7-2021

Online review solicitations reduce extremity bias in online review distributions and increase their representativeness

Hülya KARAMAN
Singapore Management University, hkaraman@smu.edu.sg

Citation

KARAMAN, Hülya. Online review solicitations reduce extremity bias in online review distributions and increase their representativeness. (2021). *Management Science*. 67, (7), 4420-4445.
Available at: https://ink.library.smu.edu.sg/lkcsb_research/6586

informs
http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 67, No. 7, July 2021, pp. 4420–4445
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Online Review Solicitations Reduce Extremity Bias in Online Review Distributions and Increase Their Representativeness

Hülya Karaman[a]

[a] Department of Marketing, Lee Kong Chian School of Business, Singapore Management University, Singapore 178899
Contact: hkaraman@smu.edu.sg, https://orcid.org/0000-0003-3488-4222 (HK)

Received: February 6, 2017
Revised: August 19, 2018; April 13, 2020
Accepted: June 16, 2020
Published Online in Articles in Advance: October 29, 2020

https://doi.org/10.1287/mnsc.2020.3758

Abstract. Representative online customer reviews are critical to the effective functioning of the Internet economy. In this study, I investigate the representativeness of online review distributions to examine how extremity bias and conformity impact it and explore whether online review solicitations alter representativeness. Past research on extreme distribution of online ratings commonly relied solely on observed public online ratings. One strength of the current paper is that I observe the private satisfaction ratings of customers regardless of whether they choose to write an online review or not. I show that both extremity bias and conformity exist in unsolicited online word-of-mouth (WOM) and introduce online review solicitations as a mechanism that can partially de-bias ratings. Solicitations increase all customers' engagement in online WOM, but if solicited, those with moderate experiences increase their engagement more than those with extreme experiences. Consequently, although extremity bias still exists in solicited online WOM, solicitations significantly increase the representativeness of rating distributions. Surprisingly, the results demonstrate that without conformity, unsolicited online WOM would be even less representative of the original customer experiences. Furthermore, I document that both solicited and unsolicited reviews equally overstate the average customer experience (compared with average private ratings) despite stark differences in their rating distributions. Finally, I establish that solicitations for reviews on the company-owned website, on average, decrease the number of one-star reviews on a third-party review platform.

History: Accepted by Eric Anderson, marketing.

Funding: This paper is based on the author's dissertation research supported by the Goizueta Business School at Emory University. The author acknowledges additional funding from the Singapore Ministry of Education Academic Research Fund Tier 1 grant [Grant C207MSS18B014].
Supplemental Material: The online appendices are available at https://doi.org/10.1287/mnsc.2020.3758

Keywords: online review solicitations • reporting biases • representative reviews • extremity bias • social influence • conformity • online reviews • online word-of-mouth • third-party review platforms • TripAdvisor

1. Introduction

When choosing goods and services, customers increasingly look to online word-of-mouth (WOM) to provide accurate information about product quality. Previous research has shown that the impact of one-star reviews on sales is greater than five-star reviews, suggesting that customers pay attention to the distribution of online reviews rather than merely relying on summary statistics (Chevalier and Mayzlin 2006). However, given the voluntary nature of online WOM, it is not clear whether the distribution of online reviews is actually representative of the underlying experiences of customers.

It is widely speculated that online reviews are subject to many reporting biases (Berger 2014, Luca 2016). One type of bias—what I refer to as *extremity bias*—reflects the notion that the set of customers who write online reviews is not representative of the underlying customer population. Researchers assume that customers at the *extreme* ends of the satisfaction spectrum engage in greater online WOM (Li and Hitt 2008, Hu et al. 2017). Another type of bias—what I call *conformity*—captures the idea that the satisfaction ratings reviewers report could depend in part on their exposure to preexisting online reviews, specifically to the average rating (which tends to be displayed prominently on review websites). This suggests that customers who write online reviews make systematic mistakes and provide an inaccurate measurement of their actual experience by misreporting their satisfaction to *conform* to the average of previous ratings.

Both types of bias call into question the degree to which online WOM represents underlying customer experiences. These sorts of biases matter because they can influence not just which products are bought, but even which products are made. For instance, if an asymmetric extremity bias exists, then an unrepresentative high share of bad ratings could harm high-quality products by discouraging future purchases, even though a representative set of reviews would endorse the product. As a result, the extremity bias could reduce the premium on quality and disincentivize companies from producing high-quality goods.

In this study, I focus on the representativeness of online review distributions to examine how extremity bias and conformity impact it and explore whether online review solicitations alter representativeness. In doing so, I answer the call by Askalidis et al. (2017) and Schoenmueller et al. (2019) for research into whether unsolicited or solicited reviews provide a more representative set of reviews. Review solicitations have the potential to change the representativeness of online review distributions through extremity bias and conformity. Previous research has documented that review solicitations are very effective in stimulating online review creation (Burtch et al. 2017, Klein et al. 2018). If review solicitations successfully motivate underrepresented customers to engage in online WOM, then companies could solicit reviews to reduce extremity bias. Review solicitations could also provide companies a unique opportunity to manipulate conformity. For instance, companies can *enable* (*disable*) conformity by (not) displaying previously written online reviews to potential reviewers throughout the review solicitation procedure.

In addition to gathering a more representative set of reviews, companies could also use solicitations to impact WOM in a variety of ways. For instance, customers who would ordinarily leave an online review on a third-party website might be induced (if solicited by the company) to leave a review on the company website instead—in effect shifting online reviews from one website to another. Companies could be interested in pursuing this strategy if it resulted in moving negative online WOM away from third-party review platforms (that potentially have a greater reach than the company website) because customers typically use these websites for product comparison. Currently, it is not well understood how review solicitations for a company website affect review writing behavior in terms of reporting biases nor how they affect review writing on third-party review aggregation websites. Nevertheless, companies have begun to implement online review solicitations for their own websites. Thus, a closer look at the impact of review solicitation is both timely and warranted. Accordingly, I address three research questions:

1. Does extremity bias or conformity occur in online reviews, and if so, how do they impact representativeness?
2. Which set of reviews—solicited or unsolicited—is most representative of customer experiences?
3. Do review solicitations for a company website affect reviews on third-party aggregation websites?

To answer these questions, I generated a unique data set from a multichain hotel group that conducts customer satisfaction surveys and hosts online reviews on its own website. Four features of the current setup made it ideal to test my questions. First, even if customers did not write an online review, I was able to observe their private satisfaction ratings. These private ratings allowed me to (1) explore which customers were more likely to engage in solicited or unsolicited online WOM (i.e., extremity bias), (2) investigate how reviewers changed their private ratings when posting their public online ratings (i.e., conformity), and (3) construct a baseline rating distribution of the entire customer base against which public online rating distributions (of either solicited or unsolicited reviews) are compared to evaluate representativeness of both solicited and unsolicited reviews. The resulting analyses build on and extend previous studies that relied solely on observed public online ratings (Li and Hitt 2008, Moe and Schweidel 2012, Hu et al. 2017).

Second, the hotel group randomly solicited a fraction of satisfaction survey takers to write online reviews. This random intervention allowed me to establish the causal effect of online review solicitations. Reporting has shown that other potential data sets (such as from TripAdvisor's Review Express service) are less than ideal, because solicitations are not random but instead are contaminated by the fraudulent activity by hotels selectively prompting only the customers they believe will give positive feedback (Guardian 2018).

A third feature of the data set that made it ideal for my purposes was that I could distinguish between two distinct groups of reviewers: those who were exposed to preexisting online ratings while writing an online review on the hotel's website and those who did not see such information throughout the review solicitation procedure. Behavioral differences between these two groups of reviewers allowed me to test the existence of conformity in online reviews.

Finally, the hotel group initially implemented review solicitations for a random subset of its hotels before rolling out solicitations to all hotels within its portfolio. This allowed me to construct treatment and control groups to examine the impact on reviewing behavior at TripAdvisor.com of online review solicitations for the hotel company's website.

The results show that, although both solicited and unsolicited customers exhibit extremity bias, two distinctions emerge within extremity bias between solicited and unsolicited reporting: (1) Extremity bias is asymmetric for unsolicited customers, with extremely dissatisfied customers engaging in greater unsolicited online WOM than extremely satisfied ones, and (2) extremity bias overall (i.e., in both directions) is weaker for solicited customers. The findings also reveal that even though solicitations significantly increase the propensity of all customers to write an online review, the exact magnitude of this solicitation effect depends on their private satisfaction rating. Solicitations increase online WOM participation of customers with moderate experiences more than those with extreme experiences, thereby attenuating extremity bias and increasing the representativeness of online WOM. Furthermore, I document the existence of conformity. Surprisingly, despite being typically viewed as a bias, conformity counteracts extremity bias and results in greater representativeness in the rating distributions of reviewers who are exposed to the preexisting online reviews during review creation. Finally, I provide evidence that review solicitations for the company website significantly shift extremely negative (i.e., one-star) reviews from third-party platforms to the company website.

Although I initially evaluate the representativeness of the online rating distributions, I recognize that customer purchasing decisions are also driven by online review valence, typically measured by the average online rating, and to a certain extent, by the online review variance (Rosario et al. 2016). Even if extremity bias exists, a disproportionately high share of extremely satisfactory and extremely unsatisfactory experiences in unsolicited online WOM could cancel each other out, leaving the average online rating unbiased. The same property applies to solicited online WOM. I show that posted online reviews equally overstate average ratings (compared with average private ratings) regardless of whether they are solicited or unsolicited. This is particularly interesting because I document considerable differences between the online rating distributions of solicited and unsolicited reviews. Remarkably, the distributional differences between them do not translate into differences in average product ratings. However, the variance between the rating distributions of solicited and unsolicited reviews does change: The variance of online reviews would be lower if all customers were solicited as opposed to unsolicited.

This study contributes to the literature on online reviews in several ways. First, whereas past work on unsolicited online WOM could only presume its existence, this study documents the presence of extremity bias in both solicited and unsolicited online WOM using observed data on customer satisfaction and online WOM engagement. Second, I provide evidence to support the hypothesis that customers adjust their private ratings to conform to the preexisting average online rating as they post them publicly on the review website. This conformity effect increases the representativeness of unsolicited online WOM. Third, the study deepens our understanding of the representativeness of unsolicited online WOM and the underlying mechanism through which representativeness is altered by online review solicitations. Fourth, to the best of my knowledge, this is the first study to document online WOM spillover effects across different review websites. This is important because the finding suggests that stimulating online reviews posted to a company website through review solicitations reduces the amount of negative online WOM at a review aggregator website, which potentially has a more extensive reach than the company website itself.

I organize the remainder of the article as follows: First, I review the previous literature. Next, I provide a formal definition of the representativeness of online review distributions, extremity bias, and the solicitation effect. I then offer a detailed description of the data and describe the empirical analysis. Finally, I present my results and robustness checks, and conclude with a discussion of this study's implications.

2. Literature Review

The current research builds on and extends the findings of previous research that studied reporting biases in online reviews. Previous research has shown that online reviews evolve systematically over time and sequence (Li and Hitt 2008, Wu and Huberman 2008, Godes and Silva 2012) and are impacted by platform interventions in which managers respond to online reviews posted (Proserpio and Zervas 2017, Chevalier et al. 2018, Wang and Chaudhry 2018). Moreover, Mayzlin et al. (2014) demonstrated that online reviews are biased not by customers but by companies themselves through the creation of promotional reviews.

This work is most relevant to research investigating one of the most prominent features of online reviews: Online reviews typically follow a J-shaped, or extreme, distribution of ratings, characterized by the tendency of extremely positive and extremely negative ratings to outnumber moderate ratings. Additionally, review ratings tend to heavily skew toward the highest possible rating, resulting in a rating distribution that resembles the shape of the letter J. Recently, Schoenmueller et al. (2019) used data from 25 online review platforms to document the high prevalence of extreme distribution in online reviews. However, the mere observation that rating distributions are extreme does not prove the existence of a reporting bias (e.g., *extremity bias*), because one plausible explanation for the extreme distribution is that the underlying distribution of experiences is itself extreme.

Researchers have explained the presence of extreme distribution in online reviews by proposing multiple alternative mechanisms. Although Brandes et al. (2019) hypothesize an attrition-based explanation (those with more extreme experiences have less probability of exiting the reviewer pool), the majority of previous work on the topic attributes extreme distributions to a utility-based mechanism (Hu et al. 2009, 2017; Schoenmueller et al. 2019). First introduced by Anderson (1998), this mechanism suggests that consumers receive greater utility from sharing extreme experiences. Using self-reported data on satisfaction and offline WOM activity, Anderson (1998) noted that the relationship between the two was U-shaped: Customers at the extreme ends of the satisfaction spectrum engaged in greater offline WOM, indicating that extremity bias exists in offline WOM. Moreover, the extremity bias was asymmetric: Extremely dissatisfied customers were more likely to engage in offline WOM than extremely satisfied ones. In line with Anderson (1998), I conjecture that customers with extreme experiences are more motivated to engage in unsolicited online WOM. Therefore, in this study I expect to find a U-shaped relationship between unsolicited online WOM and satisfaction. Furthermore, because negative emotions are more motivating than positive ones, I expect to replicate the asymmetric extremity bias reported in Anderson (1998) in the context of online WOM. The existence of such bias implies that online reviews from unsolicited customers may not be representative of the experiences of the underlying customer base.

To accelerate the creation of online WOM, both online review platforms and companies are increasingly soliciting reviews. Consistent with the actions of companies, research has shown that review solicitations are very effective in stimulating online review creation (Burtch et al. 2017, Klein et al. 2018). For instance, the leading travel website company TripAdvisor (2018) reports that hotels using its review solicitation service, Review Express, achieve an average of 28% increase in the number of TripAdvisor reviews. In a recent study, Askalidis et al. (2017) demonstrated that unsolicited reviews are more negative than solicited reviews. However, it is not a priori obvious whether certain customers would be more responsive to review solicitations. Solicitations provide an additional motivation to customers to participate in online WOM. For that reason, I expect solicitations to have the greatest impact in terms of increasing online WOM participation (compared with the baseline unsolicited WOM) on customers who lack the intrinsic motivation (e.g., moderate experiences) to post unsolicited reviews the most. For that reason, I suspect that solicitations will increase the engagement in online WOM by customers with moderate experiences compared with those with extreme experiences and therefore attenuate extremity bias. However, it remains an empirical question whether extremity bias is completely eliminated in solicited online WOM.

A second major source of potential bias in online WOM—the tendency of reviewers to modify the reporting of their experiences because of other reviews they have read—is a form of social influence I refer to as *conformity*. In their review article, Lerner and Tetlock (1999) report that individuals are inclined to conform especially when they must explain their opinions to an audience with known views. Because reviewers are expected to provide evidence that justifies their ratings, they may adjust their internal ratings based on seeing the average of existing online reviews while they are composing their own reviews. This results in conformity to the preexisting average online rating. Although many social dynamics have been shown to exist in online reviews (Li and Hitt 2008, Moe and Trusov 2011, Godes and Silva 2012, Moe and Schweidel 2012, Sridhar and Srinivasan 2012, Muchnik et al. 2013), internal rating adjustments by reviewers as a result of conformity have not been documented (mainly because of data limitations).

To establish the conformity effect, one must observe two ratings (the private internal rating of the reviewer and the eventual public online rating he posts). The current study attempts to provide evidence for conformity by comparing rating adjustment behaviors of two groups of reviewers: The first group had no exposure to the average of the preexisting online reviews before submitting their reviews, whereas the second group did. If conformity exists, the online reviews written by the second group should shift from the private rating toward the preexisting online average rating more than those written by the first group. Consequently, an important implication of the conformity effect is that it could potentially impact the representativeness of online review distributions.

3. Representativeness of Online Rating Distributions

In this section, I formally define the reporting bias in online ratings of unsolicited reviewers and discuss how it could be alleviated or exacerbated in online ratings of solicited reviewers. Additionally, I introduce a measure to evaluate the representativeness of online rating distributions. For clarity, in this section, I assume away the possibility that customers can post an online rating that differs from their private rating (i.e., ignore the effects of conformity). Later in my empirical analysis, I relax this assumption.

3.1. Reporting Bias in Unsolicited Online Ratings

Assume that there are N customers who purchased a given product and each one has a private satisfaction rating, s, which they can publicly post as an online rating on the review website. Let's denote the number of customers whose private ratings are s with N_s, that is, $N = \sum_s N_s$, and their original proportion in the entire customer population with $\alpha_s = N_s/N$. Then the proportion of reviewers with an online rating of s on the online review website is

$$\alpha_s^{unsol} = \frac{NR_s}{NR} = \frac{N_s \times p_s}{N \times p_{avg}} = \alpha_s \times \frac{p_s}{p_{avg}}, \tag{1}$$

where NR_s is the number of unsolicited online reviews posted with an online rating of s, NR is the total number of online ratings posted, p_s is the probability that a customer with a private rating of s posts an unsolicited online rating (i.e., NR_s/N_s), and p_{avg} is the average posting probability in the original customer population (i.e., $NR/N = \sum_s \alpha_s \times p_s$). The final online representation of customers whose private ratings are s is a function of three factors: (1) their original representation in the underlying customer population α_s; (2) the probability that they post an unsolicited online rating p_s; and (3) the average probability that a given customer in the original population posts an unsolicited online rating p_{avg}.

I define the reporting bias in unsolicited online ratings as

$$\frac{\alpha_s^{unsol}}{\alpha_s} = Reporting\ Bias_s^{unsol} = \frac{p_s}{p_{avg}}. \tag{2}$$

If $Reporting\ Bias_s^{unsol} > 1$, then customers with a rating of s will be overrepresented (i.e., $\alpha_s^{unsol} > \alpha_s$) on the online review website, and if $Reporting\ Bias_s^{unsol} < 1$, then they will be underrepresented. Simply put, any group will be underrepresented (overrepresented) if its members are less (more) likely to post an online rating than an average customer in the original customer population. I formally define *extremity bias* as the U-shaped relationship between $Reporting\ Bias_s$ and private ratings s, indicating that customers who are at the extreme ends of the private rating scale engage in online WOM disproportionately more than those with moderate private ratings.

The proportion of customers with an online rating of s on the review website will be equal to their proportion in the original population if and only if $Reporting\ Bias_s^{unsol} = 1$. Therefore, if this condition is satisfied, then the online ratings will be representative of the experiences of the original customers with a private rating of s. This relationship only applies to a specific private rating of s. To evaluate the representativeness of the unsolicited online rating distribution, I introduce the well-known Kullback-Leibler divergence (Kullback 1959) measure and label it as *Representativeness Score*. It is calculated as

$$\begin{aligned} Representativeness\ Score \\ &= D_{KL}\left(\alpha \| \alpha^{unsol}\right) = -\sum_s \alpha_s \times \log\left(\frac{\alpha_s^{unsol}}{\alpha_s}\right) \\ &= -\sum_s \alpha_s \times \log\left(Reporting\ Bias_s^{unsol}\right). \end{aligned} \tag{3}$$

Although the Kullback-Leibler divergence is not symmetric (i.e., $D_{KL}(P\|Q) \neq D_{KL}(Q\|P)$) and thus is not a true metric, it is a measure of how one probability distribution is different from a second reference probability distribution. Its appeal lies in both its popularity (Anderson and Simester 2014) and the fact that it captures the distance between two probability distributions using a single parameter value. Typically, $D_{KL}(P\|Q)$ is interpreted as the amount of information lost when Q is used to approximate P, which usually presents the true distribution of data. A representativeness score of 0 indicates that the two distributions are identical. Furthermore, the closer the representative score to 0, the more representative the rating distribution is of α. Consequently, I can determine whether a solicited online rating distribution is more representative than an unsolicited one by evaluating whether the unsolicited representativeness score defined in Equation (3) is greater than and statistically significantly different from its solicited counterpart.

3.2. Solicitation Effect

If the firm solicits all its customers to post an online rating on the website, the proportion of reviewers with a solicited online rating of s on the website is calculated as

$$\alpha_s^{sol} = \frac{NR_s^{sol}}{NR^{sol}} = \frac{N_s \times p_s^{sol}}{N \times p_{avg}^{sol}} = \alpha_s \times \frac{p_s^{sol}}{p_{avg}^{sol}}, \tag{4}$$

where $p_{avg}^{sol} = \sum_s \alpha_s \times p_s^{sol}$. The reporting bias in solicited ratings is correspondingly determined by $Reporting\ Bias_s^{sol} = p_s^{sol}/p_{avg}^{sol}$. Let $Lift_s = p_s^{sol}/p_s$ and $Lift_{avg} = p_{avg}^{sol}/p_{avg}$. I can then rewrite the equation in (4) as

$$\alpha_s^{sol} = \alpha_s \times \frac{p_s}{p_{avg}} \times \frac{\frac{p_s^{sol}}{p_s}}{\frac{p_{avg}^{sol}}{p_{avg}}} = \alpha_s \times Reporting\ Bias_s^{unsol} \times \frac{Lift_s}{Lift_{avg}}. \tag{5}$$

As a result, if there is any reporting bias in unsolicited online ratings, it is counteracted by the ratio of $Lift_s/Lift_{avg}$ in solicited online ratings. I define this ratio as the solicitation effect:

$$Solicitation\ Effect_s = \frac{Lift_s}{Lift_{avg}}, \tag{6}$$

where $Lift_s$ measures the magnitude by which solicitations increase the probability of review writing for customers with a private rating of s, and $Lift_{avg}$ is the weighted average of all $Lift_s$'s with a weight of $\frac{\alpha_s \times p_s}{\sum_s \alpha_s \times p_s}$ for each rating s. After soliciting all customers, the proportion of customers with a public rating of s on the review website will be equal to their original proportion in the entire population if and only if $Reporting\ Bias_s^{unsol} \times Solicitation\ Effect_s = 1$. Therefore, if $Solicitation\ Effect_s$ is equal to 1, then the reporting bias in unsolicited online ratings of s will be identical to the one in solicited online ratings of s. However, if $Reporting\ Bias_s^{unsol} < 1$ and $Solicitation\ Effect_s > 1$ ($Reporting\ Bias_s^{unsol} > 1$ and $Solicitation\ Effect_s < 1$), then underreporting (overreporting) bias in unsolicited online ratings of s will be more than that in solicited online ratings of s. Consequently, solicitations have the potential to either mitigate or exacerbate the extremity bias in unsolicited online ratings, and the solicitation effect defined in Equation (6) reveals the ultimate outcome.

4. Data

4.1. Company Background

Individual-level satisfaction survey data used in this study were provided by a major hotel group that wishes to remain anonymous. The hotel group outsources the administration of satisfaction surveys to a global market research company and uses surveys to track the performance of more than 4,000 hotels in its portfolio. Every month, the hotel group administers anywhere between 10 and 100 surveys for each hotel, depending on the hotel's size. The average number of surveys conducted per hotel was 40 surveys per month from January of 2012 to May of 2015. The data include all surveys collected during this timeframe.

The hotel group manages a loyalty program with three tiers. Typically, loyalty program members collect points for each qualifying stay and redeem them for free stays in the future. Members move between tiers depending on their purchasing activity and accumulated number of points. An individual who signs up for the loyalty program is assigned a membership ID that enables the company to track the individual's behavior over time at the hotel group. The data include stay behavior of all loyalty program members between January 2011 and May 2015. Unfortunately, it is not feasible for the company to track behaviors such as purchases, satisfaction surveys, and online reviews of those customers who are not loyalty program members. For that reason, the data are restricted to observations from loyalty program members.[1] It is also difficult for the hotel group to reach out to nonmembers because their contact information is not readily available. This is also reflected by their low participation rate in surveys. Only 20% of all surveys filled out were from nonmembers, whereas the remaining 80% came from loyalty program members (57% from the basic tier 1, 12% from the higher tier 2, and 11% from tier 3). Satisfaction survey invitations are sent via emails, and the average response rate is approximately 2%.

4.2. Timing of Surveys

All customers used in the analysis completed a customer satisfaction survey. In the data, 89% of all email invitations for surveys were sent exactly two days after the guest's checkout. The next most probable two scenarios were sending the survey request three (6%) or four days (2%) following the guest's stay. All remaining survey requests were emailed within eight days of checkout. On average, guests took 4.3 days to complete their survey. More than 92% of collected surveys were completed within 13 days of receiving the email request.

4.3. Solicitation Procedure

In 2012, the hotel group enabled its guests to leave online reviews on its website. The majority (78%) of its hotels had at least one review on the hotel's website by December 2012. In an attempt to increase the number of reviews posted on its website, the hotel group in 2014 started soliciting randomly selected survey takers to post their hotel experience as an online review on completion of their survey. I label the act of asking survey takers to submit an online review on the hotel's website as *solicitation* and divide survey takers into two groups: (1) *solicited survey takers* (SST), survey takers who were solicited to write an online review; and (2) *unsolicited survey takers* (UST), customers who took the survey but were not solicited to write an online review. I define a *reviewer* as any survey taker (either solicited or unsolicited) who writes an online review on the hotel website. I refer to an SST who writes an online review following the solicitation procedure as *a solicited reviewer* (SR), whereas *a solicited nonreviewer* (SNR) is an SST who does not write an online review despite being solicited.

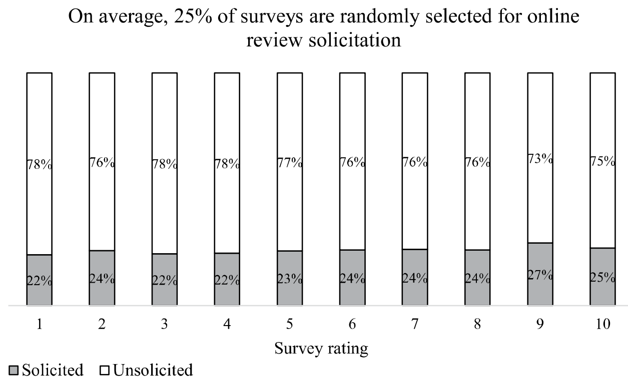
SSTs could post their experiences as online reviews following the steps illustrated in Online Appendix B.[2] To streamline the process of posting an online review through solicitation, the hotel group automatically generates online ratings based on SSTs' survey ratings. Customers report survey ratings on a scale from 1 to 10. Survey ratings are only observed by the firm and the researcher and are considered to be private ratings of survey takers (I therefore use the terms *survey ratings* and *private ratings* interchangeably in this research). When converting survey ratings into online ratings out of five, the hotel group simply divides the survey rating by two and rounds it up to the nearest integer value. For example, a survey rating of 5 (or 6) is automatically converted into an online rating of 3. The hotel group presents converted ratings to SSTs, and reviewers are allowed to change their online ratings if they disagree with the converted ratings provided by the hotel group, as illustrated in Step 2 in Online Appendix B.[3] Online reviews are made public on the hotel website (the terms *online ratings* and *public ratings* are also used interchangeably throughout this paper).

The rest of the reviewers are *unsolicited reviewers* (UR) who also received an email from the hotel and took the survey but were not solicited to write an online review at the conclusion of the survey. Nevertheless, these customers went to the hotel website and wrote a review on their own initiative (those URs that did not write a review I refer to as *unsolicited nonreviewers* or UNRs). In Online Appendix C, I present steps URs have to follow on the hotel's website to submit an online review.[4] URs and SRs are required to answer the same set of questions in the same order and are subject to the same requirements for submitting an online review (e.g., in both settings, review title and at least 50 characters of review text are required).

However, there is a stark difference between these two groups of reviewers: whether they were exposed to reviews already posted on the hotel website. To submit an online review, URs must first navigate to the website of the specific hotel for which they want to submit the review. The website displays summary statistics of previously posted reviews right next to the *write a review* button. These summary statistics include the average rating of previously posted reviews and the total number of reviews posted to date. Additionally, the same web page displays the ratings and full texts of the last eight reviews. After clicking the *write a review* button, the same summary statistics are displayed in the top left corner of the form, as shown in Online Appendix C. These summary statistics are always shown in the top left corner, even when URs scroll down to fill out the entire online review form. On the other hand, as documented in Online Appendix B, SRs are not exposed to preexisting online reviews at any point during the solicitation procedure. Although it cannot be assumed that SRs are completely unaware of preexisting online reviews (because they may have navigated to the hotel's website on their own or may have seen these reviews at other instances, such as at the time of booking), I argue that they are less likely to have seen preexisting online reviews at the time of posting because such information is not readily provided to them in the process of review creation.

One might suspect that the hotel group may be interested in soliciting online reviews primarily from extremely satisfied survey takers, but I do not observe that in the data. Figure 1 shows that, regardless of the survey rating, on average, 25% of surveys are selected for review solicitation. Additionally, given that the satisfaction survey administration is outsourced to a third-party market research company, it is harder for the hotel group to engage in such strategic behavior through review solicitation. Table 1 presents descriptive statistics for SSTs and USTs separately. These statistics confirm that differences between these two groups are also minimal in characteristics other than survey ratings, offering additional reassurance regarding the effectiveness of randomization.

Figure 1. Review Solicitation Is Random



4.4. Timeline of Events

I use two cohorts of survey takers in my analysis. The first cohort consists of those who completed a survey in 2013 just before the hotel group started its review solicitation initiative (Figure 2). Cohort 1 was only used to estimate *posting potential* models, which are described later. That is the only way their data are used. Cohort 2 consists of those who took surveys administered between January 2014 and May 2015. Survey takers in Cohort 2 were subject to random review solicitations. I draw on Cohort 2 survey takers to measure the impact of solicitations and conformity.

Cohort 2 consists of 389,789 survey takers that can be divided into the four groups (SR, SNR, UR, and UNR) described in Section 4.3. Figure 3 summarizes all four of them.

Researchers have found that some online reviews are submitted by customers with no record of ever purchasing the product they are reviewing (Anderson and Simester 2014). However, such reviews are less likely to exist in the current study because the hotel group confirms guests before they can submit a review by requiring a reservation identification code, guest name, and check-in and checkout dates. As a caveat, other self-promotion tactics, which I am unaware of, could exist on the company website. Data on customer satisfaction surveys, solicitation requests, purchase histories, and loyalty program membership were provided by the hotel group, and I scraped all online reviews (approximately 1.5 million) posted on the hotel group's website. I found that 82% of these

Table 1. Descriptive Statistics

Variable	No. of observations	SSTs average	Standard deviation	No. of observations	USTs Average	Standard deviation
Ln(room rate paid)	96,646	5.17	1.13	293,143	5.16	1.17
Ln(no. of nights)	96,646	1.04	0.45	293,143	1.03	0.45
Ln(no. of guests)	96,646	0.94	0.30	293,143	0.95	0.31
Ln(no. of months since joining the loyalty program)	96,646	1.57	1.39	293,143	1.34	1.35
Ln(no. of unique hotels stayed until the current stay)	96,646	1.04	0.57	293,143	0.97	0.52
Ln(no. of unique chains stayed until the current stay)	96,646	0.86	0.29	293,143	0.82	0.27
Ln(no. of previous stays at the current hotel)	96,646	0.27	0.60	293,143	0.24	0.57
Ln(days between survey sent and completed)	96,646	1.36	0.83	293,143	1.32	0.85
Posting potential (all satisfaction levels)	96,646	0.0124	0.0063	293,143	0.0112	0.0059
Posting potential (satisfaction levels less than 6/10)	7,099	0.0122	0.0063	23,999	0.0112	0.0059
Percentage of members in tier 1	96,646	88%		293,143	85%	
Percentage of members in tier 2	96,646	7%		293,143	8%	
Percentage of members in tier 3	96,646	5%		293,143	7%	

online reviews were posted by loyalty program members. I observed that 50% of all reviewers on the hotel group's website belonged to membership tier 1, 14% to tier 2, and 17% to tier 3. These statistics are very similar to their survey participation rate. I constructed my comprehensive data set by matching scraped online reviews to company-provided surveys using unique combinations of membership ID and reservation identification codes. I excluded from my analysis all survey takers whose membership enrollment date is unknown or whose stay was paid by their employer.

4.5. Model-Free Evidence

In this section, I provide model-free evidence on how survey ratings and review solicitations influence survey takers' probability of writing an online review and their online ratings. I present survey ratings across different types of survey takers in Figure 4. Given random solicitation, the distribution of survey ratings across SSTs and USTs is almost identical. However, I observe that survey ratings of both SRs and SNRs (URs and UNRs) differ considerably from their original SST (UST) customer base. For instance, the proportion of highly dissatisfied customers (those with a survey rating of less than or equal to 4 out of 10) is much higher in URs (10%) than in USTs (5%). Similarly, the proportion of extremely satisfied customers (those with a survey rating of 10 out of 10) is much higher in SRs (43%) than in SSTs (35%). These observations suggest that the online review writing probability of highly dissatisfied USTs is much higher than the average online review writing probability across all USTs (i.e., $p_1, p_2, p_3, p_4 > p_{avg}$) and that the solicitation acceptance rate of highly satisfied SSTs is higher than the average solicitation acceptance rate across all SSTs (i.e., $p_{10}^{sol} > p_{avg}^{sol}$). These probabilities are presented in Figure 5. Regardless of solicitation, a U-shaped relationship clearly exists between survey ratings and online WOM activity.

Figure 2. Timeline of Events

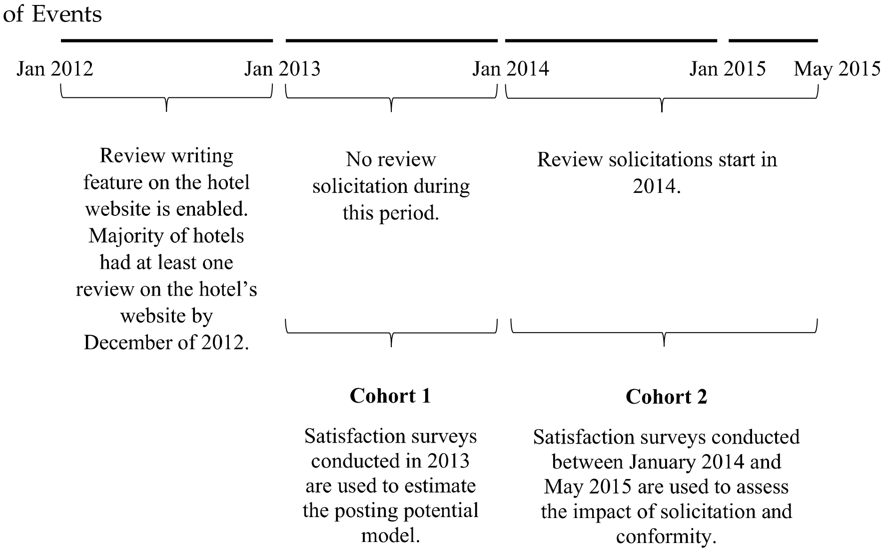
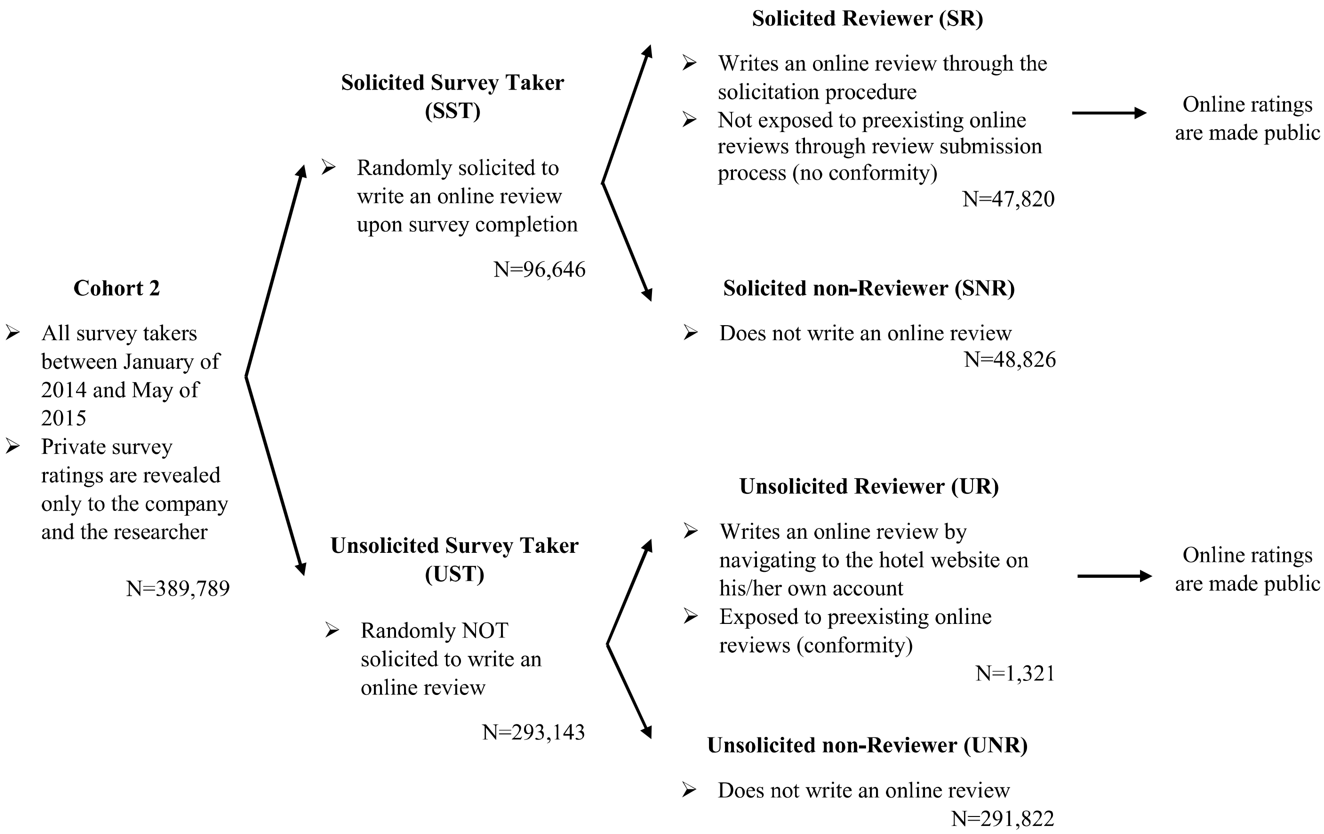


Figure 3. Different Types of Reviewers



Next, I document differences in rating adjustments made by SRs and URs. Figure 6 shows the percentage of URs and SRs at each rating level who downgraded, upgraded, or maintained their survey rating when posting it online. By visual inspection, it is clear that URs adjusted their ratings in both directions much more often than SRs. More specifically, Figure 6 demonstrates that, compared with URs, SRs are less likely to post an online rating different from their survey rating. For instance, consider reviewers whose survey rating is 2. Figure 6 shows that SRs submit an online rating of 1 (i.e., maintain their survey rating) 96% of the time, whereas URs only maintain the same rating 60% of the time. There is a 40% chance that an UR will submit an online rating of more than or equal to 2 (upgrade), whereas SRs only upgrade 4% of the time.

One obvious explanation for this behavioral difference between SRs and URs could be timing. Review solicitations are made right after survey completion, so for SRs the time between survey and online review is very short. On the other hand, 59% of URs carried out these two actions on separate days. Conceivably, URs could misremember their survey ratings and this could potentially explain their rating behavior observed in Figure 6. Unfortunately, I do not observe the exact hour when a survey is completed, or an online review is written. Instead, I have a record of the date reviewers carried out these two actions. Given this available information, I reproduced Figure 6 using observations from URs who filled out their survey and online review on the same day. Figure 7 illustrates that timing only slightly contributes to their rating adjustments reported in Figure 6. Next, I provide details of the empirical analysis implemented.

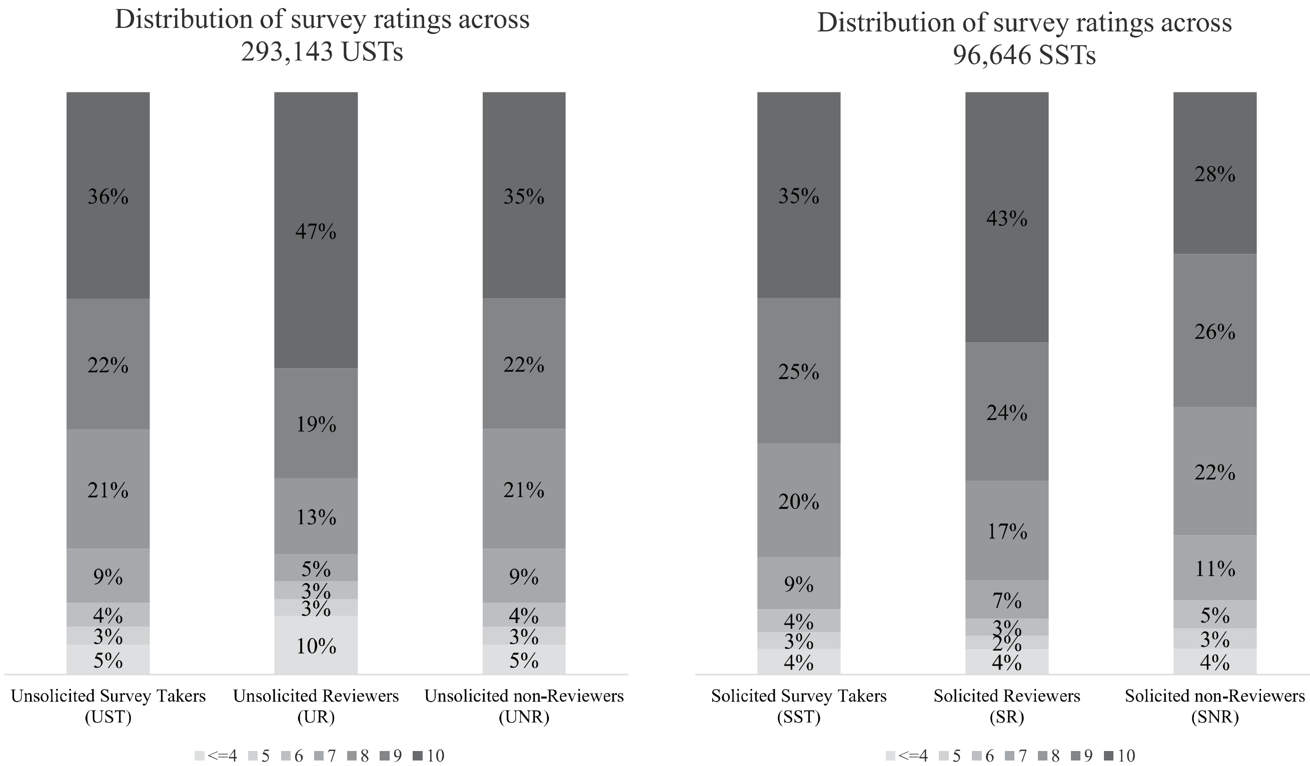
5. Analysis and Results

5.1. Model Specification

Consistent with my research focus, I examined two outcome variables of interest: (1) online review writing incidence, a binary outcome variable that indicates whether a survey taker writes an online review on the hotel website, and (2) rating adjustment, an ordinal outcome variable that measures the internal adjustment that a survey taker makes to a survey rating while posting a public online rating (and that is only observed conditional on writing an online review). I model online review writing incidence (binary outcome) and rating adjustment (ordinal outcome) as two separate but related processes using a Heckman model as follows.

5.1.1. Online Review Writing Incidence Model. The binary outcome variable $p_{iht} = 1$ indicates that a survey taker writes an online review on the hotel website for

Figure 4. Survey Ratings Across Different Types of Reviewers



their stay at hotel h at time t. I model this variable using a probit model as follows:

$$p_{iht} = 1(Z_{iht}\beta + v_{iht} > 0), \tag{7}$$

where Z_{iht} consists of the covariates used to model the online review writing process, $1(\cdot)$ is an indicator function, and v_{iht} is an idiosyncratic error term following a standard normal distribution with a mean of 0.

Figure 5 demonstrates that the relationship between survey rating and probability of writing an online review is nonlinear. To capture this potentially asymmetric U-shaped relationship, I include a dummy variable for each level of survey rating, s_{iht}, as covariates in Z_{iht}. Furthermore, given my interest in understanding the causal impact of solicitations on online review writing behavior, I also include solicitation dummy variable in Z_{iht}. I allow the impact of solicitations to be different for each survey rating by including interaction terms between survey rating dummy variables and the solicitation dummy variable in Z_{iht}.

Although solicitations are random, I recognize that I am working with field data and should control for as many observed factors (that could impact online review writing incidence) as possible in a parsimonious way. To this end, I construct a variable called *posting potential* and include it in Z_{iht}. I create the posting potential variable by estimating a separate probit model on Cohort 1, where the dependent variable is

Figure 5. Relationship Between Survey Rating and Probability of Writing an Online Review

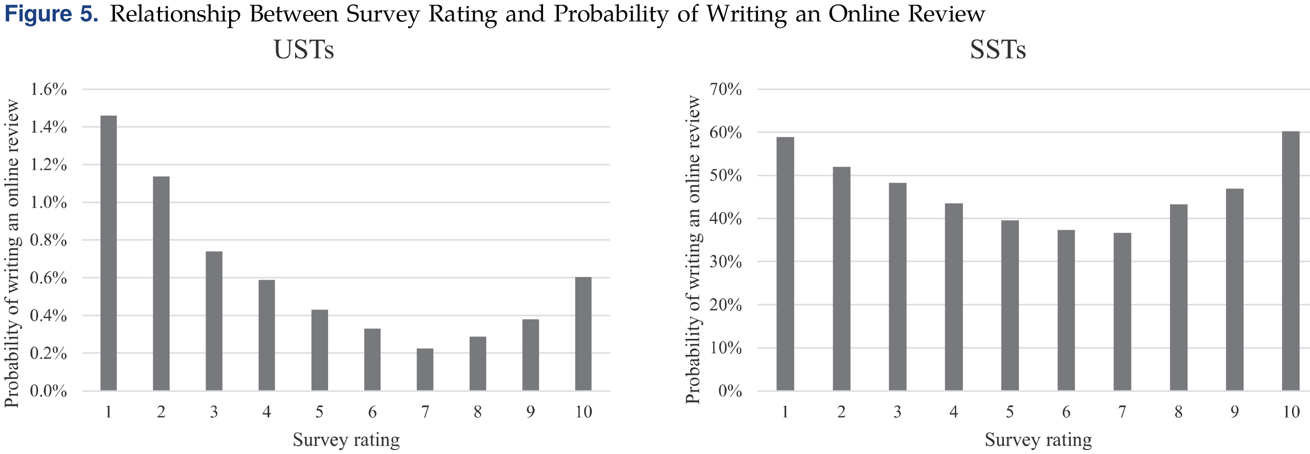
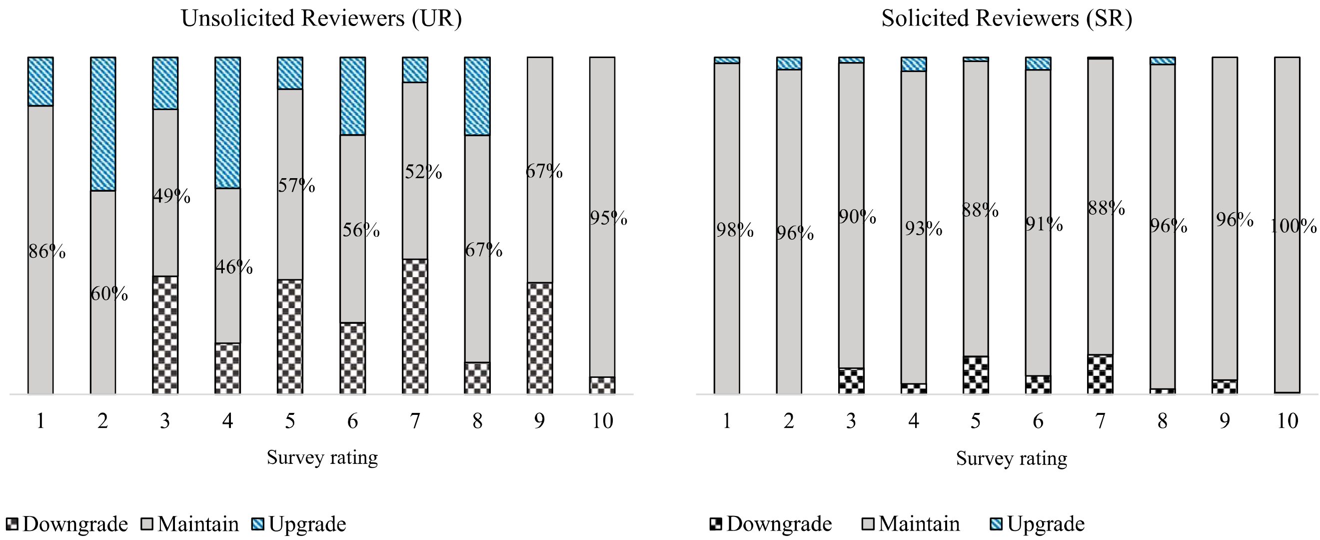


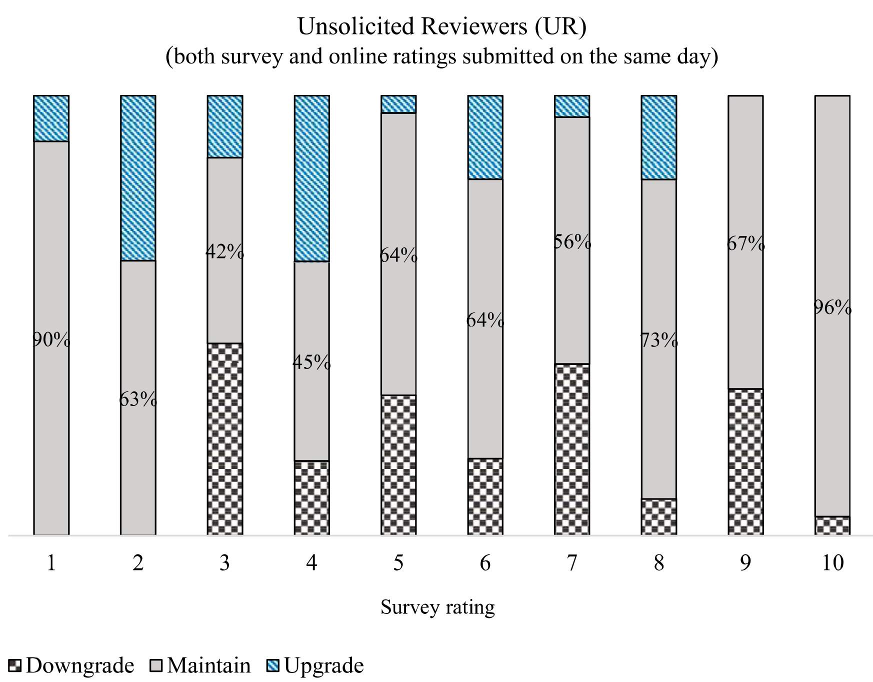
Figure 6. (Color online) Rating Adjustments



Notes. URs are more likely to post an online rating that differs from their survey rating. On the other hand, SRs submit online ratings that are very similar to their survey ratings. For instance, conditional on a survey rating of 6, SRs submit an online rating of 3 (maintain) 91% of the time, whereas the same probability is 56% for URs. There is a 23% chance that URs will post an online rating of 4 (upgrade), whereas the same is observed only 4% of the time for SRs.

whether Cohort 1 survey takers write an online review, and the independent variables are those I expected would predict whether a hotel guest would write an online review aside from my main independent variables of interest. These variables include guest characteristics such as gender and loyalty program membership, hotel characteristics such as chain scale, and stay characteristics such as nightly price paid (see Online Appendix D for a complete list of independent variables and the estimation results). Because these same variables are available for Cohort 2, I use Cohort 1 parameter estimates to predict the likelihood that the survey taker would write an online review in the absence of a review solicitation. This prediction, labeled *posting potential*, is included as an additional covariate in the Cohort 2 probit model. I do

Figure 7. (Color online) URs Who Filled Out Their Survey and Online Review on the Same Day



Notes. Timing only slightly explains rating adjustments observed in Figure 6.

not expect the inclusion of the posting potential variable to change any of the parameter estimates of interest for the online review writing incidence model (Montaguti et al. 2016), but I report results with and without it to be thorough.[5]

5.1.2. Rating Adjustment Model. The ordinal outcome variable, a_{iht}, captures the difference between s_{iht}, survey rating by reviewer i for hotel h at time t, and o_{iht}, the online rating posted by the same guest on the hotel h's website. More specifically, for a given reviewer i following their stay at hotel h at time t,

$$a_{iht} = \begin{cases} -1 & \text{if } o_{iht} - \dfrac{s_{iht}}{2} < 0 \\ 0 & \text{if } o_{iht} - \dfrac{s_{iht}}{2} = 0, \\ 1 & \text{if } o_{iht} - \dfrac{s_{iht}}{2} > 0 \end{cases} \tag{8}$$

where $s_{iht} \epsilon \{2,4,6,8,10\}$ is the survey rating reported privately to the hotel group (on a scale from 1 to 10) and $o_{iht} \epsilon \{1,2,3,4,5\}$ is the online rating posted publicly on the hotel website (on a scale from 1 to 5).[6] Conditional on online review writing incidence, a negative (positive) a_{iht} value signifies that the reviewer made a downward (upward) adjustment to their survey rating while posting an online rating. Odd survey ratings are excluded from Heckman model estimation because of the automatic rounding employed by the firm. The reasoning is as follows: consider a SR and an UR whose survey rating is 6.8 out of 10. They would both report a survey rating of 7 out of 10. Subsequently, the SR's online rating would automatically be generated as 4 out of 5 (because of the automatic rounding up rule), which he is very likely to post out of inertia. However, the UR would mentally convert her experience to 3.4 out of 5 and most probably post an online rating of 3 out of 5. This example illustrates the potential bias that an automatic rounding rule introduces in rating adjustments and how it could skew a_{iht} values upward for SRs. For this reason, I only rely on observations with even survey ratings because the rounding problem described is only present for odd numbers.

If a_{iht} takes the value of zero, the reviewer maintains their survey rating as their online rating. Therefore, a nonzero value of a_{iht} indicates a reporting bias, which I am capturing by the rating adjustment model. I model rating adjustments using an ordered probit specification as follows:

$$\Pr(a_{iht} = k | p_{iht} = 1) = \Pr\big(\mu_{k+1} < X_{iht}\beta + u_{iht} \leq \mu_{k+2}\big), \quad k \in \{-1, 0, 1\}, \tag{9}$$

where $p_{iht} = 1$ indicates that an online review is posted following the hotel stay (and $p_{iht} = 0$ otherwise), X_{iht} is a vector of covariates, and u_{iht} is the random error following a standard normal distribution with a mean of 0. Error terms (u_{iht}, v_{iht}) follow a bivariate normal distribution with mean zero and variance matrix $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. Consequently, the probability that an adjustment of k is made can be represented by

$$\Pr(a_{iht} = k | p_{iht} = 1) = \begin{cases} \Phi\big(\mu_0 - X_{iht}\beta\big), & k = -1 \\ \Phi\big(\mu_1 - X_{iht}\beta\big) - \Phi\big(\mu_0 - X_{iht}\beta\big), & k = 0, \\ 1 - \Phi\big(\mu_1 - X_{iht}\beta\big), & k = 1 \end{cases} \tag{10}$$

where μ are cutoff points for the ordered probit model and $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. I allow for two factors to influence a reviewer's rating adjustment decision: (1) solicitations and (2) conformity. To capture conformity, I create an independent variable, which measures the difference between a reviewer's survey rating and the hotel's preexisting average online rating on the hotel website on the day that she writes her review. I label this variable *Deviation from others* and calculate it as

$$\textit{Deviation from others}_{iht} = \frac{s_{iht}}{2} - \textit{PreexistingAvgOnlineRating}_{iht}. \tag{11}$$

I would like to emphasize the difference between the ordinal dependent variable (rating adjustment) and the independent variable (deviation from others). The former measures the internal rating adjustment that a reviewer makes to its private rating while posting it publicly, whereas the latter quantifies how much a reviewer's private rating differs from the already-existing average online rating of others on the hotel website.[7] A negative (positive) deviation from others indicates that the reviewer's personal experience at the hotel is worse (better) than an average experience reported on the hotel website. I estimated model parameters using maximum likelihood approach.

To test for the existence of conformity, I adopt a *difference in differences* type specification and include the solicitation dummy variable, the deviation from others variable, and their *interaction* in X_{iht}. With this specification, the evidence for the existence of conformity reduces to testing the statistical significance of the interaction term. If the impact of deviation from others is different between URs and SRs, this provides strong evidence for conformity. This identification strategy is justified for several reasons.

First, the likelihood of exposure to online reviews of others is different for SRs and URs. On the one hand, an UR is definitely exposed to online reviews of others because she has to navigate to the company website and click on a button that is right next to others' reviews to submit her online review. On the other hand, a SR is less likely to be exposed to online reviews of others because he submits his online review through the automated process, which omits any information about the online reviews of others. I cannot completely rule out that he did not check others' ratings by navigating to the company website on his own account, but clearly this action requires additional effort, and therefore a portion of SRs may not go to any lengths to find out about others' ratings. However, a portion of them may do so, but this possibility only makes it harder for me to find a statistically significant interaction term.

Second, solicitation was carried out at random. Random assignment ensures that there are no differences between USTs and SSTs, and both URs and SRs self-select into writing a review from the unsolicited and solicited survey taker groups, respectively. Additionally, the effort required to write an online review is similar for both groups, because they are asked the same set of questions in the same order and are subject to the same requirements while creating their online review. Last, I control for any remaining self-selection differences that may exist between URs and SRs by specifically modeling online review incidence, and the Heckman type model allows me to control for selection on unobservable characteristics because I allow for error terms (u_{iht}, v_{iht}) to be correlated.

5.2. Relationship Between Private Ratings and Probability of Writing an Online Review

I report the probit model parameter estimates in Table 2 and illustrate the relationship between survey ratings and online review incidence in Figure 8. I find that online review posting behavior of both SSTs and USTs exhibits extremity bias. This contributes to the overrepresentation of extreme experiences in both solicited and unsolicited online ratings. For instance, the estimated reporting bias for a survey rating of 10 in the SST group is 1.2, which indicates that the proportion of private ratings of 5 in SRs is 20% more than the proportion of private ratings of 10 in the SST group. This finding is consistent with data patterns observed in Figure 4, which shows that the proportion of SSTs with a survey rating of 10 is 35% in private ratings, whereas the proportion of SRs with the same survey rating is 43% in online ratings, indicating an approximately 20% increase in proportion.

Although both USTs and SSTs display extremity bias, there are stark differences in their reporting biases. First, the extremity bias is asymmetric in the UST group. Consequently, the overrepresentation of extremely dissatisfactory experiences in online reviews by URs is higher than the overrepresentation of dissatisfied SRs in online reviews. Second, the results show that the extremity bias is stronger among USTs. This is also reflected in both group's representativeness scores. The representativeness score for the unsolicited private rating distribution (0.078) is significantly different from its solicited counterpart (0.016; $p < 0.001$), meaning that the private rating distribution of SRs is

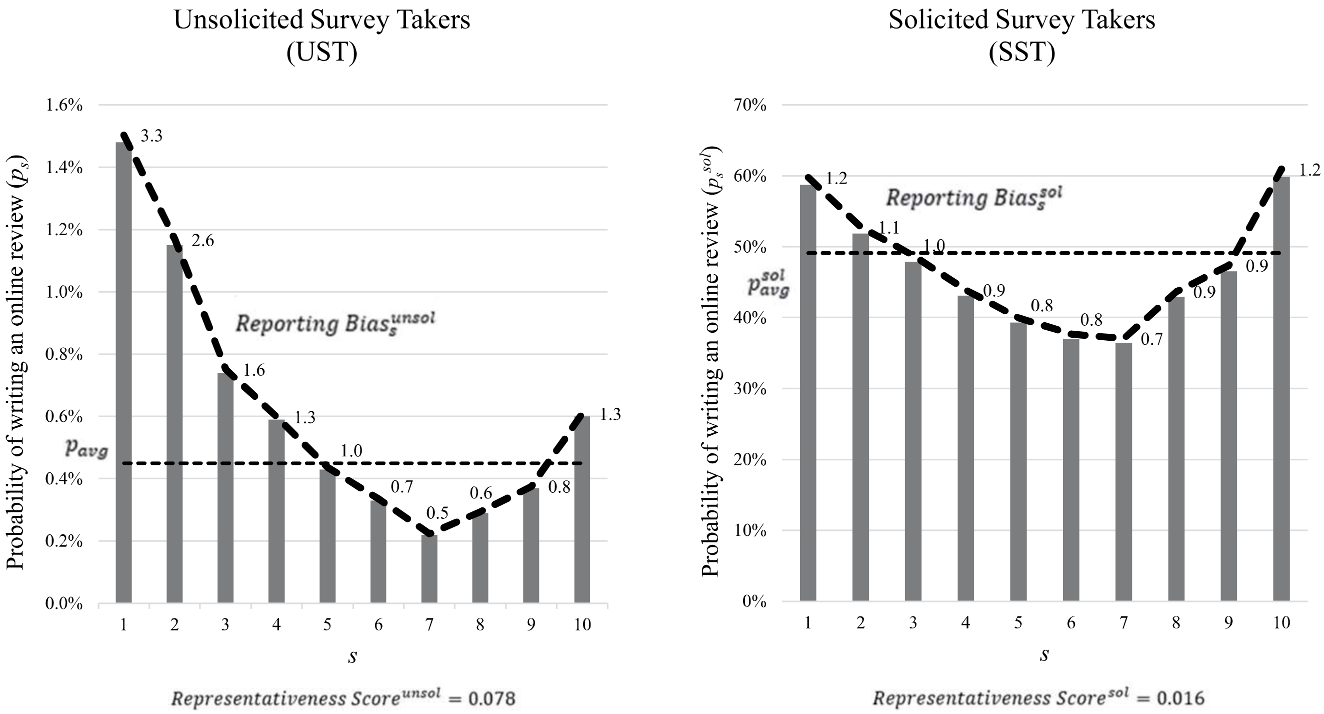
Table 2. Impact of Survey Rating and Solicitation on Probability of Writing an Online Review

Variables	Probit Model 1	Probit Model 2
I(s = 1)	0.66***	0.66***
	(0.07)	(0.07)
I(s = 2)	0.56***	0.56***
	(0.08)	(0.08)
I(s = 3)	0.40***	0.40***
	(0.07)	(0.07)
I(s = 4)	0.32***	0.32***
	(0.08)	(0.08)
I(s = 5)	0.21***	0.21***
	(0.07)	(0.07)
I(s = 6)	0.13*	0.13*
	(0.07)	(0.07)
I(s = 8)	0.08*	0.08*
	(0.05)	(0.05)
I(s = 9)	0.17***	0.17***
	(0.05)	(0.05)
I(s = 10)	0.33***	0.33***
	(0.04)	(0.04)
Solicited	2.50***	2.50***
	(0.04)	(0.04)
I(s = 1) × Solicited	−0.10	−0.10
	(0.09)	(0.09)
I(s = 2) × Solicited	−0.17*	−0.17*
	(0.10)	(0.10)
I(s = 3) × Solicited	−0.11	−0.11
	(0.08)	(0.08)
I(s = 4) × Solicited	−0.15*	−0.14*
	(0.08)	(0.08)
I(s = 5) × Solicited	−0.14*	−0.14*
	(0.07)	(0.07)
I(s = 6) × Solicited	−0.11	−0.11
	(0.07)	(0.07)
I(s = 8) × Solicited	0.09*	0.09*
	(0.05)	(0.05)
I(s = 9) × Solicited	0.09*	0.09*
	(0.05)	(0.05)
I(s = 10) × Solicited	0.27***	0.27***
	(0.05)	(0.05)
Posting Potential		9.74***
		(0.60)
Constant	−2.84***	−2.96***
	(0.04)	(0.04)
Number of observations	389,789	389,789
Log-likelihood	−73,878	−73,744

Notes. I chose survey rating of 7/10 to be the baseline category based on Figure 5, which shows that these customers are the least likely to write an unsolicited online review.

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Figure 8. Impact of Private (Survey) Ratings on Probability of Writing an Online Review



Notes. An estimated reporting bias of 1.3 indicates that the group's proportion in online public ratings will be 30% more than their original proportion in the customer base. Similarly, an estimated reporting bias of 0.5 indicates that the group's proportion in online public ratings will be half of their original proportion in the customer base.

more representative of the underlying customer experiences than that of URs. Next, I discuss why the extremity bias is less severe in solicited online ratings than unsolicited ones.

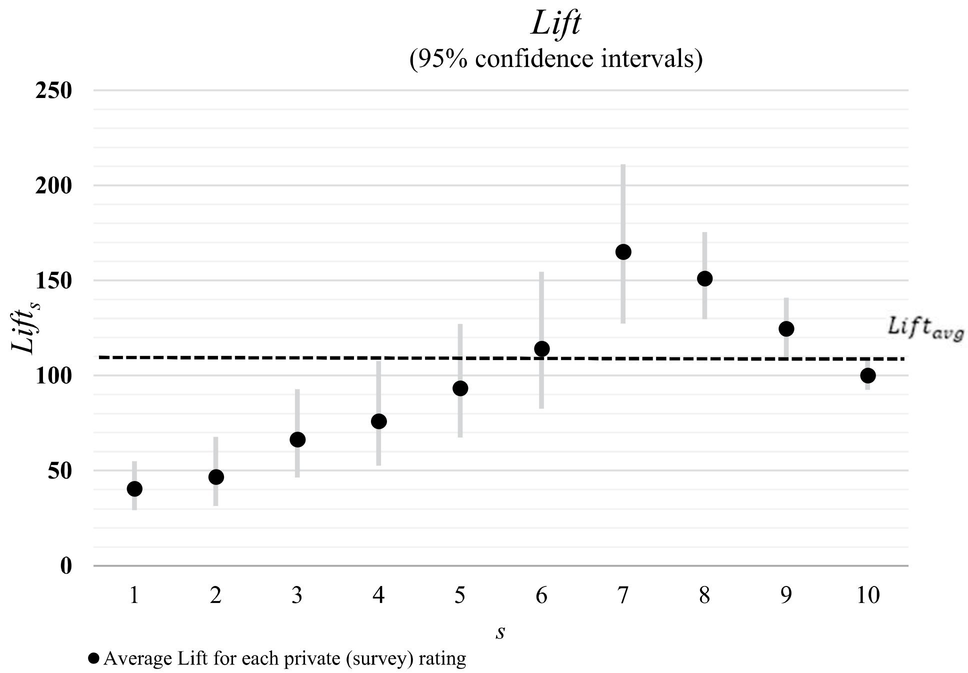
5.3. Impact of Solicitation on Probability of Writing an Online Review

The probabilities in Figure 8 demonstrate that solicitations significantly increase online WOM creation. I am certain that there were no incentives tied to an online review solicitation. However, I acknowledge that the magnitude of the solicitation effect seems rather large, especially compared with what is reported in previous work (Burtch et al. 2017). The well-known foot-in-the-door technique provides an explanation for the observed high solicitation acceptance rate. The seminal investigation of Freedman and Fraser (1966) of the foot-in-the-door technique demonstrated that individuals are more likely to comply with a large request after responding affirmatively to a smaller request. More specifically, they reported that 52.8% of individuals complied with a large request after completing a smaller request. This compliance rate is consistent with the solicitation acceptance rate I report.[8] In my setup, SSTs first agreed to fill out a survey from the company, and it is plausible that their basic desire for consistency drove them to greatly comply with the subsequent online review solicitation (Cialdini and Goldstein 2004).

Solicitation has an impact on review writing, but this impact may not be even across the satisfaction spectrum. To show the solicitation effect, defined in Equation (6), I created Figure 9 to illustrate the 95% confidence intervals for each $Lift_s$ and the $Lift_{avg}$. It demonstrates that the average solicitation effect is highest for moderate experiences with survey ratings of 7 and 8. Based on Figure 8, their reporting bias in unsolicited online WOM is 0.5 and 0.6, respectively. These customers are highly unlikely to engage in unsolicited online WOM activity despite comprising 30% of the original customer base and are therefore underrepresented in private ratings of URs. However, both $Lift_7$ and $Lift_8$ are statistically significantly different from $Lift_{avg}$ ($p < 0.01$). This suggests that the solicitation effects for these ratings are higher than 1, and their underrepresentation in private ratings of SRs is much less compared with URs.

Similarly, $Lift_1$, $Lift_2$, $Lift_3$, $Lift_4$, and $Lift_{10}$ are statistically different from $Lift_{avg}$ ($p < 0.02$). The solicitation effects for these ratings are lower than 1, suggesting that solicitations significantly reduce their overrepresentation in private ratings of SRs. In summary, the solicitation effect counteracts the extremity bias documented in private ratings of URs, greatly improving the representativeness of the private ratings of SRs.

Figure 9. Solicitation Effect on Probability of Writing an Online Review



Notes. Solicitation effect increases online WOM participation of those customers with moderate experiences more than those with extreme experiences, attenuating the extremity bias. I refer readers who are interested in finding out which $Lift_s$'s are statistically significantly different from each other to Table F1 in Appendix F.

5.4. Impact of Solicitation on Rating Adjustments

Table 3 reports parameter estimates of the Heckman model. I focus on the impact of solicitations on rating adjustments because results pertaining to review writing incidence are identical to what was discussed previously. The coefficient estimate for the solicited dummy variable measures the marginal effect of solicitations when (the customer's private survey rating) deviation from (the average online rating of reviews posted by) others is zero. I report the 95% confidence intervals of this effect in Figure 10, which shows that the effect is strongest and positive on the probability to maintain (survey ratings). It suggests that reviewers are slightly more likely to maintain the same survey rating as their online rating when they are solicited for a review. This is not surprising given that the company automatically converts survey ratings into corresponding online ratings, and reviewers may succumb to inertia instead of altering the rating. However, this effect is statistically insignificant. Therefore, I conclude that URs and SRs are equally likely to downgrade, maintain, or upgrade their survey rating when their experience is consistent with the preexisting average experience reported on the hotel website.

5.5. Impact of Conformity on Rating Adjustments

The marginal effect of solicitations on rating adjustment *is* statistically significant when deviation from others is nonzero. Figure 11 shows that the marginal effect of solicitation on the probability to upgrade is, on average, −13% when the survey rating is 6 and deviation from others is −1.7. This means that a SR with a survey rating of 6 is 13% less likely to upgrade than an UR with the same rating when the preexisting average rating on the hotel website is 4.7/5. Similarly, an extremely satisfied SR with a survey rating of 10 is, on average, 9% less likely to downgrade than an UR with the same experience when the preexisting average on the hotel website is 3.7/5.

As explained previously, to establish the existence of conformity, I am ultimately interested in the interaction effect. It is widely known that interaction effects in nonlinear models cannot be interpreted simply by looking at the estimated interaction coefficient. I derive the correct interaction effect and its standard deviation by using the method described in Ai and Norton (2003) and report them in Figure 12 (see Online Appendix G for derived formulas and figures depicting remaining interaction effects). The statistically significant interaction effects confirm the existence of conformity in rating adjustments; URs, who are more likely to be exposed to preexisting reviews than SRs, alter their ratings to a greater degree than SRs as their private ratings deviate from the average online ratings. On a related note, consistent with the finding in Section 5.4, marginal effects of solicitations

Table 3. Impact of Conformity and Solicitation on Rating Adjustment

	Coefficient estimate (standard error)	
	Heckman Model 1	Heckman Model 2
Online review writing incidence		
I(s = 2)	0.44***	0.44***
	(0.09)	(0.09)
I(s = 4)	0.20**	0.19**
	(0.08)	(0.08)
I(s = 8)	–0.05	–0.05
	(0.06)	(0.06)
I(s = 10)	0.20***	0.20***
	(0.05)	(0.05)
Solicited	2.39***	2.39***
	(0.06)	(0.06)
I(s = 2) × Solicited	–0.07	–0.06
	(0.10)	(0.10)
I(s = 4) × Solicited	–0.04	–0.04
	(0.09)	(0.09)
I(s = 8) × Solicited	0.20***	0.20***
	(0.06)	(0.06)
I(s = 10) × Solicited	0.38***	0.38***
	(0.06)	(0.06)
Posting Potential		8.79***
		(0.76)
Constant	–2.72***	–2.82***
	(0.05)	(0.05)
Rating adjustment (conditional on review writing incidence)		
Solicited	0.02	0.01
	(0.13)	(0.13)
Deviation from others	–0.77***	–0.77***
	(0.05)	(0.05)
Deviation from others × Solicited	0.68***	0.68***
	(0.05)	(0.05)
Correlation coefficient (ρ)	0.03	0.03
	(0.05)	(0.05)
Number of observations (surveys)	243,142	243,142
Number of selected observations (online reviews)	32,103	32,103
Log likelihood	–49,447	–49,380

Notes. Only surveys reporting an even rating are used for this estimation. Survey rating of 6/10 is the baseline category.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

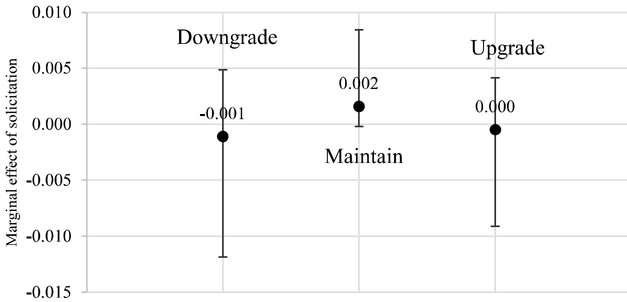
are primarily driven by the interaction effect between deviation from others and whether the review was solicited.

5.6. Robustness Checks

5.6.1. Previous Stay Histories. Previous stay histories raise two concerns for rating adjustment results: (1) customers may adjust their online ratings based on their previous experiences, and (2) frequent customers may misremember the hotel they are reviewing. If URs were disproportionately more likely to have previous stays than SRs, then these two factors could explain why they also tend to change their survey ratings more than SRs. Bccause I am able to track all survey takers starting from their first-ever hotel stay at the hotel group, I re-estimated the model using observations restricted to first-ever hotel stays. All previous findings are replicated by Model 2.1 presented in Table 4.

5.6.2. Impact of Online Reviews at the Time of Booking. The data allow me to alleviate concerns about the impact of online reviews at the time of booking because I can differentiate between online and offline bookings, although I cannot tell which website is used for online bookings. Thus, I repeated the same analysis using observations from offline bookings only. The idea is that, because these guests did not book online, they are much less likely to be impacted by online ratings at the time of their booking.

Figure 10. Marginal Effect of Solicitation on Rating Adjustment When Deviation from Others Is Zero (95% Confidence Interval)



These results are presented in Table 4 under Model 2.2. Furthermore, I observe the exact date that bookings are made. I performed an additional robustness check by using offline bookings that were made at least 10 days, the median number of days between booking and checkout, before the guest's checkout. Requiring the booking date to be at least 10 days before the checkout date means that it is harder for offline customers to recall online ratings if they were somehow exposed to them. Consequently, this condition is more stringent than the previous one. These results are presented in Table 4 under Model 2.3. Models 2.2 and 2.3 both confirm my previous findings on rating adjustments.

5.6.3. Management Response. I observe managers responding to online reviews submitted at the hotel website. The data come from a period after all hotels had responded to at least one online review. The average response rate to satisfied customers is 50%. Managers are more responsive to unsatisfied customers, responding to all online reviews with a rating of 3 or less (out of 5). Managers' responses to online reviews have been shown to impact subsequent reviews (Proserpio and Zervas 2017, Chevalier et al. 2018). For that reason, ideally the data should come from an environment where management response is not undertaken.

In light of findings from Chevalier et al. (2018), management response could provide an alternative explanation for highly dissatisfied customers' greater propensity to write an online review. If customers see that managers are listening, then they may be more likely to voice their frustration. Additionally, exposure to managers' responses is likely to be different between SRs and URs, and this difference may explain observed rating adjustments by URs. To show that the effects identified in this study cannot be attributed to managers' responses, I used the variation in observability of management response similar to Wang and Chaudhry (2018). Only the most recent eight reviews are displayed on the first page where potential reviewers have to navigate to post an online review for a given hotel. I thus restricted my analysis to observations from survey takers for whom there were no managerial responses to any of the last eight reviews displayed on the website at the time of their checkout. These survey takers were very unlikely to observe that managers had been responding to online reviews. I present these results in Table 4 under Model 2.4. These results replicate my previous findings on review

Figure 11. Marginal Effect of Solicitation on Rating Adjustment

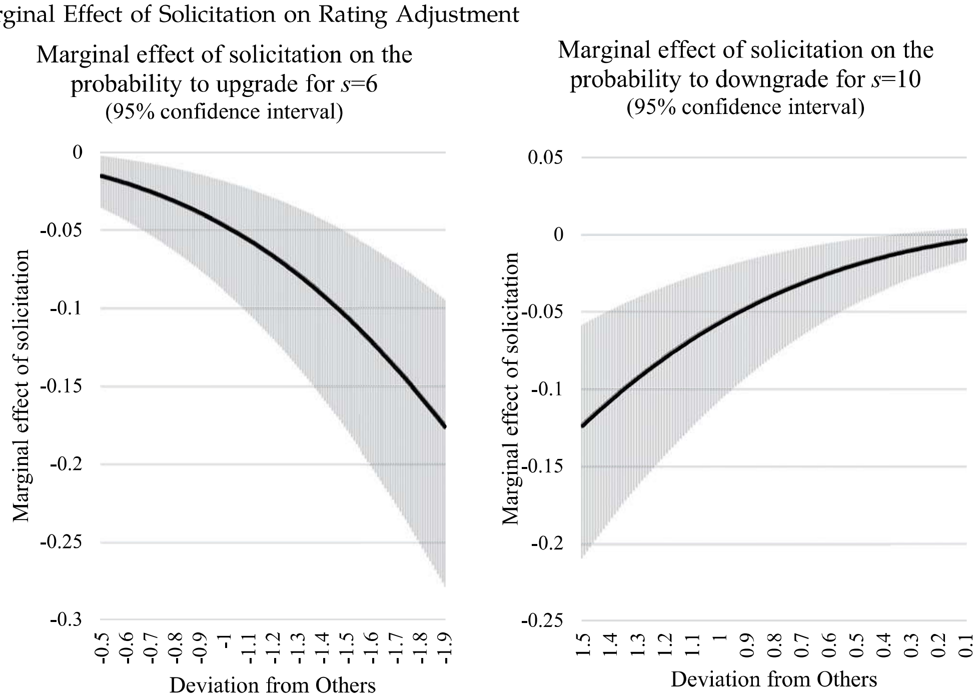
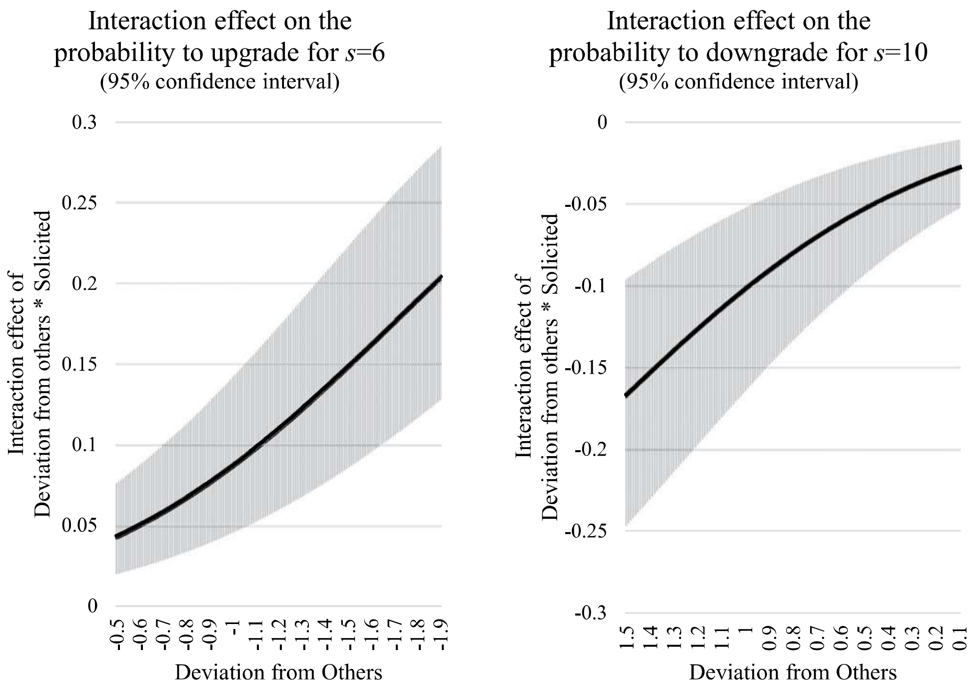


Figure 12. Evidence for Conformity on Rating Adjustment



writing incidence and rating adjustment. (See Table 5 for a summary of robustness checks for alternative explanations considered.)

5.7. Overall Impact of Solicitations on Public Ratings for a Typical Hotel

In this section, I describe how simulated public ratings of reviewers differ from their simulated private ratings and the private ratings of the original customer base. Figure 13 summarizes these differences. The distribution in Figure 13(a) illustrates the private ratings of a typical hotel's original customer base in the data; its average private rating is 4.41, and the variance of its private ratings is 0.69. I regard this distribution as the *original private rating distribution*. Next, using the parameter estimates in Table 3, I simulate the *expected private rating distributions* of reviewers under two scenarios: (1) all individuals in the original customer base are solicited to write an online review (Figure 13(b), top) and (2) none of them are solicited (Figure 13(b), bottom). In other words, these distributions show the private ratings of SRs and URs when the original customer base is solicited and unsolicited, respectively. The difference between the distribution in Figure 13(a) and distributions in Figure 13(b) captures the extremity bias phenomenon. Subsequently, I present in Figure 13(c) the *simulated public online rating distributions* for when all reviewers were either solicited (Figure 13(c), top) or not solicited (Figure 13(c), bottom).[9] I assess the impact of conformity by comparing the rating distributions in Figure 13(b) to those in Figure 13(c).

5.7.1. Impact on Representativeness. In line with my findings in Section 5.2, I see that the simulated private rating distribution of SRs is more representative of the original customer experiences than that of URs ($p < 0.001$). Surprisingly, conformity statistically significantly increases the representativeness of public online rating distribution of URs, decreasing the representativeness score in Equation (3) from 0.066 to 0.033 ($p < 0.001$). However, despite this increase, the public rating distribution of URs remains less representative than that of SRs ($p = 0.04$).

5.7.2. Impact on Valence and Variance of Ratings. In this section, I compare rating distributions based on two important online review metrics, valence and variance, that are shown to drive product sales (Rosario et al. 2016). First, I observe that the average simulated private (4.52) and public (4.51) ratings of SRs are statistically significantly different from the average private rating (4.41) of the original customer base ($p < 0.001$). The same holds for URs ($p < 0.02$). Although the extremity bias exists, it does not lead to decreased average ratings. On the contrary, it increases average ratings because most customers have extremely satisfactory experiences, and they are overrepresented in private (and public) ratings of reviewers. Because extremely satisfied customers are

Table 4. Robustness Checks

	Model 2.1	Model 2.2	Model 2.3	Model 2.4
Online review writing incidence				
I(s = 2)	0.51***	0.61***	0.62***	0.38**
	(0.14)	(0.14)	(0.18)	(0.19)
I(s = 4)	0.12	0.22	0.09	0.29*
	(0.14)	(0.14)	(0.20)	(0.16)
I(s = 8)	0.08	0.12	0.12	0.09
	(0.09)	(0.10)	(0.13)	(0.12)
I(s = 10)	0.34***	0.36***	0.36***	0.29***
	(0.09)	(0.10)	(0.12)	(0.11)
Solicited	2.59***	2.62***	2.53***	2.44***
	(0.09)	(0.10)	(0.13)	(0.11)
I(s = 2) × Solicited	−0.15	−0.30**	−0.30	0.05
	(0.16)	(0.15)	(0.20)	(0.21)
I(s = 4) × Solicited	−0.002	−0.11	0.05	−0.13
	(0.15)	(0.15)	(0.21)	(0.18)
I(s = 8) × Solicited	0.05	0.05	0.05	0.06
	(0.10)	(0.10)	(0.13)	(0.12)
I(s = 10) × Solicited	0.21**	0.21**	0.22**	0.30***
	(0.09)	(0.10)	(0.11)	(0.11)
Posting Potential	14.89***	5.84***	7.95***	7.38***
	(1.49)	(1.08)	(1.45)	(1.29)
Constant	−3.00***	−3.03***	−3.00***	−2.86***
	(0.09)	(0.09)	(0.12)	(0.11)
Rating adjustment (conditional on review writing incidence)				
Solicited	−0.10	0.17	0.20	−0.17
	(0.18)	(0.20)	(0.26)	(0.22)
Deviation from others	−0.83***	−0.75***	−0.80***	−0.88***
	(0.08)	(0.07)	(0.10)	(0.10)
Deviation from others × Solicited	0.69***	0.66***	0.66***	0.76***
	(0.08)	(0.08)	(0.11)	(0.11)
Correlation coefficient *(ρ)*	−0.02	0.01	−0.04	−0.03
	(0.07)	(0.08)	(0.11)	(0.09)
Number of observations	135,889	154,363	77,741	69,976
Number of selected observations	17,050	18,374	9,615	11,168
Log likelihood	−25,380	−27,960	−15,104	−16,858
Observations	First-time guests only	Offline bookings only	Offline booking and at least 10 days between booking and checkout	No management response

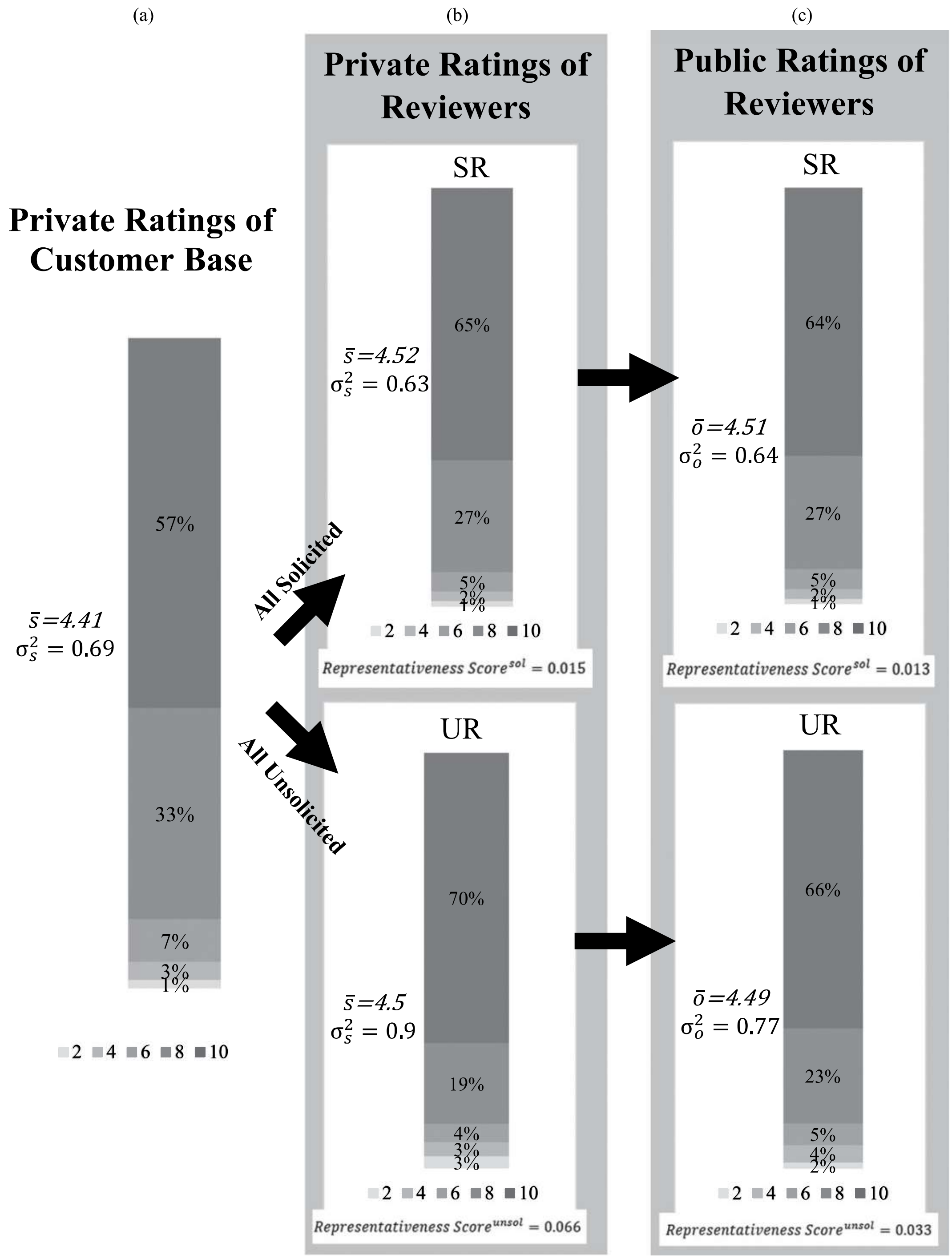
Notes. Only even survey ratings are used for this estimation. Survey rating of 6/10 is the baseline category.
***$p < 0.01$; ** $p < 0.05$; *$p < 0.1$.

overrepresented to a similar degree in both SRs and URs, I do not find any statistically significant difference between the average simulated ratings of SRs (4.52, 4.51) and URs (4.50, 4.49; $p > 0.45$). Both solicited and unsolicited averages are equally overstated relative to the average private rating of the original

Table 5. Robustness Checks for Alternative Explanations

Alternative explanations	Robustness check
Previous stay histories	Observations only from first ever hotel stay of the individual at the hotel group (Model 2.1 in Table 4)
The role of online reviews at the time of booking	1. Observations from offline bookings only (Model 2.2 in Table 4) 2. Observations from offline bookings where the number of days between booking and reviewing is at least 10 days (Model 2.3 in Table 4)
Management response	Observations from survey takers who were very unlikely to observe that managers were responding (Model 2.4 in Table 4)

Figure 13. Illustrating Bias in Public Online Ratings for a Typical Hotel



Notes. Panel (a) is the private rating distribution of the original customer base. Panel (b) is the simulated private rating distribution of solicited (top) and unsolicited (bottom) reviewers. Panel (c) is the simulated public rating distribution of solicited (top) and unsolicited (bottom) reviewers.

customer base. This is particularly interesting because Figure 13 demonstrates considerable differences between the simulated private (and public) rating distributions of SRs and URs.

Second, although I do not find statistically significant differences in average private or public ratings of SRs and URs, their rating distributions exhibit stark differences, as shown in Figure 13. These differences manifest themselves in variances of private and public ratings of two groups. The variance of simulated private ratings of SRs (0.63) is statistically significantly lower than that of URs (0.90; $p < 0.01$).

This difference is driven by the fact that extremity bias is more severe when customers are not solicited. As expected, the conformity effect reduces the variance of simulated public ratings of URs to 0.77. However, the variance of simulated public ratings of SRs (0.64) remains statistically significantly different from that of URs (p = 0.04). I conclude that the variance of solicited public ratings would be lower than its unsolicited counterpart for a typical hotel in the data if all its customers were solicited to write an online review.

5.8. Impact of Solicitations on Third-Party Review Platforms

In Section 5.3, I demonstrated that solicitations significantly increase online review writing behavior on the company website. A natural next question that follows the preceding analysis is *where does an additional online review induced by a solicitation come from*? There are two potential sources: (1) Additional online reviews are written by individuals who would not write any online review on any website in absence of solicitation, or (2) they are written by individuals who would leave an online review on another website but not on the company website. If additional reviews come from the second source, then solicitations could potentially shift online reviews from other review websites to the company website if individuals tend to leave only one review per experience. In this section, I delve deeper into the impact of solicitations on online review writing behavior at third-party review platforms.

TripAdvisor.com is the leading third-party review website for hotels. Unfortunately, I am not aware of any technology that would allow researchers to trace TripAdvisor behavior back to an individual customer. However, the experiment conducted by the hotel group creates an opportunity to investigate whether solicitations impact reviewing behavior on TripAdvisor.com by using hotel-level aggregate data. The hotel company first tested out solicitations by implementing them on a small scale. From January 1 to January 7, 2014, the hotel group randomly selected approximately 250 hotels within its portfolio to experiment with soliciting reviews during these seven days before the portfolio-wide rollout of solicitations was implemented on May 28, 2014. This natural experiment allowed me to construct two sets of hotels: (1) control hotels for which review solicitations started on May 28 and (2) treated hotels for which review solicitation experimentation occurred between January 1 and 7.

I analyze the number of online reviews posted on TripAdvisor.com to identify the causal effects of review solicitations on online review writing behavior on a third-party review platform. I scraped online reviews from TripAdvisor.com only for hotels that are located in the United States. Availability of online reviews only for U.S. hotels reduces the size of the treatment group to 151 hotels. The control group includes the remaining 2,140 hotels located in the United States. The dependent variable of my analysis is the total number of reviews posted on TripAdvisor.com within 120 days. Although the selection process was random, there are small and statistically insignificant baseline differences between treated and control hotels. This is not unusual given that data are at the aggregate level. Therefore, I use Difference-in-Differences (DID) analysis to account for statistically insignificant baseline differences. Figure 14 summarizes my research setup.

Because I am investigating whether the impact of solicitations differs depending on the review valence, I model the number of reviews for each five available ratings separately. Furthermore, because the dependent variable is a count variable and the Poisson model assumptions are violated—because mean and variance of the dependent variable (number of reviews posted) differs greatly in data—I use negative binomial regression (NB2) to model it using the following DID specification. The expected

Figure 14. Control vs. Treated Hotels and Before vs. After Periods

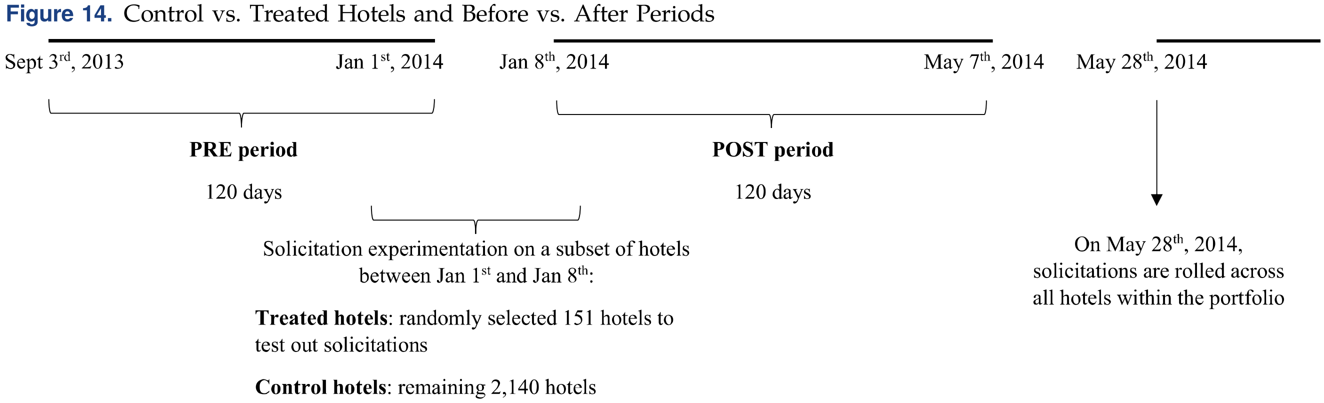
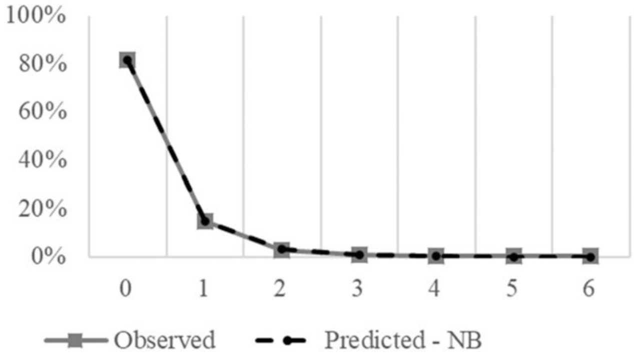


Figure 15. Expected and Observed Probabilities for the Number of One-Star Reviews



number of online reviews on TripAdvisor.com, y, is modeled as

$$E[y|x,\beta] = e^{(\beta_0+\beta_1\times Treated+\beta_2\times Post+\beta_3\times Treated\times Post)}. \quad (12)$$

The model parameters are estimated by maximizing the negative binomial log-likelihood function. I provide expected and observed probabilities for one-star reviews in Figure 15 (see Online Appendix F for expected and observed probabilities for all other star levels in Figure F1). NB2 model recovers observed data patterns well. I am interested in the interaction effect, which measures the *treatment effect on the treated*. I derive the correct interaction effect and its standard deviation using the method described in Ai and Norton (2003) (see Online Appendix F for derived formulas for NB2) and present my results in Table 6.

The results suggest that solicitations primarily shift one-star reviews from TripAdvisor to the company website. I do not find statistically significant effects for any other rating. Note that solicitations are launched to the entire hotel portfolio 120 days after the start of the experiment, and if the impact of solicitations on TripAdvisor takes longer than 120 days to materialize, then I may be missing some of the effects of solicitations. Figure 16 provides 95% confidence intervals for the percentage change in number of reviews for each rating. It shows that solicitations, on average, decrease the number of one-star reviews on the TripAdvisor website by 48%. This result suggests that online review solicitations for the company website could possibly increase the representativeness of online review distribution on a third-party review website if the asymmetric extremity bias also exists in such platforms.

One plausible explanation as to why the effect exists only for one-star reviews could be that a reviewer who writes a one-star online review is only motivated to write this review because of that particular one-time awful experience, whereas a reviewer who writes a five-star online review could be interested in reviewing for the sake of reviewing (e.g., as a hobby). In other words, one-star reviewers do not gain any pleasure from the act of writing an online review, whereas five-star reviewers enjoy writing online reviews. Once a one-star reviewer writes an review on the company website, they do not post additional reviews because they do not derive any intrinsic utility from posting them, whereas a five-star reviewer is more likely to post on multiple websites. I provide some evidence for this explanation in Table 7.

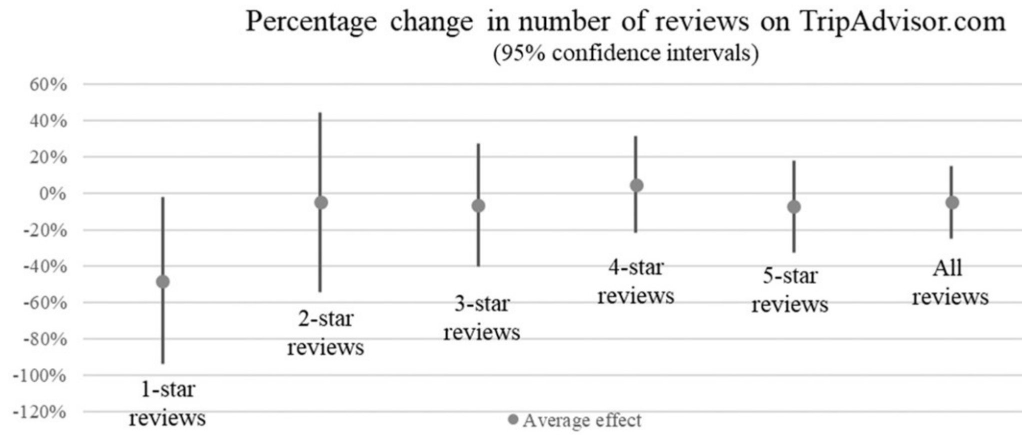
Table 7 provides summary statistics from more than 1.6 million online reviews from the TripAdvisor website. To make sure that membership timing cannot explain away the statistics provided in Table 7, I focus on reviewers who posted an online review on the TripAdvisor website at least 400 days after and at most 500 days after joining the website. Table 7 shows

Table 6. Impact of Solicitations on Third-Party Review Platforms

	All reviews	One-star reviews	Two-star reviews	Three-star reviews	Four-star reviews	Five-star reviews
Treated	0.05	0.29	0.01	0.01	0.03	0.06
	(0.08)	(0.19)	(0.18)	(0.13)	(0.10)	(0.10)
Post	−0.08***	0.12	−0.04	−0.10**	−0.11***	−0.08
	(0.03)	(0.07)	(0.07)	(0.05)	(0.04)	(0.04)
Treated × *Post*	−0.05	−0.60**	−0.05	−0.07	0.06	−0.08
	(0.10)	(0.29)	(0.26)	(0.18)	(0.14)	(0.14)
Constant	1.89***	−1.52***	−1.16***	−0.20***	0.74***	1.16***
	(0.02)	(0.05)	(0.05)	(0.03)	(0.03)	(0.03)
α	0.68***	1.30***	1.33***	1.08***	0.85***	1.02***
	(0.02)	(0.17)	(0.13)	(0.06)	(0.03)	(0.03)
Number of observations	4,582	4,582	4,582	4,582	4,582	4,582
Log likelihood	−13,295	−2,720	−3,268	−5,591	−8,706	−10,407
Corrected	−0.34	−0.14**	−0.02	−0.05	0.11	−0.24
Interaction term	(0.71)	(0.07)	(0.08)	(0.14)	(0.29)	(0.43)

Notes. In all models, the Likelihood-Ratio test of $\alpha = 0$ is rejected, indicating negative binomial model to be more appropriate than Poisson model.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Figure 16. Treatment Effect of Review Solicitations for Company Website on Posting Behavior at a Third-Party Review Platform



two pieces of supporting evidence. First, the average number of online reviews (with any rating) written by an individual who wrote at least one one-star review is 22, whereas an individual who wrote at least one five-star review, on average, writes 32 online reviews. This observation lends support to the argument that one-star reviewers may derive much less intrinsic utility from the act of writing an online review (and therefore, on average, write fewer online reviews) compared with two-, three-, four-, and five-star reviewers. Otherwise, on average, they would also be posting similar number of reviews. Second, consistent with the same argument, Table 7 shows that a higher proportion (7.7%) of one-star online reviews are the only review ever written by an individual. Thus, reviewers who leave one-star online reviews are less likely to be habitual reviewers and more likely to be uniquely motivated to review by a particularly negative experience. Nonetheless, it remains a topic for future research to pinpoint the exact mechanism for why solicitations primarily shift one-star online reviews away from the third-party website.

6. Discussion

In this paper, I studied the representativeness of online WOM, primarily focusing on how extremity bias, conformity, and online review solicitations impact it. The findings can be split into four components. First, I showed that extremity bias exists in unsolicited online WOM. Customers who are extremely satisfied or extremely dissatisfied are more likely to engage in unsolicited online WOM than those with moderate experiences. Extremity bias thus reduces the representativeness of unsolicited online WOM. This finding offers a justification for companies to *manipulate* their online WOM to generate a more representative set of reviews and suggests that not all manipulation should be seen as fraudulent, because some could be used as a means to increase representativeness of online WOM.

I documented a second set of findings that focuses on two mechanisms that significantly increase the representativeness of online review distributions: online review solicitations and conformity. First, although the extremity bias persists in solicited online WOM,

Table 7. Reviewing Behavior on TripAdvisor.com

Number of observations				
One-star	Two-star	Three-star	Four-star	Five-star
41,657 reviews	72,081 reviews	222,905 reviews	560,740 reviews	732,863 reviews
Percentage of reviewers who write only one review conditional on review rating				
One-star	Two-star	Three-star	Four-star	Five-star
7.7%	4.6%	2.1%	2.1%	3.3%
Average number of reviews per reviewer conditional on review rating				
One-star	Two-star	Three-star	Four-star	Five-star
22 reviews/reviewer	30 reviews/reviewer	44 reviews/reviewer	47 reviews/reviewer	32 reviews/reviewer

Notes. Of 41,657 one-star reviews observed, 7.7% of them (3,224 reviews) are posted by individuals who posted only once. Of 732,863 five-star reviews observed, 3.3% of them (24,210 reviews) are posted by individuals who posted only once. The average number of reviews (with any rating) posted by individuals who posted at least a one-star review is 22. Statistics are calculated using reviewers who posted a review on the TripAdvisor website at least 400 days after and at most 500 days after joining the website to make sure that membership timing cannot explain away observed data patterns.

I showed that it is greatly reduced relative to the extremity bias in unsolicited online WOM because the solicitation effect is greatest for moderate experiences and lowest for extreme experiences. Hence, random online review solicitations have the potential to de-bias online reviews to a certain extent. Second, the results indicate that conformity could counteract the extremity bias. Exposing prospective reviewers to preexisting online reviews could therefore yield a more accurate representation of customers' underlying experiences.

These findings have implications for review aggregation platforms that seek to accurately reflect underlying customer experiences. For instance, YouTube and Netflix switched their rating systems to thumbs up/down and like/dislike ratings, respectively, perhaps partially because of the extreme distribution of reviews seen in their previous five-point rating systems. Clearly, these binary rating systems (thumbs up/down or like/dislike) do not provide a platform for more moderate (i.e., neither like nor dislike) opinions, whereas reviewers with extreme opinions can still indicate their love (via like or thumbs up) or hate (via dislike or thumbs down) toward a product or service. Yet a *silent majority* of customers with moderate experiences exists, but they are not self-motivated enough to engage in unsolicited online WOM. This study shows that to collect a more representative set of opinions, review aggregation platforms could use review solicitations instead of abandoning five-point rating systems to reduce the reporting biases of such systems. Platforms should realize that the silent majority can be motivated to join online conversations and express their *middle of the road* opinions through solicitations and that exposing all reviewers to the ratings of others can further dampen the extremity bias.

Third, I calculated the impact of online review solicitations on two key online review metrics: valence and variance of online ratings. Average online ratings of both solicited and unsolicited reviews overstate the average customer experience. Moreover, the variance of online ratings would be significantly lower if all customers were solicited to write an online review (as opposed to not solicited). The latter finding has implications for sellers of high-quality products. Previous research establishes that a lower variance of ratings increases sales for products with higher average ratings (Sun 2011). Therefore, soliciting online reviews could be especially beneficial for companies offering high-quality products (presumably with higher average ratings) because the current findings show that soliciting reviews decreases the variance of online ratings. Hence, when promoting high-quality products, companies should consider online review solicitation strategies to manage the variance of their online ratings.

These results also have implications for customers. Customers should be mindful of the possibility that the average online rating inflates the average customer experience. Contrary to conventional wisdom, I find that this inflation is not greater for solicited reviews than for unsolicited ones. However, customers should be aware that solicited online reviews could present an inaccurate picture of product-mismatch risk because the variance of solicited online ratings is significantly reduced. Because variance serves as a proxy about the divergence of opinions regarding product quality, if variance is low, then customers perceive the mismatch risk to be low. Therefore, although the average rating is not affected by solicited reviews (compared with unsolicited reviews), they do provide an inaccurate sense of the extent of opinion divergence about product quality, thus downplaying the potential product-mismatch risk. This underrepresentation could be consequential for companies as well because if their products do not match customer expectations (based on online reviews), then actions taken by disappointed customers (e.g., returning the product, writing negative reviews) could hurt their profitability.

My final set of findings explains the impact of online review solicitations for the company website on review writing behavior on a third-party review platform. Solicitations, on average, decrease the number of one-star reviews on a third-party review platform by 48%. I do not find statistically significant effects for any other rating. This finding has important implications for companies that aspire to reduce the amount of negative online WOM for their products on a third-party review platform. These companies should seriously consider the tradeoff between the cost of implementing review solicitations on their own website and the benefit of decreased negative online WOM on a third-party review platform, which potentially has a greater reach.

Although observing customers' private ratings through satisfaction surveys is a strength of the current setup, it also poses a limitation in terms of the study's generalizability. All customers in this study completed a customer satisfaction survey and self-selected into doing so. Therefore, whether the effects identified here would replicate using data from non-survey takers remains a question for future research. Although it sounds impractical to measure customers' private ratings without some form of a survey, I hope that future research could overcome this challenge through clever research design.

The review solicitations in this study did not include any incentives. However, solicitations could involve financial (Stephen et al. 2012, Klein et al. 2018) or social (Chen et al. 2010, Burtch et al. 2017) incentives. I leave it to future research to determine whether different types of solicitations would have similar consequences. Future research could also examine whether solicitations could reduce many other biases identified in the literature (Berger 2014).

A final observation: it is plausible for solicited and unsolicited reviewers to have different motivations in mind while posting online reviews. For example, solicited reviewers could be motivated to help managers improve service, whereas unsolicited reviewers could be driven by their desires to help other customers. In this study, I do not differentiate between different types of motivations because these differences would be reflected in review text as opposed to review ratings. However, review text differences between solicited and unsolicited reviews could be an interesting avenue for future research.

Acknowledgments

The author thanks Anandhi Bharadwaj, Douglas Bowman, Sandeep Chandukala, Christilene Du Plessis, Ryan Hamilton, Balázs Kovács, Gaël Le Mens, Stephanie Lin, Giacomo Negro, Ernst C. Osinga, Marko Pitesa, Jagdish Sheth, Anand Swaminathan, Maciej Szymanowski, Necati Tereyağoğlu, Vilma Todri, Kapil Tuli, the editors, and three anonymous reviewers for insightful and thoughtful comments that significantly improved the paper.

Endnotes

[1] Online review rating and survey rating distributions of loyalty program members bear close resemblance to that of nonmembers. These distributions are provided in Online Appendix A.

[2] In the final step of the solicitation procedure, SSTs are allowed to opt out of posting their review. The data do not allow me to differentiate between SSTs who rejected the solicitation and those who opted out of submitting their review in the final step of the solicitation procedure. Both types of SSTs are recorded as not posting an online review in the data.

[3] I recognize that many SSTs may post these converted ratings out of inertia even though they are given the opportunity to change them; I take this point into consideration in my analysis.

[4] Guidelines provided to URs are exactly the same as SRs. Very minor differences observed (between guidelines seen in Online Appendices A and B) are because of the different timing of documentation.

[5] Moreover, results do not change when I directly include all factors as control variables in Z_{iht}. These results are available from the author upon request.

[6] Alternatively, I can set $a_{iht} = o_{iht} - \frac{s_{iht}}{2}$. This alternative specification results in nine distinct values ranging from −4 to 4 for the ordinal outcome variable, a_{iht}. Results from this alternative specification are identical and are presented in Online Appendix E, Table E2. For exposition purposes, I chose to include the more parsimonious specification in the paper. This decision was also based on the observation that reviewers typically upgrade or downgrade their survey ratings by 1 or maintain it. I include this transition matrix in Online Appendix E, Table E1, as well.

[7] For example, consider a reviewer whose survey rating is 6/10 and online rating is 4/5 for a hotel whose average rating on the hotel website is 4.5/5. In this instance, the deviation from others is $(\frac{6}{2} - 4.5) = -1.5$, and the ordinal dependent variable is +1 (since $4 - \frac{6}{2} > 0$). The reviewer's personal experience at the hotel was worse than the average experience reported on the hotel website, yet he upgrades his survey rating from 6/10 to 4/5 while posting his online rating.

[8] I argue that asking to post an online review is a bigger request than asking to fill out a satisfaction survey because online reviewers are required to provide review text (as well as a review title) and are expected to provide justifications for their online ratings. On the other hand, it is optional for survey takers to provide comments in the survey, and they do not have to defend their opinions publicly.

[9] The average online rating of a typical hotel in the data were 4.2/5 on December 31, 2013, right before review solicitations were implemented. Therefore, I use 4.2 as the preexisting average online rating in estimating public ratings of reviews.

References

Ai C, Norton EC (2003) Interaction terms in logit and probit models. *Econom. Lett.* 80(1):123–129.

Anderson EW (1998) Customer satisfaction and word of mouth. *J. Service Res.* 1(1):5–17.

Anderson ET, Simester DI (2014) Reviews without a purchase: Low ratings, loyal customers, and deception. *J. Marketing Res.* 51(3):249–269.

Askalidis G, Kim SJ, Malthouse EC (2017) Understanding and overcoming biases in online review systems. *Decision Support Systems* 97:23–30.

Berger J (2014) Word of mouth and interpersonal communication: A review and directions for future research. *J. Consumer Psych.* 24(4):586–607.

Brandes L, Godes D, Mayzlin D (2019) What drives extremity bias in online reviews? Theory and experimental evidence. Working paper, University of Lucerne, Lucerne, Switzerland.

Burtch G, Hong Y, Bapna R, Griskevicius V (2017) Stimulating online reviews by combining financial incentives and social norms. *Management Sci.* 64(5):2065–2082.

Chen Y, Harper FM, Konstan J, Li SX (2010) Social comparisons and contributions to online communities: A field experiment on movielens. *Amer. Econom. Rev.* 100(4):1358–1398.

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* 43(3):345–354.

Chevalier JA, Dover Y, Mayzlin D (2018) Channels of impact: User reviews when quality is dynamic and managers respond. *Marketing Sci.* 37(5):688–709.

Cialdini RB, Goldstein NJ (2004) Social influence: Compliance and conformity. *Annual Rev. Psych.* 55:591–621.

Freedman JL, Fraser SC (1966) Compliance without pressure: The foot-in-the-door technique. *J. Personality Social Psych.* 4(2):195.

Godes D, Silva JC (2012) Sequential and temporal dynamics of online opinion. *Marketing Sci.* 31(3):448–473.

Guardian (2018) Meriton fined $3M for manipulating TripAdvisor hotel reviews. Accessed September 3, 2020, https://www.theguardian.com/travel/2018/jul/31/meriton-fined-3m-for-manipulating-tripadvisor-hotel-reviews.

Hu N, Pavlou PA, Zhang J (2017) On self-selection biases in online product reviews. *MIS Quart.* 41(2):449–A17.

Hu N, Zhang J, Pavlou PA (2009) Overcoming the J-shaped distribution of product reviews. *Comm. ACM* 52(10):144–147.

Klein N, Marinescu I, Chamberlain A, Smart M (2018) Online reviews are biased. Here's how to fix them. *Harvard Business Review* (March 6), https://hbr.org/2018/03/online-reviews-are-biased-heres-how-to-fix-them.

Kullback S (1959) *Information Theory and Statistics* (Wiley, New York).

Lerner J, Tetlock P (1999) Accounting for the effects of accountability. *Psychol. Bull.* 125:255–275.

Li X, Hitt LM (2008) Self-selection and information role of online product reviews. *Inform. Systems Res.* 19(4):456–474.

Luca M (2016) User-generated content and social media. Anderson SP, Stromberg D, Waldfogel J, eds. *Handbook of Media Economics* vol. 1B (Elsevier, Amsterdam), 563–592.

Mayzlin D, Dover Y, Chevalier J (2014) Promotional reviews: An empirical investigation of online review manipulation. *Amer. Econom. Rev.* 104(8):2421–2455.

Moe WW, Schweidel DA (2012) Online product opinions: Incidence, evaluation, and evolution. *Marketing Sci.* 31(3):372–386.

Moe WW, Trusov M (2011) The value of social dynamics in online product ratings forums. *J. Marketing Res.* 48(3):444–456.

Montaguti E, Neslin SA, Valentini S (2016) Can marketing campaigns induce multichannel buying and more profitable customers? A field experiment. *Marketing Sci.* 35(2):201–217.

Muchnik L, Aral S, Taylor SJ (2013) Social influence bias: A randomized experiment. *Science* 341(6146):647.

Proserpio D, Zervas G (2017) Online reputation management: Estimating the impact of management responses on consumer reviews. *Marketing Sci.* 36(5):645–665.

Rosario B, Ana FS, de Valck K, Bijmolt THA (2016) The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *J. Marketing Res.* 53(3):297–318.

Schoenmueller V, Netzer O, Stahl F (2019) The extreme distribution of online reviews: Prevalence, drivers and implications. Working paper, Universita Bocconi, Milan, Italy.

Sridhar S, Srinivasan R (2012) Social influence effects in online product ratings. *J. Marketing* 76(5):70–88.

Stephen A, Bart Y, Du Plessis C, Goncalves D (2012) Does paying for online product reviews pay off? The effects of monetary incentives on content creators and consumers. Gürhan-Canli Z, Otnes C, Zhu RJ, eds. *NA–Advances in Consumer Research*, vol. 40 (Association for Consumer Research, Duluth, MN), 228–231.

Sun M (2011) How does the variance of product ratings matter? *Management Sci.* 58(4):696–707.

TripAdvisor (2018) Get more feedback with Review Express. Accessed September 3, 2020, https://www.tripadvisor.com/ReviewExpress.

Wang Y, Chaudhry A (2018) When and how managers' responses to online reviews affect subsequent reviews. *J. Marketing Res.* 55(2):163–177.

Wu F, Huberman BA (2008) How public opinion forms. Papadimitrou C, Zhang S, eds. *Internet and Network Economics* (Springer, New York), 334–341.