Disclosure in Incentivized Reviews: Does It Protect Consumers?

Sungsik Park
University of South Carolina

Woochoel Shin
University of Florida

Jinhong Xie
University of Florida

July 18, 2023

Sungsik Park (sungsik.park@moore.sc.edu) is an Assistant Professor at the Darla Moore School of Business, University of South Carolina, Columbia, SC, 29201, Woochoel Shin (wshin@ufl.edu) is Brian R. Gamache Professor and Associate Professor, and Jinhong Xie (xiej@ufl.edu) is Professor and J.C. Penney Eminent Scholar Chair at the Warrington College of Business Administration, University of Florida, Gainesville, FL 32611.

Disclosure in Incentivized Reviews: Does It Protect Consumers?

Abstract

The well-documented rating inflation of incentivized reviews (IRs) can mislead consumers into choosing a product that they would otherwise not buy. To protect consumers from this undesirable influence, the U.S. Federal Trade Commission recommends that reviewers conspicuously disclose any material connection they may have with sellers. In theory, such disclosures safeguard consumers by motivating reviewers to be truthful and inducing consumers to discount inflated IR ratings. Our research finds, however, that IR disclosure accomplishes neither. Specifically, our empirical analysis of consumer reviews on Amazon reveals that, even with disclosure, (1) rating inflation of IRs remains, and (2) this inflation boosts sales at consumers' expense. Finally, we propose an alternative approach to eliminate rating inflation of IRs and empirically demonstrate its effectiveness. These findings have important implications for consumers, firms, and on-going policy discussions around IRs.

Keywords: information disclosure, consumer protection, online reviews, incentivized reviews, user-generated information, incentive compatibility

1 Introduction

> *"Online shopping runs on reviews. ... If 500 other people have bought something and say it works, you can have a lot more confidence. But what if those people were paid to leave those positive reviews?...."*[1]
>
> Lina M. Khan, Chair, Federal Trade Commission, Oct. 20, 2022

Chair Khan's statement above refers to the pressing problem of incentivized reviews (IRs), whereby a firm provides a material incentive (e.g., free product, bonus, or cash) to motivate consumers to post online reviews of its products. Although such incentives increase the volume of a product's user-generated information (Shi and Wojnicki 2014, Burtch et al. 2018), research has shown that they markedly increase the favorability of the reviews that the product receives (e.g., Cabral and Li 2015, Khern-am-nuai et al. 2018). This rating inflation puts shoppers at risk of selecting products they otherwise would not have bought. Because online product reviews have become a key channel for consumers to gather product information, and because IRs are so prevalent, the social welfare implications of IR bias can be substantial.

In the US, the Federal Trade Commission (FTC) is charged with discouraging deceptive trade practices. To address the problem of bias in IRs, the FTC recently issued guidance that IRs be accompanied by a conspicuous disclosure of the material relationship between the reviewer and the seller (see, for example, FTC's "Featuring Online Customer Reviews: A Guide for Platforms," January 2022).[2] The FTC's recommendation that IRs incorporate disclosure has been embraced by prominent advertisers[3] and product review platforms.[4]

Intuitively, incorporating disclosure into incentivized reviews has the potential to help consumers in two ways. First, from the reviewer's perspective, because public disclosure draws a spotlight on the reviewer's conflict-of-interest, the reviewer might be motivated to be more truthful (Snyder 1974). This effect could be especially strong for reviewers who care about their reputation (Pinch and Kesler 2011). Second, from the consumer's perspective, disclosure serves an

---

[1] https://www.ftc.gov/system/files/ftc_gov/pdf/P214504ChairStatementFakeReviewsANPR.pdf (accessed March 18, 2023).

[2] https://www.ftc.gov/business-guidance/resources/featuring-online-customer-reviews-guide-platforms (accessed March 18, 2023).

[3] BBB National Advertising Division, "5 Tips for Truthful and Transparent influencer Marketing and Product Reviews" (accessed March 18, 2023).

[4] Smiley360's disclosure statement provides an example of such encouragement: "Disclosure is an important part of being a Smiley member. Disclosing that you have received a free product means you're following guidelines set by the Federal Trade Commission (we wouldn't want you to get in trouble!) Following these guidelines will help you to enjoy your Smiley membership and all that it brings."

information function, i.e., alerting shoppers that a particular review may be biased and inducing them to adjust their assessment of it.

As regulators and market stakeholders build momentum for instituting disclosure as the solution to IR bias, neither the ability of disclosure to curb incentivized reviewers' rating inflation nor its effectiveness in helping consumers discount such inflation has been carefully examined in real markets. This paper empirically investigates the effectiveness of IR disclosure on reviewers and consumers. We also propose and test an alternative approach to IRs that eliminates rating inflation by avoiding the underlying conflict-of-interest.

We begin by examining whether IR disclosure can effectively eliminate rating inflation. Specifically, we compare reviews with an "incentive disclaimer" (i.e., "disclosed IRs" or DIRs) to reviews with a "verified purchase" tag (i.e., "verified reviews" or VRs) on Amazon. After controlling for review- and reviewer-specific characteristics, we find that, for the same product, DIRs have higher average ratings than do VRs. Note that some VRs in our data can be undisclosed IRs with more favorable ratings (i.e., written by reviewers who received an incentive but failed to disclose their conflict of interest). This suggests that the average rating of organic reviews written by users with no financial incentive may be even lower than that of VRs, implying that the rating inflation of DIRs can be even greater than we report. More troublingly, we also find that the rating difference between DIRs and VRs is larger for low-quality than for high-quality products.

Our finding that ratings associated with DIRs exceed those of VRs is important because it shows that having IR disclosure in place is insufficient to eliminate the rating inflation of IRs. Our finding that DIR rating inflation for low-quality products is higher than for high-quality products is especially bad news for consumers because more harm is caused when an inflated rating induces the purchase of a low-quality product.

Given the inability of disclosure to eliminate IR rating inflation, it is crucial to examine whether disclosure can serve an effective information function to prevent consumers from being misled by rating bias. To do so, we exploit an abrupt policy change made by Amazon on October 3, 2016. For several years prior to this date, Amazon allowed IRs if reviewers disclosed their financial incentive. After this date, however, Amazon prohibited the posting of new incentivized reviews and gradually deleted existing ones. This policy change provides an opportunity to compare the pre- and post-policy periods for the same product and estimate the effect of DIRs (or, more precisely, the effect of their removal).

Our difference-in-differences (DID) estimates, based on Amazon data surrounding the date of its policy change, lead to three important findings. First, DIRs have a positive sales effect, suggesting that when incentivized reviewers disclose their conflict of interest, the inflated reviews can still effectively induce more consumers to buy a product. Second, such a positive sales effect is stronger for products with higher IR rating inflation, confirming that, even with disclosure, rating inflation directly contributes to the sales increase induced by IRs. Third, consumers are more likely to give poor post-purchase product evaluations (e.g., one-star ratings) in the presence than in the absence of DIRs, suggesting that, even with disclosure, IRs lower consumers' post-purchase product satisfaction. Taken together, these findings provide evidence that IR disclosure is not effective in helping consumers discount IR rating inflation. Despite disclosure, inflated ratings continue to enhance sales and consumers remain vulnerable to a disappointing consumption experience. We replicate these findings via an alternative cross-platform DID design, employing data from two platforms that implemented the same policy change at two different points in time: Amazon US and Amazon UK.

Finally, given that disclosure neither eliminates DIR rating inflation nor enables consumers to effectively discount inflated ratings, we propose and test a different way to remove rating inflation of IRs. Based on incentive compatibility between the two parties engaged in the creation of an IR, i.e., the firm and the reviewer, we conjecture that offering rewards to reviewers leads to inflated ratings when IRs are commissioned by a firm that benefits directly from sales of the reviewed products. Such bias can be removed, however, if IRs are instead commissioned by an independent platform (that benefits from sales of many products from various suppliers). We empirically test our conjecture by formally defining two types of IR, (a) firm-initiated, and (b) platform-initiated, which differ only in who commissions them. Examples of the latter include Amazon's Vine program and Home Depot's Seeds Program, in which the platform distributes many brands of free products to reviewers. We test our conjecture using Amazon's Vine program, under which IRs are disclosed with a badge, i.e., "Vine Customer Review of Free Product." Results of our empirical examination of Amazon's Vine program show that, unlike firm-initiated DIRs, platform-initiated DIRs have the same ratings as VRs. These results suggest that moving from a firm-initiated to a platform-initiated IR system is an effective way to reduce IR rating inflation. This provides empirical evidence of the advantage of a platform-initiated IR system.

Our research contributes by advancing the academic literature, alerting consumers, and offering insights for public policy. First, as an important societal concern, IRs have garnered increasing academic attention in recent years. Some studies have adopted a reviewer's perspective, investigating how incentives influence review characteristics (such as rating, sentiment, length, quality) and reviewers' future review participation (e.g., Cabral and Li 2015, Burtch et al. 2018, Khern-am-nuai 2018, Qiao et al. 2020). Other studies have taken the firm's perspective, exploring the sales effect of IRs and the characteristics of firms that sponsor them (e.g., Li et al. 2020, Fradkin and Holtz 2022). Still, other studies have used game theory to model incentives, identify conditions that promote IR bias, and explore the impact of IR bias on social welfare (Dellarocas 2006, Mostagir and Siderius 2023).[5] Our work enhances the literature on IRs by focusing on the effectiveness of IR *disclosure*, which is a commonly prescribed regulatory remedy for IR bias.

Second, rating inflation by IRs is a problem that all online shoppers can relate to. Our analysis directly benefits consumers by drawing their attention to IR bias, the failure of disclosure to counter that bias, and the benefits of platform-initiated IRs.

Finally, our investigation is directly relevant to the FTC's active rulemaking process on the regulation of reviews and endorsements, including IRs.[6] Our research provides compelling empirical evidence that disclosure neither eliminates IR rating inflation nor prevents consumer harm. Further, we propose and empirically demonstrate that moving from a firm-initiated to a platform-initiated IR system can eliminate rating inflation by removing the conflict-of interest underlying it. We note that it might be more beneficial to consumers if policymakers foster the growth of platform-initiated review systems, rather than nudging consumers to pay attention to disclosures embedded in firm-initiated IRs.

2 Empirical Context

2.1 Background

Our study examines IRs on Amazon, which, as a platform provider, has taken several proactive measures to keep reviews informative and unbiased. In September 2009, Amazon started displaying a "verified purchase" tag on reviews from purchases made through its own site

[5] Past research has also investigated fake reviews, which differ substantially from IRs (e.g., see page 899 of He et al. 2022).
[6] See FTC Trade Regulation Rule on the Use of Reviews and Endorsements (16 CFR Part 465), and Advance Notice of Proposed Rulemaking, Request for Public Comment, Federal Register Vol. 87, No. 215, November 8, 2022.

(Anderson and Simester 2014). The verified-purchase tag provides consumers with information that helps them decide which reviews are more trustworthy. More recently, Amazon announced a more stringent policy: From October 3, 2016 onward, Amazon banned IRs (see Figure 1 for Amazon's announcement) and subsequently allowed no newly posted IRs. Moreover, since this announcement, Amazon has deleted all existing IRs from its website.

In this context, we examine two types of reviews. First, *disclosed IRs* (DIRs) are reviews with a disclaimer (of the material connection) in the text. Note that DIRs are truly incentivized, since there is no reason for a reviewer to write a disclaimer without having a material connection with a firm.[7] Second, as a benchmark, we also examine *verified reviews* (VRs), i.e., reviews with a verified purchase tag. By definition, VRs are posted by consumers who actually purchased the product on Amazon. While we expect that most VRs are authentically produced, it is possible that some might have been incentivized by firms. We classify a VR as a DIR whenever it contains a disclaimer, but we cannot completely rule out the existence of undisclosed incentivized reviews among VRs. To the extent that we distinguish between the two types of reviews, however, their existence strengthens our results: when we compare DIRs with VRs (in Section 3), the comparison becomes more conservative.

2.2 Data

We combine two sources of data: (1) sales data and (2) review data. We collect the former from the website *keepa.com* and the latter directly from Amazon. First, our sales data collection largely follows the literature (e.g., Park et al. 2023). Specifically, since sales data on Amazon are not publicly available, we collect sales-rank data to proxy sales. This approach has been well substantiated in the literature (e.g., Chevalier and Goolsbee 2003, Sun 2012). Moreover, the sales rank serves our purposes especially well because, in our study of the sales impact, we are interested in the *change*, rather than the absolute level of sales.

Second, our review data were collected twice. In September 2016, we collected from the Amazon website all reviews of 5,273 products, for which at least one DIR was posted by top 1000-ranked reviewers (see Section A1 of the Appendix for details). This review dataset contains 654,463 reviews, including 241,831 DIRs. After Amazon's announcement of its IR policy change

[7] In our study, we identify DIRs using a set of disclaimers as well as their variations, such as "provided a free sample," "received a free sample," "received this for free," "received this product for free," "received this at a discount," "received a sample," and "received this product at a discount."

(October 2016), we again collected all reviews for the same set of products, specifically in March 2017. For our analysis in Section 3, we use reviews on all products in the first data set that were not affected by Amazon's policy change. For the analysis of product sales in Section 4, however, we use only the matched products in both the review and the sales datasets. Since *keepa.com* only keeps track of products above a certain sales level, we observe 4,147 matched products.

Besides our *main* dataset, we construct additional datasets for various analyses in the paper. The detailed construction of all the datasets used in the paper is reported in Section A1 of the Appendix. Table 1 lists theses datasets, along with their key construction procedures and the analyses in which they are used.

3 Impact of IR Disclosure on Reviewers

In this section, we examine whether the disclosure requirement for IRs can change incentivized reviewers' posting behavior. If the disclosure requirement disciplines reviewers to be impartial, we expect no difference in the ratings of DIRs from those of VRs. However, different rating levels would imply that the requirement is not sufficient to curb the incentive to distort reviews. Thus, we investigate how DIRs fare against VRs in terms of their ratings.

To begin, a simple mean comparison shows that the average rating of DIRs is substantially higher than that of VRs (4.73 vs. 4.08; $p < 0.01$). Moreover, as reported in Table 2, most DIRs are positive, with 95.7% of them exhibiting a rating of 4 or 5 (compared to 75.3% of VRs). However, this comparison may have been driven by the characteristics of products, reviewers, or posting timing of DIRs that are distinct from those of VRs. Thus, we formally test the ratings difference by considering the following regression of individual ratings of both DIRs and VRs with a set of control variables:

$$Rating_{ij} = \alpha \cdot DIR_{ij} + \boldsymbol{\beta}\mathbf{X}_{\mathbf{ij}} + \boldsymbol{\gamma}_{\mathbf{i}} + \varepsilon_{ij}, \quad (1)$$

where $Rating_{ij}$ is the rating of review j for product i, DIR_{ij} is a dummy variable indicating review j is a DIR (as opposed to a VR), $\mathbf{X}_{\mathbf{ij}}$ is a vector of controls, $\boldsymbol{\gamma}_{\mathbf{i}}$ are product-level fixed effects, and ε_{it} is the error term. Note that, since we absorb cross-product variation with $\boldsymbol{\gamma}_{\mathbf{i}}$, our estimation only utilizes within-product variation in ratings. Moreover, by including a set of individual-review-specific controls ($\mathbf{X}_{\mathbf{ij}}$), we account for reviewer- as well as review-specific characteristics. Specifically, we include reviewer ranking and the number of reviews written by the authoring reviewer to account for reviewer heterogeneity, as well as *Time* (i.e., the number of days since the first review was posted) and *Order* (i.e., the position of the review in the sequence of reviews of the product) of the review to account for temporal and sequential dynamics (Godes and Silva 2012). The inclusion of product-level fixed effects enables us to control for both observed- and unobserved time-invariant product characteristics. As noted earlier, to fully utilize our sample of DIRs, we employ data collected prior to the Amazon's policy change (i.e., September 2016) where we restrict our attention to products with at least one DIR and one VR. Lastly, we cluster the standard errors at the product level to account for correlation patterns in errors (Bertrand et al. 2004).

In equation (1), the coefficient of our primary interest is α, which captures the ratings difference between DIRs and VRs. As reported in Column (1) of Table 3, we find the estimate of α is positive and highly significant ($\hat{\alpha} = 0.48,\ p < 0.01$), suggesting that DIRs are, on average, 0.48 stars higher than VRs for the same product. This substantial difference in ratings suggests that the disclosure requirement does not provide a sufficient incentive for incentivized reviewers to remove rating inflation from their reviews. This finding can be attributed to psychological mechanisms described in behavioral research. For example, reviewers may exhibit both "moral licensing" and "strategic exaggeration" (Cain, et. al. 2010, Loewenstein et. al., 2011). Under moral licensing, a reviewer's disclosure may relieve him of guilt over writing a biased review. Under strategic exaggeration, a reviewer may consciously increase the bias in his review to counteract any discounting that disclosure may cause.

In our dataset, 98 percent of reviewers wrote only one review. This raises the possibility that the estimated rating difference might be driven by reviewer selection. While we have already controlled for the reviewer-specific characteristics, we can rule out such a possibility by conducting an additional within-reviewer analysis. Specifically, among Amazon's top 1000-ranked reviewers, we identify 620 reviewers who have written one or more reviews of both types (i.e., DIR and VR) and compare their ratings of DIRs and VRs while incorporating reviewer-level fixed effects and clustering the standard errors at the reviewer level. We find that even within the same reviewer, the ratings for DIRs consistently appear higher than those for VRs by 0.17 stars ($p < 0.01$). This reaffirms the inability of the disclosure to curb the distortion incentive of reviewers (for details of this analysis, see Section A2 of the Appendix).

Finally, one may wonder whether the rating inflation of DIRs may be more pronounced for products of a certain quality. To answer this question, we constructed the *Consumer Reports* dataset by collecting quality scores for all products in 48 categories of appliances and electronics in *Consumer Reports*. For each product, we also collect review data from Amazon but since we use within-product variation in this analysis, we focus on 105 products which received one or more reviews of both types (DIRs and VRs) on Amazon at the time of data collection (i.e., August 2016).[8]

We estimate equation (1) by including an interaction term $IR_{ij} \times Q_i$ where Q_i is a percentile of product i's quality score within a category. We report the estimation result in Column (2) of Table 3. Results show that the rating inflation of DIRs *decreases* as the product's quality *increases*. In other words, DIR rating inflation is systematically higher for lower-quality products.[9] Greater rating distortion at the lower end is particularly bad for consumers because more harm is caused when an inflated rating induces the purchase of a low-quality product. Taken together, our findings suggest that a mere disclosure requirement may not prevent information distortion by IRs.

4 Impact of IR Disclosure on Consumers

In the previous section, we show that the disclosure requirement is not sufficient to remove reviewers' incentives to distort their reviews. It remains unclear, however, whether purchase decisions are influenced by such inflated ratings, because consumers may discount the information contained in DIRs when they see a disclosure. In this section, we examine whether disclosure helps consumers sufficiently discount rating inflation by investigating whether DIRs may fail to alter product sales. In the case of sales increase, we also seek evidence that consumers are misled by rating inflation.

To estimate the sales impact of DIRs, we exploit the fact that Amazon did not delete all IRs immediately after their policy change in October 2016. Specifically, we observe that 35 percent of DIRs in our data set had not been removed six months after the policy change (March 2017). This implies that there are two groups of products in our data: One group for which some (or all) DIRs had been removed and another group for which none had yet been removed.[10] By comparing

[8] For more details of the *Consumer Reports* dataset construction, see Section A1 of the Appendix. Also, note that this dataset includes a distinct set of products compared to our *main* dataset.

[9] It is useful to note that we do not find evidence that the likelihood of DIR posting significantly differs across quality levels (see Section A3 of the Appendix for details). Thus, our results are not driven by the concentration of DIRs on products of a certain quality level.

[10] After a significant amount of time, we no longer observe the second group of products and find that all DIRs had been deleted as of December 2018.

these two groups, we can identify the impact of DIRs, or precisely, the impact of DIR removal. More specifically, the first group of products (having lost some DIRs) serves as our treatment group, while the second group (having lost no DIRs) comprises the control group. We estimate the effect of DIR removal by measuring the temporal change in the sales of the treatment group relative to that of the control group, via a difference-in-differences (DID) approach.

We consider the following difference-in-differences specification, using pre-period data from April to September 2016 and post-period data from October 2016 to March 2017, with the unit of analysis being product (i) × day (t):

$$\ln(S_{it}) = \lambda \cdot (Treat_i \times Post_t) + \boldsymbol{\beta}\mathbf{X_{it}} + \boldsymbol{\gamma}_\mathbf{i} + \mathbf{M_t} + \varepsilon_{it} \tag{2}$$

where $\ln(S_{it})$ is the log sales rank for product i on day t, $Treat_i$ is a binary indicator for products in the treatment group, $Post_t$ is a binary indicator for observations after the policy change, $\mathbf{X_{it}}$ is the vector of control variables, $\boldsymbol{\gamma}_\mathbf{i}$ is the product-level fixed effect, $\mathbf{M_t}$ is the year-month fixed effect, and ε_{it} is the error term. As controls, we include average ratings, log number of reviews, log price, and the category-specific log product age (the number of days posted on the platform).[11] Note that the inclusion of product-level fixed effects not only controls for product characteristics but also helps us to reduce the concern for potential selection in the group assignment. As shown in Section A4 of the Appendix, the group assignment can be predicted by the number of pre-existing DIRs, which, if controlled, makes the group assignment largely idiosyncratic to the dependent variable. In our formulation, product-level fixed effects account for the number of pre-existing DIRs. Finally, to account for correlation in the errors, we cluster standard errors at the product level (Bertrand et al. 2004).

An important identifying assumption of our DID approach is that time trends in the outcome variable would have been similar for both groups if DIRs had not been removed. To explore the validity of this assumption, we estimate the month-specific group difference on equation (2) by replacing $Post_t$ with year-month dummies. As shown in Figure 2(a), we find no evidence of differential trends for the outcome variable before the policy change, thus lending support for our identifying assumption.

[11] Note that both average ratings and log number of reviews are calculated from all reviews that are not DIRs. Including DIRs into the calculation would bias the estimation of the DIR removal effect by making these control variables partially capture the effect of DIR removal.

Given this, we discuss our estimation result. Under the above specification, the product-level fixed effects capture the underlying difference in $\ln(S_{it})$ across products, while the year-month fixed effects capture the common time trends in $\ln(S_{it})$ before and after the policy change. Then, the coefficient of $Treat_i \times Post_t$ captures the difference in changes in the log sales rank before and after the policy change between treatment and control groups. As reported in Column (1) of Table 4, the estimated coefficient of $Treat_i \times Post_t$ is positive and highly significant ($\hat{\lambda} = 0.087, p < 0.01$). This result shows that, after the policy change, the sales rank of products in the treatment group increased 9.1% ($= e^{0.087} - 1$) more than those in the control group, indicating that the removal of DIRs has a negative impact on sales, given the inverse relationship between sales and sales rank. Conversely, this result suggests that the presence of DIRs increases product sales, thus establishing a positive impact of DIRs on sales.

Notably, the sales increase occurs even when the material connection between firms and reviewers is disclosed, implying that disclosure does not motivate consumers to fully discount DIR rating inflation. Psychological mechanisms that could hamper consumers' response to disclosure include their failure to notice the disclaimers because of information overload, inability to fully discount the biased advice (Tversky and Kahneman 1974), and/or increased trust of reviewers if the disclosure is interpreted as a sign of honesty (Pearson et al. 2006).

Given the above result, we next explore whether inflated ratings of DIRs have indeed contributed to the estimated sales change. For this purpose, we operationalize the rating inflation as the difference in the average ratings with and without DIR removal and denote it by ΔR. Figure 3 shows that in our data, ΔR is heterogeneous across products in the treatment group. Thus, we divide the treatment group into two groups using a median split and examine whether the DIR-removal effect is greater in the high ΔR group (i.e., products that experienced a larger reduction in average ratings). Specifically, we separately estimate equation (2) for the two groups while keeping the control group the same and compare their estimated effects.

We report our estimation results in Table 4. The estimate of λ is 0.078 ($p < 0.05$) for the low-ΔR group and 0.109, ($p < 0.01$) for the high-ΔR group (Columns 2 and 3), with the estimate being larger on the high-ΔR group. To formally test whether the high-ΔR group experienced a greater reduction in sales compared to the low-ΔR group after the policy change, we re-estimate equation (2) using the low-ΔR group as the control group and the high-ΔR group as the treatment group. In this "within treatment group" approach, we find that the estimated λ is positive and significant ($\hat{\lambda} = 0.051, p < 0.05$), suggesting that the sales of the high-ΔR group decreased significantly more than those of the low-ΔR group (Column 4). These results confirms that even with disclosure, rating inflation directly contributes to the sales change induced by DIRs removal.

Our findings thus far suggest that DIRs increase product sales by misleading consumers through their inflated ratings. To provide direct evidence that consumers are misled by inflated ratings of DIRs, we look into the number of one-star verified reviews (VRs). If DIRs had misled some consumers and made them unsatisfied, DIR removal should decrease the number of one-star VRs. We specifically estimate the following equation:

$$\ln\left(1 + N_{iT}^{One\ star}\right) = \lambda \cdot (Treat_i \times Post_t) + \boldsymbol{\gamma}_{\mathbf{i}} + \mathbf{M}_{\mathbf{t}} + \varepsilon_{it}, \quad (3)$$

where $N_{iT}^{One\ star}$ is the number of VRs with one-star rating posted for product i in month T. First, as shown in Figure 2(b), we find no evidence of differential trends in the outcome variable between the two groups before the policy change. Given this, we find that the estimate of λ is negative and significant ($\hat{\lambda} = -0.043$; $p < 0.01$), suggesting that the number of one-star VRs decreased by 4.3-percent due to the DIR removal (Column (1) of Table 5).[12] We interpret this finding as evidence that inflated ratings of DIRs mislead some consumers into making unsatisfactory purchases. In Section A5 of the Appendix, moreover, we show that the magnitude of this effect is higher for high-ΔR group than low-ΔR group, implying that DIRs with greater rating inflation are even more misleading to consumers. These results provide further evidence that the disclosure in IRs does not help consumers sufficiently discount their rating inflation.

Robustness Checks

So far, we identify the effect of DIRs on sales by comparing products from which DIRs were removed with those whose DIRs were retained in the first six months after the policy change. We now show the robustness of our results by adopting a different design, which compares the sales of the same products across two platforms, one with a policy change and the other with no change. This alternative design exploits the fact that Amazon UK did not change its IR policy until November 22, 2016 (see Figure 4 for the policy announcement date on Amazon UK). Thus, focusing on the period before Nov 22, 2016, we identify the sales impact of DIR removal by the cross-platform difference in the sales change of the same products before and after the Amazon

[12] One may wonder whether the decrease in the number of one-star reviews is driven by the sales decrease from DIR removal. However, we find qualitatively the same results ($\hat{\lambda} = 0.031, p < 0.05$), even after controlling for log sale rank in equation (3).

US policy-change date. Note that this design is not subject to the potential selection discussed earlier on our main design, since we now compare the sales of the same products on the two platforms.

For the analysis, we construct the Amazon UK dataset. We specifically collect review and sales data on a subset of products from the *Main* dataset that are also available from Amazon UK.[13] Using observations prior to Nov 22, 2016, we estimate the following equation:

$$\ln\left(S_{it}^{k}\right) = \lambda \cdot \left(Treat^{k} \times Post_{t}\right) + \boldsymbol{\beta}\mathbf{X}_{\mathbf{it}}^{k} + \boldsymbol{\gamma}_{\mathbf{i}}^{k} + \mathbf{M}_{\mathbf{t}} + \varepsilon_{it}^{k}, \tag{4}$$

where $\ln\left(S_{it}^{k}\right)$ is the log sales rank for product i on date t and platform $k \in \{\text{US}, \text{UK}\}$, $Treat^{k}$ is an indicator for the platform with DIR removal, taking the value of one for Amazon US and zero for Amazon UK, $Post_{t}$ is the indicator for observations after the policy change (i.e., one for Oct 3 to Nov 22, 2016 and zero for April 1 to Oct 2, 2016), $\mathbf{X}_{\mathbf{it}}^{k}$ is the vector of the same set of controls used in equation (2), and $\boldsymbol{\gamma}_{\mathbf{i}}^{k}$ are product-platform-specific fixed effects. Note that $\boldsymbol{\gamma}_{\mathbf{i}}^{k}$ absorb any potentially differential consumer preferences across the two platforms for the same product. Lastly, standard errors are clustered at the product-platform level.

The coefficient of our primary interest is λ, which captures the sales impact of DIR removal from the Amazon US platform. Consistent with our main result, we find that the estimate of λ is positive and significant ($\hat{\lambda} = 0.230, p < 0.05$), suggesting that products on Amazon US experienced a measurable sales reduction after the policy change compared to the same products on Amazon UK, which experienced no policy change (Table 6). As in our main model, we check the validity of the parallel-trends assumption by replacing $Post_{t}$ with year-month dummies in equation (4). As shown in Panel (A) of Figure 5, we find no evidence of differential trends in the outcome variable before the policy change.

Similarly, we check for the robustness of the impact of DIRs on the number of one-star verified reviews using the cross-platform design. We specifically, estimate the following equation:

$$\ln\left(1 + N_{iT}^{One\ star,k}\right) = \lambda \cdot \left(Treat^{k} \times Post_{t}\right) + \boldsymbol{\gamma}_{\mathbf{i}}^{k} + \mathbf{M}_{\mathbf{t}} + \varepsilon_{it}^{k}. \tag{5}$$

Consistent with our main result, as reported in Column (2) of Table 5, we find that the estimate for λ is negative and significant ($\hat{\lambda} = -0.072, p < 0.01$). This implies that the removal of DIR led to a decrease in the number of one-star verified reviews on Amazon US, in comparison to Amazon

[13] For details of the construction of the Amazon UK dataset, see Section A1 of the Appendix.

UK where DIR was not removed. Note that we find no evidence of differential trends in the outcome variable before the policy change as shown in Panel (B) of Figure 5.

Finally, our results are robust to alternative samples. We specifically show the robustness on (1) a common-support subsample (to rule out the effect of group difference) and (2) a subsample of products losing only DIRs after the policy change (to rule out the effect of undisclosed IR removal). We provide details of these robustness checks in Section A6 of the Appendix.

5 An Alternative Solution: Platform-Initiated IRs

Our empirical results so far reveal that the disclosure neither reverses reviewers' incentive to distort their IR ratings nor motivates consumers to fully discount the ratings inflation. Thus, the disclosure requirement is not sufficient to protect consumers from being misled by IRs. These findings motivate us to raise an important question: Is there an alternative way to eliminate reviewers' incentives to distort IR ratings? In this section, we examine another type of IR to answer this question.

According to our interviews with active incentivized reviewers, in an IR campaign, firms reach out to reviewers who are likely to post positive and informative reviews, as judged by their past IRs. Reviewers agree to post IRs in exchange for rewards from the firm. This suggests that inflated ratings result from compatible incentives between the reward providers (firms) and receivers (reviewers): Firms encourage inflated ratings to achieve higher demand for their products, and reviewers write overly positive IRs to obtain greater future rewards. Given this reasoning, we propose that reviewers would have little incentive to inflate their ratings if IRs were commissioned by a third party that benefits from fair (rather than inflated) reviews. Uninflated reviews do not necessarily destroy firms' motivation to participate in IRs, since they increase review volume, which is one of the two key word-of-mouth metrics (i.e., valance and volume) having a strong positive impact on sales (e.g., Liu 2006, Gong et al. 2017).

An online platform (e.g., Amazon, eBay, Airbnb, TripAdvisor) is a good candidate for being such a third party. One of the platform's main goals is to facilitate transactions between two parties (i.e., firms and consumers). Thus, the platform benefits from fair reviews for all participating products/services as these reviews allow consumers to make better-informed decisions (Dellarocas 2003). Thus, we formally define two types of IR system: firm-initiated and platform-initiated, which differ in who commissions the IRs (thus far, all our reported empirical findings have been based on firm-initiated IRs).

In what follows, we provide an empirical analysis of the platform-initiated IRs (PIR henceforth) using Amazon's Vine program as our empirical context. In their Vine program, Amazon posts a set of free products to its website. From this list of products, selected reviewers (called Vine Voices) pick one or two items to review each month.[14] Reviewers are invited to the program by Amazon while firms voluntarily participate by providing free products. Reviewers remain anonymous to the firms, however, since Amazon distributes products to reviewers at its discretion. We identify Vine reviews using the "Vine reviews" tag attached by Amazon. For the analyses in this section, we use the *Consumer Reports* dataset (introduced in Section 2.2). To exploit within-product variation, we restrict our attention to 432 products with at least one Vine and one verified review.

To examine whether PIRs have inflated ratings, we test for the rating difference between Vine reviews and verified reviews by estimating equation (1), while replacing DIR_{ij} with the PIR dummy (PIR_{ij}), which is one for Vine reviews and zero for verified reviews. The coefficient for PIR_{ij} captures the difference in ratings between verified reviews and PIRs. According to the estimation results reported in Column (1) of Table 7, the estimated coefficient of PIR_{ij} is not significant ($\hat{\alpha} = -0.035, p > 0.10$). In other words, we find no evidence of rating difference between verified and Vine reviews.

While Vine-review ratings are overall uninflated, it is possible that low-quality products may have inflated Vine ratings. To examine whether this may be the case, we estimate equation (1) by replacing DIR_{ij} with PIR_{ij} and adding an interaction term $PIR_{ij} \times Q_i$, where Q_i is the percentile of product i's quality score within a category. As reported in Column (2) of Table 7, we continue to find that the ratings are not systematically different between platform-initiated IRs and VRs. Importantly, we find no evidence for the moderating role of product quality. This result implies that PIRs are not significantly inflated even for lower-quality products and that reviewers of PIRs remain largely unaffected by financial gain from the free products they receive, regardless of the product quality.

[14] For more details see: https://www.npr.org/templates/transcript/transcript.php?storyId=247833514 (accessed February 21, 2023).

6 Conclusion

Online reviews are a credible source of information for consumers. To take advantage of this credibility, many firms incentivize reviewers to post reviews for their products, especially in the early stages of the product life cycle given the consequential impact of earlier reviews (Park et al. 2021, Kuksov and Xie 2010). The concern is that due to financial ties between firms and reviewers, IRs are rarely impartial and if undisclosed, likely to mislead consumers' purchase decisions. The current FTC guideline recommends that material connections should be disclosed in IRs. In this paper, we assess the efficacy of this recommendation by examining whether disclosure of IRs is effective as a self-discipline mechanism for reviewers and a warning signal to consumers. We also aim to suggest an alternative way to remove the bias in IRs while keeping their volume of information.

Our empirical investigation of DIRs reveals that their ratings are inflated compared to VRs, suggesting that the disclosure cannot discipline reviewers to be strictly impartial in their IRs. Moreover, our analysis shows that inflated ratings of DIRs increase product sales, implying that disclosure cannot help consumers to fully discount the rating inflation. We further find that DIRs propagate one-star reviews, indicating that some consumers are indeed misled by IRs even in the presence of the disclosure. Together, our analyses suggest that disclosure is not an effective measure to prevent distorted reviews or to protect consumers from being misled by inflated ratings.

We find, however, that an alternative IR system can restore the impartiality of IRs. Note that the rating distortion arises because inflated ratings align firms' and reviewers' incentives. When the platform provides incentives to reviewers, however, inflated ratings no longer align incentives of the two parties because the platform, as the marketplace benefiting from fair transactions of all products, does not gain from IRs' inflated ratings for any particular product. Hence, switching to platform-initiated IRs can remove rating distortion. Empirically, we find a consistent result: the ratings are not systematically different between Amazon's Vine and verified reviews.

The results of our study have important implications. First, our research calls for consumers' attention to IR disclaimers and their awareness of the differences between incentivized and organic reviews, as well as between firm- and platform-initiated IRs. Second, our results suggest that merely requiring a disclosure of an incentivized review is insufficient. We find that the disclosure can neither keep the review system free of bias nor protect consumers from such potential bias.

This calls for stronger platform or government policies than simply requiring disclosure of financial relationships. Third, review platforms should carefully design the review ecosystem while providing specific guidelines for IRs. We show that, even with disclosure, IRs may mislead consumers into suboptimal choices because of overly positive review content. In this process, we also underscore the importance of considering the incentives of self-interested reviewers (see also Kim et al. 2019 and Chung et al. 2020 for how reviewer incentives influence their ratings). Platforms may consider adopting platform-initiated IRs, which could alter reviewers' incentives.

A few avenues for further research follow. First, we primarily use Amazon as our empirical context; however, our findings for this platform may not perfectly translate to other online platforms. Thus, it would be worthwhile to extend the study to other contexts (e.g., Fradkin and Holtz 2022). Second, while we suggest one alternative (i.e., platform-initiated IRs) to the current policy, our research also calls for more academic research on many other measures that can keep the review ecosystem impartial and protect consumers from the potential bias in IRs. When the results of the current paper were presented to FTC's Division of Advertising Practices, its consumer protection attorneys indicated that FTC is preparing a new recommendation related to the IR practice. Academic research can provide useful inputs in this process. We leave these for future research.

References

Anderson, E. T., & Simester, D. I. (2014). Reviews without a purchase: Low ratings, loyal customers, and deception. *Journal of Marketing Research*, 51(3), 249-269.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics*, 119(1), 249-275.

Burtch, G., Hong, Y., Bapna, R., & Griskevicius, V. (2018). 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.

Cain, D. M., Loewenstein, G., & Moore, D. A. (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *The Journal of Legal Studies*, 34(1), 1-25.

Chevalier, J., and Goolsbee, A. (2003). Measuring prices and price competition online: Amazon. com and BarnesandNoble. com. *Quantitative Marketing and Economics*, 1(2), 203-222.

Chung, K., Kim, K., & Lim, N. (2020). Social structures and reputation in expert review systems. *Management Science*, 66(7), 3249-3276.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10), 1407-1424.

Dellarocas, C. (2006). "Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms," *Management Science*, 52 (10), 1577–93.

Fradkin, A., & Holtz, D. (2022). Do Incentives to Review Help the Market? Evidence from a Field Experiment on Airbnb. *Marketing Science,* forthcoming.

Godes, D., and Silva, J. C. (2012). Sequential and temporal dynamics of online opinion. *Marketing Science*, 31(3), 448-473.

Gong, S., Zhang, J., Zhao, P., & Jiang, X. (2017). Tweeting as a marketing tool: A field experiment in the TV industry. *Journal of Marketing Research*, 54(6), 833-850.

He, S., Hollenbeck, B., & Proserpio, D. (2022). The market for fake reviews. *Marketing Science*, 41(5), 896-921.

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.

Kim, K., Chung, K., & Lim, N. (2019). Third-party reviews and quality provision. *Management Science*, 65(6), 2695-2716.

Kuksov, D., & Xie, Y. (2010). Pricing, frills, and customer ratings. *Marketing Science*, 29(5), 925-943.

Li, L., Tadelis, S., & Zhou, X. (2020). Buying reputation as a signal of quality: Evidence from an online marketplace. *The RAND Journal of Economics*, 51(4), 965-988.

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

Loewenstein, G., Cain, D. M., & Sah, S. (2011). The limits of transparency: Pitfalls and potential of disclosing conflicts of interest. *American Economic Review*: *Papers and Proceedings*, 101(3), 423-428.

Mostagir, M., & Siderius, J. (2023). Strategic reviews. *Management Science*, 69(2), 904-921.

Park, S., W. Shin, and J. Xie (2021). The Fateful First Consumer Review, *Marketing Science*, 40(3), 481-507.

Park, S., Xie, M., & Xie, J. (2023). Frontiers: Framing price increase as discount: A new manipulation of reference price. *Marketing Science*, 42(1), 37-47.

Pearson, S. D., K. Kleinman, and D. Rusinak (2006). A Trial of Disclosing Physicians' Financial Incentives to Patients, *Archives of Internal Medicine*, 166,n 623–28.

Pinch, T., & Kesler, F. (2011). How aunt Ammy gets her free lunch: A study of the top-thousand customer reviewers at amazon. com. *Managing Overflow in Affluent Societies*.

Qiao, D., Lee, S. Y., Whinston, A. B., & Wei, Q. (2020). Financial incentives dampen altruism in online prosocial contributions: A study of online reviews. *Information Systems Research*, 31(4), 1361-1375.

Shi, M., & Wojnicki, A. C. (2014). Money talks… to online opinion leaders: What motivates opinion leaders to make social-network referrals?. *Journal of advertising research*, 54(1), 81-91.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, 30(4), 526.

Sun, M. (2012). How Does the Variance of Product Ratings Matter? *Management Science*, 58(4), 696-707.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157), 1124-1131.

Figure 1. Amazon's Policy Change on Incentivized Reviews

Innovation

Update on customer reviews

By Chee Chew on October 03, 2016

Customer reviews are one of the most valuable tools we offer customers for making informed purchase decisions, and we work hard to make sure they are doing their job. In just the past year, we've improved review ratings by introducing a machine learned algorithm that gives more weight to newer, more helpful reviews; applying stricter criteria to qualify for the Amazon verified purchase badge; and suspending, banning or suing thousands of individuals for attempting to manipulate reviews.

Our community guidelines have always prohibited compensation for reviews, with an exception – reviewers could post a review in exchange for a free or discounted product as long as they disclosed that fact. These so-called 'incentivized reviews' make up only a tiny fraction of the tens of millions of reviews on Amazon, and when done carefully, they can be helpful to customers by providing a foundation of reviews for new or less well-known products.

Today, we updated the community guidelines to prohibit incentivized reviews unless they are facilitated through the Amazon Vine program. We launched Vine several years ago to carefully facilitate these kinds of reviews and have been happy with feedback from customers and vendors. Here's how Vine works: Amazon – not the vendor or seller – identifies and invites trusted and helpful reviewers on Amazon to post opinions about new and pre-release products; we do not incentivize positive star ratings, attempt to influence the content of reviews, or even require a review to be written; and we limit the total number of Vine reviews that we display for each product. Vine has important controls in place and has proven to be especially valuable for getting early reviews on new products that have not yet been able to generate enough sales to have significant numbers of organic reviews. We also have ideas for how to continue to make Vine an even more useful program going forward. Details on that as we have them.

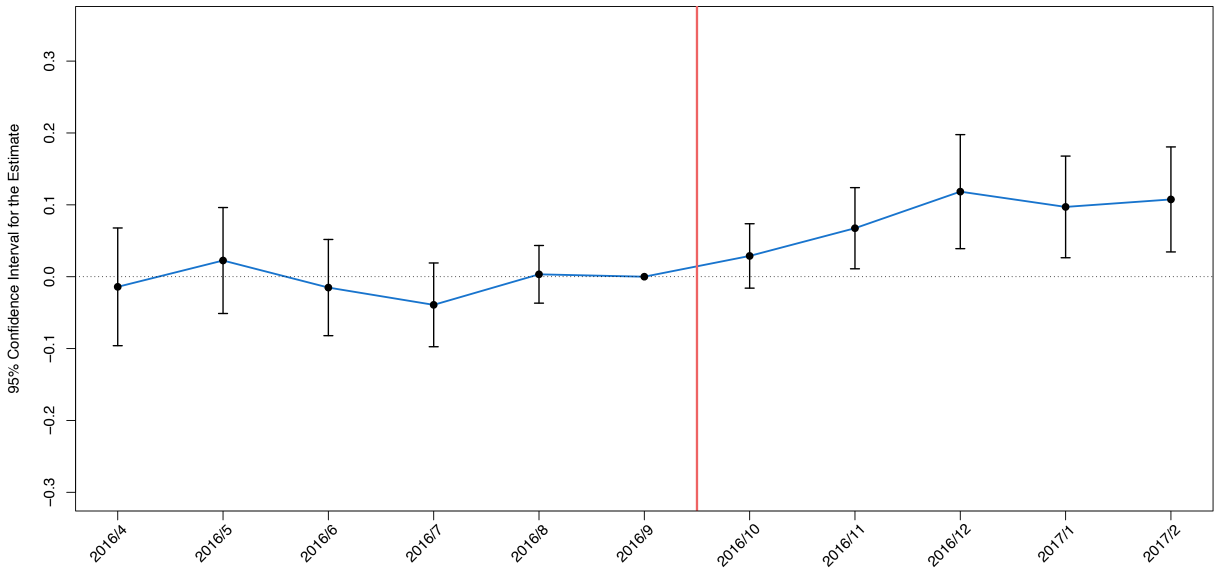
The above changes will apply to product categories other than books. We will continue to allow the age-old practice of providing advance review copies of books.

Thank you.

Source: https://www.aboutamazon.com/news/innovation-at-amazon/update-on-customer-reviews (accessed February 21, 2023)

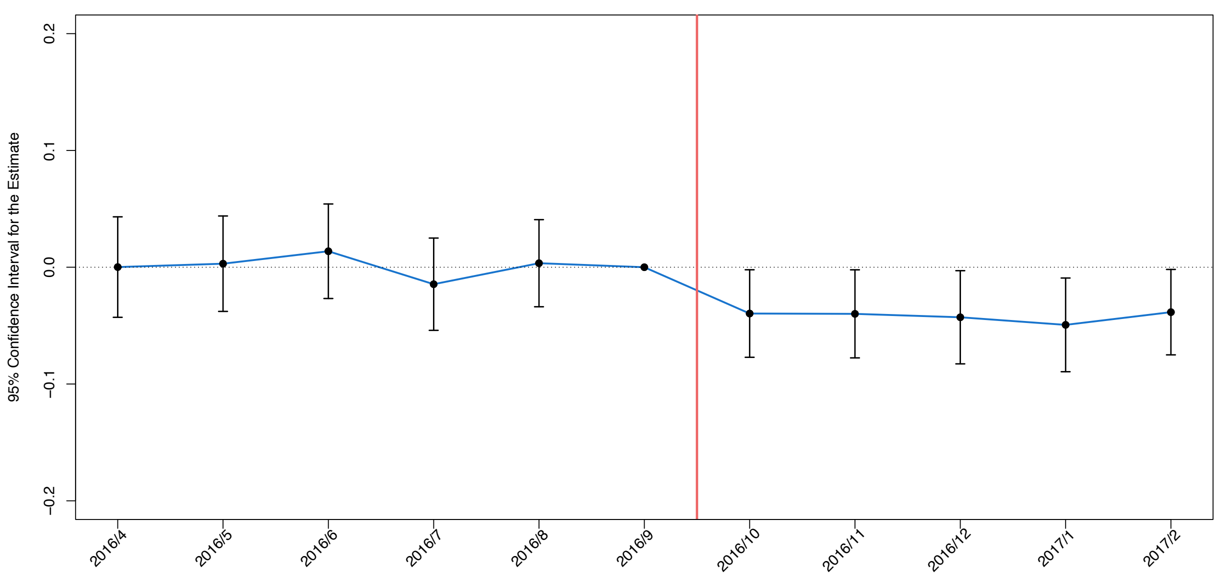
Figure 2. Impact of the Removal of DIRs

Panel (A) Log Sales Rank



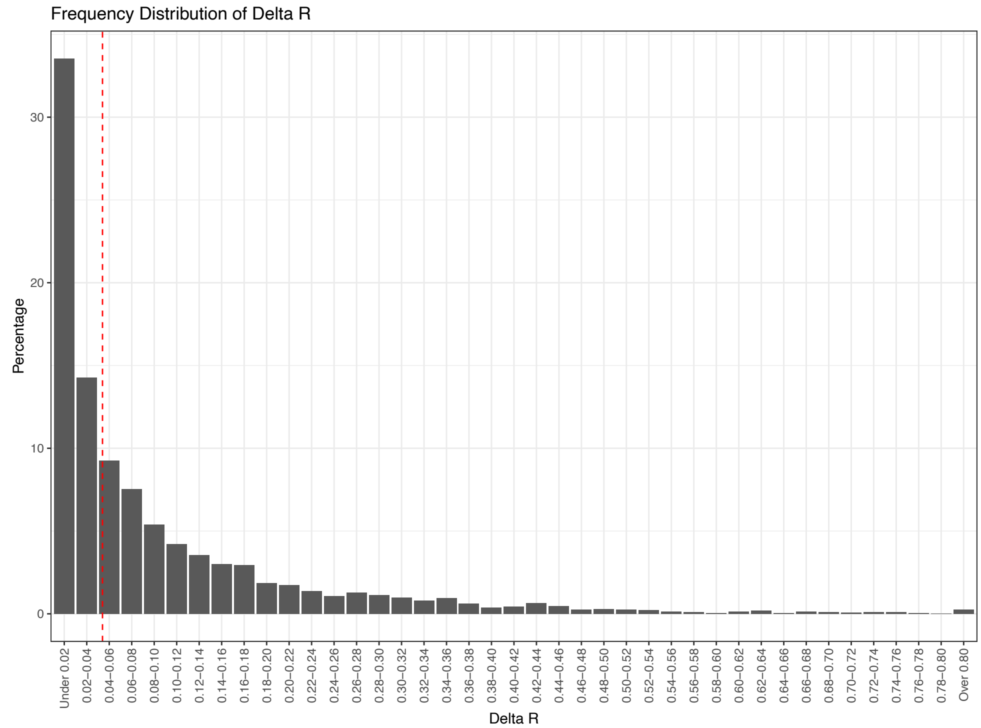
Notes. In this figure, we report the month-specific estimates of **λ** along with 95-percent confidence intervals.

Panel (B) Log Number of One-Star Verified Reviews



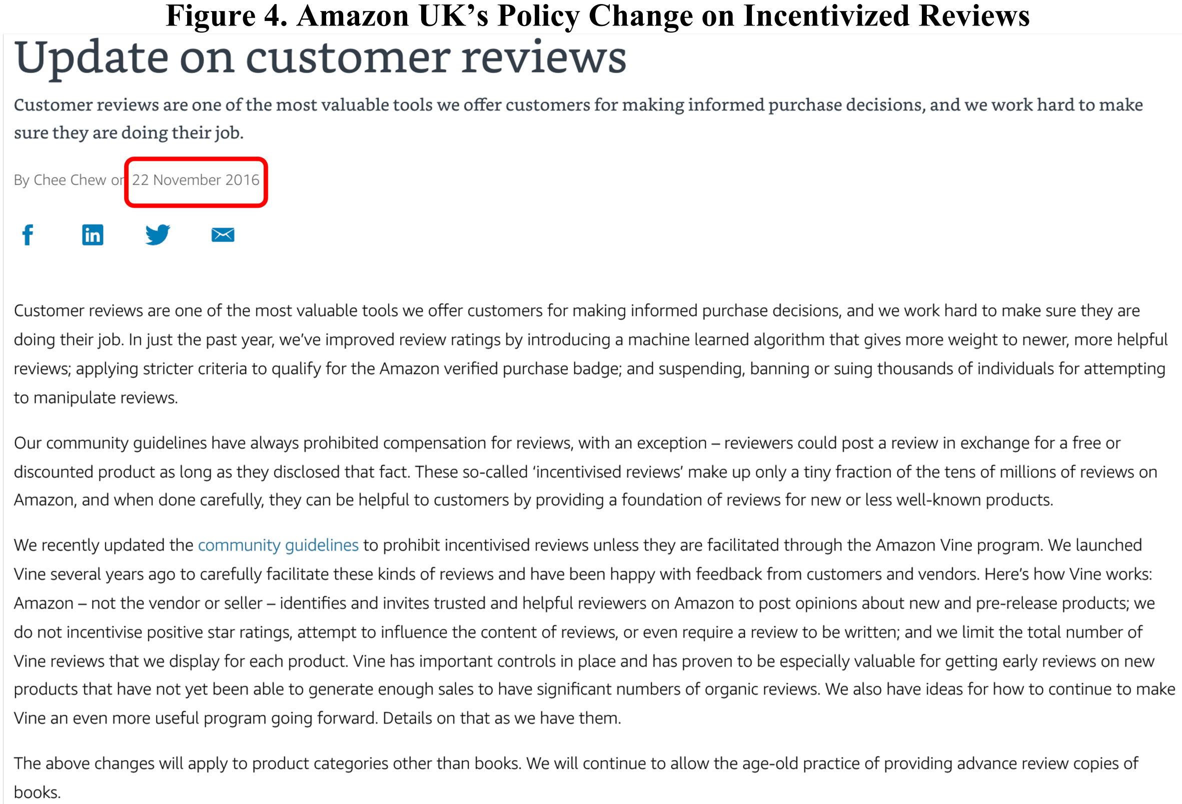
Notes. In this figure, we report the month-specific estimates of **λ** along with 95-percent confidence intervals.

Figure 3. Distribution of ΔR



Notes. A vertical dotted line represents the median.

Figure 4. Amazon UK's Policy Change on Incentivized Reviews



Update on customer reviews

Customer reviews are one of the most valuable tools we offer customers for making informed purchase decisions, and we work hard to make sure they are doing their job.

By Chee Chew on 22 November 2016

Customer reviews are one of the most valuable tools we offer customers for making informed purchase decisions, and we work hard to make sure they are doing their job. In just the past year, we've improved review ratings by introducing a machine learned algorithm that gives more weight to newer, more helpful reviews; applying stricter criteria to qualify for the Amazon verified purchase badge; and suspending, banning or suing thousands of individuals for attempting to manipulate reviews.

Our community guidelines have always prohibited compensation for reviews, with an exception – reviewers could post a review in exchange for a free or discounted product as long as they disclosed that fact. These so-called 'incentivised reviews' make up only a tiny fraction of the tens of millions of reviews on Amazon, and when done carefully, they can be helpful to customers by providing a foundation of reviews for new or less well-known products.

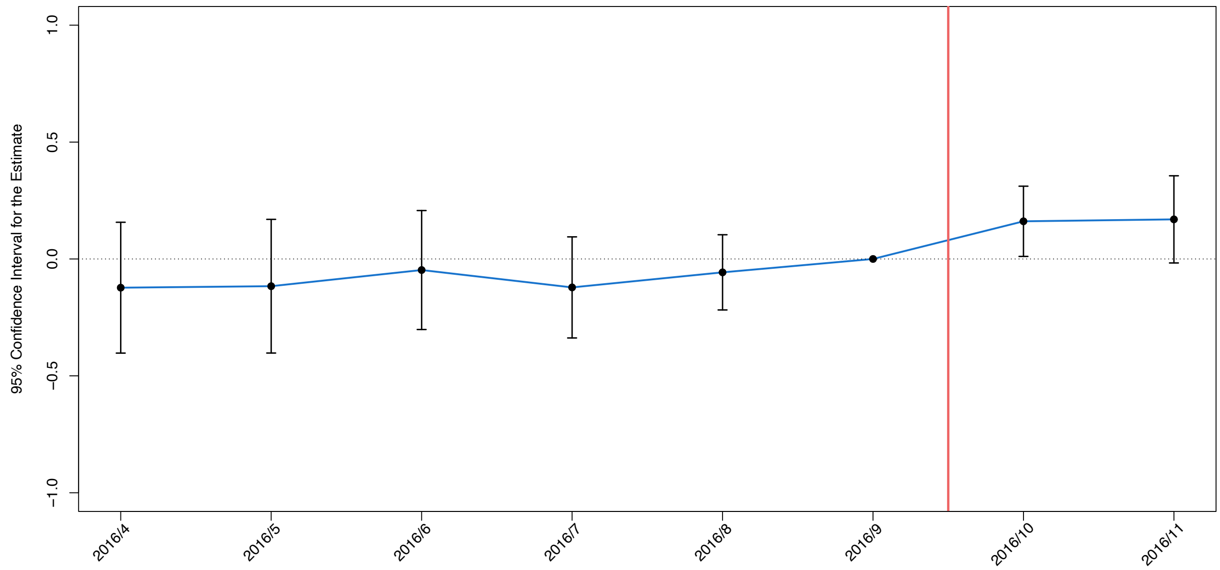
We recently updated the community guidelines to prohibit incentivised reviews unless they are facilitated through the Amazon Vine program. We launched Vine several years ago to carefully facilitate these kinds of reviews and have been happy with feedback from customers and vendors. Here's how Vine works: Amazon – not the vendor or seller – identifies and invites trusted and helpful reviewers on Amazon to post opinions about new and pre-release products; we do not incentivise positive star ratings, attempt to influence the content of reviews, or even require a review to be written; and we limit the total number of Vine reviews that we display for each product. Vine has important controls in place and has proven to be especially valuable for getting early reviews on new products that have not yet been able to generate enough sales to have significant numbers of organic reviews. We also have ideas for how to continue to make Vine an even more useful program going forward. Details on that as we have them.

The above changes will apply to product categories other than books. We will continue to allow the age-old practice of providing advance review copies of books.

Source: https://blog.aboutamazon.co.uk/shopping-and-entertainment/update-on-customer-reviews (accessed on December 21, 2021)

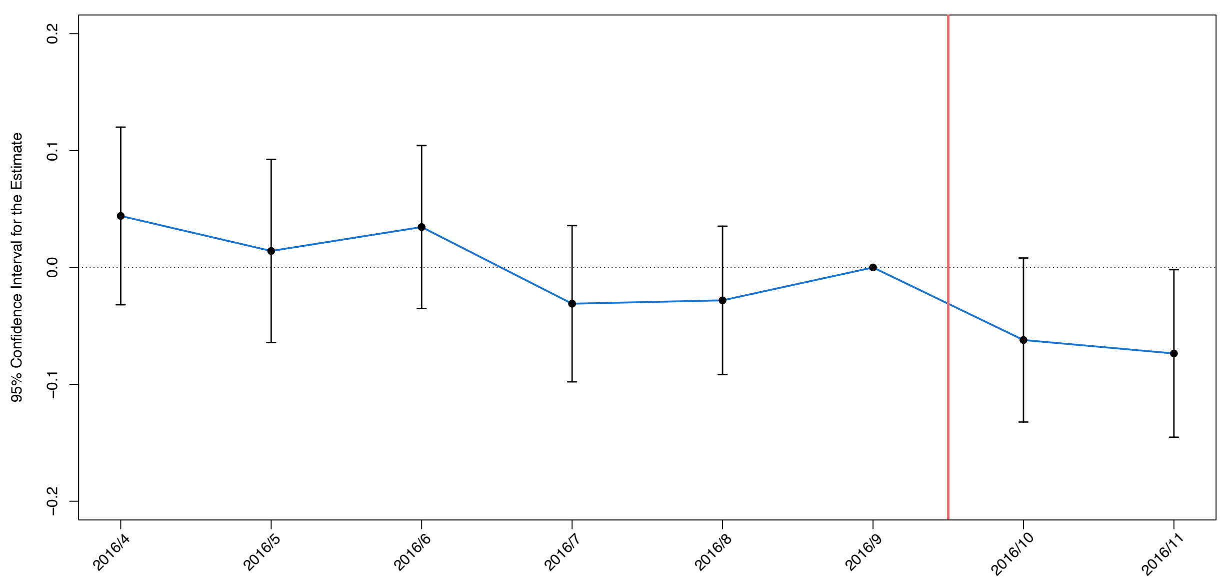
Figure 5. Impact of the Removal of DIRs: Amazon US – Amazon UK Design

Panel (A) Log Sales Rank



Notes. In this figure, we report the month-specific estimates of **λ** along with 95-percent confidence intervals.

Panel (B) Log Number of One-Star Verified Reviews



Notes. In this figure, we report the month-specific estimates of **λ** along with 95-percent confidence intervals.

Table 1. Datasets

Dataset	Construction	Number of Products	Analysis
Main	• Products with at least one DIR • Reviews: Before and after DIR removal	• Before Amazon US policy change: 5,273 • After Amazon US policy change: 5,101	• Difference in ratings: DIRs vs. VRs (Section 3) • The impact of DIR removal on sales (Section 4) • The impact of DIR removal on the number of one-star VRs (Section 4)
Consumer Reports	• Products with quality scores from *Consumer Reports* in appliances and electronics	• 1,723 with quality score • 1,616 (among 1,723) had at least one review from Amazon	• Difference in ratings between DIRs and VRs: moderation by quality (Section 3) • Difference in ratings: Vine reviews vs. VRs (Section 5) • Difference in ratings between Vine reviews and VRs: moderation by quality (Section 5)
Amazon UK	• A subset of products in the main dataset matched with products from Amazon UK.	• 616 matched products with at least one review from Amazon UK	• The impact of DIR removal on sales: cross-platform analysis (Section 4) • The impact of DIR removal on the number of one-star VRs: cross-platform analysis (Section 4)
Top Reviewers	• All reviews posted by Amazon's top-1,000 ranked reviewers as of April 2016.	• 708,091 products with at least one review from top-1,000 ranked reviewers.	• Difference in ratings between DIRs and VRs: within-reviewer analysis (Section 3)

Notes. In the Main dataset, 172 products became unavailable when we collected the review dataset after the policy change.

Table 2. Rating Distributions: DIRs and VRs

Rating	DIRs	Verified Reviews
Rating: 1	0.49%	11.11%
Rating: 2	0.84%	5.89%
Rating: 3	2.94%	7.75%
Rating: 4	17.09%	14.58%
Rating: 5	78.64%	60.67%
Average Ratings	4.73	4.08
N	241,831	310,580

Notes. For this Table, we use DIRs and VRs posted to 5,273 products in the *Main* dataset collected before the policy change.

Table 3. Difference in Ratings: DIRs and VRs

	(1)		(2)	
	Estimate	S.E.	Estimate	S.E.
DIR	0.480***	0.009	0.486***	0.086
DIR × Quality			-0.367**	0.140
Log Reviewer Ranking	4.23E-04	1.26E-03	9.64E-03	1.05E-02
Log Number of Reviews of Reviewer	0.071***	0.003	0.102***	0.013
Time	-1.41E-04***	2.70E-05	-1.63E-04***	5.27E-05
Order	-2.95E-04***	6.00E-05	2.15E-07	6.77E-07
Product Fixed Effects	Yes		Yes	
R^2	0.204		0.043	
N	544,014		389,175	

Notes. Standard errors are clustered at the product level. The unit of analysis is an individual review. For Column (1), we use DIRs and VRs posted to a subset of products in the *Main* dataset. Specifically, we focus on 5,191 products with at least one DIR and one VR to exploit the within-product variation. For Column (2), we use DIRs and VRs posted to products in the *Consumer Reports* dataset. Note that all 105 products used in Column (2) have quality scores and each product received at least one DIR and one VR.
**Significantly different from zero at $p<0.05$. ***Significantly different from zero at $p < 0.01$.

Table 4. The Impact of the DIR Removal on Sales

	(1) Full Sample		(2) Low-ΔR Group		(3) High-ΔR Group		(4) Within Treatment Group	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Treat×Post	0.087***	0.027	0.078**	0.031	0.109***	0.032	0.051**	0.023
Log Number of Reviews	-0.238***	0.030	-0.265***	0.041	-0.218***	0.040	-0.220***	0.031
Average Ratings	-0.473***	0.054	-0.468***	0.086	-0.471***	0.066	-0.438***	0.056
Log Price	0.519***	0.031	0.525***	0.045	0.486***	0.040	0.541***	0.031
Product Fixed Effects	Yes		Yes		Yes		Yes	
Year-Month Fixed Effects	Yes		Yes		Yes		Yes	
R^2	0.818		0.829		0.816		0.812	
N	982,999		537,231		574,336		854,431	

Notes. The unit of analysis is Product × Day. The Standard errors are clustered at the product level. The log number of reviews and the average ratings are computed based on all reviews but DIRs. In Columns (2) and (3), we use the low-ΔR and high- ΔR groups as the treatment groups, respectively, while keeping the control group the same. In Column (4), the low- ΔR group serves as the control group, and the high- ΔR group serves as the treatment group.

**Significantly different from zero at p<0.05. ***Significantly different from zero at p <0.01.

Table 5. The Impact of the DIR Removal on the Number of One-Star Verified Reviews

	(1)		(2) Cross-Platform Analysis	
	Estimate	S.E.	Estimate	S.E.
Treat×Post	-0.043***	0.012	-0.073***	0.022
Product Fixed Effects	Yes			
Product-Platform Fixed Effects			Yes	
Year-Month Fixed Effects	Yes		Yes	
R^2	0.402		0.411	
N	44,858		4,016	

Notes. In Column (1), the unit of analysis is a Product×Month. Standard errors are clustered at the product level. We focus on products that received at least one VR during the analysis period. In Column (2), the unit of analysis is a Product×Platform×Month. Standard errors are clustered at the Product×Platform level. We use products that received at least one VR on both Amazon US and Amazon UK during the analysis period.

***Significantly different from zero at p <0.01.

Table 6. The Impact of the DIR Removal on Sales: Cross-Platform Analysis

	Estimate	S.E.
Treat×Post	0.230**	0.093
Log Number of Reviews	0.046	0.092
Average Ratings	0.065	0.144
No-review dummy	0.654	0.679
Log Price	0.747***	0.109
Product-Platform Fixed Effects	Yes	
Year-Month Fixed Effects	Yes	
R^2	0.832	
N	33,550	

Notes. The unit of analysis is a Product×Platform×Day. Standard errors are clustered at the Product×Platform level. The log number of reviews and the average ratings are computed based on all reviews other than DIRs. The average rating is coded as zero when there is no review. In this Table, we restrict our attention to products that lost at least one DIR from Amazon US to measure the DIR-removal effect. We use products for which sales rank and price data are available from *Keepa* on both Amazon US and Amazon UK during the analysis period.

**Significantly different from zero at p<0.05. ***Significantly different from zero at p <0.01.

Table 7. Difference in Ratings: Platform-Initiated IRs and Verified Reviews

	(1)		(2)	
	Estimate	S.E.	Estimate	S.E.
Platform-Initiated IRs	-0.035	0.032	-0.027	0.049
Platform-Initiated IRs ×Quality			-0.015	0.068
Log Reviewer Ranking	0.051***	0.004	0.051***	0.004
Log Reviewer Number of Reviews	0.163***	0.010	0.163***	0.010
Time	-1.21E-04***	4.10E-05	-1.21E-04***	4.10E-05
Order	-8.51E-06	1.40E-05	-8.52E-06	1.40E-05
Product-level Fixed Effects	Yes		Yes	
R^2	0.087		0.087	
N	180,966		180,966	

Notes. Standard errors are clustered at the product level. For this Table, we use PIRs and VRs posted to products in the *Consumer Reports* dataset. We restrict our attention to products with at least one PIR and one VR to exploit the within-product variation.
***Significantly different from zero at p <0.01.