2021

Uncertainty-reduction or reciprocity? Understanding the effects of a platform-initiated reviewer incentive program on regular review generation

Uncertainty-Reduction or Reciprocity? Understanding the Effects of a Platform-Initiated Reviewer Incentive Program on Regular Review Generation

Abstract

To stimulate product reviews, many e-commerce platforms have launched reviewer incentive programs in which free product samples are provided to reviewers in exchange for their ratings of the samples. This study focuses on an unexplored aspect of reviewer incentive programs—the impact of participating in such programs on reviewers' ratings of products they purchased normally (i.e., regular ratings). We find that after reviewers join the program and receive free product samples, their average regular rating increases by 2.25% (i.e., 0.093 more stars on the five-star scale). Our follow-up analyses indicate that the observed regular-rating increase can be attributed to an *uncertainty-reduction effect* evoked by the free product samples, as opposed to a *reciprocity effect*. We further delve into the underlying mechanism by analyzing the reviewers' regular ratings at a granular, product-category level. Consistent with our theorization of the uncertainty-reduction effect, our findings reveal that reviewers' regular-rating increase is driven by improved assessment and knowledge about products sharing common attributes with the sampled products, resulting in better post-purchase outcomes. Our results demonstrate that apart from motivating the feedback for the sampled products, free product sampling can reduce reviewers' product uncertainty and trigger evident change in their regular ratings for the purchased products.

Keywords: *Product review ratings, reviewer incentive program, uncertainty reduction, reciprocity*

1. Introduction

Online reviews have become an important information source. 66% of customers read product reviews before making purchasing decisions (Nielson 2014), and 90% of those who read reviews state that their purchasing decisions are influenced by the reviews (Dimensional Research 2013). Owing to this influence on consumers, e-commerce platforms have begun using reviewer incentive programs to stimulate product reviews through free product sampling. For example, Amazon.com (henceforth, Amazon) introduced a reviewer incentive program, *Vine,* to which experienced reviewers were invited. Participants can receive free product samples through the program, and they need to provide their opinions for the received samples; Walmart.com launched a similar program, known as *Spark*, to facilitate consumer feedback; Qunar.com and Ctrip.com, two of the biggest travel booking platforms in China, provided free hotel stays to consumers in exchange for their feedback; and Christianaudio.com developed a similar reviewer incentive program to generate reviews for their audiobooks. The model for incentive programs is similar across platforms. It is usually most trusted and experienced reviewers who are invited to such programs. Once they have joined the program, participants can choose and receive free product samples, for which they are required to provide feedback, since the purpose of such programs is to generate reviews for the sampled products. In other words, reviewer incentive programs are originally designed to accelerate the generation of consumer reviews for the sampled products.

Extant studies find that receiving free products from independent *sellers* can trigger a reciprocity effect and motivate receivers to post higher ratings for the sampled products (Lin et al. 2019), and significantly reduce the likelihood of negative feedback (Cabral and Li 2015). Concerned that such seller-initiated incentives can bias the review system, platforms start to restrict its use. For instance, starting from October 2016, Amazon stops all the seller-initiated reviewer incentive practice and prohibits independent sellers from sending out free samples in exchange for favorable reviews, and instead encourages reviewers to join its *Vine* program.[1] Moreover, platforms emphasize that the purpose of their reviewer incentive

[1] See https://blog.aboutamazon.com/innovation/update-on-customer-reviews (accessed Feb 26th, 2020)

programs is to generate organic feedbacks to products, and the reviewers need to provide their honest opinions of the sampled products. For example, Amazon specifies that reviewers' ratings are irrelevant to their eligibility to its *Vine* program.[2] Thanks to these policies, the seller-initiated reviewer incentive practice has been gradually phased out, and the platform-initiated reviewer incentive programs become the dominant reviewer incentive programs on e-commerce platforms.

However, despite the significant influence of platform-initiated reviewer incentive programs, scant attention has been paid to the fact that the reviewers continue to generate reviews for their purchased products after joining such programs. In this study, we refer to the ratings for the free product samples as the *incentivized ratings* and the ratings for the purchased products as the *regular ratings*. While existing studies have focused on the effect of seller-initiated incentives on the *incentivized ratings* (Cabral and Li 2015, Lin et al. 2019), whether and how receiving free product samples from the platform-initiated programs can impact the reviewers' *regular ratings* is not well understood despite its economic significance.

The impact of the platform-initiated programs on regular ratings is not trivial, since the participants in such programs are often trusted and experienced reviewers, and their opinions are influential on consumers' affinity and purchase decisions (Chen et al. 2008, Cone 2011, Nielson 2014). The regular ratings also outnumber the incentivized ratings in volume, and customers trust the regular ratings more (Kokkodis and Lappas 2016). It is thus critical for both practitioners and researchers to understand the effects of the reviewer incentive programs on participants' regular ratings. Therefore, we pose our main research question—*How do a reviewer's regular ratings change after he or she starts to receive free product samples from a reviewer incentive program*?

The answer to this question is not straightforward, as reviewers' regular ratings can increase or remain unchanged after they join the program and receive free product samples. Reviewers' regular ratings result from their post-purchase outcomes (e.g., Ho et al. 2017) and their pursuit of benefits (e.g., Shen et al. 2015). On the one hand, experiencing product samples provides reviewers with more concrete product

[2] https://www.amazon.com/gp/vine/help (accessed Feb 26th, 2020)

information (Mark and Kamins 1988) and reduces their uncertainty on related products that share similar attributes with the samples, leading to better purchase decisions and thus higher regular ratings. Besides, reviewers may perceive free samples as monetary gains and attempt to reciprocate the platform by posting higher regular ratings. On the other hand, however, reviewers' regular ratings can remain unchanged. As reviewers' purchased products are not usually identical to the sampled ones, it is unclear to what extent their purchase uncertainty can be alleviated by their experience of product samples. Besides, reviewers can repay the platform through different means (e.g., more purchases), and whether they will reciprocate the platform by intentionally increasing their regular ratings is an open question.

An empirical challenge to answer the proposed question is the possible endogeneity issue arising from self-selection into such a reviewer incentive program. While a randomized controlled experiment where reviewers are randomly assigned to the treatment group (i.e., the reviewer incentive program) or the control group is ideal, it is not feasible in our context because the regular ratings result from reviewers' experience with their purchased products, and it can be prohibitively challenging to mimic consumers' real purchase behaviors in an experimental setting. Therefore, we rely on observational data from the *Vine* program, Amazon's reviewer incentive program. For the identification purpose, we use a difference-in-differences method combined with a propensity score matching technique. Specifically, we estimate a two-way fixed effects model with both time and reviewer fixed effects to control for the unobserved time trend and the reviewer attributes.

Our results indicate that after reviewers receive free product samples from the *Vine* program, their regular ratings increase by 2.25% (0.093 more stars on the five-star scale). Following the existing literature (e.g., Jung et al. 2019), we confirm the result with a rigorous falsification test, in which we show that the rating change can only be observed around the actual program participating time but not around the arbitrarily assigned time points. Moreover, we use a relative time model and confirm that in the pre-participation period, there is no significant difference between the regular ratings posted by the reviewers who join the program and the ratings posted by the reviewers who do not.

We further delve into the underlying mechanism of the increased regular ratings after reviewers receive free product samples from the platform-initiated reviewer incentive programs. Our finding implies that receiving free product samples from the platform may bias the recipients' evaluations for the purchased products and that platforms can use free product sampling to "buy" positive ratings. We explore two possible mechanisms driving the regular-rating increase, namely, the *uncertainty-reduction effect* and the *reciprocity effect*. Under the uncertainty-reduction effect, the free product samples help the reviewers better understand the products and reduce their uncertainty about products similar to the sampled product, leading to improved post-purchase outcomes such as higher regular ratings. Under the reciprocity effect, after receiving free products (benefit) from the platform (benefit provider), reviewers (benefit recipients) reciprocate with higher ratings on the platform (Gouldner 1960).

Differentiating the two mechanisms may result in very different managerial decisions for e-commerce platforms. If the former is the case, that receiving free samples biases reviewers' evaluations is less of concern. That is, the regular-rating increase is mainly due to improved assessment and knowledge about products sharing common attributes with the sampled products, resulting in better post-purchase outcomes. However, if the observed regular-rating increase is due to a reciprocity effect, platforms need to be cautious in implementing the reviewer incentive program because free products can be considered financial gains by reviewers and potentially bias their evaluations of purchased products.

To explore these two possibilities, we investigate reviewers' regular rating changes at a fine-grained, product-category level, and results consistently indicate that the uncertainty-reduction mechanism is at play in our context. Specifically, we find that (i) reviewers' regular ratings increase only in the categories where the product samples are available, (ii) reviewers' regular-rating increase is more pronounced in product categories where products are less differentiated (or more homogenous) such that they are more conducive to product assessment in the same category, and (iii) reviewers' regular ratings in one category are positively associated with the number of product samples they receive from the same category, but not with the number of samples from other categories. We further confirm these results by using products' full category information and verifying our results at more granular, subcategory and sub-subcategory levels.

Our study demonstrates that a platform-initiated reviewer incentive program not only encourages incentivized reviews but also elevates the recipients' regular ratings for the purchased products. Given the economic significance of regular ratings, platforms need to keep abreast of these economic consequences and design their reviewer incentive programs accordingly.

In the next section, we review the related literature. In Section 3, we discuss theoretical backgrounds of this study. Section 4 describes our research context and dataset. Section 5 specifies our empirical models and results. We discuss the underlying mechanism in Section 6. Section 7 conducts a series of robustness checks. Section 8 discusses scholarly and practical implications and concludes.

2. Literature Review

2.1 Regular Rating Generation

There are two streams of literature investigating reviewers' rating generation. The first stream studies the relationship between consumers' post-purchase outcomes and their regular ratings. For instance, consumers are more likely to post ratings when they have extreme post-purchase utility, leading to a bimodal distribution of ratings (Hu et al. 2006). Moreover, Li and Hitt (2008) indicate that consumers who like the product are more likely to post reviews early in a product's life cycle, and the product's average rating tends to decrease over time. Li and Hitt (2010) demonstrate that reviewers determine their ratings based on the overall purchase utility, thus the post-purchase value of the product decreases as product prices rise. Ho et al. (2017) show the effect of disconfirmation, the difference between the expectation at the pre-purchase stage and realized assessment of the same product at the post-purchase stage, on consumers' rating behaviors. Similarly, Lin and Heng (2015) indicate that extremely high ratings at the early stage of the product cycle are more likely to result in negative reviews subsequently because of the large disconfirmation in a later stage of the product cycle.

The second stream investigates the relationship between social and economic factors and reviewers' product ratings. Some studies focus on social benefits, such as the attention of their peers. For example, Shen et al. (2015) demonstrate that the competition between reviewers for attention affects the reviewers' ratings, as reviewers tend to post ratings different from the existing ones. Goes et al. (2014) show that consumers' ratings are affected by the number of incoming social ties from other consumers. Lee et al. (2015) report the impact of prior ratings by strangers is different from that by friends on a subsequent individuals' rating provision. Other studies examine the impact of financial incentives offered by platforms or sellers on reviewers' regular ratings. Cabral and Li (2015) show that when the sellers offer a promotion, such as a cash rebate, the likelihood of negative feedback is significantly reduced due to reciprocity. Similarly, Lin et al. (2019) document that reviewers increase their rating of sampled product provided by a seller on Taobao.com as they reciprocate the seller with higher ratings.

To the best of our knowledge, our study is the first to investigate the effects of the platform-initiated reviewer incentive program on participants' regular ratings for their purchased products.

2.2 Product Sampling

Free product sampling is a common promotional tool. The free samples help consumers understand the product better and make more informed purchase decisions (Biswas et al. 2010). Free product sampling has been reported more effective than other marketing strategies (e.g., Smith and Swinyard 1983, Mark and Kamins 1988, Smith 1993) and result in improved economic outcomes in terms of product sales (Hahn et al. 1994), brand sales (Bawa and Shoemaker 2004), and firm profits (Cheng and Liu 2012).

The premise of free product sampling is that sampling allows consumers to learn about a product and thus reduce their perceived uncertainty associated with its quality and value (Rothschild and Gaidis 1981, Goering 1985, Jamieson and Bass 1989). For example, Mark and Kamins (1988) show that belief and attitudinal confidence for certain products are higher whey consumers experience and sample them. Mano and Oliver (1993) show that the purchase satisfaction is affected by the accuracy of product evaluation, which can be enhanced by product sampling. In general, product sampling helps consumers form a more accurate expectation (Goering 1985, Cheng and Liu 2012) and belief (Mark and Kamins 1988) of a product.

In this study, we focus on an emerging marketing practice, free product sampling, to stimulate user product ratings. Our study contributes to the burgeoning literature that investigates the relationship between free product sampling and consumers' product ratings (e.g., Lin et al. 2019). However, unlike extant literature that investigated the impact of product sampling on the incentivized rating (Lin et al. 2019) or on the sampled products (Lee and Tan 2013, Mo and Li 2018), we aim to explore the impact of free product sampling on regular ratings.

3. Theorization of the Impact on Regular Ratings

In this section, we build theoretical foundations for the effect of free product sampling on reviewers' regular ratings. To this end, we draw upon two streams of prior work which have documented that reviewers' regular ratings are related to their post-purchase outcomes (e.g., Lin and Heng 2015, Ho et al. 2017) and their pursuit of benefits (e.g., Shen et al. 2015), respectively.

First, receiving free product samples can result in higher regular ratings by helping consumers make more informed purchase decisions. Consumers have uncertainties about products they have not experienced, and such uncertainties become more pronounced in e-commerce contexts (Hong and Pavlou 2014). Before they make purchases, consumers conduct an investigation to better understand the product (Murray 1991), especially when monetary cost is involved (Gu et al. 2012). Studies have found that with product samples, consumers can not only learn more about products (Mark and Kamins 1988) but also form more accurate expectation of them (Goering 1985). Free product sampling enables reviewers to improve their assessment and knowledge about the sampled product as well as other products sharing similar features, leading to an uncertainty-reduction effect. Therefore, we posit that extended knowledge and experiences from product sampling can lead to better post-purchase outcomes in terms of higher regular ratings. Alternatively, it is also possible that receiving free product samples may not help consumers make better purchase decisions if there is little or no common attribute between their sampled and purchased products.

Second, after joining the program, reviewers can change their regular ratings due to their pursuit of benefits. In this light, reviewers may post higher regular ratings to the platform because of reciprocity (Gouldner 1960). Recognizing a platform's interest in seeking higher product ratings (e.g., Gu et al. 2012), participants in the platform-initiated reviewer incentive program may reciprocate the platform in the form of higher regular ratings. Alternatively, it is also possible that reviewers do not necessarily increase their regular ratings but reciprocate the platform through other means, such as more purchases and greater effort on review generation.

Drawing upon two aforementioned theoretical grounds, we hypothesize that participating in a platform-initiated reviewer incentive program increases, or at least does not decrease, regular ratings, which we empirically examine in the next section.

4. Research Context and Data

We collect data from Amazon to investigate these issues empirically. In August 2007, Amazon launched the *Vine* program, a reviewer incentive program providing participants with free product samples. Amazon invited reviewers to join the program based on how experienced they are (e.g., the overall helpfulness of their reviews, review experience in different product categories, and review volume). The participants in this program are thus experienced reviewers. After joining the program, reviewers choose and receive product samples without paying,[3] and they are required to provide feedback for those samples. The reviewers can maintain their status in *Vine* as long as they provide feedback within 30 days after receiving a free product. Amazon specifies that reviewers' ratings (i.e., our interested aspect) don't affect their status in the program.[4]

On the review page, the badge "Vine Customer Review of Free Product" indicates that the reviewer receives the product as a free sample from Amazon. In this study, we treat the reviews for the free Vine product samples as the incentivized reviews (or *Vine* reviews) and the rest as the regular reviews. Figure 1 shows examples of these two types of reviews.

[3] Reviewers receive a newsletter every month listing the free products they can choose. See https://en.wikipedia.org/wiki/Amazon_Vine (accessed Feb 26th, 2020).
[4] See https://www.amazon.com/gp/vine/help (accessed Feb 26th, 2020).

Figure 1. Screenshot of Two Types of Review on Amazon

To understand how reviewers' regular ratings change after they join the *Vine* program and start to receive free product samples, we need to check all reviews posted by a group of sampled reviewers to identify who have generated *Vine* reviews and when. We therefore collect all reviews posted by the top experienced reviewers and the corresponding product information for each review.[5] We choose the top experienced reviewers as our sampled reviewers because the platforms generally invite experienced reviewers into the program. Moreover, to examine the impact of the *Vine* program on the participants' regular ratings and correct the sample selection issue, we conduct sample matching before our empirical estimation. Focusing on reviewers who are all top experienced helps us make a fair comparison.

In this study, we examine reviewers' ratings between January 1, 2007 and December 31, 2009 for several reasons. First, Amazon initiated the *Vine* program in August 2007, so targeting this period helps us better understand reviewers' rating differences before and after the existence of the program. Second, during our study period, the policies of the *Vine* program remained constant. Once participating in the program, reviewers can stay in the program as long as they provide feedback on time after receiving free product samples. In other words, whether they are eligible to be in the program is not contingent on their regular ratings. Finally, this sample period is before reviewers start to receive free products from independent sellers on Amazon; thus, our estimations are not affected by those activities.[6, 7]

---

[5] We collect the top experienced reviewers from Amazon's Top Customer Reviewers page, which lists the top 10,000 reviewers based on their experience, such as total reviews posted and the total helpfulness votes received.

[6] This practice started in 2010 and was fully banned by Amazon in 2016. See https://reviewmeta.com/blog/analysis-of-7-million-amazon-reviews-customers-who-receive-free-or-discounted-item-much-more-likely-to-write-positive-review/ (accessed Feb 26th, 2020).

[7] We also analyze the reviewers' regular ratings in different sampled time periods, such as January 2007 to December 2013, and the results are qualitatively consistent.

We focus on the active reviewers who have written at least one review before the study period and continue to post reviews throughout this period. Consequently, our sample includes 3,487 reviewers, of whom 1,673 reviewers participated in the *Vine* program. Also, our review dataset encompasses 336,899 reviews, including 46,162 incentivized reviews and 290,737 regular reviews. For each review, our dataset also includes the reviewer ID, product ID, product category, product name, product price, review timestamp, review text, rating, and review helpfulness vote.

5. Econometric Model and the Impact on Regular Ratings

We employ a difference-in-differences (DID) approach to test how participating in a platform-initiated reviewer incentive program affects reviewers' regular ratings. The basic idea is to compare a group of treated reviewers in the pre- and post-treatment periods (i.e., before and after joining the program and receiving product samples), with a group of untreated reviewers to control for the unobserved time-variant factors. Figure 2 illustrates our DID design. The treatment group includes the reviewers who generate at least one incentivized review and the control group includes other reviewers. The treated reviewers generate regular reviews before joining the program, and they provide both regular and incentivized reviews after starting to receive the product samples. Reviewers in the control group only generate regular reviews during the entire period. It is worth noting that in this study, we only focus on reviewers' regular ratings. Although the treated reviewers generate both types of reviews after joining the program, we do not consider their incentivized ratings when we construct the dependent variable.

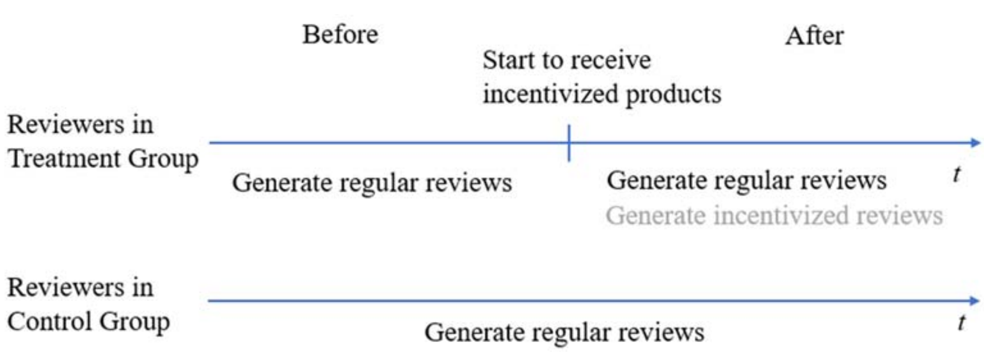

Figure 2. Review Activities of Reviewers in Different Groups

We employ a propensity score matching to correct for a potential sample selection bias due to the observable differences between treated reviewers and untreated reviewers. Specifically, we generate a sample of untreated reviewers who are similar to the treated reviewers. In other words, each treated reviewer is paired with an untreated reviewer who is similar in terms of the probability of participating in the *Vine* program (the detailed matching procedures are in Appendix A). From our calculation of the propensity score and Amazon's official announcement, we learn that reviewers' propensity for joining the program is associated with factors such as the overall helpfulness of the reviews, reviewers' experience in different product categories, and review volume. We use one-to-one matching with replacement, under a caliper size 0.2 times the standard deviation of the propensity score (Xu et al. 2016). After matching, we compare the treated reviewers with the reviewers in the control group, and we find the reviewers are not statistically different in two groups (Appendix A, Table A.2). We further compare the distributions of the propensity scores of the matched and unmatched samples and check the covariate balance before and after matching using a Kolmogorov–Smirnov test. The results of the covariate balance check and the distribution test jointly reveal that we have a strong match between the two groups of reviewers. Figure 3 shows that the propensity score distributions of the treated and untreated reviewers are similar after matching.

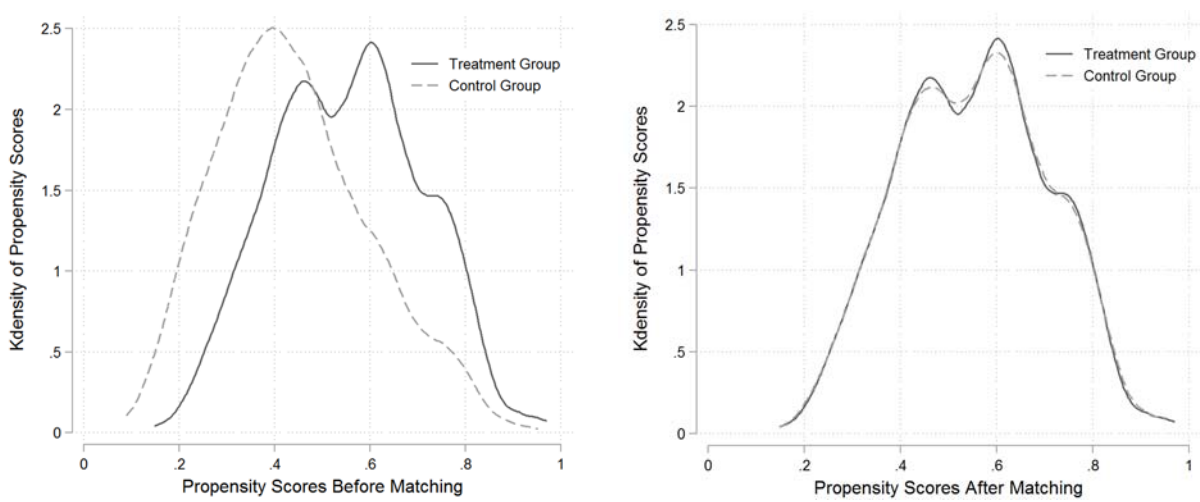

Figure 3. Distributions of Propensity Scores for Treatment Group and Control Group

After constructing the reviewer sample, we build a panel dataset at the reviewer-month (*it*) level, where *i* and *t* indicate a reviewer and a month, respectively. In our setting, reviewers' program participation is a staggered adoption (Athey and Imbens 2018). Specifically, (i) the observations are in multiple periods, (ii) reviewers' program participation time varies, and (iii) reviewers remain in the program after participating. With this setting, following the existing literature (e.g., Xu et al. 2016, Callaway and Sant'Anna 2019), we utilize a two-way fixed effects regression model to estimate the average treatment effects. Each observation describes reviewer *i*'s average regular rating at month *t*. Combining the matching strategy with a DID approach, our empirical model is as follows:

$$AvgRegRating_{it} = \beta_0 + \beta_1 VineReviewer_i + \beta_2 VineMonth_{it} + \beta_3 VineReviewer_i \times VineMonth_{it} + X_{it} + \tau_t + \delta_i + \epsilon_{it} \quad (1)$$

where $AvgRegRating_{it}$ is the dependent variable, representing reviewer *i*'s average regular rating at month *t*. τ_t is the time fixed effects to control for unobserved time trend and δ_i is the reviewer-level fixed effects to control for the unobserved reviewer-level characteristics. $VineReviewer_i$ is a binary variable that equals 1 if reviewer *i* is an incentivized reviewer, and 0 otherwise. $VineMonth_{it}$ is a binary variable indicating the post-participation period for the treated reviewer *i* and his/her matched counterpart. For instance, if a reviewer joins the *Vine* program in December 2007, $VineMonth_{it}$ equals 1 for that month and the months after and equals 0 before that month. The matched reviewer for this treated reviewer will have the same value for $VineMonth_{it}$ throughout the sample period. The coefficient of the interaction term β_3 is the DID estimator that captures how the reviewers' regular ratings change after they join the *Vine* program—our main interest. Table 1 summarizes the description and statistics of the main variables in our matched dataset.

In addition to our main covariates, we also control for a series of observed covariates, X_{it}. We first include reviewers' past review behaviors in each product category. The ratio of the reviews on a specific product category to the total generated reviews represents the reviewer's review experience in that category. Specifically, we include $BookPastRatio_{it-1}$, $MoviesPastRatio_{it-1}$, $CDsPastRatio_{it-1}$, $ElectronicPastRatio_{it-1}$, $HomePastRatio_{i-1t}$, $GroceryPastRatio_{it-1}$, and $HealthPastRatio_{it-1}$ to represent the ratio of total reviews from the categories “Books,” “Movies & TV,” “CDs & Vinyl,” “Electronics,” “Home & Kitchen,” “Grocery & Gourmet Food,” and “Health & Personal Care,” respectively, to the total number of reviews reviewer *i* generated by the end of month *t-1*. We choose those seven product categories because they are the most popular ones, and we use the ratio for the rest of the product categories as our baseline.

Table 1. Description of Variables

Variable	Description	Mean	Std. Dev.	N
$VineReviewer_i$	A binary variable indicating whether reviewer *i* participated in the *Vine* program between our study period: 1 = Reviewer *i* participated 0 = Reviewer *i* did not participate	0.424	0.494	120,240
$VineMonth_{it}$	A binary variable indicating whether reviewer *i* participated in the *Vine* program at month *t*: 1 = Reviewer *i* participated at month *t* 0 = Reviewer *i* did not participate at month *t*	0.500	0.500	120,240
$AvgRegRating_{it}$	The average rating of regular reviews reviewer *i* generated at month *t*	4.142	0.920	54,668

Note: The number of observations for *AvgRegRating* is less than the number of observations for the other variables. Because our observation is on the reviewer-month level, *AvgRegRating* is missing values for the months that reviewers did not generate any regular reviews. *AvgRegRating* is on a scale of 1 to 5.

We further account for the possibility that reviewers’ ratings may vary by the reviewed product category. To this end, we include $BookRatio_{it}$, $MoviesRatio_{it}$, $CDsRatio_{it}$, $ElectronicRatio_{it}$, $HomeRatio_{it}$, $GroceryRatio_{it}$, and $HealthRatio_{it}$, representing the ratio of regular reviews from the categories of “Books,” “Movies & TV,” “CDs & Vinyl,” “Electronics,” “Home & Kitchen,” “Grocery & Gourmet Food,” and “Health & Personal Care,” respectively, to the total number of regular reviews generated by reviewer *i* at month *t*. Moreover, we use the log-transformed number of months reviewer *i* has been on Amazon by month *t* to represent reviewer *i*’s experience on Amazon. We cluster the error terms at the reviewer level to account for autocorrelation over time (Sun and Zhu 2013, Xu et al. 2016). Last, β_0 is an intercept and ϵ_{it} is a mean-zero random error term.

We present the estimation for the impact on reviewers’ regular ratings in Table 2. β_3 is positive and statistically significant (p-value<0.001). After reviewers join the reviewer incentive program and start to receive free product samples, their regular ratings increase by 0.093 stars on average, equivalent to an increase of 2.25% (0.093/4.142*100%). Such an increase in ratings is also managerially significant, especially when consumers are more likely to trust regular reviews and feedback from experienced reviewers (Chen et al. 2008). Furthermore, review systems normally use discrete rating systems (one star, two stars, etc.), so a small increase in reviewers' ratings can translate into a one-star difference in the actual rating displayed to consumers (Goes et al. 2014).

Table 2. Impact of Participating in *Vine* Program on Regular Ratings

	(1)
VineReviewer (β_1)	-
VineMonth (β_2)	-0.017 (0.017)
VineReviewer× *VineMonth* (β_3)	0.093*** (0.018)
Control Variables	YES
Time Fixed	YES
Reviewer Fixed	YES
R-squared	0.315
Observations	54,668

Note. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at the user-month level. *Significant at 10%; **significant at 5%; ***significant at 1%. The results are qualitatively consistent if we don't include the reviewer or time fixed effects.

6. Mechanism Explorations

In this section, we explore several possible explanations to delve into the key underlying mechanism at play. Results demonstrate that (i) the increased regular ratings from the platform-initiated reviewer incentive program are not caused by reviewers' selection of the highly rated products, (ii) the number of product samples is the key driver of the observed effect, and (iii) the free product samples affect participants' regular ratings through an *uncertainty-reduction effect*, instead of a *reciprocity effect*.

6.1 Driver Identification of Regular-Rating Changes

One possible concern is that the observed regular-rating increase is simply due to other changes in reviewers' review generation. For instance, it is possible that after starting to produce the incentivized reviews, reviewers decrease their regular-review volume and focus only on the high-rated products. To rule out this alternative explanation, we construct two dependent variables and test them with Equation (1); one is the reviewer *i*'s regular review volume at month *t*, and the other is the average rating difference between reviewer i's regular rating to a product and the product's average rating. The analysis of the first dependent variable discerns whether reviewers are actively engaged in regular-review generation, and the analysis of the second variable indicates whether the observed regular-rating increase is simply caused by reviewers' selection of the high-rated products.

Table 3 presents the results. When the dependent variable is the regular review volume (Model 1), β_3 is positive and significant, indicating that after joining the program, reviewers tend to produce more regular reviews. This finding further supports the conclusion that the magnitude of the regular-rating increase is meaningful. Moreover, when the dependent variable is the average rating difference between reviewer i's regular rating of a product and the product's average rating (Model 2), the estimation of β_3 is qualitatively consistent with that in Table 2. The result shows that after reviewers join a reviewer incentive program, they tend to provide higher regular rating relative to the product's average rating. This indicates reviewers' regular-rating increase is not simply due to reviewers' selection of the high-rated products.

Table 3. The Managerial Importance of the Regular-Rating Increase

Dependent Variable	Review Volume	Rating Difference
	(1)	(2)
VineReviewer (β_1)	-	-
VineMonth (β_2)	0.031** (0.013)	0.020 (0.012)
VineReviewer× *VineMonth* (β_3)	0.190*** (0.019)	0.013*** (0.005)
Control Variables	YES	YES
Time Fixed	YES	YES
Reviewer Fixed	YES	YES
R-squared	0.567	0.283
Observations	120,240	50,425

Note. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at the user-month level. *Significant at 10%; **significant at 5%; ***significant at 1%.

Our analysis shows that after reviewers join the program and start to receive product samples, their regular ratings increase. The observed rating change could potentially be caused by a participation effect or by receiving product samples (i.e., *Vine* products). The participation effect happens if reviewers are affected by simply joining the program, and their regular ratings increase as soon as they join the program.

However, if the reviewers' rating increase is related to the product samples, their rating increase should vary by their engagement with the product sampling. To formally check whether the participation effect alone plays a role, we first conduct subsample analyses to examine the effects of the *Vine* program on the reviewers who receive a small number of sampled products and on the reviewers who receive a large number of sampled products. In operationalizing this, we first choose reviewers who receive the lowest quartile of aggregated sampled products: 418 incentivized reviewers are in this quartile, and on average these reviewers receive only 0.54 *Vine* products per month. We then consider reviewers who receive the highest quartile of aggregated sampled products: 426 incentivized reviewers are in this quartile, and they receive 3.32 free products per month on average. If reviewers' rating increase is solely due to the participation effect, the average regular rating should increase for the reviewers from both groups. In contrast, if the received product samples are the key driver for reviewers' rating increase, we should observe the rating increase only for the reviewers in the highest quartile.

Table 4 presents the results of the subsample analyses. After reviewers join the *Vine* program, their regular ratings do not significantly change if they receive a small number of free products (Model 1). However, for the reviewers who do receive a large number of free products, regular ratings significantly increase after they start to receive the *Vine* products (Model 2). The results suggest that the participation effect is not, by itself, enough to motivate reviewers to provide higher regular ratings. In other words, the reviewers' rating increase is closely related to the received product samples.

We then alter Equation (1) to analyze the effect of product samples on reviewers' regular ratings. Specifically, we substitute the DID term $VineReviewer_i \times VineMonth_{it}$ with $VineProducts_{it-1}$, which is the log-transformed number of *Vine* products reviewer i receives up to the end of month $t\text{-}1$. The DID term in Equation (1) captures the average treatment effect on regular ratings, and the examination of $VineProducts_{it-1}$ reveals the relationship between reviewers' regular ratings and the total number of *Vine* products they receive. The results (Table 4, Model 3) indicate that reviewers' average regular rating is positively associated with the number of *Vine* products they receive from the program.

Table 4. Impact of Volume of *Vine* Products

Specifications	Lowest Quartile	Highest Quartile	Vine Volume
	(1)	(2)	(3)
VineReviewer	-	-	-
VineMonth	0.064 (0.036)	-0.008 (0.033)	0.030** (0.013)
VineReviewer× *VineMonth*	0.037 (0.037)	0.099*** (0.048)	-
VineProducts	-	-	0.014** (0.006)
Control Variables	YES	YES	YES
Time Fixed	YES	YES	YES
Reviewer Fixed	YES	YES	YES
R-squared	0.311	0.316	0.315
Observations	12,842	13,733	54,668

Note. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at the user-month level. *Significant at 10%; **significant at 5%; ***significant at 1%.

6.2 Theorization of Two Possible Mechanisms

Our analyses thus far paint the following picture. After participating in the *Vine* program, reviewers increase their regular ratings (Table 2). The regular-rating increase is managerially significant, and it cannot be explained by reviewers' selection of highly rated products (Table 3). Moreover, receiving a sufficient number of free samples is essential for the rating increase, and the reviewers' regular ratings are positively associated with the number of samples they receive (Table 4). The analyses demonstrate the impact of the *Vine* program on reviewers' regular ratings and the key driver—free product samples. Here, we explore two possible underlying mechanisms by which platform-initiated reviewer incentive programs elevate regular product ratings, namely, the *uncertainty-reduction effect* and the *reciprocity effect*.

First, receiving free product samples can result in higher regular ratings through an *uncertainty-reduction effect*. Consumers' understanding of a product category can help them make better purchase decisions in that category (Johnson and Russo 1984). Product samples help consumers learn more about product features (Mark and Kamins 1988) and increase their assessment and knowledge about other products sharing common attributes (e.g., products within the same category) (Biswas et al. 2010), leading to enhanced post-purchase outcomes. By the same token, the uncertainty-reduction theory does not predict or explain increased regular ratings of products that don't share common attributes with the sampled products (e.g., products in other categories). Moreover, reviewers' regular ratings are negatively associated with disconfirmation between the pre-purchase expectation and post-purchase experience (Lin and Heng 2015, Ho et al. 2017). Such expectation disconfirmation can be mitigated if consumers understand the product better and form more accurate expectations at the pre-purchase stage. Therefore, we posit that increased knowledge from product samples can lead to better post-purchase outcomes in terms of higher regular ratings.

Second, receiving free products may stimulate a *reciprocity effect* as reviewers may perceive the free products as financial benefits and reciprocate the platform with higher ratings. This is because individuals receiving benefits from others may feel obligated to behave in a more friendly way to the provider of the benefits (Gouldner 1960). Moreover, benefit receivers determine their reciprocal behaviors based on their understanding of the purpose of benefits providers (Falk and Fischbacher 2006). In the platform-initiated reviewer incentive program, the provider of the free samples is the platform; and the platform's ultimate goal is to sell products and product sales are positively associated with product ratings (Duan et al. 2008, Forman et al. 2008, Gu et al. 2012). In this light, reviewers may choose to repay the platform with higher ratings because they believe doing so can discharge their obligation to the platform. Additionally, the platform-initiated reviewer incentive program publicly announces that the participants should rate the product samples in an unbiased manner. The theory of reciprocity states that receiving benefits can place the recipients in an uncomfortable state of tension ("indebtedness" in Greenberg 1980). As a result, reciprocity occurs when the recipient wants to repay the received benefits to reduce this discomfort. When an opportunity to reciprocate is limited, people tend to find alternative ways to reduce this discomfort (Shumaker and Brownell 1984, p. 14). Therefore, when repaying the platform with higher incentivized ratings is restricted, it is plausible reviewers repay the platform with higher regular ratings as an alternative. Unlike the uncertainty-reduction effect, the reciprocity effect does predict or explain increased regular ratings of products across all product categories, regardless of whether they belong to the category of sampled products, which is useful in our empirical validation in the next section.

6.3 Empirical Tests for Two Possible Mechanisms

To empirically check whether the *uncertainty-reduction effect* or the *reciprocity effect* is at play, we construct a panel dataset at the reviewer-category-month level (*ikt*) where *i*, *k*, and *t* represent a reviewer, a product category, and a month, respectively. Each observation describes reviewer *i*'s behaviors for product category *k* at month *t*. This model is also based on the matched reviewer samples, and further allows us to better describe the received *Vine* products. Instead of using the portion of the products from each category as controls in Equation (1), we now introduce the product category fixed effects to control for the category heterogeneity. Specifically, we use the following model:

$$\begin{aligned} AvgRegRating_{ikt} = \quad & \beta_0 + \beta_1 VineReviewer_i + \beta_2 VineMonth_{it} + \beta_3 VineReviewer_i \times \\ & VineMonth_{it} + X'_{it} + \tau_t + \delta_i + \rho_k + \epsilon_{it} \end{aligned} \tag{2}$$

where the dependent variable $AvgRegRating_{ikt}$ represents reviewer *i*'s average regular rating for product category *k* at month *t*, and ρ_k is the product-category fixed effects. We estimate this model and confirm that the result (Table 5, Model 1) is consistent with that in Table 2.

Furthermore, with the panel dataset at the reviewer-category-month level, we distinguish the *uncertainty-reduction effect* and the *reciprocity effect* by differentiating the product categories based on several standards. The following analyses consistently demonstrate that reviewers' regular-rating increase can be explained by the *uncertainty-reduction effect* of product sampling.

6.3.1 *Vine* and Non-*Vine* Categories

We differentiate the product categories based on the availability of *Vine* products. Specifically, *Vine* products are not always available in all product categories. We conduct separate subsample analyses for reviewers' regular ratings in the product categories where *Vine* products are available (i.e., *Vine* categories) and in the product categories where *Vine* products are not (i.e., Non-*Vine* categories). If an uncertainty-reduction effect manifests and reviewers learn from the sampled products, we expect the available product samples to be related to the purchased products. That is, if reviewers' regular ratings increase owing to enhanced product information from *Vine* product sampling, their regular ratings are likely to increase only in the *Vine* categories, not in the Non-*Vine* categories. In contrast, if reviewers intentionally increase their regular ratings to repay the platform, their regular ratings should increase in both the *Vine* and Non-*Vine* categories indiscriminately. The results show that reviewers' regular ratings increase only in the *Vine* categories (Table 5, Model 2), not in the Non-*Vine* categories (Table 5, Model 3), suggesting that the availability of the related product samples plays an important role in reviewers' regular-rating increase for the purchased products.

Table 5. Impact of the *Vine* Program in Vine and Non-Vine Product Categories

Specifications	Overall	Vine Categories	Non-Vine Categories
	(1)	(2)	(3)
VineReviewer (β_1)	-	-	-
VineMonth (β_2)	-0.004 (0.016)	-0.007 (0.016)	1.020* (0.545)
VineReviewer× *VineMonth* (β_3)	0.067*** (0.017)	0.071*** (0.017)	-0.523 (0.581)
Control Variables	YES	YES	YES
Time Fixed	YES	YES	YES
Reviewer Fixed	YES	YES	YES
R-squared	0.212	0.224	0.789
Observations	85,496	84,736	760

Note. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at the user-month-category level. *Significant at 10%; **significant at 5%; ***significant at 1%.

6.3.2 Low- and High-Heterogeneity Categories

We differentiate the product categories based on the level of product heterogeneity within a category. Specifically, we differentiate the product categories based on whether the products within the category are closely related to each other. The awarded *Vine* products and purchased products are not identical, and if reviewers can understand more about the purchased products by using the sampled ones, the *Vine* products that reviewers receive can better inform reviewers on related products they purchase. In this light, for the categories where products are highly differentiated from each other, we conjecture that the uncertainty-reduction effect will be minimal because reviewers are less able to infer helpful information from the sampled products due to heterogeneity in products within the same category. Accordingly, we divide *Vine* categories into high-heterogeneity and low-heterogeneity product categories. Specifically, we treat the product categories that are about entertainment and media as high-heterogeneity product categories, since creating highly heterogeneous content is normally the goal of this industry (e.g., Hanson and Xiang 2011, Hennig-Thurau and Houston 2019, p. 205). For instance, we treat the “Book” and “Kindle Store” as the high-heterogeneity product categories. Reading one sampled book does not necessarily help consumers understand more about the qualities of other books to be purchased due to their inherent differences in genre, style, appeal, and other characteristics. We treat the rest categories as low-heterogeneity product categories. Table 6 provides the detailed category classifications.

We conduct subsample analyses and estimate reviewers’ regular-rating changes in the high- and low-heterogeneity categories, respectively. Results in Table 7 show that both the magnitude and the significance level of reviewers’ regular-rating increases in the low-heterogeneity categories are higher than those in the high-heterogeneity ones. Specifically, after joining the program, reviewers’ average regular rating increases by 0.091 stars in the high-heterogeneity categories (β_3 in Model 1, p-value=0.091), compared to 0.161 stars in the low-heterogeneity categories (β_3 in Model 2, p-value<0.001). The results demonstrate that the positive effect of *Vine* products on reviewers’ regular ratings is more pronounced for the categories where the products are not less differentiated, confirming that the uncertainty-reduction effect of the free samples plays a key role.

Table 6. Classifications of Product Categories

Classifications	Product Categories
Vine Categories	**High-Heterogeneity Categories** Books; Video Games; Movies & TV; CDs & Vinyl; Kindle Store
	Low-Heterogeneity Categories Arts, Crafts & Sewing; Automotive; Baby Products; Beauty; Cell Phones & Accessories; Clothing, Shoes & Jewelry; Electronics; Grocery & Gourmet Food; Software; Health & Personal Care; Home & Kitchen; Industrial & Scientific; Office Products; Patio, Lawn & Garden; Pet Supplies; Power & Hand Tools; Sports & Outdoors; Tools & Home Improvement; Toys & Games
Non-*Vine* Categories	Appliances; Collectibles & Fine Art; Grills & Outdoor Cooking; Kitchen & Dining; Lawn Mowers & Outdoor Power Tools; Lights & Lighting Accessories; Magazine Subscriptions; Medical Supplies & Equipment; Mobility & Daily Living Aids; Musical Instruments; Outdoor Cooking Tools & Accessories; Power Tool Parts & Accessories; Safety & Security; Small Appliance Parts & Accessories; Sports & Fitness

Table 7. Impact of Product Categories of Different Heterogeneity

Specifications	High-Heterogeneity Categories	Low-Heterogeneity Categories
	(1)	(2)
VineReviewer (β_1)	-	-
VineMonth (β_2)	-0.008 (0.017)	-0.026 (0.038)
VineReviewer× *VineMonth* (β_3)	0.091* (0.019)	0.161*** (0.040)
Control Variables	YES	YES
Time Fixed	YES	YES
Reviewer Fixed	YES	YES
R-squared	0.270	0.152
Observations	53,027	31,709

Note. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at the user-month-category level. *Significant at 10%; **significant at 5%; ***significant at 1%.

6.3.3 Within- and Cross-Category *Vine* Products

We divide the product samples based on the product category. Specifically, we investigate how reviewers' regular ratings in a product category are associated with the number of *Vine* products from the same category (i.e., focal category) and the number of *Vine* products from the other categories. If the reviewers' regular-rating increase is related to an uncertainty-reduction effect, their regular ratings in a category should be positively associated only with the number of *Vine* products from the same category, but not with the number of *Vine* products from the other categories. This is because experiencing product samples can help reviewers better understand closely related products, the products that share common attributes with the samples (e.g., from the same category). In contrast, if the reviewers' regular-rating increase is stimulated by a reciprocity effect, we can expect that the number of *Vine* products they receive, no matter which categories they come from, is positively associated with their regular ratings. In other words, receiving *Vine* products would increase the regular ratings regardless of the similitude between the *Vine* products and their purchased ones, if the product samples are treated as benefits.

After receiving product samples, if the reviewers' regular ratings increase due to an uncertainty-reduction effect, their regular ratings for one product category should be positively associated with the number of product samples from that same category, but not samples from other categories. Moreover, we expect this pattern to be significant in the low-heterogeneity product categories but not in the high-heterogeneity product categories. On the contrary, if the free samples increase the reviewers' regular ratings through a reciprocity effect, their regular ratings in one product category would be positively associated with the number of product samples from both the focal and other categories. Similarly, this positive association would then be found in all categories (i.e., both high- and low-heterogeneity product categories).

To test our conjecture, we alter Equation (2) and substitute $VineReviewer_i \times VineMonth_{it}$ with $VineProductsFocal_{ikt-1}$ and $VineProductsOther_{ikt-1}$, which are the log-transformed number of *Vine* products reviewer *i* respectively receives from category *k* and other categories up to the end of month *t-1*. Table 8 presents the results. For the high-heterogeneity categories, reviewers' regular ratings are not associated with the number of products they receive (Model 1). In contrast, for the low-heterogeneity categories, reviewers' regular ratings in one category are positively associated only with the number of *Vine* products from the focal category, but not with the number of *Vine* products from other categories (Model 2). These results suggest that reviewers' regular-rating increase primarily stems from an uncertainty-reduction effect caused by the received *Vine* products, rather than a reciprocity effect.

Table 8. Impact of Volume of Within- and Cross-Category Vine Products

Specifications	High-Heterogeneity Categories	Low-Heterogeneity Categories
	(1)	(2)
VineReviewer	-	-
VineMonth	0.009 (0.014)	0.087*** (0.025)
VineProductsFocal	-0.004 (0.007)	0.042** (0.017)
VineProductsOther	0.006 (0.007)	0.008 (0.012)
Control Variables	YES	YES
Time Fixed	YES	YES
Reviewer Fixed	YES	YES
R-squared	0.310	0.221
Observations	53,027	31,709

Note. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at the user-month-category level. *Significant at 10%; **significant at 5%; ***significant at 1%.

We also construct the panel dataset at a higher level of granularity. Amazon classifies the products into a category tree. There are several subcategories within each category and several sub-subcategories within each subcategory. Each product's full category information is "Category › Subcategory › Sub-subcategory," as illustrated by a tree-shaped graph in Figure 4. We construct the panel dataset at the reviewer-subcategory-month level, where the dependent variable is reviewer *i*'s average regular rating for a subcategory on month *t*. We divide *VineProductsFocal* from the previous analyses into *VineProductsSubFocal* and *VineProductsSubOther*, which are the log-transformed numbers of *Vine* products reviewer *i* gets from the focal subcategory and other subcategories, respectively, up to the end of month *t-1*. The results (Table 9, Model 1) indicate that reviewers' regular ratings in one subcategory are positively associated with the number of *Vine* products they receive from the focal subcategory but not with the number of *Vine* products from other subcategories or other categories.

We further construct the panel dataset at the reviewer- 'sub-subcategory'- month level, where the dependent variable is reviewer *i*'s average regular rating for a sub-subcategory on month *t*. We divide the *VineProductsSubFocal* into *VineProductsSubSubFocal* and *VineProductsSubSubOther*, which are the log-transformed number of *Vine* products reviewer *i* gets from the focal sub-subcategory and other sub-subcategories, respectively, up to the end of month *t-1*. The results show that reviewers' regular ratings in one sub-subcategory are positively associated with the number of *Vine* products they receive from the focal sub-subcategory but not with the number of *Vine* products from other sub-subcategories, other subcategories, or other categories (Table 9, Model 2).

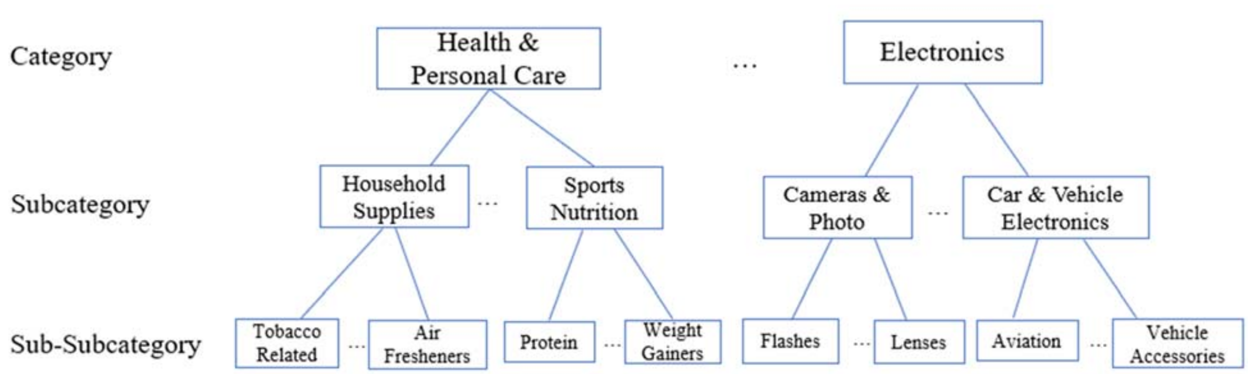

Figure 4. Category Tree in Amazon

Table 9. Impact of Volume of Same- and Cross-Category Vine Products (Higher Granularity)

Specifications	Subcategory Level	Sub-Subcategory Level
	(1)	(2)
VineReviewer	-	-
VineMonth	0.078*** (0.023)	-0.008 (0.033)
VineProductsSubFocal	0.052** (0.026)	-
VineProductsSubSubFocal	-	0.071** (0.034)
VineProductsSubSubOther	-	0.040 (0.028)
VineProductsSubOther	0.023 (0.017)	0.023 (0.016)
VineProductsOther	0.003 (0.011)	0.003 (0.011)
Control Variables	YES	YES
Time Fixed	YES	YES
Reviewer Fixed	YES	YES
R-squared	0.205	0.197
Observations	37,156	41,021

Note. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at the user-month-subcategory level in column (1) and the user-month-sub-subcategory level in column (2). *Significant at 10%; **significant at 5%; ***significant at 1%.

7. Robustness Checks

We perform a series of robustness checks, including: (i) classic DID and falsification tests, (ii) alternative matching approaches, and (iii) a relative time model. We find that our main results remain consistent across these robustness checks.

7.1 Classic DID and Falsification Tests

In our context, reviewers in the treatment group join the platform-initiated reviewer incentive program in different time periods, and after joining the program, reviewers remain in the program during our study period. For this staggered adoption, we have utilized a two-way fixed effects DID model coupled with propensity score matching in our main analyses (e.g., Xu et al. 2016, Callaway and Sant'Anna 2019). Existing studies also use a classic DID approach to estimate the treatment effect (e.g., Jung et al. 2019, Kumar et al. 2019). In a classic DID, researchers use a fixed number of observations (e.g., m) for each user before and after the treatment and construct the panel dataset. To check the robustness of our results, we conduct a classic DID with the same propensity score matching specification used in our model. That is, supposing one treated reviewer joins the program in month t_0, we would include the observations from t_{0-m} to t_{0+m} of the matched pair in the panel dataset and estimate Equation (1).

Following Jung et al. (2019), we also adopt a falsification test to re-estimate the same DID model with all the same reviewer samples, model parameters, estimation model, propensity score matching equation, and caliper size except that we now shift $VineMonth_{it}$, the binary variable indicating the program participation time, *m* months earlier. That is, the $VineMonth_{it}$ variable (described in our DID model) will equal 1 in $t_{0-m} \sim t_0$, a time period before reviewer *i*'s actual program participation time, not after. Naturally, due to the time shift, $VineMonth_{it}$ will equal 0 during $t_{0-2m} \sim t_{0-m}$ and the total time periods for the falsification test will be from t_{0-2m} to t_0. In this falsification test, the shifted $VineMonth_{it}$ arbitrarily points to the periods before the actual program joining time, not after. The reviewers have not actually joined the *Vine* program yet. Therefore, the same model that we use to show the statistically significant impact should produce a non-significant estimation for the DID term because no treatment effect has actually occurred during the pre-treatment period. Figure 5 illustrates the sampled time periods for the classic DID and falsification tests.

We check the classic DID and falsification test with m = 2, 3 and 4, and the results are qualitatively consistent with one another. Table 10 reports the results. For the classic DID test without the $VineMonth_{it}$ indicator being shifted, the estimation of the DID term is positive and significant (Models 1, 3, and 5). Consistent with the result in Table 2, classic DID estimation detects the significant regular-rating increase after reviewers join the *Vine* program. However, for the falsification test in which $VineMonth_{it}$ is shifted *m* months before the actual program joining time, the DID estimator turns out to be insignificant (Models 2, 4, and 6), confirming our expectation.

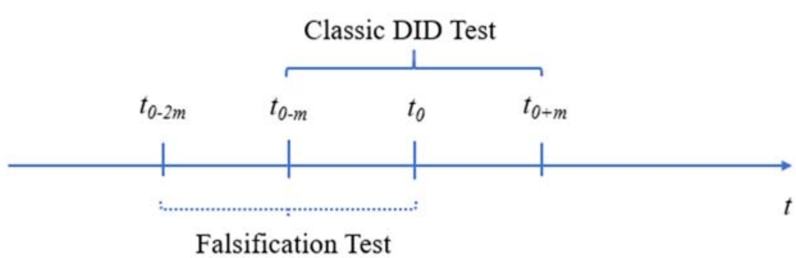

Figure 5. Time Periods for Classic DID and Falsification Test

Table 10. Classic DID and Falsification Test

	Panel A (m = 2)		Panel B (m = 3)		Panel C (m = 4)	
	DID	Falsification	DID	Falsification	DID	Falsification
	(1)	(2)	(3)	(4)	(5)	(6)
VineReviewer (β_1)	-	-	-	-	-	-
VineMonth (β_2)	-0.129**	-0.062	-0.048	0.061	-0.049	0.010
	(0.060)	(0.064)	(0.045)	(0.051)	(0.038)	(0.043)
VineReviewer×	0.126**	-0.016	0.094**	-0.037	0.114***	-0.052
VineMonth (β_3)	(0.051)	(0.054)	(0.039)	(0.042)	(0.034)	(0.037)
Control Variables	YES	YES	YES	YES	YES	YES
Time Fixed	YES	YES	YES	YES	YES	YES
Reviewer Fixed	YES	YES	YES	YES	YES	YES
R-squared	0.628	0.657	0.550	0.560	0.490	0.507
Observations	6,611	5,945	9,797	8,915	12,963	11,566

Note. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at the user-month level. *Significant at 10%; **significant at 5%; ***significant at 1%.

7.2 Alternative Matching Approaches

We also employ different matching algorithms to show our results are robust. We use one-to-one matching without replacement and the nearest three neighbors matching algorithms to further validate the results. Moreover, to correct for potential selection biases arising from the unobservable differences between our treatment and control groups, we employ a look-ahead matching technique (Manchanda et al. 2015, Xu et al. 2016, Jung et al. 2019). To this end, we restrict the reviewers in the control group as non-participating reviewers at the time of matching who are going to participate in the *Vine* program at a future period. Since both sets of users do eventually join the program, this methodology accounts not just for the observed characteristics, but also for unobserved time-invariant attributes that affect whether reviewers receive treatment. To implement this matching approach, we choose the reviewers who joined the reviewer incentive program after December 2009 to be matching candidates in the control group, and we match them to the treated reviewers who joined the program before December 2009. Table 11 reports the results with different matching specifications. The results are consistent with the results in Table 2, providing more evidence that reviewers' regular ratings significantly increase after joining the *Vine* program.

Table 11. Alternative Matching Approaches

	One-to-One (Without Replacement)	Nearest Three Neighbors (With Replacement)	Look-Ahead Matching (With Replacement)
	(1)	(2)	(3)
VineReviewer (β_1)	-	-	-
VineMonth (β_2)	-0.014	-0.006	-0.026
	(0.022)	(0.011)	(0.018)
VineReviewer×	0.087***	0.080***	0.136***
VineMonth (β_3)	(0.023)	(0.014)	(0.019)
Control Variables	YES	YES	YES
Time Fixed	YES	YES	YES
Reviewer Fixed	YES	YES	YES
R-squared	0.300	0.341	0.277
Observations	37,030	98,896	48,505

Note. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at the user-month level. *Significant at 10%; **significant at 5%; ***significant at 1%.

7.3 A Relative Time Model

It is possible that the treated reviewers gradually increase their regular ratings for unknown reasons during the pre-treatment period, but untreated reviewers do not. In this case, the treatment effect identified in our DID estimation will be due to the differences in their pre-existing trends in regular ratings, prior to participation in the *Vine* program, rather than to reviewers' engagement in reviewer incentive program. Our falsification test helps us rule out the difference in pre-existing trends between the reviewers in two groups at some level. Here, we adopt a relative time model to further show that the regular ratings of the participating reviewers are not different from those of other reviewers at the pre-treatment periods.

Following Jung et al. (2019), we check the existence of an unusual pre-existing trend with a two-way fixed effects model, including a group of dummy variables for the relative time periods before and after the treatment:

$$AvgRegRating_{it} = \alpha_0 + \sum_{j=T-5}^{T+3} \alpha_j (s_i \varphi) + X_{it} + \tau_t + \delta_i + \epsilon_{it} \quad (3)$$

where $AvgRegRating_{it}$ is the dependent variable, representing reviewer *i*'s average regular rating at month *t*. τ_t is time fixed effects, and δ_i is reviewer fixed effects. s_i is a binary variable indicating whether reviewer *i* participates in the program in our study period and φ is the relative time dummies representing the chronological distance between an observation period and reviewer *i*'s participation time. We set the relative time periods to range from minus five to three, representing the periods five months before and three months after program participation. Additionally, to avoid a dummy variable trap, following Greenwood and Wattal (2017) and Jung et al. (2019), we choose one month before the reviewer's participating month as the baseline and omit the dummy variable for that month.

Table 12 presents the results of the relative time analysis. None of the estimation with the relative time dummies during the pre-treatment periods is significant (i.e., for $t-5$, $t-4$, $t-3$, and t - 2), indicating pre-treatment homogeneity between the participating reviewers and their counterparts in the control group. Moreover, treated reviewers post higher regular ratings than the untreated reviewers in the post-treatment periods, a result consistent with the estimations of our main analyses.

Table 12. A Relative Time Model

	(1)
t - 5	-0.053 (0.041)
t - 4	-0.010 (0.041)
t - 3	-0.028 (0.037)
t - 2	-0.035 (0.039)
t - 1	Omitted
t + 0	0.096*** (0.034)
t + 1	0.089*** (0.034)
t + 2	0.059* (0.035)
t + 3	0.075** (0.036)
Control Variables	YES
Time Fixed	YES
Reviewer Fixed	YES
R-squared	0.437
Observations	16,029

Note. Robust standard errors clustered by each reviewer are in parentheses. The unit of analyses is at user-month level. *Significant at 10%; **significant at 5%; ***significant at 1%.

8. Implications and Conclusions

Many e-commerce platforms have launched reviewer incentive programs that provide free product samples to reviewers in exchange for their reviews on the sampled products. The practice provides platforms with an unprecedented opportunity to facilitate product feedback. While the extant literature focuses on the effects of seller-initiated incentives on the *incentivized* ratings (Cabral and Li 2015, Lin et al. 2019), to the best of our knowledge, this study is the first to quantify the impact of a platform-initiated reviewer incentive program on reviewers' *regular* ratings.

Our results indicate that reviewers' regular ratings significantly increase after they join the program and start to receive product samples. We also confirm the result with a series of robustness checks and falsification tests to rule out alternative explanations. In addition, reviewers' regular ratings increase only after they receive a sufficient number of free products, and their regular ratings are positively associated with the number of free samples they receive.

Our follow-up analyses reveal that reviewers' regular-rating increase is mainly due to an uncertainty-reduction effect instead of a reciprocity effect. In other words, the observed regular-rating increase is driven by improved assessment and knowledge about products sharing attributes with the sampled products. Specifically, we find that (i) reviewers' regular ratings increase only in the categories where the product samples are available, (ii) reviewers' regular rating increases are more pronounced in product categories where products are less differentiated (or more homogenous) such that they are more conducive to product assessment in the same category, and (iii) reviewers' regular ratings in one category are positively associated with the number of product samples they receive from the same category, but not with the number of samples from other categories.

This study contributes to the extant literature in several fields. First, our study contributes to the research stream on regular rating generation where prior work focuses on the effect of seller-initiated incentives on the incentivized ratings (e.g., Cabral and Li 2015, Lin et al. 2019). Our results suggest that the platform-initiated free samples drive increased regular ratings by extending the reviewers' knowledge and experiences about products sharing similar features as sampled products.

Second, this study improves our understanding of marketing strategies to accelerate product word-of-mouth (WOM) (e.g., Cabral and Li 2015, Lin et al. 2019). Financial incentives, such as monetary rewards (Burtch et al. 2017, Khern-am-nuai et al. 2018), free product samples (Lin et al. 2019), and cash rebates (Cabral and Li 2015) are widely used marketing strategies to facilitate review generation and consumer referrals (Jung et al. 2020). We contribute to this stream of literature by showing how a platform's sponsorship of reviewers influences the recipients' regular WOM behaviors for non-sponsored products.

Lastly, our paper contributes to the literature that analyzes the relationship between free product sampling and consumers' review generations. The existing literature focuses on either the impact of product sampling on the incentivized rating (Lin et al. 2019) or the sampled products (Lee and Tan 2013, Mo and Li 2018). This study, however, specifically investigates the effects of free product sampling on the sample recipients' regular ratings of purchased products. Moreover, this study broadly extends our understanding of the reciprocity effect in the platform-reviewer relationship. The theory of reciprocity states that receiving benefits can place the recipients in an uncomfortable state of tension ("indebtedness" in Greenberg 1980). When opportunities to reciprocate are limited, people tend to find alternative ways to reduce this discomfort (Shumaker and Brownell 1984, p. 14). In our research context, after receiving free products from the platform, reviewers are not able to repay the platform with higher incentivized ratings because the platform-initiated program publicly prohibits such behavior. In this context, reviewers are able to contribute to the platform via different activities, such as more purchases, higher effort on review generation, and review ratings. However, reciprocity is not found to affect reviewers' regular ratings in this study.

Our study has important managerial implications for e-commerce platforms and online review systems as online platforms attempt to stimulate user activities such as review ratings and comments via reviewer incentive programs. Our findings demonstrate that a platform-initiated reviewer incentive program not only encourages incentivized reviews but also elevates the recipients' regular ratings for the purchased products. Platforms need to keep this unintended consequence in mind when they making policies for their reviewer incentive programs, as many important aspects of the platform, such as product sales and sellers' strategy, are closely associated with the product ratings (e.g., Duan et al. 2008). Moreover, our results speak to policymakers and consumers concerned that platforms' incentive may bias reviewers' evaluation. Our results reveal that the reviewer incentive program does not bias reviewers' evaluation of products they purchased owing to reciprocity, rather, the observed regular-rating increase is due to an uncertainty-reduction effect.

Our study is subject to several limitations that warrant future research. First, we do not analyze other aspects of review-generation behaviors than reviewers' regular ratings. Future research could investigate how reviewers' review effort and review writing styles change after they join the program. Second, we do not investigate activities other than regular ratings by which reviewers can reciprocate the platform, such as purchasing more products or making greater efforts in review generation. Future studies could investigate such means by which reviewers can repay the platform. Finally, we uncover the uncertainty-reduction effect by examining the relationship between reviewers' regular ratings in one product category and the availability of product samples from the same category as well as the level of improved product knowledge brought by the product samples. We use the product category information as a proxy to describe whether the sampled and purchased products share common attributes. Future researcher could also apply advanced text mining techniques onto product descriptions to capture how product samples help reviewers understand the purchased products. Despite these limitations, our analyses document an important consequence of a platform-initialized reviewer incentive program on the reviewers' regular ratings and provide useful implications for e-commerce platforms.

Reference

Athey, S., & Imbens, G. W. (2018). Design-based Analysis in Difference-in-differences Settings with Staggered Adoption. Working Paper. Available at NBER: https://www.nber.org/papers/w24963.

Bawa, K., & Shoemaker, R. (2004). The Effects of Free Sample Promotions on Incremental Brand Sales. *Marketing Science*, 23(3), 345–363.

Biswas, D., Grewal, D., & Roggeveen, A. (2010). How the Order of Sampled Experiential Products Affects Choice. *Journal of Marketing Research*, 47(3), 508-519.

Burtch, G., Hong, Y., Bapna, R., & Griskevicius, V. (2017). Stimulating Online Reviews by Combining Financial Incentives and Social Norms. *Management Science*, 64(5), 2065-2082.

Cabral, L. & Li, L. (2015). A Dollar for Your Thoughts: Feedback-Conditional Rebates on eBay. *Management Science*, 61(9), 2052-2063.

Callaway, B., & Sant'Anna, P. H. (2019). Difference-in-differences with Multiple Time Periods. Working Paper. Available at SSRN: https://ssrn.com/abstract=3148250.

Chen, P. Y., Dhanasobhon, S., & Smith, M. D. (2008). All Reviews are not Created Equal: The Disaggregate Impact of Reviews and Reviewers at Amazon.com. Working paper. Available at SSRN: https://ssrn.com/abstract=918083.

Cheng, H. K., & Liu, Y. (2012). Optimal Software Free Trial Strategy: The Impact of Network Externalities and Consumer Uncertainty. *Information Systems Research*, 23(2), 488-504.

Cone (2011). Online Influence Trend Tracker. Last accessed on http://www.conecomm.com/research-blog/2011-online-influence-trend-tracker.

Dimensional Research (2013). The Impact of Customer Service on Customer Lifetime Value. 2013. Last Accessed on https://www.zendesk.com/resources/customer-service-and-lifetime-customer-value/

Duan, W., Gu, B., & Whinston, A. B. (2008). Do Online Reviews Matter?—An Empirical Investigation of Panel Data. *Decision Support Systems*, 45(4), 1007-1016.

Falk, A., & Fischbacher, U. (2006). A Theory of Reciprocity. *Games and Economic Behavior*, 54(2), 293–315.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. *Information Systems Research*, 19(3), 291–313.

Greenberg, M. S. (1980). A Theory of Indebtedness. In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), Social Exchange: Advances in Theory and Research (pp. 3-26). New York: V. H. Winston and Sons.

Goering, P. A. (1985). Effects of Product trial on Consumer Expectations, Demand, and Prices. *Journal of Consumer Research*, 12(1), 74-82.

Goes, P.B., Lin, M., & Au Yeung, C.M. (2014). “Popularity Effect” in User-Generated Content: Evidence from Online Product Reviews. *Information Systems Research*, 25(2), 222-238.

Gouldner, A. W. (1960). The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review*, 25(2), 161–178.

Gu, B., Park, J., & Konana, P. (2012). Research Note: The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products. *Information Systems Research*, 23(1), 182-196.

Hahn, M., Park, S., Krishnamurthi, L., & Zoltners, A. A. (1994). Analysis of New Product Diffusion Using a Four-Segment Trial-Repeat Model. *Marketing Science*, 13(3), 224–247.

Hanson, G., & Xiang, C. (2011). Trade Barriers and Trade Flows with Product Heterogeneity: An Application to US Motion Picture Exports. *Journal of International Economics*, 83(1), 14-26.

Hennig-Thurau, T. & Houston, B. M. (2019). Entertainment Science: Data Analytics and Practical Theory for Movies, Games, Books, and Music. Switzerland AG: Springer International Publishing.

Ho, Y. C. C., Wu, J., & Tan, Y. (2017). Disconfirmation Effect on Online Rating Behavior: A Structural Model. *Information Systems Research*, 28(3), 626-642.

Hong, Y., & Pavlou, P. A. (2014). Product Fit Uncertainty in Online Markets: Nature, Effects, and Antecedents. *Information Systems Research*, 25(2), 328-344.

Hu, N., Pavlou, P. A., & Zhang, J. (2006). Can Online Reviews Reveal a Product's True Quality?: Empirical Findings and Analytical Modeling of Online Word-of-mouth Communication. In Proceedings of the 7th ACM Conference on Electronic Commerce (pp. 324–330). New York, NY, USA: ACM.

Jamieson, L. F., & Bass, F. M. (1989). Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods. *Journal of Marketing Research*, 26(3), 336-345.

Johnson, E. J., & Russo, J. E. (1984). Product Familiarity and Learning New Information. *Journal of Consumer Research*, 11(1), 542-550.

Jung, J., Bapna, R., Ramaprasad, J., & Umyarov, A. (2019). Love Unshackled: Identifying the Effect of Mobile App Adoption in Online Dating. *MIS Quarterly*, 43(1), 57-72.

Jung, J., Bapna, R., Golden, J., & Sun, T (2020). Words Matter! Towards Pro-social Call-to-Action for Online Referral: Evidence from Two Field Experiments. *Information Systems Research*, Forthcoming.

Khern-am-nuai, W., Kannan, K., & Ghasemkhani, H. (2018). Extrinsic Versus Intrinsic Rewards for Contributing Reviews in an Online Platform. *Information Systems Research*, 29(4), 871-892.

Kokkodis, M., & Lappas, T. (2016). The Relationship between Disclosing Purchase Information and Reputation Systems in Electronic Markets. ICIS 2016 Proceedings.

Kumar, A., Mehra, A., & Kumar, S. (2019). Why Do Stores Drive Online Sales? Evidence of Underlying Mechanisms from a Multichannel Retailer. *Information Systems Research*, 30(1), 319-338.

Lee, Y. J., Hosanagar, K., & Tan, Y. (2015). Do I Follow My Friends or the Crowd? Information Cascades in Online Movie Ratings. *Management Science*, 61(9), 2241–2258.

Lee, Y. J., & Tan, Y. (2013). Effects of Different Types of Free Trials and Ratings in Sampling of Consumer Software: An Empirical Study. *Journal of Management Information Systems*, 30(3), 213-246.

Li, X., & Hitt, L. M. (2008). Self-Selection and Information Role of Online Product Reviews. *Information Systems Research*, 19(4), 456–474.

Li, X., & Hitt, L. M. (2010). Price Effects in Online Product Reviews: An Analytical Model and Empirical Analysis. *MIS Quarterly*, 34(4), 809–831.

Lin, Z., & Heng, C. S. (2015). The Paradoxes of Word of Mouth in Electronic Commerce. *Journal of Management Information Systems*, 32(4), 246-284.

Lin, Z., Zhang, Y., & Tan, Y. (2019). An Empirical Study of Free Product Sampling and Rating Bias. *Information Systems Research*, 30(1), 260-275.

Marks, L. J., & Kamins, M. A. (1988). The Use of Product Sampling and Advertising: Effects of Sequence of Exposure and Degree of Advertising Claim Exaggeration on Consumers' Belief Strength, Belief Confidence, and Attitudes. *Journal of Marketing Research*, 25(3), 266-281.

Manchanda, P., Packard, G., & Pattabhiramaiah, A. (2015). Social Dollars: The Economic Impact of Customer Participation in a Firm-Sponsored Online Customer Community. *Marketing Science*, 34(3), 367–387.

Mano, H., & Oliver, R. L. (1993). Assessing the Dimensionality and Structure of the Consumption Experience: Evaluation, Feeling, and Satisfaction. *Journal of Consumer Research*, 20(3), 451-466.

Mo, J., & Li, Z. A. (2018). Is Product Sampling Good for Brands Online Word of Mouth?. ICIS 2018 Proceedings.

Murray, K. B. (1991). A Test of Services Marketing Theory: Consumer Information Acquisition Activities. *Journal of Marketing*, 55(1), 10-25.

Nielson (2014). The Role of Content Report. 2014. Last Accessed on http://www.inpwrd.com/nielsen.

Rao, A. R., & Monroe, K. B. (1988). The Moderating Effect of Prior Knowledge on Cue Utilization in Product Evaluations. *Journal of Consumer Research*, 15(2), 253-264.

Rothschild, M. L., & Gaidis, W. C. (1981). Behavioral Learning Theory: Its Relevance to Marketing and Promotions. *Journal of Marketing*, 45(2), 70-78.

Shen, W., Hu, Y. J., & Rees Ulmer, J. (2015). Competing for Attention: An Empirical Study of Online Reviewers’ Strategic Behaviors. *MIS Quarterly*, 39(3), 683-696.

Shumaker, S. A., & Brownell, A. (1984). Toward a Theory of Social Support: Closing Conceptual Gaps. *Journal of Social Issues*, 40(4), 11–36.

Smith, R. E. (1993). Integrating Information from Advertising and Trial: Processes and Effects on Consumer Response to Product Information. *Journal of Marketing Research*, 30(2), 204-219.

Smith, R. E., & Swinyard, W. R. (1983). Attitude-behavior Consistency: The Impact of Product Trial versus Advertising. *Journal of Marketing Research*, 20(3), 257-267.

Sun, M., & Zhu, F. (2013). Ad Revenue and Content Commercialization: Evidence from Blogs. *Management Science*, 59(10), 2314–2331.

Xu, K., Chan, J., Ghose, A., & Han, S. P. (2017). Battle of the Channels: The Impact of Tablets on Digital Commerce. *Management Science*, 63(5), 1469-1492.

Appendix A. Matching Strategies

We perform matching using propensity score approaches. First, we identity a list of reviewer characteristics that affect a reviewer's participation of the *Vine* program, and we find that consistent with Amazon's official announcement, the factors representing one reviewer's review experience are significantly related to the reviewer's program participation: the overall helpfulness of his or her reviews, review experience in different product categories, and review volume. Table A.1 shows the variable description. Second, we predict the propensity score of a reviewer at the month when the program launched (i.e. August 2007) and match the participating reviewer with the non-participating reviewer of the similar propensity score. Finally, we compare the distribution of propensity scores and ensure the balance of all variables. Table A.2 shows that there is no significant difference between the two groups across most variables after matching. This indicates that our matching is successful.

Table A.1 Variables Used to Match Reviewers

Variable Category	Variable Description	Variables
The reviewer's general experience	The log-transformed total number of months since the reviewer joins Amazon.	Log (review tenure+1)
The reviewer's reviewed products	The portion of the number of reviews from the specific category to the whole reviews generated by the reviewer since the reviewer joins Amazon.	The portion of the number of reviews from category "Books," "Movies & TV," "CDs & Vinyl," "Electronics," "Home & Kitchen," "Grocery & Gourmet Food," and "Health & Personal Care" to the total number of reviews generated by reviewer
The reviewer's recent review intensity	The reviewer's monthly average reviews in the past five months	The average number of reviews (Regular and Incentivized) in the past five months The average number of regular reviews in the past five months
The reviewer's review quality	The average helpfulness ratio of the reviews produced by the reviewer since the reviewer joins Amazon.	Average helpfulness ratio
The reviewer's willingness to be contacted	The binary variables to indicate whether the reviewer provides the contact or personal information.	Three binary variables indicating whether reviewer disclose Contact, Description, and Location information

Table A.2 T-test of the Variables Before and After Matching

Variable	Sample	Mean		T-test	
		Treated	Control	t	p > |t|
Log (review tenure+1)	Unmatched	3.908	3.659	10.83	0.000
	Matched	3.908	3.943	-1.59	0.111
Category: Books	Unmatched	0.309	0.281	2.76	0.006
	Matched	0.309	0.326	-1.65	0.098
Category: Electronics	Unmatched	0.079	0.097	-2.87	0.004
	Matched	0.079	0.067	2.19	0.028
Category: Health & Personal Care	Unmatched	0.008	0.011	-1.72	0.086
	Matched	0.008	0.005	2.55	0.011
Category: CDs	Unmatched	0.082	0.113	-4.32	0.000
	Matched	0.082	0.076	1.09	0.274
Category: Movies & TV	Unmatched	0.114	0.107	1.02	0.309
	Matched	0.114	0.118	-0.74	0.459
Category: Home & Kitchen	Unmatched	0.068	0.062	1.05	0.292
	Matched	0.068	0.067	0.18	0.858
Category: Grocery & Gourmet Food	Unmatched	0.007	0.002	4.61	0.000
	Matched	0.007	0.006	0.63	0.531
Average Number of Reviews (Regular and Incentivized) in the Past 5 Months	Unmatched	2.607	1.475	5.31	0.000
	Matched	2.607	2.364	1.04	0.300
Average Number of Regular Reviews in the Past 5 Months	Unmatched	0.653	0.382	6.33	0.000
	Matched	0.653	0.677	-0.44	0.658
Average Helpfulness Ratio	Unmatched	0.844	0.828	3.68	0.000
	Matched	0.844	0.838	1.67	0.096
Contact Information	Unmatched	0.280	0.125	11.56	0.000
	Matched	0.280	0.246	2.24	0.025
Description information	Unmatched	0.565	0.352	12.79	0.000
	Matched	0.565	0.576	-0.63	0.529
Location information	Unmatched	0.900	0.878	1.98	0.048
	Matched	0.900	0.908	-0.88	0.377