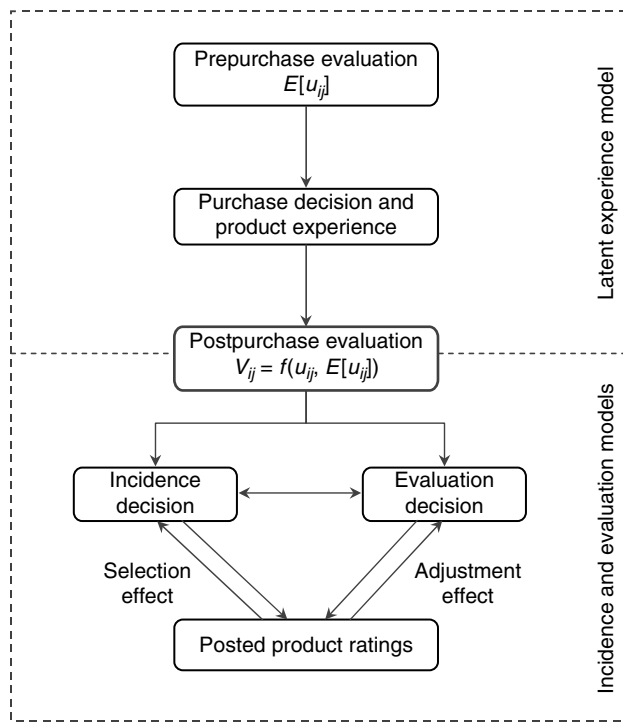
To illustrate this dynamic, we simulate ratings environments arising from different customer bases, each defined by their underlying distribution of product evaluations. From these simulations, we show how posted opinions evolve over time as a core group of active individuals become more prevalent among the posting population. The behavior of this core group can significantly shape the direction of expressed opinions. This dynamic is most pronounced when the customer base is highly polarized. When customers are polarized, posted opinions are more negative and exhibit a stronger downward trend when compared with a neutral customer base with the same median opinion. Posted ratings from polarized customer bases tend to be dominated by individuals with extremely negative opinions while underrepresenting individuals with more positive opinions. As a result, the opinions expressed online are not representative of the entire customer base, highlighting the caution that must be exercised by both marketers and consumers in drawing conclusions from a cursory view of posted product opinions.

In the next section, we discuss factors that influence an individual's posting behavior, both in terms of his or her posting incidence decision and his or her posted evaluation. We subsequently present a joint modeling framework for the incidence and evaluation decisions and examine the robustness of the results to variations in model specification. Using the empirical results of this model, we conduct a series of simulations to demonstrate how product opinions evolve and the effect of customer base composition on this evolution.

Why Do People Post?

In this section, we discuss the process in which a consumer formulates, modifies, and ultimately expresses

¹ We use the term *rating* to refer to the quantitative product evaluation provided by the consumer and the term *review* to refer to the textual content of his or her posted evaluation. We use the term *opinion* to refer to the more general construct that encompasses both ratings and reviews. In this paper, we use a consumer's posted product rating as a quantitative metric of that individual's product opinion.

Figure 1 Conceptual Model of the Consumer Ratings Process

his or her opinion about a product (see Figure 1). These product opinions can be expressed in the form of a verbal/textual review or as a numerical rating. Although publicly posted opinions (in our case, product ratings) are typically provided after the consumer purchases and experiences the product, the process in which an opinion is formulated can start well before. Berinsky (2004) proposed a framework that separates an individual's response to political polling into two separate stages: opinion formation and opinion expression. In the context of product opinion, we similarly distinguish between a consumer's underlying product evaluation and his or her posted product rating.

With respect to a consumer's underlying product evaluation, past researchers have differentiated between prepurchase and postpurchase product evaluations, two constructs separated in time by the consumer's purchase of and direct experience with the product (Kuksov and Xie 2010, Anderson and Sullivan 1993).² Prepurchase evaluations are often formulated on the basis of publicly available information, such as observable product attributes, marketing mix activities, word of mouth, etc., and drive

the consumer's purchase decision. These prepurchase evaluations reflect the consumer's expected utility for the product, $E[u_{ij}]$, and as a result, they provide a benchmark against which the actual product experience will be compared (Anderson and Sullivan 1993). In the postpurchase stage, the consumer has access to new private information obtained from his or her own experience with the product. This new information contributes to customer satisfaction and the formulation of a postpurchase evaluation. Specifically, Anderson and Sullivan (1993) have shown that customer satisfaction is a function of both the individual's experienced utility u_{ij} and how this experienced utility compares with the expected utility $E[u_{ij}]$. Together, these constructs contribute to the individual's postpurchase evaluation of the product, V_{ij} .

In this paper, our focus is not directly on the antecedents of a consumer's postpurchase evaluation. Instead, our objective is to develop a model that estimates V_{ij} , accounting for the fact that not all individuals in the consumer population have necessarily purchased or experienced the product. Our particular interest in this research is the processes that follow once a consumer purchases/experiences the product and develops a postpurchase evaluation—namely, his or her decision of *whether* to post a product rating (incidence decision) and *what* rating to post (evaluation decision).

The Incidence Decision: What Influences Participation?

Across a variety of online contexts, an overwhelming majority of individuals engaged with a site tend to be "lurkers" (i.e., those who read the comments of a small population of posters but do not provide posts themselves). Whereas some studies have examined the behavioral differences between posters and lurkers (Schlosser 2005), few have looked at the factors that encourage (or discourage) a given individual to post (or be silent).

In the off-line word-of-mouth literature, Anderson (1998) showed that individuals who are extremely dissatisfied are more likely to engage in word-of-mouth activities. As a consequence, off-line word of mouth tends to be disproportionately negative. However, for reasons still unknown, a very different dynamic exists in the online environment. Across several studies, researchers have observed overwhelmingly positive product ratings being posted online (see Chevalier and Mayzlin 2006 for an example). For this reason, Dellarocas and Narayan (2006) proposed that individuals with extreme opinions, both positive and negative, will be more likely to post an opinion online than those with more moderate opinions. Whereas different effects have been shown across contexts, it is clear that an individual's postpurchase product evaluation affects his or her decision to

² Whereas some consumers may experience the product without purchasing it (e.g., if the product were received as a gift), this constitutes a small proportion of consumer experiences. Our modeling methodology is not dependent on observing purchasing behavior. We use "postpurchase" simply for ease of exposition and to be consistent with existing literature.

engage in word-of-mouth activities, be it off-line or online. Therefore, we will explicitly examine the effect postpurchase evaluations have on posting incidence.

In addition, the posting incidence decision may also be subject to environmental factors. Political scientists have long known that the results from opinion polls can affect election turnout (see McAllister and Studlar 1991 for a review). The direction of these effects have been the subject of extensive discussion, and researchers have debated the presence of a “bandwagon” effect, where opinion polls influence voter behavior in favor of the candidate leading in the polls (McAllister and Studlar 1991, Marsh 1984), versus an “underdog” effect, where the candidate trailing in the polls is favored (Gartner 1976, Straffin 1977). Other studies have shown that the declaration of a clear winner in opinion polls can depress overall voter turnout because voters perceive their votes to be inconsequential (e.g., Epstein and Strom 1981, Dubois 1983, Jackson 1983, Delli Carpini 1984, Sudman 1986). Taken together, these studies show that the opinions of others can influence an individual’s decision of whether or not to voice his or her own opinion. In this research, we also examine how the ratings environment, as characterized by previously posted opinions, can affect an individual’s posting incidence decision and refer to these covariates as having a *selection effect*.

The Evaluation Decision: What Influences Posted Ratings?

Should an individual choose to post an opinion, the decision of *what* to post can also be subject to a number of factors. In theory, an individual posts a rating that is reflective of his or her postpurchase evaluation of the product. However, a number of recent studies have emerged documenting the presence of noticeable opinion dynamics in online product ratings (Godes and Silva 2012, Li and Hitt 2008, Moe and Trusov 2011). These studies contribute to a larger body of work (including the political science literature mentioned previously) suggesting that an individual’s publicly expressed opinion can be influenced by the opinions of others and does not necessarily mirror the individual’s socially unbiased and independent product evaluation.

In a controlled experimental setting, Schlosser (2005) showed that an individual poster has a tendency to adjust his or her posted product evaluation after viewing what others have posted. She demonstrates a differentiation effect where some posters, particularly those who consider themselves “experts,” try to differentiate themselves from others by posting more negative opinions. This is in contrast to studies that have shown that individuals can be subject to bandwagon effects and adopt the opinion of the majority (McAllister and Studlar 1991,

Marsh 1984). The conclusion we take from these results is that individuals are heterogeneous and may be subject to either differentiation or bandwagon effects. Studies have also shown that individuals moderate their expressed opinions in the presence of an audience with high opinion variance (Fleming et al. 1990). Overall, these studies highlight the fact that previously posted opinions can influence the individual’s decision of *what* to post. Therefore, we consider a number of covariates that characterize the ratings environment in terms of the previously expressed opinions of others and examine their effects on the individual’s evaluation decision. We refer to these covariates as having an *adjustment effect* on posting behavior, allowing for differentiation effects, bandwagon effects, and the effect of consensus (or dissension).

The Composition of the Posting Population

Although observed ratings dynamics have been attributed to the aforementioned adjustment effect (Schlosser 2005), few researchers have considered the additional impact that selection effects have on posted product ratings. Specifically, factors affecting the incidence decision have the potential to systematically alter the composition of the posting population. Overall, the composition of the posting population is determined by (1) the composition of the larger customer base from which it is drawn and (2) the selection effect of covariates. In this paper, we explicitly consider both factors and examine the impact of each on the ratings environment.

Understanding the factors that influence the composition of the posting population can have significant implications for managers and consumers. If the posting population were randomly drawn from the full customer base (i.e., there is no selection bias), the posted opinions would resemble the opinions held by the customer base as a whole. As a result, online product ratings could provide an informative source of product feedback for both managers and potential buyers. However, differences across both consumers and ratings environments may influence the incidence decision, consequently skewing the composition of the posting population such that it no longer resembles the overall customer base.

Data

Our data, provided by BazaarVoice, consist of product ratings that were contributed by customers of an online retailer of bath, fragrance, and home products.³

³ This retailer primarily sells products under its own brand name. No national brands are represented in this data set. The products carrying this retailer’s name are sold only through its stores. Marketing is limited to activities that promote the retailer’s overall brand; the retailer does not engage in product-specific marketing efforts.

For each product, consumers can provide an overall product rating on a discrete five-star scale along with review text, a ratings format that is very common in the online environment. The average overall rating is displayed prominently in the middle of the product page immediately below the manufacturer's description of the product. Individual ratings and reviews are presented (by default) in reverse chronological order as one scrolls down the page, making it easy for consumers to ascertain the number of reviews that have been posted. For the purposes of this research, we consider each consumer's posted star rating as a metric of his or her product opinion.⁴

The data span a six-month time period in 2007 when the ratings functionality was first introduced to the site. All ratings provided in this initial six-month period are included in our data. During this time, 4,974 unique individuals posted product ratings, resulting in a total of 10,460 ratings across 1,811 products. Approximately 18% of the raters posted ratings for multiple products. The data also indicate the time at which ratings were posted, facilitating identification of the set of products for which an individual provides evaluations in each of his or her rating sessions.

Our objective is to model the behavior of each individual across a variety of products. With this data set, this would necessitate the construction of a $4,974 \times 1,811$ matrix of ratings. Because of the computational constraints associated with a matrix of such size, we sample 200 products from this data set. However, to maintain a sufficient number of observations for each individual rater in the sample, we draw a systematic sample as follows. We include the 100 most rated products at the site, which provide a large base of individual raters, many of whom post ratings for multiple products. To ensure variation across products and hence ratings environments, our sample also includes 100 additional products that were chosen at random. Our sampling results in a data set that includes 2,436 individual raters who provide a total of 3,681 product ratings.

We supplement our ratings data with data collected from Google Trends. Google Trends data provide an index based on the number of searches on Google that include specific search terms relative to the total volume of searches conducted using Google. The data from Google Trends have been linked to short-term

economic trends (Varian and Choi 2009) and disease outbreaks (Carneiro and Mylonakis 2009). We use these data as a proxy for consumer interest in the different product categories represented in our ratings data. Specifically, we queried Google Trends for the number of searches that included mention of each product category. Weekly data were obtained to correspond with our ratings data. In the next section, we discuss how they are incorporated into our model to serve as a control variable.

Model Development

Consistent with the conceptual framework presented above, we model ratings incidence and evaluation behavior as two separate but related processes. Central to our modeling framework is the role that individual i 's postpurchase evaluation for product j , denoted as V_{ij} , has on both components of posting behavior. In our model, the postpurchase evaluation V_{ij} is the primary driver of the individual's posted rating (evaluation decision) and simultaneously affects the decision of whether or not to post (incidence decision) in a nonlinear manner.

Methodologically, our approach is similar to previous models that have employed parsimonious latent constructs to model multiple outcomes, and hence it allows for relationships among the outcomes (e.g., Kamakura et al. 2003, Park and Bradlow 2005, Li et al. 2005). Our specification builds on previous models of product ratings (e.g., Ansari et al. 2000, Ying et al. 2006) by flexibly linking the incidence and evaluation decisions, a desirable feature given the variety of relationships that have been documented between word-of-mouth activity and customer satisfaction (Anderson and Sullivan 1993, Dellarocas and Narayan 2006).

To capture selection and adjustment effects, we include a set of covariates that characterize the ratings environment in terms of the previously posted ratings of others. We allow these covariates to differentially influence both the incidence and evaluation decisions. These covariates take advantage of the longitudinal nature of our data, as the ratings environment varies both across products and over time during the six-month period.

Finally, we consider two methodological assumptions. First, we condition the incidence decision on the occurrence of a rating session defined as any day in which an individual posts *at least one* rating. This assumption provides model tractability given our longitudinal data.⁵ Our second assumption is that

⁴ This particular retail site also allows consumers to rate specific product attributes as defined by the retailer. These attributes vary across products depending on the nature of the product. The attribute-specific evaluations and textual content are interesting to consider; however, they raise methodological complications without contributing directly to our research objective. Furthermore, product pages prominently feature the average overall product rating. Our focus on the overall product rating as the key metric of consumer opinion is consistent with this emphasis.

⁵ The alternative is to model incidence and evaluation decisions for each moment in time. If we aggregated time to the daily level, this would result in 445,788 ($2,436 \text{ individuals} \times 183 \text{ days}$) observations per product compared with the 2,960 observations per product when we condition on a rating session.

individuals will only consider rating a product if he or she has had experience with the product. Because product experience or purchase is not observable to us at the individual level, we incorporate a *latent* measure of experience in our incidence model.

We begin our model development by presenting first the evaluation model, which is composed of a latent postpurchase evaluation (V_{ij}) and adjustment effects from the ratings environment. We then describe the incidence model that includes a latent experience component, an effect from one's postpurchase evaluation, and selection effects resulting from the ratings environment.

Evaluation Model: What to Rate a Product

Assuming that an individual decides to post a product rating, the expressed opinion is dependent on both the underlying postpurchase evaluation of the product V_{ij} and the ratings environment. Therefore, conditional on a rating being contributed, we model the posted product rating using an ordered probit model (Ying et al. 2006).

First, we assume that the rating contributed by i is driven in large part by his or her postpurchase evaluation V_{ij} :

$$V_{ij} = \gamma_{i0} + \kappa_j, \quad (1)$$

where γ_{i0} allows for different levels of baseline positivity (or negativity) across customers, and κ_j allows for variation across products (perhaps as a result of differences in product quality) such that $\kappa_j \sim N(0, \sigma_\kappa^2)$. The respondent, however, may demonstrate adjustment effects because of the nature of the ratings environment. As such, we denote the net of the postpurchase evaluation and the adjustment effect as V_{ijk}^* :

$$V_{ijk}^* = V_{ij} + \gamma_{i,1:N} X_{j,k(i)}, \quad (2)$$

where $X_{j, \cdot}$ is an $N \times 1$ vector of covariates that describe the ratings environment for product j at the time of rating session k by individual i . The $1 \times N$ vector $\gamma_{i,1:N}$ captures the impact that the ratings environment may have on an individual's posted rating, which may result in the posted evaluation differing from an individual's postpurchase evaluation.

As ratings are submitted on a five-point scale, we model posted product ratings as follows:

$$\Pr(y_{ijk} = r | z_{ijk} = 1) = \Pr(\mu_{i,r-1} < V_{ijk}^* + \varepsilon_{ijk} < \mu_{i,r}), \quad r = 1, 2, \dots, 5, \quad (3)$$

where y_{ijk} is the rating contributed by i for product j in rating session k , $z_{ijk} = 1$ indicates that a rating is contributed (and $z_{ijk} = 0$ otherwise), and ε_{ijk} is the idiosyncratic error with a mean of 0. Under the assumption that ε_{ijk} follows a standard normal distribution, the probability with which an r -star rating is

contributed is represented by the following ordered probit specification:

$$\Pr(y_{ijk} = r | z_{ijk} = 1) = \begin{cases} \Phi(-V_{ijk}^*), & r = 1, \\ \Phi(\mu_{i1} - V_{ijk}^*) - \Phi(-V_{ijk}^*), & r = 2, \\ \Phi(\mu_{i2} - V_{ijk}^*) - \Phi(\mu_{i1} - V_{ijk}^*), & r = 3, \\ \Phi(\mu_{i3} - V_{ijk}^*) - \Phi(\mu_{i2} - V_{ijk}^*), & r = 4, \\ 1 - \Phi(\mu_{i3} - V_{ijk}^*), & r = 5, \end{cases} \quad (4)$$

where $\mu_{i, \cdot}$ are individual-specific cutpoints for the ordered probit model, and $\Phi(\cdot)$ denotes the standard normal cumulative distribution function (c.d.f.). Although our empirical context examines a five-star rating scale, our evaluation model can be generalized with ease to other rating formats. In the case of ordinal scales with a different number of response options, this would simply require modifying Equation (4) to accommodate the number of options available. For a continuous rating scale, the reported rating can be assumed to follow a normal distribution, in which case the reported rating y_{ijk} can be modeled directly using a linear model (Ansari et al. 2000) with mean V_{ijk}^* .

Incidence Model: Whether or Not to Rate a Product

We model an individual's decision to submit a product rating based on four components. First, the decision to contribute a product rating is contingent on an individual having experience with the product. Second, individuals may vary in their baseline tendencies to submit ratings for those products for which they have experience. Third and central to this research, the incidence decision may depend on the current state of the ratings environment at the time of the rating session. Last, we consider the impact of an individual's postpurchase evaluation for a particular product on his or her incidence decision.

We begin by discussing the importance of conditioning the incidence decision on a latent experience measure:

$$\Pr(z_{ijk} = 1) = \Pr(\text{experience}_j = 1) \cdot \Pr(z_{ijk} = 1 | \text{experience}_j = 1), \quad (5)$$

where $\Pr(z_{ijk} = 1)$ represents the probability that i contributes a rating for product j on rating session k . The inclusion of a latent experience term serves an important conceptual role. In a given rating session, we may observe products for which an individual does not post ratings. This may be a deliberate decision, perhaps as a result of the selection effects of the ratings environment, or it may arise from an individual not having experienced the product and hence not having the requisite knowledge to contribute a rating.

The latent experience component allows for the likelihood of the latter. In Equation (5), we conceptualize experience as a product-specific construct. As a result, our estimate of this latent construct would represent the average consumer's experience with the product and is analogous to market penetration parameters in models of product sales (Fourt and Woodlock 1960, Eskin 1973, Hardie et al. 1998, Moe and Fader 2001). Although we cannot identify the latent experience model at the individual level, future researchers with observable measures of experience can easily incorporate such data as covariates in this model component.⁶

Our main interest lies in the way in which incidence behavior may vary across individuals as a function of postpurchase evaluation and the ratings environment. Conditional on having experienced product j , individual i submits a rating for the product if

$$\beta_{i0} + \beta_{i,1:N}X_{j,k(i)} + \beta_{i,N+1}GT_{j,k(i)} + \delta_1 V_{ij} + \delta_2 V_{ij}^2 + \omega_{ijk} > 0, \quad (6)$$

where ω_{ijk} is idiosyncratic error with a mean of 0. Assuming that ω_{ijk} follows a standard normal distribution, this results in the following probit model with conditional probability:

$$\begin{aligned} \Pr(z_{ijk} = 1 \mid \text{experience}_j = 1) \\ = \Phi(\beta_{i0} + \beta_{i,1:N}X_{j,k(i)} + \beta_{i,N+1}GT_{j,k(i)} + \delta_1 V_{ij} + \delta_2 V_{ij}^2), \end{aligned} \quad (7)$$

where $\Phi(\cdot)$ denotes the standard normal c.d.f.

The first term (β_{i0}) allows for variation across individuals in their baseline propensities to submit product ratings. The vector $\beta_{i,1:N}$ captures the effect of covariates characterizing the ratings environment on the incidence decision (i.e., the selection effects). As public opinion and political science research has demonstrated, we expect the current ratings environment to affect an individual's decision to contribute product ratings. However, the nature and direction of these effects are empirical questions to be answered with the above-specified model. The term $\beta_{i,N+1}$ is a coefficient for the Google Trends control variable GT that serves as a proxy for general category-level interest and allows for differences in posting propensity across product categories related to the level of interest. Berger and Schwartz (2011) demonstrated how product characteristics relate to word of mouth, suggesting that the volume of opinion expression may

vary across product categories; here, we include the Google Trends variable in the posting incidence decision and use this as an exclusion restriction in our selection model.

The coefficients δ_1 and δ_2 allow for variation in incidence behavior based on i 's postpurchase evaluation of product j . Research has suggested that individuals holding extreme opinions (either positive or negative) will be more likely to contribute online opinions than individuals with moderate opinions (Dellarocas and Narayan 2006). We therefore expect that the postpurchase evaluation V_{ij} will affect the decision to post a rating in a nonlinear manner, allowing us to link the incidence and evaluation decisions in a flexible fashion.

The probability that i contributes a rating for product j in rating session k ($z_{ijk} = 1$) is given by the following unconditioned probability statement:

$$\Pr(z_{ijk} = 1) = \Phi(v_j) \cdot \Phi(\beta_{i0} + \beta_{i,1:N}X_{j,k(i)} + \beta_{i,N+1}GT_{j,k(i)} + \delta_1 V_{ij} + \delta_2 V_{ij}^2), \quad (8)$$

where $\Phi(v_j)$ is a function representing the probability that an average individual has experience with product j , $v_j = \theta + \tau_j$, θ is a parameter to be estimated, and $\tau_j \sim N(0, \sigma_\tau^2)$.⁷ Note that by simultaneously estimating the incidence and evaluation models, along with the inclusion of the GT covariate, we can distinguish among the effects of postpurchase evaluation V_{ij} (with coefficient vector δ), baseline incidence behavior (β_{i0}), and response to the ratings environment ($\beta_{i,1:N}$) in the posting incidence model.⁸

Linking the Incidence and Evaluation Models

Postpurchase evaluation, V_{ij} , is a key component in both the incidence and evaluation models. Whereas V_{ij} directly affects the evaluation model, its impact on posting incidence is governed by the parameters δ_1 and δ_2 . Thus, if $\delta_1 \neq 0$ or $\delta_2 \neq 0$, the incidence and evaluation models are not independent of each other. Our model allows for a flexible relationship to exist between the incidence and evaluation models. Consider first a linear relationship between postpurchase product evaluation and ratings incidence such

⁷ The model specified in Equation (8) assumes that latent experience and posting incidence are governed by two separate probit processes. As an alternative, we considered a single probit process that incorporates the components in Equation (8). This results in a more restrictive specification of the model that has a lower log marginal density but yields the same substantive findings as our proposed model.

⁸ Although we do not explicitly incorporate state dependence, we allow for behavior across products and rating sessions to be related through the model parameters, because an individual will exhibit the same behavioral tendencies (e.g., baseline propensity to post and response to covariates) across products and rating sessions.

⁶ Because we assume our latent experience measure to be product-specific and constant across individuals, we do not model the effects of an individual's experience on his or her incidence decision. We leave this for future researchers with access to individual-level experience data.

that individuals with higher postpurchase evaluations for a given product will be more likely to contribute a product rating. This would emerge from parameter values such that $\delta_1 > 0$ and $\delta_2 = 0$. However, the relationship between ratings incidence and postpurchase evaluation may not be monotonic. In fact, as noted previously, we anticipate that individuals with extreme postpurchase evaluations (either positive or negative) will be more likely to contribute product ratings ($\delta_2 > 0$).

To illustrate the associations between the incidence and evaluation components of our modeling framework, we assume a common distribution of postpurchase evaluations (V_{ij}) and consider three sets of values for the parameters δ_1 and δ_2 , yielding the three distinct distributions of *posted* ratings presented in Figure 2. Whereas the scenarios presented in the first two panels are accommodated by a model specification in which the idiosyncratic error terms for the evaluation and incidence models (ε_{ijk} and ω_{ijk} , respectively) are correlated (e.g., Heckman 1979, Ying et al. 2006), such a model does not permit the nonmonotonic relationship presented in the third panel.⁹

Model Estimation

The incidence model developed thus far has been conditional on a rating session occurring. However, to observe a rating session, at least one product rating must be posted; that is, it is not possible for $z_{ijk} = 0$ for all j . As a consequence, the joint likelihood of observing the vector of posting decisions $z_{i \cdot k}$ and the set of ratings observed on i 's k th rating session $y_{i \cdot k}$ is given by

$$L^*(z_{i \cdot k}, y_{i \cdot k} | X_{\cdot k(i)}, \beta_{i \cdot}, \gamma_{i \cdot}, \delta, \theta, \kappa, \tau) = \frac{\prod_{j \in \{z_{ijk}=1\}} \Pr(y_{ijk} | z_{ijk}=1) \Pr(z_{ijk}=1) \prod_{j \in \{z_{ijk}=0\}} (1 - \Pr(z_{ijk}=1))}{1 - \prod_{j=1}^J (1 - \Pr(z_{ijk}=1))}, \quad (9)$$

where $\Pr(y_{ijk} | z_{ijk}=1)$ is given by Equation (4) and $\Pr(z_{ijk}=1)$ is given by Equation (8). For those products for which ratings are posted ($z_{ijk}=1$), the likelihood is composed of both the likelihood that a rating is posted (incidence) and the likelihood associated with the particular rating posted (evaluation). For those products for which ratings are not

posted ($z_{ijk}=0$), the likelihood is composed only of the likelihood associated with the incidence model. The denominator accounts for the fact that at least one rating must be contributed during a rating session.

The joint likelihood for individual i who has K_i rating sessions is then given by

$$L((z_{i \cdot 1}, y_{i \cdot 1}), (z_{i \cdot 2}, y_{i \cdot 2}), \dots, (z_{i \cdot K_i}, y_{i \cdot K_i}), | X, \beta_{i \cdot}, \gamma_{i \cdot}, \delta, \theta, \kappa, \tau) = \prod_{k=1}^{K_i} L^*(z_{i \cdot k}, y_{i \cdot k} | X_{j, k(i)}, \beta_{i \cdot}, \gamma_{i \cdot}, \delta, \theta, \kappa, \tau). \quad (10)$$

To fit the proposed model, we use a hierarchical Bayes procedure. To allow for heterogeneity across individuals, as well as consider the correlation that may exist among the individual-level response parameters for the evaluation (γ) and incidence models (β), we assume that

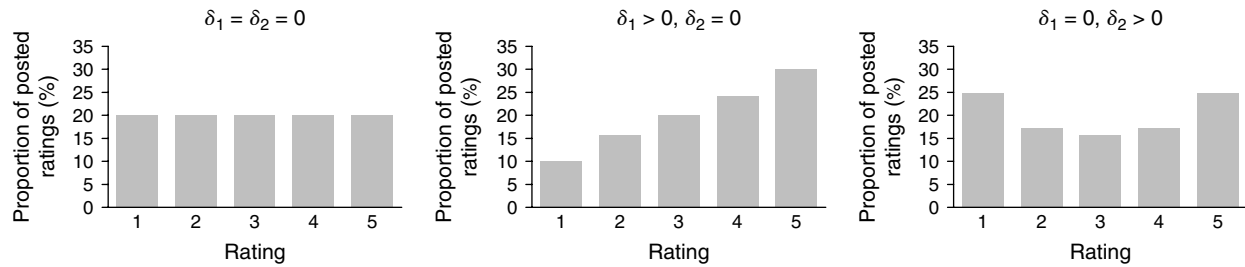
$$\begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix} \sim \text{MVN} \left(\begin{bmatrix} \bar{\beta} \\ \bar{\gamma} \end{bmatrix}, \Sigma \right). \quad (11)$$

Allowing for correlation between the parameters governing the incidence (β) and evaluation models (γ) generalizes the assumptions of heterogeneity made in prior analysis of online ratings. Ying et al. (2006) assumed that the individual-level parameters governing the incidence decision are correlated with each other and that those governing the evaluation decision are correlated with each other, but they do not consider the relationship that may exist across these two sets of parameters.

To complete our hierarchical model specification, we assume diffuse normal priors for the mean effects of the individual-level parameters in the incidence and evaluation models ($\bar{\beta}$ and $\bar{\gamma}$, respectively), the mean latent experience measure (θ), and the influence of postpurchase evaluation on the incidence decision (δ_1 and δ_2). For Σ , we employ an Inverse Wishart prior. For the cutpoints of the ordered probit evaluation model, we take the logs of the difference of adjacent cutpoints (Ying et al. 2006) and assume that the vector $[\log(\mu_{i1}), \log(\mu_{i2} - \mu_{i1}), \log(\mu_{i3} - \mu_{i2})] \sim \text{MVN}(\eta, T)$. We assume a diffuse normal prior for the elements of η and an Inverse Wishart prior for T . To make inferences under the proposed model, a Markov chain Monte Carlo sampler was run for 20,000 iterations, which served as a burn-in period. We then obtained inferences from posterior samples from the next 20,000 iterations. Details of the estimation procedure are presented in an electronic companion, available as part of the online version that can be found at <http://mktsci.journal.informs.org/>.

To assess model robustness, we also considered a number of alternative model specifications. These allowed for (1) differences across product categories

⁹ In our model specification, we assume that the errors in the incidence and evaluation models are independent. However, we also estimated a correlated error model like the one proposed by Ying et al. (2006) for comparison purposes. Like Ying et al. (2006), we find a small positive correlation between the two decision stages. Furthermore, the parameter estimates were not substantively different. Because of the computational demand of estimating this model and the substantively similar findings, we present the results from the model using independent errors and conduct further model comparisons using this specification.

Figure 2 Illustrative Distributions of Posted Ratings

and (2) changes over time. In both cases, the results were not meaningfully different from the proposed model.¹⁰

Covariate Specification

Extant research in the ratings literature has converged on a set of metrics that best describe previously posted ratings (see Dellarocas and Narayan 2006). These metrics have focused on the valence, variance, and volume of posted product ratings. Valence is typically represented by average rating, variance has been measured using statistical variance measures as well as other dispersion measures such as entropy, and volume is simply captured as the number of posted product ratings. However, the interpretation of these metrics can be problematic when each metric is considered separately. Consider a ratings environment with a single five-star rating compared to another with multiple five-star ratings. The valence and variance of ratings in both cases are the same. The effect of valence, however, may depend on the volume of ratings that have been contributed. For instance, a ratings environment with multiple five-star ratings may be perceived as more positive than one with a single five-star rating, whereas a ratings environment with multiple one-star ratings may be seen as more negative than one with just a single one-star rating. In a similar fashion, the impact of the ratings variance on the incidence and evaluation decisions may also depend on the volume of postings.

In addition to the main effects of valence, variance, and volume of previously posted ratings, we also consider the two-way interactions among these metrics. However, for the sake of parsimony and to eliminate potential collinearity in our modeled covariates,¹¹ we performed a factor analysis on the set of 36,600 daily ratings environments (200 products \times 183 days) as described by the main effects and interactions (see Mason and Perreault 1991 and Lehmann et al. 1998

Table 1 Rotated Component Matrix Resulting from Factor Analysis

Component	F_1	F_2
Valence	0.023	0.861
Variance	0.979	0.083
Volume	0.227	0.932
Valence \times Variance	0.981	0.145
Valence \times Volume	0.177	0.955
Variance \times Volume	0.936	0.238

for discussions of using factor scores as independent variables). The factor analysis results in two underlying constructs that explain 92% (61% by the first factor and 31% by the second factor) of the observed variation in daily ratings environments (see Table 1).

Although interpretation of the resulting factors is a concern when using factor analysis (Mason and Perreault 1991), our resulting factor scores offer a straightforward interpretation. The first factor (F_1) is strongly related to the variance of posted ratings and the variance's interaction with volume, whereas the second factor (F_2) is influenced by valence, volume, and their interaction. In other words, F_1 reflects the degree of consensus or dissention in the ratings environment with higher values associated with increased dissention, whereas higher values of F_2 reflect the overall positivity of posted product ratings, accounting for the number of ratings that have contributed to this positivity.¹² To assess the sensitivity of our analysis and the robustness of our findings, we considered an analysis in which we directly employed the valence and variance measures (omitting volume as a variable because of its collinearity with valence) and found that the model performed worse in terms of the log marginal density while yielding the same substantive results.

¹⁰ Details and results of these alternative models are available from the authors upon request.

¹¹ Measures of valence, variance, volume, and their interaction terms are highly correlated. A complete correlation matrix is available from the authors upon request.

¹² An examination of the scatterplots comparing the relationships between the raw metrics and the factor scores supports this interpretation of the factors. The scatterplots are available from the authors upon request.

Results

Parameter Estimates

Table 2 provides estimates of the mean effects and random effects from our model estimation. To demonstrate the extent of heterogeneity across individuals and products, we present the square root of the diagonal element from the corresponding covariance matrix.

We observe considerable heterogeneity across individuals in their incidence decisions. In addition to variation in the baseline propensity to post, we find variation in the nature and direction of selection effects (see Figure 3). We see that individuals vary in terms of their preference for posting in environments that exhibit consensus. Although some may be more prone to provide a product rating when previous posters were in agreement ($\beta_{i1} < 0$), others may abstain from posting in such environments and instead exhibit a preference for contributing product ratings when there is increased dissension ($\beta_{i1} > 0$). In contrast to the variation in how individuals respond to consensus (or the lack thereof), virtually all individuals exhibited a preference for posting in more positive ratings environments ($\beta_{i2} > 0$). In other words,

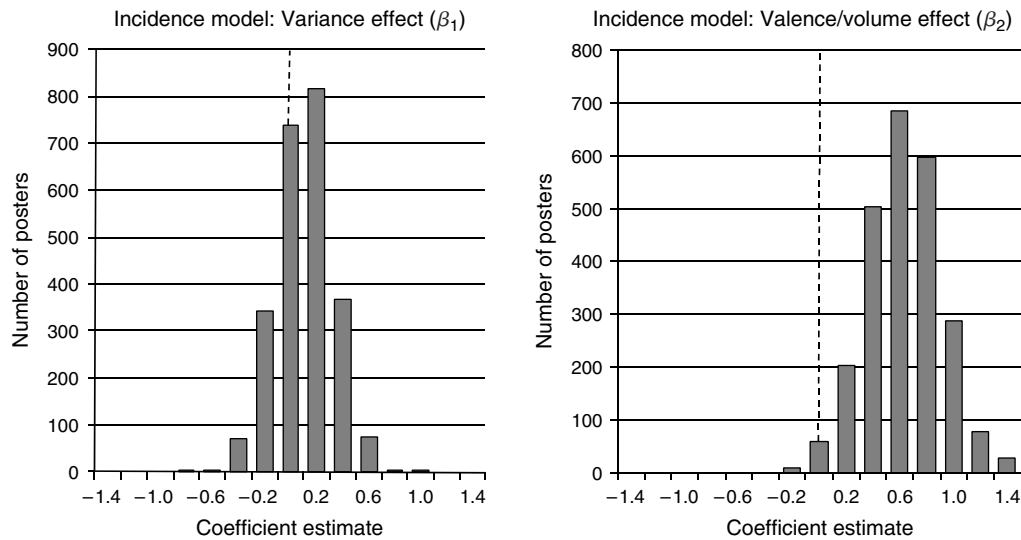
consumers are more likely to post an opinion when the ratings already posted are more positive.

To examine adjustment effects, we turn our attention next to the results of the evaluation model (see Figure 4). Compared with the variation seen across individuals' responses to opinion variance on the incidence model (as shown in Figure 3), we find less heterogeneity in the same effect in the evaluation model. In contrast, our results indicate a wider range of adjustment effects in the evaluation decision associated with the positivity of the ratings environment. Whereas some adjust their ratings upward in more positive ratings environments ($\gamma_{i2} > 0$), others adjust their ratings downward ($\gamma_{i2} < 0$) by lowering their reported product evaluations. In other words, we observe substantial heterogeneity in individuals' propensities to exhibit bandwagon versus differentiation effects in the evaluation stage.

An individual's posting decisions are also influenced by his or her postpurchase evaluation for product j , V_{ij} . It is tautological that an individual's postpurchase evaluation for a product will affect his or her posted evaluation of it; however, V_{ij} also affects the incidence decision. Figure 5 shows the relationship

Table 2 Model Estimation Results

Component	Parameter	Description	Mean effect (S.E.)	diag(Σ) ^{1/2} (S.E.)
Incidence model	v_j	Latent experience	−2.24 (0.02)	0.34 (0.02)
	β_{i0}	Incidence intercept	−2.53 (0.15)	1.75 (0.03)
	β_{i1}	Effect of F_1	0.0031 (0.01)	0.26 (0.00)
	β_{i2}	Effect of F_2	0.53 (0.01)	0.31 (0.01)
	β_{i3}	Effect of GT	0.047 (0.02)	0.22 (0.02)
	δ_1	Postpurchase evaluation (linear)	−0.44 (0.07)	—
	δ_2	Postpurchase evaluation (quadratic)	0.074 (0.01)	—
Evaluation model	κ_j	Product heterogeneity	—	0.61 (0.03)
	γ_{i0}	Evaluation intercept	3.04 (0.12)	0.62 (0.07)
	γ_{i1}	Effect of F_1	−0.024 (0.06)	0.38 (0.05)
	γ_{i2}	Effect of F_2	−0.11 (0.06)	0.50 (0.06)
	$\log(\mu_{i1})$	Difference in cutoff for rating = 2	−1.13 (0.17)	0.60 (0.10)
	$\log(\mu_{i2} - \mu_{i1})$	Difference in cutoff for rating = 3	−0.81 (0.10)	0.40 (0.08)
	$\log(\mu_{i3} - \mu_{i2})$	Difference in cutoff for rating = 4	−0.68 (0.11)	0.53 (0.10)

Figure 3 Posterior Means from the Incidence Model

between the postpurchase evaluation (V_{ij}) and the argument of the probit incidence model in Equation (7), as governed by the parameters δ_1 and δ_2 . We see that individuals are more likely to submit ratings when their postpurchase evaluation is high, consistent with empirical findings in the literature that show a strong positivity bias in online product ratings (Dellarocas and Narayan 2006). We also observe an increased likelihood of posting incidence when postpurchase evaluation is low, consistent with research showing that consumers are more likely to engage in word-of-mouth activities when they are dissatisfied (Anderson 1998). This empirical finding illustrates the potentially nonlinear relationship between the incidence and evaluation decisions and is in contrast with the monotonic relationship that has been assumed in previous research (e.g., Ying et al. 2006).

Relationships Among Incidence and Evaluation Behavior

To further examine the interdependencies between the incidence and evaluation decisions, we present the correlation matrix between the incidence model parameters (β) and the evaluation model parameters (γ) resulting from posterior estimates of the covariance matrix Σ in Table 3. The correlation coefficients indicate a number of interesting relationships, particularly with respect to the incidence model intercept, which reflects the frequency with which an individual posts a rating. The correlation coefficients indicate that frequent raters are more likely to post in dissident environments ($r = 0.37$) and more positive environments ($r = 0.45$). In terms of their evaluation behavior, these frequent raters also tend to be more

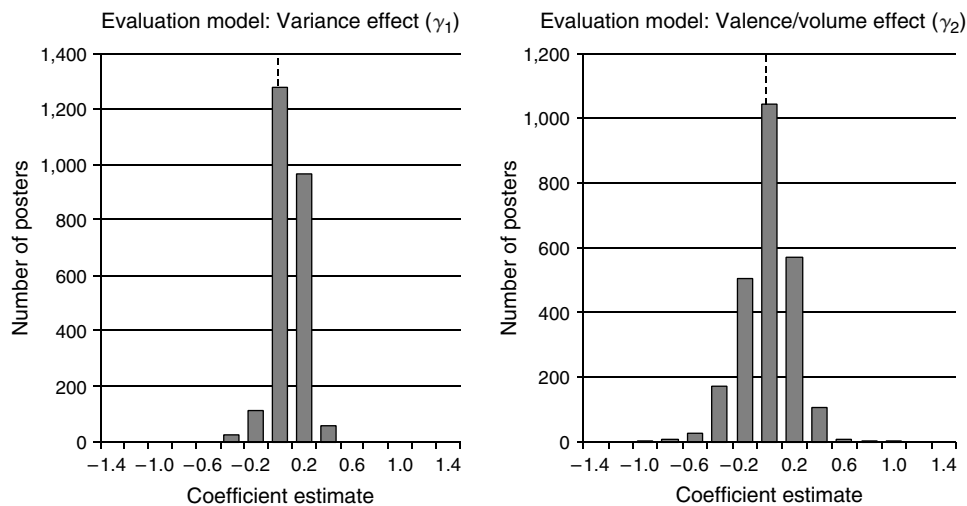
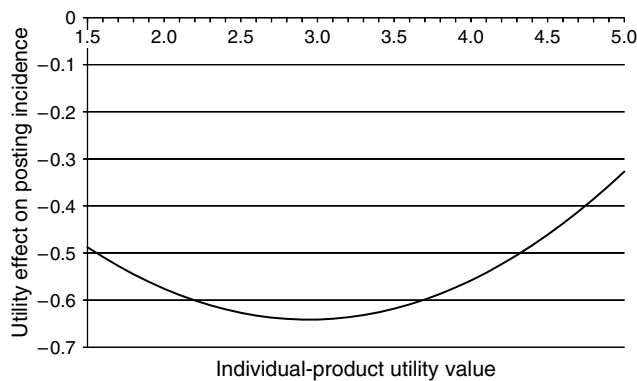
Figure 4 Posterior Means from the Evaluation Model

Figure 5 Role of Postexperience Evaluation in Rating Incidence



negative in their ratings ($r = -0.68$) and tend to differentiate their posted ratings from others ($r = -0.34$).

To further highlight the differences across posters, Table 4 categorizes individuals into low-, moderate-, and high-frequency posters based on posterior estimates of their incidence baseline (β_0). Our results indicate noticeably different posting behavior between frequent and infrequent posters. Whereas all individuals are more prone to post in more positive environments, the impact of consensus varies. Less active posters are more prone to post in consensus environments ($\beta_1 = -0.0932$), whereas the most active posters are more likely to contribute when there is higher opinion variance ($\beta_1 = 0.116$).

A closer examination of evaluation behaviors further distinguishes active posters from less active posters. The least active posters tend to be more positive ($\gamma_0 = 3.496$) than more active posters. Furthermore, they exhibit bandwagon effects ($\gamma_2 = 0.0749$) by adjusting their posted evaluations upward when previous ratings have been more positive. They are also more positive when previously posted opinions vary ($\gamma_1 = 0.0497$). To some extent, these individuals can be characterized as “low-involvement.” Not only are they less engaged in terms of their level of posting activity, but they are also easily influenced by others through bandwagon effects.

Table 3 Correlation Matrix

	Incidence model			Evaluation model		
	β_0	β_1	β_2	γ_0	γ_1	γ_2
Intercept (β_0)	1					
Variance (β_1)	0.37*	1				
Valence – Volume (β_2)	0.45*	0.45*	1			
Intercept (γ_0)	–0.68*	–0.25*	–0.13*	1		
Variance (γ_1)	–0.17	0.10	–0.02	0.06	1	
Valence – Volume (γ_2)	–0.34*	–0.21*	–0.27*	0.12	0.12	1
Google Trends (β_3)	–0.12*	–0.01	–0.10	0.19	–0.02	0.02

*Indicates 0 is not contained in the 95% HPD interval.

Table 4 Comparing Low-Involvement Contributors to Activists

	Low-involvement	Moderates	Activists
Incidence model			
Intercept (β_0)	–4.426	–2.490	–0.566
Variance (β_1)	–0.0932	0.00251	0.116
Valence – Volume (β_2)	0.379	0.534	0.685
Evaluation model			
Intercept (γ_0)	3.496	3.043	2.591
Variance (γ_1)	0.0497	–0.0284	–0.0965
Valence – Volume (γ_2)	0.0749	–0.109	–0.298

The most active posters stand in stark contrast to the low-involvement posters. In general, they are more negative ($\gamma_0 = 2.591$) and exhibit differentiation behavior (as opposed to bandwagon behavior) by adjusting their evaluations downward in more positive ratings environments ($\gamma_2 = -0.298$). Their posted opinions are even more negative when there is increased opinion variance ($\gamma_1 = -0.0965$). These behaviors are consistent with previous research showing that individuals who want to be perceived as “experts” often try to differentiate themselves by contributing more negative opinions (Schlosser 2005, Amabile 1983). Overall, these individuals appear to be “activists” who are highly engaged (as indicated by their high posting frequency) and may try to establish themselves in the community by offering opinions designed to attract the attention of others that are both differentiated and more negative.

Discussion of Empirical Results

Our results show that both the incidence and evaluation decisions are affected by an individual’s postpurchase evaluation in a nonmonotonic fashion. Individuals with either high or low postpurchase evaluations are more likely to contribute ratings, whereas individuals with moderate postpurchase evaluations are less likely to contribute ratings. By decomposing ratings behavior into distinct but related incidence and evaluation stages, we see that the effects of the ratings environment differ across stages. Although prior research has documented the role of the ratings environment on the evaluation decision (Schlosser 2005), the literature has not discussed the effect that the ratings environment can have on incidence. We find that the ratings environment significantly affects an individual’s decision of whether or not to contribute a rating, resulting in a selection effect that affects the composition of the posting population. Our analysis further reveals that incidence and evaluation behaviors are linked and that the opinions of those who frequently contribute ratings differ from those who do not. These dynamics result in the distribution of posted ratings not necessarily resembling the underlying sentiment of the full customer base—an important caveat for marketers who turn to such

online forums to gauge customer opinion and for consumers looking for unbiased product evaluations.

The Evolution of Product Opinion

In this section, we simulate ratings environments based on the results obtained from our model in an effort to more closely examine the drivers of observed evolutionary patterns. Our simulation procedure is as follows:

Step 1. Based on the model parameters, we simulate a population of 10,000 individuals.

Step 2. From the population of 10,000 individuals, we draw a sample of 1,000 individuals to represent the product's customer base.

Step 3. We categorize each individual according to his or her baseline posting incidence.

Step 4. We simulate the postpurchase evaluation V_{ij} .

Step 5. Based on V_{ij} , we simulate the incidence decision according to Equation (7) and the posted rating conditional on a rating being posted according to Equation (4).

Step 6. We repeat Steps 4 and 5 until 50 ratings are posted.

Step 7. We repeat Steps 1–6 for 5,000 iterations and average the results across iterations.

Simulation 1: Overall Evolutionary Patterns

We begin by randomly drawing a customer base from the total population and examining their posted ratings over time. Figure 6(a) plots the average rating and ratings variance, and Figure 6(b) illustrates the composition of the poster population. As more ratings arrive, the average rating decreases and the variance across ratings increases, a dynamic driven by a shift in the composition of the posting population. As more ratings are posted, the ratings environment is likely to exhibit more variation in opinions, thereby attracting activists who are more negative in their evaluations (and deterring lower involvement/more positive posters) and further contributing to a gradual decline in average ratings. Although this trend has been observed in prior research (Li and Hitt 2008, Godes and Silva 2012), our analysis and simulation suggest that this trend can (at least partially) be explained by a shift in the composition of the posting population. As such, the selection effect is central to the way in which expressed opinions evolve.

Simulation 2: Diversity of Customer Opinions

We next compare a ratings environment resulting from a highly polarized customer base to one that is relatively homogeneous and representative of the median opinion. To construct the highly polarized

Figure 6 Simulated Ratings Evolution Resulting from a Representative Customer Base

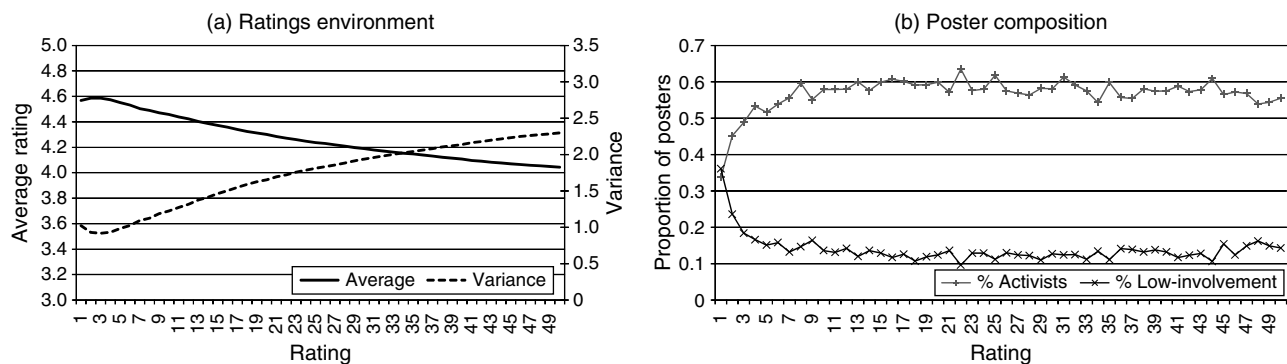
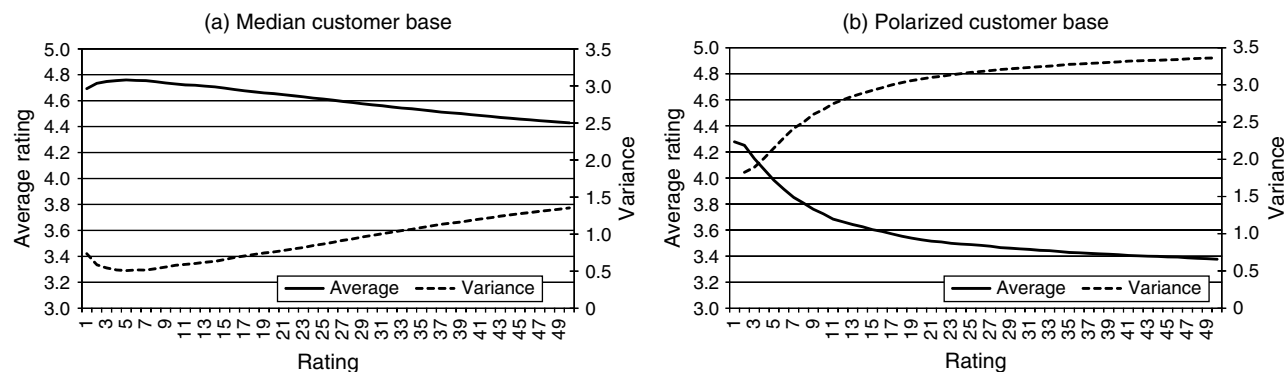


Figure 7 Simulated Ratings Evolution Resulting from a Median vs. Polarized Customer Base



customer base, we sample the 500 individuals with the lowest evaluation model intercept (γ_0) and the 500 individuals with the highest evaluation model intercept. To construct the median customer base, we sample the 1,000 individuals in the middle of the distribution. Although the variance differs substantially between these two hypothetical customer bases, the average evaluation is the same.

Figure 7 describes the simulated ratings resulting from the median customer base and the polarized customer base. As expected, the posted ratings resulting from the homogeneous median customer base are relatively uniform and exhibit a slight downward trend. In contrast, the posted ratings from the polarized customer base are more negative and exhibit a stronger downward trend. This dynamic is driven by the more negative baseline evaluation and the differentiation behavior of the activists who represent half of the customer population in the polarized customer base but are absent from the median customer base. The contributions of this core group of posters eventually result in posted ratings that are dominated by extreme negative opinions and do not reflect the distribution of opinions from the overall customer base.

Discussion and Conclusions

Although previous work has studied the effects of online product ratings on consumers' purchase decisions at the product level, research to date has not explored dynamics affecting the individual-level decisions of *whether* to post an opinion or *what* to post. We present a modeling framework to examine the effects of previously posted content on both posting incidence and evaluation decisions. The evaluation decision is subject to adjustment effects that alter the content of future postings, and the incidence decision is subject to selection effects that can shape the composition of the posting population. This latter dynamic has been largely ignored in the extant research on online product opinions.

We show empirically that (1) positive environments increase posting incidence, whereas negative environments discourage posting; and (2) less frequent posters are more positive and exhibit bandwagon behavior, whereas more active posters are more negative and exhibit differentiation behavior. Differentiating between low-involvement posters and activists reveals systematic differences in their incidence and evaluation behaviors. Overall, we find that online opinions are dominated by activists who offer opinions that are more negative and differentiated from previously expressed opinions. Moreover, participation by these activists increases over time, whereas participation by low-involvement individuals decreases. This shift in the composition of the posting population can substantially affect the overall tone of posted opinions.

Additionally, the composition of the customer base can exert a substantial influence on the manner in which posted online opinions evolve. Because of selection and adjustment effects, the content posted may not necessarily reflect the customer base's overall opinion of the product. Rather, a vocal subset of the customer base may dominate the ratings environment, consequently steering the subsequently posted evaluations and deterring some customers from contributing to the environments. Marketers and consumers alike must consequently exercise caution in drawing inferences from posted product ratings and reviews, because the opinions they observe may not provide an accurate gauge of the overall customer base's perceptions.

Although we considered a number of extensions to our general modeling framework, a number of directions remain open for future research. For example, with additional information our framework can be generalized to investigate the prepurchase and postpurchase processes simultaneously. Browsing behavior or past purchases, for instance, could provide additional information as to the set of products for which an individual has the requisite experience to post ratings, as well as distinguish between the prepurchase expectation and postpurchase evaluation. This would allow for the development of an individual-level experience model, which may be related to subsequent incidence and evaluation decisions. We believe such an integrated model of the pre- and postpurchase processes can offer additional insight into individuals' rating decisions and offer further guidance to marketers who must decide how to react to online evaluations of their products or brands.

Although our empirical application considers product ratings, it is worthwhile to examine the way in which other types of online forums are shaped by such opinion dynamics. In addition to the generalizations for ordinal and continuous scales noted in our discussion of the model, future research may consider extending our framework to contexts that use multi-item scales or textual comments. Research may also explore more flexible discussion forums to further our understanding of user-generated content and how this increasingly important factor in the consumer decision process is created. Whatever opinion format is being studied, both selection and adjustments effects must be taken into account in examining the dynamics of the forum.

As our research suggests, future work must recognize that the direction of the conversation may discourage the participation of some individuals and consequently no longer represent the general customer base. As the online marketplace becomes increasingly interactive, consumers play a larger role

in the creation of content that can influence the success or failure of a product. Because of this, it is critical that marketers understand who is contributing their opinions to forums, their motives for doing so, and what influences their behavior. This paper contributes to this larger effort.

Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://mktsci.journal.informs.org/>.

Acknowledgments

The authors thank the Wharton Customer Analytics Initiative and the Marketing Science Institute for supporting this research. The authors also thank BazaarVoice for providing the data used in this research.

References

- Amabile, T. M. 1983. Brilliant but cruel: Perceptions of negative evaluators. *J. Experiment. Soc. Psych.* **19**(2) 146–156.
- Anderson, E. W. 1998. Customer satisfaction and word of mouth. *J. Service Res.* **1**(1) 5–17.
- Anderson, E. W., M. W. Sullivan. 1993. The antecedents and consequences of customer satisfaction for firms. *Marketing Sci.* **12**(2) 125–143.
- Ansari, A., S. Essegai, R. Kohli. 2000. Internet recommendation systems. *J. Marketing Res.* **37**(3) 363–375.
- Berger, J., E. Schwartz. 2011. What drives immediate and ongoing word of mouth? *J. Marketing Res.* **48**(5) 869–880.
- Berinsky, A. J. 2004. *Silent Voices: Public Opinion and Political Participation in America*. Princeton University Press, Princeton, NJ.
- Carneiro, H. A., E. Mylonakis. 2009. Google Trends: A Web-based tool for real-time surveillance of disease outbreaks. *Clinical Infectious Diseases* **49**(10) 1557–1564.
- Chevalier, J. A., D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* **43**(3) 345–354.
- Dellarocas, C., R. Narayan. 2006. A statistical measure of a population's propensity to engage in post-experience online word-of-mouth. *Statist. Sci.* **21**(2) 277–285.
- Delli Carpini, M. X. 1984. Scooping the voters? The consequences of the networks' early call of the 1980 presidential race. *J. Politics* **46**(3) 866–885.
- Dubois, P. L. 1983. Election night projections and voter turnout in the West: A note on the hazards of aggregate data analysis. *Amer. Politics Res.* **11**(3) 349–364.
- Epstein, L. K., G. Strom. 1981. Election night projections and West Coast turnout. *Amer. Politics Res.* **9**(4) 479–491.
- Eskin, G. J. 1973. Dynamic forecasts of new product demand using a depth of repeat model. *J. Marketing Res.* **10**(2) 115–129.
- Fleming, J. H., J. M. Darley, J. L. Hilton, B. A. Kojetin. 1990. Multiple audience problem: A strategic communication perspective on social perception. *J. Personality Soc. Psych.* **58**(4) 593–609.
- Fourt, L. A., J. W. Woodlock. 1960. Early prediction of market success for new grocery products. *J. Marketing* **25**(2) 31–38.
- Gartner, M. 1976. Endogenous bandwagon and underdog effects in a rational choice model. *Public Choice* **25**(1) 83–89.
- Godes, D., J. C. Silva. 2012. Sequential and temporal dynamics of online opinion. *Marketing Sci.* **31**(3) 448–473.
- Hardie, B. G. S., P. S. Fader, M. Wisniewski. 1998. An empirical comparison of new product trial forecasting models. *J. Forecasting* **17**(June/July) 209–229.
- Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica* **47**(1) 153–161.
- Jackson, J. E. 1983. Election night reporting and voter turnout. *Amer. J. Political Sci.* **27**(4) 615–635.
- Kamakura, W. A., M. Wedel, F. de Rosa, J. A. Mazzon. 2003. Cross-selling through database marketing: A mixed data factor analyzer for data augmentation and prediction. *Internat. J. Res. Marketing* **20**(1) 45–65.
- Kuksov, D., Y. Xie. 2010. Pricing, frills, and customer ratings. *Marketing Sci.* **29**(5) 925–943.
- Lehmann, D. R., S. Gupta, J. H. Steckel. 1998. *Marketing Research*. Addison-Wesley, New York.
- Li, S., B. Sun, R. T. Wilcox. 2005. Cross-selling sequentially ordered products: An application to consumer banking services. *J. Marketing Res.* **42**(2) 233–239.
- Li, X., L. M. Hitt. 2008. Self-selection and information role of online product reviews. *Inform. Systems Res.* **19**(4) 456–474.
- Marsh, C. 1984. Back on the bandwagon: The effect of opinion polls on public opinion. *British J. Political Sci.* **15**(1) 51–74.
- Mason, C. H., W. D. Perreault Jr. 1991. Collinearity, power, and interpretation of multiple regression analysis. *J. Marketing Res.* **28**(3) 268–280.
- McAllister, I., D. T. Studlar. 1991. Bandwagon, underdog, or projection? Opinion polls and electoral choice in Britain, 1979–1987. *J. Politics* **53**(3) 720–741.
- Moe, W. W., P. S. Fader. 2001. Modeling hedonic portfolio products: A joint segmentation analysis of music compact disc sales. *J. Marketing Res.* **38**(3) 376–385.
- Moe, W. W., M. Trusov. 2011. The value of social dynamics in online product ratings forums. *J. Marketing Res.* **48**(3) 444–456.
- Park, Y.-H., E. T. Bradlow. 2005. An integrated model for bidding behavior in Internet auctions: Whether, who, when, and how much. *J. Marketing Res.* **42**(4) 470–482.
- Schlosser, A. E. 2005. Posting versus lurking: Communicating in a multiple audience context. *J. Consumer Res.* **32**(September) 260–265.
- Stratton P. D., Jr. 1977. The bandwagon curve. *Amer. J. Political Sci.* **21**(4) 695–709.
- Sudman, S. 1986. Do exit polls influence voting behavior? *Public Opinion Quart.* **50**(3) 331–339.
- Varian, H. R., H. Choi. 2009. Predicting the present with Google Trends. *Google Research* (blog), April 2, <http://googleresearch.blogspot.com/2009/04/predicting-present-with-google-trends.html>.
- Ying, Y., F. Feinberg, M. Wedel. 2006. Leveraging missing ratings to improve online recommendation systems. *J. Marketing Res.* **43**(3) 355–365.