What Drives Extremity Bias in Online Reviews? Theory and Experimental Evidence*

Leif Brandes†
University of Lucerne

David Godes‡
University of Maryland

Dina Mayzlin§
University of Southern California

September 17, 2019

Abstract

In a range of studies across platforms, online ratings have been shown to be characterized by distributions with disproportionately-heavy tails. We focus on understanding the underlying process that yields such "j-shaped" or "extreme" distributions. We develop a simple analytical model to capture the most-common explanations: differences in utility or differences in base rates associated with posting extreme versus moderate reviews. We compare the predictions of these explanations with those of an alternative theory based on differential rates of attrition from the potential reviewer pool across people with moderate versus extreme experiences. The attrition rate, by assumption, is higher for moderate reviews. The three models yield starkly different predictions with respect to the impact on the relative prevalence of extreme versus moderate reviews of a review solicitation email: while existing theories predict a relative increase in extreme reviews, our attrition-based model predicts a decrease. Our results from a large-scale field experiment with an online travel platform clearly support the predictions from the attrition-based explanation, but are inconsistent with those from the utility-based and base-rate explanations alone.

Keywords: microeconomics, field experiments, electronic commerce, online reviews, online word of mouth

*The authors are thankful to the anonymous participating platform and for helpful comments from seminar participants at University of Lucerne, Emory, Duke, Wharton, Yale, John Hopkins, the Marketing Science Conference (Rome, Italy), Interactive Marketing Research Summit (Amsterdam, the Netherlands), ZEW 9th Conference on the Economics of Information and Communication Technologies (Mannheim, Germany), and the Social@IDC Conference (Herziliya, Israel). Flavia Tinner provided excellent research assistance. Part of this research was completed while the first author was a faculty member at Warwick Business School, University of Warwick, UK. The authors are listed in alphabetical order and contributed equally to the manuscript.

†leif.brandes@unilu.ch

‡dgodes@rhsmith.umd.edu

§mayzlin@marshall.usc.edu

INTRODUCTION

Word of mouth plays an important role in driving consumer decisions. In particular, existing research has shown that online reviews have a significant causal impact on purchases (Chevalier and Mayzlin, 2006; Chintagunta et al., 2010). In addition, there is a growing literature that examines the antecedents, content and consequences of online reviews.[1] One interesting robust empirical finding in the existing literature is the disproportionate prevalence of "extreme" distributions of online reviews. That is, relative to moderate review scores, extremely negative and positive review scores are posted more often, with the greatest tendency to post very high review scores. Since the highest possible rating score is often the mode, the resulting distribution has the shape of the letter "J." As documented by Schoenmüller et al. (2018), this "J-shaped," or extreme, distribution is pervasive in a variety of categories and platforms.[2] However, the underlying mechanism behind these J-shaped distributions is not well-understood.

Why do we so often observe extreme online review distributions? One possible explanation is that the underlying distribution of experiences is disproportionately extreme.[3] We call this the *base rate* explanation since it suggests that the cause of extreme online review distributions is the relatively-high base rate of extreme experiences. An alternative explanation could be that there is selection at the review provision stage. Indeed, the most-frequently used extant explanation for extreme distributions is that consumers receive greater utility from sharing extreme opinions. For example, Anderson (1998) proposes a model of word of mouth as a function of satisfaction, where more-extreme experiences increase the utility of engaging in word of mouth. We call this the *utility-based* explanation since it relies on differences in utility from providing reviews of extreme versus moderate experiences.

[1] For example, see Berger (2014), Babic Rosario et al. (2016)

[2] Moe et al. (2017) recently called extreme online review distributions 'one of the most robust findings in product reviews' (p. 484).

[3] Hu et al. (2009) point out that there may be selection at the product purchase stage since consumers who choose to purchase the product have higher expected utility than non-purchasers. Note that this explains the higher incidence of very positive reviews, but not the higher incidence of extreme negative reviews.

We propose a new explanation for the prevalence of extreme review distributions and test the implications of each of these theories in a field experiment. We define the potential pool of reviewers in any given period as those who have completed their experience (for example, they have finished reading a book or have returned from their vacation) and have neither written a review in a previous period nor exited the reviewer pool through attrition. This *attrition-based* explanation is based on the idea that those with more-extreme experiences exit the pool via attrition with lower probability. As a result, customers with more-extreme experiences have relatively more time to write about them before they become inactive.

Understanding the mechanism behind review distributions is important for theoretical and managerial reasons. First, it helps us understand the extent to which the distribution of online reviews is representative of the underlying distribution of consumer experiences. That is, while the base rate explanation implies that reviews are an accurate reflection of underlying experiences, the utility-based and the attrition-based explanations imply that there is selection at the review provision stage. Moreover, the attrition-based and the utility-based theories have very different implications for the best way to de-bias the reviews in order to obtain an accurate sense of the distribution of consumer purchase experiences. For example, if the utility-based explanation is the main driver, paying customers for reviews should decrease the bias by increasing the utility from posting moderate reviews. However, if the main driver is the attrition-based explanation, direct marketing interventions that reduce attrition may be more effective.

To empirically test the different theories, we proceed in three steps. First, we develop a simple analytical model that allows us to derive clear testable predictions for the different mechanisms. In our model, customers arrive in each period and then decide whether to write a review of their experience. Our key novel assumption is that, each period, some proportion of consumers exogenously leave the pool of potential reviewers. We separately vary the review utility, base rates, and attrition rates for customers with and without extreme experiences. We show that, while the attrition-based and the base rate explanations imply extreme review distributions uniformly across time, the utility-based explanation predicts extreme-review distributions only for the early periods after consumption. Since extreme distributions are so commonly-observed, this seems to suggest that the utility-based explanation in itself may not be consistent with the observed data.

Second, we examine analytically the impact of a review solicitation email, or "reminder" email, sent to all customers who have not yet written a review for their most recent experience.[4] We show that, under certain assumptions, the competing theories make markedly-different predictions for the effect of the reminder on the distribution of posted reviews: the attrition-based explanation predicts a relative *decrease* in the posting of extreme experiences after a reminder, while the base rate and the utility-based explanations predict a relative *increase*. This enables us to empirically test the relative explanatory power of the different models.

Finally, we report the findings from a large-scale field experiment that we designed in cooperation with a major European online travel portal where customers can book and review hotel trips. We randomly assigned customers to four different conditions that differed in the length of the time interval between the end of the customer's vacation and the reminder email. Specifically, while some customers received the email on the first day after the end of travel, others received it on the second, fifth, and ninth day. This design allows us to compare the distribution of provided reviews following a reminder email with that of a control group that did not yet receive a reminder. Our results show that reminders lead to a relative *decrease* in the posting of extreme experiences. The effect sizes are considerable: 10 percent fewer extreme reviews are written in the treatment conditions, in which a reminder has already been sent, relative to the control conditions, in which the reminder has not yet been sent. Importantly, this comparison holds constant, across conditions, the number of elapsed days since the end of travel. Accordingly, our results cannot be explained by previous work that suggests that extreme experiences for hedonic

[4] Throughout this paper, we are going to use the terms "review solicitation email" and "reminder email" interchangeably.

goods, such as hotels, may become more moderate over time (e.g., Moore (2012)). Further analyses demonstrate that the reminder email affects review extremity on both ends of the rating scale, and shows the robustness of our results when controlling for differences in review text characteristics across treatment and control groups, and to the use of alternative measures of review extremity. Overall, our empirical results provide strong support for our attrition-based explanation.

This paper makes several substantial contributions to the literature on online word of mouth. First, we advance existing knowledge by identifying a novel mechanism – attrition – that drives the commonly-observed extreme distribution of online reviews. Specifically, ours is the first study to present a theoretical argument and empirical evidence for the attrition-based explanation. The results strongly support this mechanism. Second, our study presents novel evidence that the distribution of reviews is not stable over time. Specifically, we show that reducing reviewer attrition can change the relative extremity of posted reviews. Third, we demonstrate that extreme distributions represent a form of bias in online ratings, and do not just reflect differences in base-rates. Fourth, we provide a general model of review provision that is able to accommodate different theoretical mechanisms and thus should be useful for future analytical and empirical work. Finally, the findings of our study have clear managerial relevance, and inform marketers who may wish to de-bias review distributions, but are hesitant to use the potentially-costly monetary incentives that have been studied in previous work (e.g., Fradkin et al. (2018)).

The rest of the paper is structured as follows. First, we review the previous literature and existing explanations for extreme distributions. Second, we present our analytical model and derive testable predictions from the alternative, theoretical explanations for extreme distributions. Third, we describe our experimental design and our identification strategy. Next, we present our experimental results. Finally, we discuss the implications of our findings and conclude.

EXTREME DISTRIBUTIONS: EVIDENCE AND EXPLANATIONS

Numerous studies have documented that online reviews distributions exhibit a disproportionate share of extreme reviews. This phenomenon appears in virtually all product and service categories, including books (Chevalier and Mayzlin (2006), Godes and Silva (2012), Hu et al. (2009)), DVDs (Hu et al., 2009), movies (Dellarocas and Narayan (2006), Liu (2006)), bath, fragrance and home products (Moe and Schweidel, 2012), home improvement products (Lafky, 2014), physicians (Gao et al., 2015), restaurants (Yelp, 2018), and accommodations (Fradkin et al., 2018).[5]

Table 1 presents an illustrative overview of studies that find extreme distributions in online reviews and reveals three important insights. First, extreme reviews account for about two thirds of posted reviews on platforms that do not allow for reciprocal rating between buyers and sellers. Second, the highest possible rating score accounts for about fifty to sixty percent of reviews on these platforms. In contrast, platforms that allow for reciprocal ratings, such as Airbnb, exhibit an even greater share of extreme reviews, and this share is exclusively driven by extremely positive reviews (Fradkin et al., 2018). Third, the table reveals that most existing research has focused on Amazon reviews. In response to this over-representation of Amazon, Schoenmüller et al. (2018) recently conducted an extensive study of extreme reviews across a wide range of platforms and product categories. They report that on all 12 studied platforms that use a five-point rating scale, such as Amazon, extreme distributions are quite prevalent. For Amazon itself, the authors find that between 84% to 98% of products from 24 product categories exhibit extreme distributions. In contrast, the prevalence of extreme distributions is considerably smaller for platforms that deviate from the frequently encountered five-point scale, and allow customers many more rating score options, such as RateBeer which uses a 20-point scale, or MovieLens, which uses a 10-point scale.

- Insert Table 1 about here -

[5]Recent evidence further demonstrates that this phenomenon is not restricted to consumption experiences. For example, Marinescu et al. (2018) study online reviews on Glasdoor.com for employers, and also find a disproportionate share of extreme reviews.

The prevalence of extreme distributions has led researchers to speculate about the underlying mechanism behind the phenomenon. The last column in Table 1 demonstrates that the utility-based explanation has by far been the most widely-applied explanation for extreme distributions. In fact, we did not come across a study that provided an explanation for these distributions and did not at least mention the idea that customers derive greater utility from sharing extreme experiences. The second most frequent explanation was the base-rate explanation, although there were relatively-few mentions of this. A third class of explanations related to platform-specific mechanisms, such as reciprocal-rating procedures between buyers and sellers on Airbnb (Fradkin et al., 2018). Finally, Schoenmüller et al. (2018) discuss evidence by Mayzlin et al. (2014) and Luca (2011) on review fraud, which suggests reviews tend to be more extreme due to the presence of either very positive or very negative promotional reviews.

While empirical evidence for the relative importance of different explanations remains scarce, existing studies mostly refer to the utility-based explanation for extreme distributions. For example, Schoenmüller et al. (2018) present empirical evidence from surveys, experiments, and secondary data, which they argue is consistent with the utility-based explanation. For example, customers who review more frequently and, hence, presumably are less selective in deciding which experiences to review, are more likely to post moderate reviews. In addition, forcing experimental participants to review their last product experience leads to less extreme review distributions than allowing them to choose any past experience to review. While the authors also report support for other drivers, such as the base rate explanation, or review fraud, these effects are found to be much smaller. Similarly, Fradkin et al. (2018) argue that the utility-based explanation is the greatest source of review bias on Airbnb. In a field experiment, they show that reminder emails with $25 coupons in return for a review reduced extreme distributions relative to reminder emails without such coupons. Since these coupons increase the utility of posting *any* travel experience, they argue that this finding is consistent with the idea that, in the absence of such incentives, posting extreme experiences yields greater utility for customers. Lafky (2014) finds in a laboratory experiment that customers are more likely to share extreme reviews when reviewing is costly than when it is free. He argues that this finding is consistent with consumers having higher intrinsic costs of reviewing moderate experiences, which is a variant of the utility-based explanation.

In summary, studies indicate that online review distributions exhibit considerable self-selection at the review-provision stage. However, our understanding of the exact mechanism behind this self-selection remains limited. Specifically, it is suggested by the extant literature that customers' decisions as to whether or not to provide a review are driven primarily by the relative utility associated with doing so. In the next Section, we propose a different explanation for this self-selection based on differential attrition rates for consumers with extreme versus moderate experiences.

THEORY

We proceed in three steps in this section. First, we provide a simple analytical model of review provision. We then show how the three focal theories can be captured in this model and explore each theory's ability to explain extreme distributions. Finally, we introduce a review solicitation email from the review platform into the model and derive, for each explanation, theoretical predictions for its effect on the relative prevalence of extreme and moderate reviews.

A Simple Model of Review Provision

Consider the following simple process. In period $t = 0$, there are N customers who have just completed a consumption experience such as seeing a movie, eating at a restaurant or traveling to some holiday destination. In the following, we use $i \in \{x, m\}$ to denote the type of a customer, where the customer type is defined by the type of experience. We assume that N_x customers had an extreme experience, and N_m customers had a moderate experience. At the beginning of each period t, each customer posts a review with probability r_i. Conditional on not having posted a review, a customer of type i leaves the pool of potential posters with probability ϕ_i. We allow ϕ_i, the "attrition rate," to differ depending on the type of experience.

Note that our modeling approach is similar to the approach undertaken in the forecasting literature (see Schweidel and Knox, 2013; Fader et al., 2010; Reinartz and Kumar, 2003).[6] In these models, the customer may transition with some probability from being "alive" to "dead" (or from "active" to "inactive"). Similarly, in our model we specify that consumers leave the pool of potential reviewers with probability ϕ_i. It is important to emphasize that, just as is the case in much of this literature, we are agnostic on what exactly drives the attrition process. There are a number of plausible explanations for attrition and for differences in attrition rates between consumers with moderate versus extreme experiences. For example, one explanation may be memory-based: extreme experiences are more arousing, surprising, emotional, and thus inherently more memorable than moderate experiences. Accordingly, consumers with extreme experiences should be less likely to forget to post a review relative to consumers with moderate experiences.[7] Another plausible explanation may relate to agenda-setting: consumers who completed the consumption experience may prioritize tasks, and posting a review may fall off the list as more urgent tasks arrive over time. Moreover, posting a review may be more

[6]We thank Eric Bradlow for his suggestion on the connection between our work and the forecasting literature.

[7]Theories of learning emphasize the role of surprises for memory encoding insofar as surprising events result in greater physiological arousal, which helps to focus attention and memory encoding, consolidation, and retrieval of such events (Kensinger, 2009). Moreover, memory operates by the principle of selective encoding, where the storage of the most relevant information and experiences enjoys priority as an energy- and resource-efficient way to learn. However, an adaptive memory system also needs to be flexible and allow for memory modification if necessary, e.g., through forgetting. Indeed, some authors have concluded that "forgetting may be important for efficient use of memory, rather than a design fault" (Ward (2015), p. 220). Similarly, emotions guide memory formation, such that emotional experiences and stimuli, both positive and negative, are more memorable than neutral stimuli (Kensinger (2009); Kensinger and Schacter (2006)). Moreover, suppose that there is a link between how memorable an experience is and the likelihood that the consumer will post a review based on the theory of spreading activation (see Collins and Loftus (1975)). That is, suppose that there is a link between the node "memory of the experience" and the node "post a review." As the memory of the experience fades, the latter node may fall below the activation threshold (Crestani, 2012). In this way, the consumer effectively forgets to post a review from then on, until this node is reactivated. Hence, since extreme experiences are more memorable, consumers should be less likely to forget to post a review about them.

likely to fall off the list for a consumer who had a moderate experience relative to a consumer with an extreme experience, because of differential salience. Note that both these explanations suggest that the attrition rate is *lower* for consumers with extreme experiences relative to consumers with moderate experiences.

Next, we define P_i^t as the number of reviews of type i posted in period t, and M_i^t as the number of potential "active" reviewers who have, as of the start of period t, neither posted a review nor left the sample of potential reviewers via attrition. At the start of period 1, then, $M_i^1 = N_i$. At the start of period 2, the expected number of active reviewers in the population is

$$E\left[M_i^2\right] = N_i\left(1 - r_i\right)\left(1 - \phi_i\right).$$

In general, we have

$$E\left[M_i^t\right] = N_i\left[\left(1 - r_i\right)\left(1 - \phi_i\right)\right]^{(t-1)}, \tag{1}$$

and the expected number of reviews of type i posted in period t is

$$E\left[P_i^t\right] = E\left[M_i^t\right] r_i = N_i r_i\left[\left(1 - r_i\right)\left(1 - \phi_i\right)\right]^{(t-1)} \tag{2}$$

Using (2), we can also calculate the total expected number of posted reviews of type i that will be posted across all periods:

$$\sum_{t=1}^{\infty} E\left[P_i^t\right] = \frac{N_i r_i}{1 - (1 - \phi_i)(1 - r_i)} \tag{3}$$

And, finally, we can calculate the expected number of reviews of type i posted in the first T periods:

$$\sum_{t=1}^{T} E\left[P_i^t\right] = \frac{N_i r_i\left[1 - \left[\left(1 - r_i\right)\left(1 - \phi_i\right)\right]^T\right]}{1 - (1 - \phi_i)(1 - r_i)} \tag{4}$$

Explaining Extreme Distributions in Reviews

In this section, we use our model to demonstrate that each of these three possible theories may explain the existence of extreme distributions. Based on the reviewed literature and the simple model introduced above, we assume that extreme and moderate reviews follow a different posting process. Specifically, we explore three possible sources of heterogeneity across the two groups: 1) *the utility-based* model: a higher utility associated with posting about an extreme experience as compared with a moderate experience, which we capture in the model as a higher conditional probability of posting an extreme review in each period ($r_x > r_m$); 2) *the attrition-based* model: a lower attrition rate by the customer with an extreme experience versus the customer with a moderate experience ($\phi_x < \phi_m$); and 3) *the base-rate* model: a larger number of customers with extreme versus moderate experiences ($N_x > N_m$).

Our strategy is to focus on one explanation at a time. For example, in the utility-based model, we assume that $r_x > r_m$, while maintaining that $\phi_x = \phi_m \equiv \phi$ and $N_x = N_m \equiv N$. This allows us to investigate the extent to which any one source of heterogeneity could be driving the general pattern that we observe in the data. While we acknowledge that it is possible, if not likely, that there are multiple sources of heterogeneity, our goal is to assess the dominant mechanism that can explain the pattern of data we observe. While the utility-based and the base rate explanations have been offered in the literature before, the attrition-based theory is a novel explanation which we propose here.

Theorem 1 *The following reflect the expected review distributions conditional on the underlying mechanism:*

1. *The utility-based explanation ($r_x > r_m, \phi_x = \phi_m \equiv \phi, N_x = N_m \equiv N$) implies that the expected number of extreme reviews posted is greater than the expected number of moderate reviews posted for low t, and the reverse for high t.*

2. *The attrition-based explanation ($r_x = r_m \equiv r, \phi_x < \phi_m, N_x = N_m \equiv N$) implies the same expected number of extreme and moderate reviews are posted*

in $t = 1$ and strictly more extreme than moderate reviews are posted for all $t > 1$.

3. *The base rate explanation ($r_x = r_m \equiv r, \phi_x = \phi_m \equiv \phi, N_x > N_m$) implies strictly more extreme than moderate reviews are posted, in expectation, for all t.*

All proofs may be found in the Appendix. Theorem 1 establishes that each of these theories might be useful in explaining the prevalence of extreme distributions. Interestingly, however, the utility-based model (which Table 1 reveals to be the most popular explanation for extreme distributions in the extant literature) actually predicts a reversal of the effect for large t. To appreciate the source of the non-monotonicity in the utility-based explanation, note first that the expected number of reviews posted in each period is a product of two different variables: 1) the posting rate, and 2) the current pool of active reviewers. For low t, the fact that the extreme reviews have a higher posting rate drives the result that we expect more of them to be posted relative to moderate reviews. However, it is also the case that, over time, the population of active users is larger for moderate reviews since there are relatively fewer of them leaving the pool due to posting. As t increases, this latter effect dominates, and more moderate reviews are posted in expectation.

In contrast, the attrition-based explanation implies that more extreme reviews will be posted for all $t > 1$. The intuition is the following: based on our assumptions, we expect to see fewer customers with extreme experiences leaving the pool of active reviewers due to attrition. This implies the posting of more extreme reviews, in expectation. Finally, the base rate explanation assumes that extreme experiences occur more frequently, which means that the pool of active reviewers and posters will be greater for extreme experiences for all t.

While Theorem 1 suggests that these three theories are each reasonable candidates to explain extreme review distributions, it offers little guidance in terms of our objective of differentiating among them. In order to distinguish precisely among these theories, we introduce an exogenous shock to the review process in the form of a review solicitation email. Next, we derive analytically the predicted impact of such a shock on review provision as a function of each of these theories.

Review Solicitation Analysis

In this section, we examine the effect on the review distribution of a review solicitation email sent by the platform, asking all customers who have not yet written a review to post a review about their experience with the product. We assume that the email brings back those customers who previously left the pool of potential posters due to attrition. Why would a review solicitation email bring back these customers? For example, if attrition were driven by forgetting, the email acts as a reminder to those consumers who left the pool of potential reviewers. Similarly, if writing a review falls off the to-do list with time, a review solicitation email makes this task more salient and hence "activates" these consumers again. Note that on this point our model differs from the forecasting literature, such as Schweidel and Knox (2013) and Fader et al. (2010), where attrition is permanent. Due to the differences between purchasing and reviewing, we have a slightly different conceptualization of the concept of attrition.

More precisely, suppose that the email arrives after the end of period T. Since we assume that the potential reviewers who previously left the pool via attrition are returned to the pool as a function of the email, the expected number of potential reviewers is the expected number of customers who have not yet posted a review:

$$E\left[M_i^{(T+1)}|\text{reminder}\right] = N_i - \sum_{t=1}^{T} E\left[P_i^t\right] = N_i\left[1 - r_i\frac{1-\left[\left(1-r_i\right)\left(1-\phi_i\right)\right]^T}{1-(1-\phi_i)(1-r_i)}\right] \quad (5)$$

We will compare the relative proportion of expected extreme reviews with and without an email under each of the three theories. These results, in turn, will form the basis for the design of our field experiment. The focal quantity in this section will be Δ_i, the difference in the expected number of reviews of type i following a reminder after period T vs. no reminder. Analytically, this quantity is given by:[8]

$$\begin{aligned}\Delta_i^T &\equiv E\left[P_i^{T+1}|\text{reminder}\right] - E\left[P_i^{T+1}|\text{no reminder}\right] \\ &= N_i r_i \left[1 - r_i \frac{1 - \left[(1-\phi_i)(1-r_i)\right]^T}{1 - \left[(1-\phi_i)(1-r_i)\right]}\right] - N_i r_i \left[(1-r_i)(1-\phi_i)\right]^T\end{aligned}$$

This expression can be further simplified to:

$$\Delta_i^T \equiv \frac{N_i r_i (1-r_i)}{1-(1-\phi_i)(1-r_i)} \phi_i \left[1 - (1-r_i)^T (1-\phi_i)^T\right] \tag{6}$$

Before proceeding to our main results, it is worth noting here that:

$$\Delta_i^T \xrightarrow{\phi \to 0} 0$$

That is, a clear implication of our model is that an assumption of no attrition at all implies that the solicitation email should have no impact at all on the review distribution. Our results to follow suggest quite clearly that there is an impact and that, therefore, attrition is an important element of the review-provision process.

The following result demonstrates that the review solicitation email allows us to generate starkly-different predictions based on our attrition-based theory as compared with both the utility-based theory and the base rates theory.

Theorem 2 *The following reflect the expected impact of a review solicitation email on review distributions conditional on the underlying mechanism:*

1. *According to the attrition-based explanation, a review solicitation email always leads to less-extreme review distributions.*

2. *According to the utility-based explanation, a sufficient condition for a review solicitation email following period T to lead to more-extreme review distribu-*

[8] We attempt to make clear through our notation that the superscript on Δ_i captures the period following which the reminder is sent while the superscripts on the posted reviews captures the periods in which the reviews arrive.

tions is:

$$r_i < \frac{1}{T+1}$$

3. *According to the base rate explanation, a review solicitation email always leads to more-extreme review distributions.*

To appreciate the intuition behind the attrition-based theory result, notice that since $\phi_m > \phi_x$, those consumers who exited the potential-reviewer pool due to attrition are disproportionately moderate. As a result, those returning to the pool and ultimately posting after the reminder are, again, more moderate. With respect to the utility-based model, for r_i not too high, the result is again fairly straightforward as the higher posting rate ($r_x > r_m$) implies that consumers with extreme experiences who are brought back to the pool have a higher likelihood of posting following the review solicitation email, as compared with those consumers with moderate experiences who left the sample but are brought back with an email. However, if the difference becomes too pronounced ($r_x >> r_m$), then the pool of those consumers who left the sample is so heavily-skewed toward moderate experiences that the review solicitation email leads to a pool of reviewers with more-moderate experiences. In order for Theorem 2 to predict a decrease in extremeness following a reminder, we require that $r < (T+1)^{-1}$. In the context of our experimental conditions with $T = 1, T = 2$ and $T = 5$, this requires that $r < .167$. It is straightforward to demonstrate that this condition is easily met in our data.[9]

[9]It is critical to highlight that r is the periodic posting rate capturing what proportion of those remaining potential reviewers will post a review *on that day*. That is, it is the hazard rate. To see that our condition is met, one can simply note that, as mentioned in the Data Section, only approximately 19% of potential reviews are written over the entire course of our study. More precisely, our model of the review-provision process implies that one can derive an estimate of r by inspecting the posted reviews by (non-treated) subjects at $t = 1$. In our dataset, of the 142,348 subjects in Conditions 2, 5 and 9, only 1,394 wrote a review at $t = 1$, which is less than 1%. This implies that our condition holds over the entire range of our experiment and up to $T = 100$.

EXPERIMENTAL DESIGN AND DATA

To test the theoretical predictions from the three different explanations, we conducted a field experiment in cooperation with a large European online travel platform. The travel platform wishes to remain anonymous.

Company Background

For more than 10 years, the partnering online travel platform has been very successful, making it one of the two largest travel platforms in its core market segment. The platform attributes much of its success to the availability of more than 7 million customer reviews for more than 700k hotels on its site and places great strategic importance on having a robust set of current customer reviews. To reflect on the dynamic quality of hotels, the platform constructs average hotel rating scores based only on reviews from within the last two years, even if older reviews are available.

The travel platform obtains review content from two different groups of travelers: its customers, i.e., those who have previously booked a vacation through the platform's travel agency, and travelers who have previously booked a vacation with a different travel platform or agent. In this way, the platform combines the approaches of similar platforms, such as Expedia (where only customers of the platform can write a review), and Tripadvisor (where all travellers can submit their reviews). To receive more review content from its own customers, the platform sends out a review solicitation email on the first day after the end of vacation to all customers who have not yet provided a review for their hotel experience.[10] This email welcomes customers home, asks them for a hotel review, and provides links to the most recent evaluation for this hotel and an online rating form. Figure 1 displays a translated, stylized example of this email.

- Insert Figure 1 about here -

If a customer clicks on the email link to the online rating form, she will be asked to answer a number of questions, such as whether she would recommend the

[10]In our constructed sample, 5.65% of reviews were provided before the end of vacation, and another 2.8% were provided on the day that the vacation ended.

hotel (Yes/No), how she would rate the hotel on a scale from 1 (very bad) to 6 (very good) overall, how she would rate different quality aspects of the hotel (e.g., location, service), and how she would rate the value for money at this hotel. The consumer then needs to provide a text description that is at least 100 characters long and is asked about some personal and travel characteristics (e.g., age, country of residence, timing and length of stay, reason for travel).[11] If a customer does not respond to this email, the travel platforms makes up to two additional attempts to solicit a review from this customer. The second and third email (if the second email did not result in a review) are sent on the fifth and ninth day after the end of vacation, respectively.[12] If no review has been provided after 9 days, the company ends its review-solicitation attempt, and waits another 14 days before sending a final email in which customers can win a 100 Euro voucher for their next booking.

It was against this background that the company agreed to implement our field experiment in which it randomly allocated customers to one of four experimental conditions that we designed to test for the effect of a review solicitation email on the share of extreme reviews. Importantly, these conditions differed only in the timing of the solicitation emails. All other aspects of the emails and review solicitation procedure remained identical across conditions.

Experimental Manipulation

Figure 2 displays the experimental design with our four experimental conditions. Condition 1 represents the previously-discussed status quo at the travel platform. In the other three conditions, we increased the amount of time between the end of travel and the day that the first review solicitation email was sent: in condition 2, the first email was sent on the second day after the end of travel, and in conditions 5 and 9, it was sent on the fifth and ninth day after the end of travel, respectively.

- Insert Figure 2 about here -

[11]Customers can choose between a long and short review format. These formats differ in whether a separate review text is required for each hotel quality dimension (long) or not (short). In our sample, less than 2% of reviews were short reviews.

[12]The form and content of the second and third email differ from that of the first email. However, as we only focus on the effect of the first email in our empirical study, we do not discuss these differences in more detail throughout this paper.

Returning customers were randomly allocated to the four different conditions as follows. On the first day after the end of a vacation, an algorithm confirmed each customer's review status, i.e., whether a review had already been provided, or not. All customers who had not yet provided a review for the vacation under study were randomly allocated to one of our four conditions. Each condition had a 25% allocation probability. To avoid unnecessary emails, the system always confirmed a consumer's review status on the scheduled day before sending a review solicitation email. Our experiment began on June 1, 2017 and concluded on September 26, 2017. In addition, we obtained detailed information on bookings and hotel characteristics, which allow us to match this information to reminder emails and hotel reviews.

Based on the previously-discussed study design, we identify six possible tests to evaluate the impact of a review solicitation email on review extremity. Table 2 provides an overview of these tests that are based on different review latency values, i.e., the number of days that have elapsed between the end of travel and the time of review provision. Test 1 uses only reviews that are provided on the first day after the end of travel and compares the share of extreme reviews across condition 1, where the reminder email has already been sent, and all other conditions, where the email has not yet been sent. As the experimental treatment in our design is receiving the review solicitation email, condition 1 serves as the treatment condition in Test 1, and the others serve as control conditions. In Test 2, we use more observations to increase the statistical power of our test. Specifically, we include all reviews that were posted within the first four days after the end of travel, and compare the share of extreme reviews across the treatment condition 1 and control conditions 5 and 9. Note that, since the email is sent on the second day in condition 2, we can no longer use this condition in our control group.

- Insert Table 2 about here -

Table 2 shows that there exist four additional tests that cleanly assign posted reviews from a given day after end of travel to treatment and control conditions, two using condition 2 as treatment, and two using condition 5 as treatment. However, as we move from left to right in the Table, there remain fewer non-treated observations left to serve as the control group. Note that by holding constant, within each test, across conditions the number of elapsed days since the end of travel and review provision, we are able to rule out common patterns across time, such as improved customer understanding of past extreme experiences (Moore (2012)), as an alternative explanation for a change in the share of extreme reviews across conditions. Ending this section, we emphasize that the focus of our study is on the effect of the *travel platform*'s reminder emails on review provision and extremity, and not on effects of the hotel management's communication with customers.[13]

The Data

Our data set is constructed by matching review-solicitation emails to bookings and reviews, each of which resided in distinct data tables. We describe the exact data construction procedure in the Appendix. We note here that, in order to ensure a balanced number of observations across each of the four experimental conditions, we exclude all bookings with end dates between September 18, 2017 and September 25, 2017. This nine-day window ensures that each of the subjects in each of the treatment conditions received their first email.[14] Based on this procedure, our final data set includes observations for 189,842 hotel bookings with 35,238 matched reviews. Accordingly, the probability to review is 18.6% in our sample.

To evaluate the effectiveness of our randomization procedure, we tested for differences in key booking characteristics across all four experimental conditions. Table 3 displays summary statistics and results from Kruskal-Wallis tests for differences across conditions. We observe that trips lasted on average about eight days, with an average price of around 1,670 Euro. Looking at customer characteristics, we see that the average trip involved 2.34 travellers (the median value was 2), that customers returned on average from one trip within our sample period (although

[13] We do not observe any such communication in our data. Moreover, any effect of the latter should be controlled via the experimental design which was characterized by randomization at the individual consumer level.

[14] For example, if we do not exclude these observations, subjects in, say, condition 5 who returned home on September 25, 2017, would not have received an email and would thus be identical to subjects in condition 9. Note that our results are unchanged if these bookings are included.

some customers had multiple bookings),[15] and that the average customer age was 41 years.[16] By looking at the data across conditions, we see very little variation, which is reassuring for the effectiveness of our randomization procedure. The results from Kruskal-Wallis tests largely support this impression and only detect a significant difference across conditions for the number of travellers per booking.

- Insert Table 3 about here -

EMPIRICAL RESULTS

We present our empirical results in five steps. First, we establish that our review data exhibit the well-known extreme distribution that we aim to explain. Second, we test for the underlying mechanism behind this distribution by identifying the effect of a review solicitation email on the share of extreme ratings. Third, we demonstrate that the email affects review extremity on both ends of the rating scale. Fourth, we explore whether the review solicitation email may have affected the review content. Finally, we report the results from a number of robustness checks.

Establishing the Extreme Distribution

Figure 3 displays the rating score distribution in our sample, and yields two important insights: first, and comparable to previous research, we observe a left skewed distribution, in which 44 percent of reviews involve the highest possible rating score (6). Second, reviews with the most negative rating score (1) are extremely rare, and account for less than two percent of all posted reviews in our sample. To make the share of extreme ratings in our sample comparable to shares of around 50 to 65 percent in previous studies (as reviewed in Table 1), we classify a review as "extreme"

[15]In the Robustness Checks Section, we demonstrate that our results are robust to the exclusion of customers with multiple bookings.

[16]Note that this is the age of the customer who booked the vacation. We constructed this information from each user's self-reported profile information on the system. However, not all profiles included the date of birth, which explains the difference in the number of observations.

if it involves a rating score of 1, 2, 3, or 6. Based on this approach, extreme reviews account for 54 percent in our sample.[17]

- Insert Figure 3 about here -

The Effect of a Review Solicitation Email on Review Extremity

In Theorem 2 we demonstrated that the attrition-based explanation predicts a *decrease* in the share of extreme reviews in response to a review solicitation email, whereas the utility-based and the base rate explanations predict an *increase*.

Table 4 presents the results on the average treatment effect of a review solicitation email on review extremity for our six theory tests. In Tests 1 and 2, we see that the share of extreme reviews is significantly lower in Condition 1 than in the other Conditions. Specifically, in Test 1, we focus on the first day after the end of travel and find that the share of extreme reviews is 55 percent in Condition 1 but 61 percent when pooling across Conditions 2, 5 and 9. Similarly, Test 2 shows that across days 1 to 4, the share of extreme reviews is 55 percent in condition 1, but 61 percent when pooling across conditions 5 and 9. Looking at Tests 3 and 4, we see that the share of extreme reviews is also significantly lower in Condition 2 on days 2 to 4 after the end of travel than in Conditions 5 and 9. The results for Tests 5 and 6 replicate this pattern for days 5 to 8 after the end of travel when comparing Condition 5 to Condition 9. Overall, the displayed results strongly support our attrition-based explanation, and contradict the predictions of the utility-based and base rate explanations. We summarize this finding as our first result.

Result 1 *After a review solicitation email, the share of extreme reviews reduces significantly. This is consistent with the predictions of the attrition-based explanation, but inconsistent with predictions of the utility-based and base-rate explanations.*

- Insert Table 4 about here -

[17] As we report in the Robustness Checks Section, our results are robust to alternative extremeness classifications, in which we either exclude rating scores of 3, or also include rating scores of 4. These two classifications imply a share of 48 and 67 percent extreme ratings, respectively.

Table 5 displays estimation results from Logit models on the likelihood of an extreme review when controlling for the number of travellers per booking. This model specification addresses the potential concern that the previously-reported differences across conditions might be a reflection of the previously-detected differences in the number of travellers per booking across conditions (as displayed in Table 3). However, the results in Table 5 clearly reject this idea and demonstrate that the likelihood of extreme reviews is consistently around 6 percent lower when travellers have just received a reminder email compared with the case in which they have not. Overall, these results confirm our previous insights and provide strong support for our attrition-based explanation as an important driver behind review posting decisions, in general, and extreme distributions, specifically.[18]

- Insert Table 5 about here -

Extremity Bias vs Positivity Bias

It is important at this stage to ask whether our results in Tables 4 and 5 really represent a reduction in review *extremity*, or whether, given the relatively-low prevalence of negative extreme reviews, they might instead be driven by a change in review *positivity*? To address this question, we report the results from two additional analyses in which we study separately the impact of a review solicitation email on positive and negative extremity relative to moderate reviews.

We begin with negative extremity. Table 6 compares the share of negative extreme reviews relative to moderate reviews across our previous six test conditions. Consistent with our previous results, we see that the share of negative extreme reviews is always lower after the email reminder than before. For example, on Day 1 after the end of vacation, the share of negative extreme reviews in Condition 1 is 19 percent, and thus significantly lower than the corresponding share of 28 percent across Conditions 2,5, and 9. The same significant pattern obtains when using Days

[18] We note that our previous tests allocated customers to the treatment conditions, whenever their posting date coincided with the date of their first email. Our data also allow us to observe whether the email or posting event occurred first on that date. We note that the difference between treatment and control conditions increases, when we exclude those customers from our analysis that actually posted before the email.

1 to 4, Day 2, and Days 2 to 5 as test conditions. However, for the test conditions that use Day 5 and Days 5 to 8, respectively, the differences are no longer statistically significant (for Days 5 to 8, $p < 0.17$). Next, we report the findings from a separate analysis of positive extreme reviews. Table 7 displays the associated proportions and test statistics. In each test condition, we observe the, by now familiar, pattern that the proportion of positive extreme reviews is significantly lower after the reminder than before the reminder. For example, on Day 1 after the end of vacation, the share of positive extreme reviews in Condition 1 is 50 percent, and thus significantly lower than the corresponding share of 54 percent across Conditions 2,5, and 9. We summarize this finding as our second result:

Result 2 *Immediately following a review solicitation email, there is a significant decrease in the share of positive and negative extreme reviews.*

- Insert Tables 6 and 7 about here -

Figure 4 illustrates the previous insights in a different way and shows the distributions of rating scores across treatment and control conditions for all six test conditions from before. In panel a), for example, the focus is on the first day after the end of travel. Accordingly, the treatment condition consists of Condition 1, and the control conditions are Conditions 2, 5, and 9. Overall, and in light of the consistently detected patterns for positive and negative extreme reviews in this section, we can thus say that the review solicitation email affects both types of review extremity, and not just positive or negative extremity.[19]

- Insert Figure 4 about here -

Does the review solicitation email affect the review content?

According to our theoretical model, the review solicitation email from the travel platform reduces the extremity of the review distribution by changing the composition of posters. To this point, we have implicitly maintained the assumption that,

[19]For the interested reader, we note that the changes in the negative and positive extremity seem to cancel each other out, yielding no effect on the average rating valence in most conditions.

conditional on one's experience, the reported rating is unchanged. In particular, we have assumed that the reduced extremeness cannot be explained by such a change in the reported ratings effected by the email. It is important to acknowledge the possibility that such an impact may have occurred and, to the extent possible, to assess the veracity of such a concern. To do so, one would ideally observe each customer's evaluation score both with and without the review solicitation email. Unfortunately, this approach is not feasible in our context, because we only observe one evaluation score (either before or after the review solicitation email) for each reviewer. Therefore, we turn to an alternative, albeit second-best, approach to identify a potential effect of the email on cognition, effort and affect: analysis of the review text.

To analyze the review text, we use the Linguistic Inquiry and Word Count (LIWC) software developed by Tausczik and Pennebaker (2010).[20] Our empirical approach follows the logic of our previous analyses: we fix the number of days after the end of vacation, and test for differences in the content of the review text across our treatment and control groups. This time, however, we conduct this analysis for each possible evaluation score (1-6) separately. We are thus able to see whether, say, a 5-star rating posted after the review solicitation email is accompanied by more-moderate review text than a 5-star rating posted before the email. We focus our analysis on three groups of word categories that seem particularly relevant for any effect of the review solicitation email: review length (word count, words per sentence), psychological processes (affective processes, positive emotion, negative emotion, anxiety, anger, sadness), and cognitive processes (insight, causation, discrepancy, tentative, certainty, cognitive mechanism). In total, we analyze 14 categories for six rating levels (1-6) and six sets of days after the end of vacation.

Our results show few systematic differences in the review text across treatment and control groups.[21] In fact, we only observe consistent differences for three measures: word count, affective processes and positive emotions. Consistent with previ-

[20] As all reviews in our sample are written in German, we actually work with the adaptation of the LIWC dictionaries for the German language (Wolf et al., 2008).

[21] The full set of results is available from the authors upon request.

ous work (Askalidis et al. (2017)), we find that reviews after the review solicitation email are significantly shorter on average. However, the post-email reviews score higher on both affective processes and positive emotions. Importantly, this latter result is inconsistent with the idea that the review solicitation email dampens the effect of affect, thereby resulting in the posting of more-moderate opinions. With respect to the decrease in word count, it is not immediately clear how to interpret the result or how it might connect with our main findings on extremity. However, to be sure that it is not an indication that some unobserved factor is driving our result, we estimate another model with the (log-transformed) word count as covariate. As Table 8 shows, our main results are robust to this alternative specification.

- Insert Table 8 about here -

Robustness Checks

In this section, we report the robustness of our main results across two alternative classifications of review extremity, as well as when focusing only on those travellers who appear only once in our sample.

Alternative measures of review extremity. In a first alternative definition, we exclude rating scores of 3 from the "extreme" category and, instead, consider these to be "moderate." Overall, this approach results in the classification of 48 percent of reviews in our sample as extreme reviews. Table 9 shows the results from a replication of our main analysis, and reveals that our previous findings are robust to the use of this more restrictive measure, although the level of statistical significance is slightly reduced. In a second alternative definition, we use a more-expansive definition of "extreme," by including rating scores of 4 in the category. Overall, this approach results in the classification of 67 percent of reviews in our sample as extreme reviews. Table 10 presents the results from a replication of our main analyses, and shows that the share of extreme reviews is still significantly higher in treatment versus control conditions for 5 out of 6 comparisons. We conclude that our main results are not driven by the specific classification of extreme and moderate reviews.

- Insert Table 9 about here -

- Insert Table 10 about here -

Results for one-time customers. Table 3 revealed that some customers had more than one trip that ended during our experimental period. As some of these travellers may have been allocated to more than just one experimental condition, we also conducted our main analyses when excluding customers with multiple bookings from our analysis. Table 11 presents the associated estimation results and shows that the use of this restricted sample does not alter our findings. We thus conclude that our main results are not driven by customers with multiple bookings in our sample.

- Insert Table 11 about here -

DISCUSSION

In this paper, we introduce a novel, attrition-based mechanism that explains the prevalence of extreme distributions, "one of the most robust findings in product reviews" (Moe et al. (2017). p. 484). Starting from a simple model of review provision, we showed analytically that, under certain conditions, this attrition-based explanation gives rise to markedly different review patterns after a review solicitation email than the utility-based and base-rate explanations on which existing studies have typically focused. Specifically, while the latter both predict an immediate relative *increase* in review extremity in response to a review solicitation email, the attrition-based explanation predicts a *decrease*. The results from a large-scale field experiment that we conducted in cooperation with a leading European travel platform showed that email reminders *decrease* the share of extreme reviews (by about 10 percent in our main specifications). Importantly, this result reflected changes in review extremity at both ends of the rating scale and did not seem to be caused by any direct effects of the review solicitation email on review content. We also demonstrated that our results are robust to the use of two alternative measures of review extremity and different estimation samples. Overall, our study thus provides first evidence for the importance of attrition-based effects for review provision in the field.

Theoretical Contribution

Our work contributes to the literature on word of mouth in four important ways. First, we introduce a novel attrition-based explanation for extreme distributions and show that this mechanism explains the observed empirical patterns better than existing explanations that focus exclusively on reviewer utility from posting or differences in customer base rates across types of experiences. Our model requires the following two assumptions: 1) consumers with more extreme experiences have a lower attrition rate, 2) those who left the sample can be brought back with a reminder. That is, our model can be thought of as a modified version of the attrition model where "dead" consumers can be brought back with a review solicitation email.

Second, our theory identifies the interaction of customer attrition and review solicitation emails as a novel driver for the instability of online review distributions over time. From a theoretical point of view, this contributes to our understanding of dynamics in online reviews. Existing work has either focused on social and temporal dynamics in review distributions (Li and Hitt, 2008; Wu and Huberman, 2008; Moe and Trusov, 2011; Moe and Schweidel, 2012; Godes and Silva, 2012), dynamics that result from (hotel) managers' responses to previous reviews (Chevalier et al., 2018; Proserpio and Zervas, 2017), or has exclusively attributed any effects of solicitation emails to changes in customer motivation (albeit without testing this mechanism (Askalidis et al., 2017)). From a methodological point of view, we emphasize that the documented change in the share of extreme reviews in the distribution prior and after reminder emails would not have been discernible from an exclusive focus on average ratings, as has been common in previous studies. Future research should thus consider a broader range of distributional measures.

Third, we contribute to the existing knowledge of biases in online reviews. While some theories, such as the base rate explanation, imply that posted reviews are an unbiased representation of customers' underlying product experiences, others, such as our attrition-based explanation, suggest otherwise. Based on the results of our field experiment, we conclude that online reviews do not represent an unbiased view of customers' actual experiences. Our results thus confirm the importance of reviewer self-selection as a source of review bias as previously discussed in the literature (e.g., Moe and Schweidel, 2012; Godes and Silva, 2012), but specify a novel selection mechanism. Interested researchers may thus consider the integration of attrition into their empirical and analytical work.

Finally, we provide a general model of review provision and demonstrate how to integrate different theoretical lenses into this model. Our results demonstrate the importance of formalizing theories for review provision as this enables clean empirical tests of their predictions. Future research could build on our work to study additional factors, such as social influence, or differences across groups of customers, such as expertise, or across different channels, on review provision.

Managerial Implications

Companies are constantly looking for ways to attract more word of mouth activity from their customers. This is evident from the considerable amounts of money that they are spending on review solicitation (Babic Rosario et al., 2016), and the increasing supply of websites that offer guidance on the best ways to get more reviews and word of mouth. Our research has two important implications for companies trying to attract more reviews from their customers. First, our identification of an attrition-based mechanism behind extreme distributions implies that managers should take previous managerial recommendations that focus exclusively on interventions to change customers' cost-benefit evaluations during review provision with a pinch of salt: unless a consumer remembers to write a review in the first place, any such improvements are necessarily likely to be ineffective. At the same time, the lack of empirical support for the utility-based explanation in our study suggests that expensive financial incentives may be unnecessary to reduce the extremity bias, and that a simple reminder email can already substantially reduce this bias.

Second, we show that firms need to carefully design an effective email management system. For example, additional analyses of our data reveal that the timing of the first review solicitation email is crucial. Figure 5 shows the likelihood to write a review across our four experimental conditions. We can see that waiting to email customers may not be a good strategy for platforms that aim to maximize the review-provision likelihood from customers: while this likelihood is 20 percent when the first email is sent on the first day after the end of travel, it monotonically decreases with additional delay in the first email to only 17 percent when the first email is sent on the ninth day. This shows that it is important to engage customers early on in review provision for travel experiences. Future research could consider the extent to which similar results can be observed for other product categories, in which the consumption experience has a clear start and end date that is observable to the firm (e.g., flights, cabs, restaurant visits, hospitalization, education), and how the optimal start time for reminders differs (if at all), when consumption starts and ends only after purchase (e.g., for books, DVDs, or household appliances) and cannot be observed by firms.

- Insert Figure 5 about here -

While not directly related to firms' attempts to generate more word of mouth for their products, our results finally demonstrate that solicitation emails are not an innocuous intervention from firms. Instead, these emails affect the share of extreme reviews and the length of reviews. This raises the practically relevant question as to whether firms need to disclose their review solicitation practices and which reviews they obtained through these practices. Other examples suggest that policy makers and consumer protection agencies might require such disclosure in the future. For instance, reviewers who received the reviewed product for free from the producer are nowadays required by law to disclose this information. One potential avenue might be to require firms and platforms to separately group reviews that were submitted before and after review solicitation emails (Askalidis et al. (2017)). However, more research on the implications of such email interventions for customer decision-making and welfare is required for policy makers to decide if they need to take action.

Limitations of this Study

Just like any other research, our study is not without limitations. First, we acknowledge that our experimental design does not allow us to rule out *any* impact of the utility-based and base rate explanations on review provision. Accordingly, we emphasize that our results should *not* be read to imply that the utility-based and base rate theories are irrelevant for real-world review provision behavior. Instead, they demonstrate that our attrition-based theory possesses unique explanatory power that extends beyond extant theories.

Second, we acknowledge that it would have been very interesting to separate the effect of merely receiving a review solicitation email (but not opening it) from the effect of receiving and opening this email. Unfortunately, our data do not allow us to study this effect. The reason is that, for customers who did not open the email, we have no way to establish the exact point in time when they first noted the reception of this email. This creates a problem for our identification approaches that focus on review provision shortly after the email was actually *sent*: for those customers that ended up writing a review on the day of the review solicitation email without having opened this email, we do not know whether they had observed the email or not. However, a closer look at our data suggests that, even with this information, the effect would have been difficult to identify as this group of customers is small, accounting for only 0.8 % of all reviews in our sample.

CONCLUSION

Extreme distributions are a persistent feature of many online review distributions. In this paper, we introduced a novel, attrition-based mechanism to explain such distributions, and demonstrated its empirical relevance in the context of a large-scale field experiment in the travel industry. Starting from a simple model of review provision, we showed how to integrate different theories for extreme distributions into the same theoretical framework and derived testable predictions from it. Based on a specifically designed field experiment, we reported the results from six alternative identification approaches to test those predictions with our data. While our attrition-based theory explained the observed empirical patterns very well, existing explanations alone were insufficient to do so. Future research should thus integrate considerations and implications of the dynamics of reviewer attrition for review provision.

REFERENCES

Anderson, E. (1998). Customer satisfaction and word of mouth. *Journal of Service Research*, 1(1):5–17.

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

Babic Rosario, A., Sotgiu, F., De Valck, K., and Bijmolt, T. H. A. (2016). The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors. *Journal of Marketing Research*, 53:297–318.

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

Chevalier, J. A., Dover, Y., and Mayzlin, D. (2018). Channels of impact: User reviews when quality is dynamic and managers respond. *Marketing Science*.

Chevalier, J. A. and Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 43(3):345–354.

Chintagunta, P. K., Gopinath, S., and Venkataraman, S. (2010). The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science*, 29(5):944–957.

Collins, A. and Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6):407–428.

Crestani, F. (2012). *Spread of Activation Theory*, pages 3175–3179. Springer US, Boston, MA.

Dellarocas, C. and Narayan, R. (2006). A statistical measure of a population's propensity to engage in post-purchase online word-of-mouth. *Statistical Science*, 21(2):277–285.

Fader, P. S., Hardie, B. G., and Shang, J. (2010). Customer-base analysis in a discrete-time noncontractual setting. *Marketing Science*, 29(6):1086–1108.

Fradkin, A., Grewal, E., and Holtz, D. (2018). The Determinants of Online Review Informativeness: Evidence from Field Experiments on Airbnb. Working Paper.

Gao, G., Greenwood, B. N., Agarwal, R., and McCullough, J. S. (2015). Vocal minority and silent majority: How do online ratings reflect population perceptions of quality. *MIS Quarterly*, 39(3):565–589.

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

Hu, N., Pavlou, P. A., and Zhang, J. (2009). Overcoming the j-shaped distribution of product reviews. *Communications of the ACM*, 52(10):144–147.

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review*, 1(2):99–113.

Kensinger, E. A. and Schacter, D. L. (2006). When the red sox shocked the yankees: Comparing negative and positive memories. *Psychonomic Bulletin & Review*, 13(5):757 – 763.

Lafky, J. (2014). Why do people rate? theory and evidence on online ratings. *Games and Economic Behavior*, 87:544–570.

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

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

Luca, M. (2011). Reviews, reputation, and revenue: The case of yelp.com. Harvard Business School Working Paper.

Marinescu, I., Klein, N., Chamberlain, A., and Smart, M. (2018). Incentives Can Reduce Bias in Online Reviews. Working Paper.

Mayzlin, D., Chevalier, J., and Dover, Y. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8):2421–55.

Moe, W. and Trusov, M. (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research*.

Moe, W. W., Netzer, O., and Schweidel, D. A. (2017). Social media analytics. In *Handbook of Marketing Analytics*, pages 483–504. Springer.

Moe, W. W. and Schweidel, D. A. (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science*, 31(3):372–386.

Moore, S. G. (2012). Some things are better left unsaid: How word of mouth influences the storyteller. *Journal of Consumer Research*, 38(6):1140 – 1154.

Proserpio, D. and Zervas, G. (2017). Online reputation management: Estimating the impact of management responses on consumer reviews. forthcoming in Marketing Science.

Reinartz, W. J. and Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of marketing*, 67(1):77–99.

Schoenmüller, V., Netzer, O., and Stahl, F. (2018). The Extreme Distribution of Online Reviews: Prevalence, Drivers and Implications. Working Paper.

Schweidel, D. A. and Knox, G. (2013). Incorporating direct marketing activity into latent attrition models. *Marketing Science*, 32(3):471–487.

Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54.

Ward, J. (2015). *The Student's Guide to Cognitive Neuroscience*. Psychology Press, third edition edition.

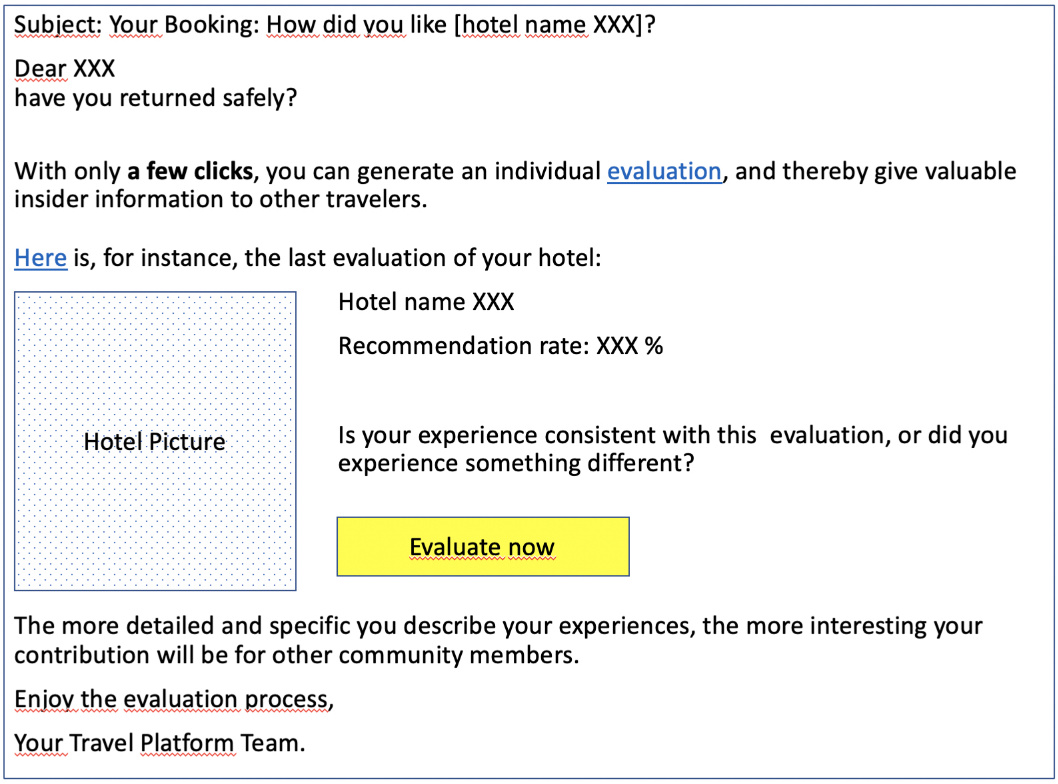
Wolf, M., Horn, A. B., Mehl, Matthias, R., Haug, S., Pennebaker, J. W., and Kordy, H. (2008). Computergestützte quantitative textanalyse: Äquivalenz und robustheit der deutschen version des linguistic inquiry and word count. *Diagnostica*, 54:85–98.

Wu, F. and Huberman, B. (2008). *Internet and Network Economics. Lecture Notes in Computer Science*. Springer.

Yelp (2018). Yelp factsheet 30, june 2018. https://www.yelp.com/factsheet. Accessed: 2018-08-09.

FIGURES

Figure 1: First Solicitation Email: Content and Form


Subject: Your Booking: How did you like [hotel name XXX]?

Dear XXX
have you returned safely?

With only **a few clicks**, you can generate an individual evaluation, and thereby give valuable insider information to other travelers.

Here is, for instance, the last evaluation of your hotel:

Hotel Picture

Hotel name XXX

Recommendation rate: XXX %

Is your experience consistent with this evaluation, or did you experience something different?

Evaluate now

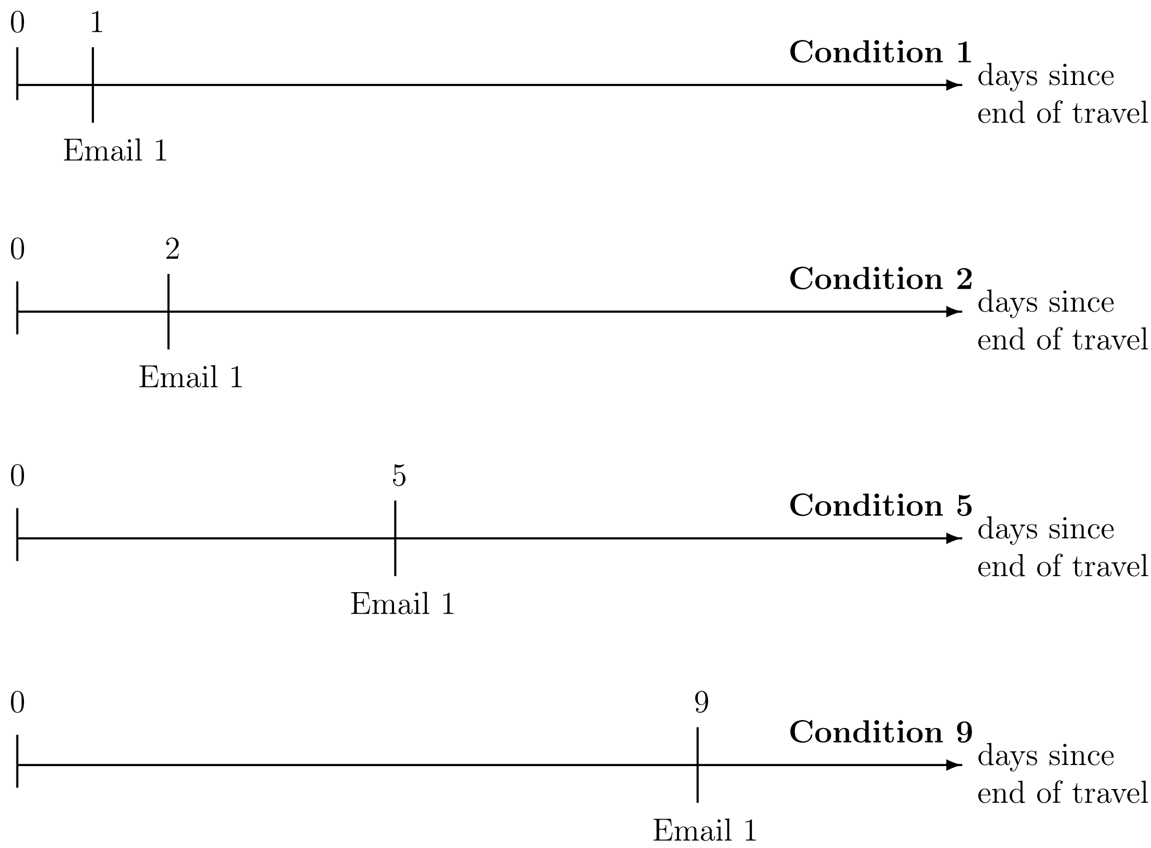
The more detailed and specific you describe your experiences, the more interesting your contribution will be for other community members.

Enjoy the evaluation process,

Your Travel Platform Team.

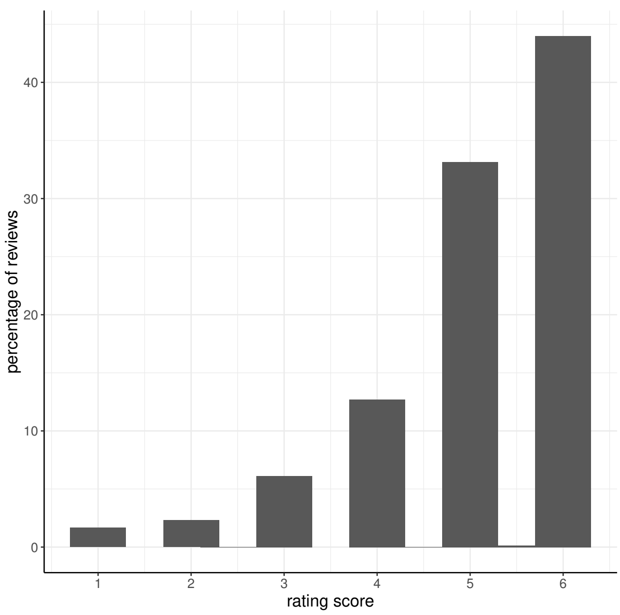
Notes: A translated, stylized example for the content and form of the first solicitation email that non-reviewers receive after the end of their vacation.

Figure 2: Experimental Design



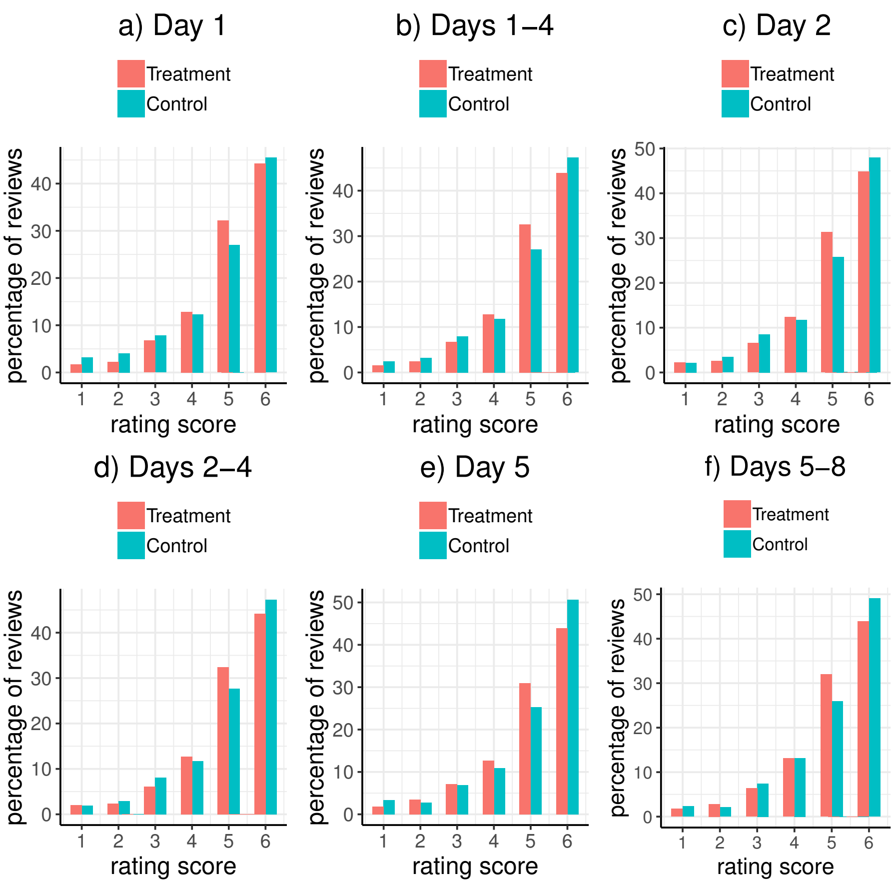
Notes: Displayed are the four experimental conditions that we used in our study. Condition 1 represents the current status quo.

Figure 3: Distribution of Rating Scores at Travel Platform



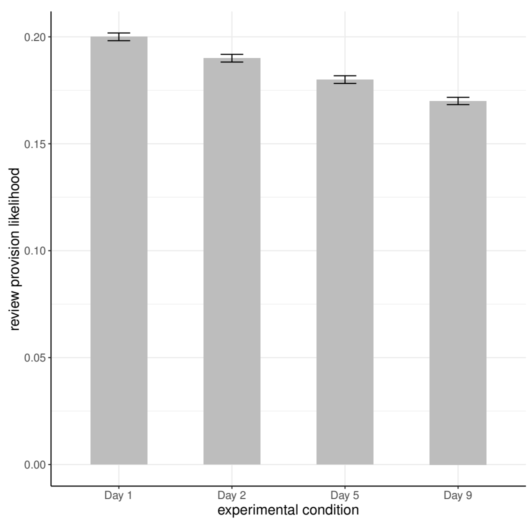
Notes: Displayed is the distribution of the overall rating score in posted reviews.

Figure 4: Distributions of Review Scores Across Set of Days and Conditions



Notes: Displayed is the rating distribution for different sets of days after consumption. Treatment indicates conditions, in which travellers had received an email reminder prior to, or on the chosen set of days. Control indicates conditions, in which travellers had not yet received an email reminder.

Figure 5: Review Provision Likelihood Across Conditions



Notes: Displayed is the review provision likelihood across the four experimental conditions.

TABLES

Table 1: Previous Studies That Report Extreme Distributions

Study	Product Category (Data Source)	% Extreme Ratings	% Highest Rating Score	Theoretical Explanations
Chevalier and Mayzlin (2006)	Books (Amazon)	60 - 70	57 - 67	none provided
Dellarocas and Narayan (2006)	Movies (Yahoo! Movies)	65	47	utility-based
Fradkin et al (2018)*	Accommodation (Airbnb)	75	74	utility-based and others
Gao et al (2015)*	Physicians (RateMDs.com)	64	59	utility-based and others
Godes and Silva (2012)	Books (Amazon)	64	56	none provided
Hu et al. (2009)*	Books, DVDs, Video (Amazon)	58 - 64	47 - 56	utility-based, base rate
Lafky (2014)**	Home Improvement (Amazon)	-	-	utility-based, base rate
Moe and Schweidel (2012)**	Bath, Fragrance and Home (anonymous retailer)	-	-	utility-based
Schoenmueller et al (2018)*	Multiple Products and Platforms	41 - 85	31 - 84	utility-based and others
Yelp (2018)	Restaurants (Yelp)	64	48	none provided
This study	Accommodation (anonymous platform)	54	44	attrition-based

Notes: To ease the comparison across studies, we re-labelled theoretical explanations as utility-based, if authors cited Anderson (1998) as a key reference for drivers behind extreme distributions, or if they argued that posting extreme experiences yields greater utility to customers. An example are Gao et al. (2015) who introduce "hyperbole effects" in rating valence to explain the prevalence of more extreme reviews. However, most of their discussion is in the spirit of Anderson (1998), and emphasizes the higher utility that individuals derive from sharing extreme experiences. For papers marked with *, shares of rating scores for extreme distributions were manually calculated from Tables and Figures in the paper. Papers marked with ** presented evidence for extreme distributions, but did not report distributions of rating scores across categories.

Table 2: Experimental Design: Treatment and Control Conditions

	Latency values included in the analysis:					
	Day 1 (Test 1)	**Days 1-4 (Test 2)**	**Day 2 (Test 3)**	**Days 2-4 (Test 4)**	**Day 5 (Test 5)**	**Days 5-9 (Test 6)**
Treatment	Condition 1	Condition 1	Condition 2	Condition 2	Condition 5	Condition 5
Control	Conditions 2,5,9	Conditions 5,9	Conditions 5,9	Conditions 5,9	Condition 9	Condition 9

Table 3: Balance Checks Across Treatment Conditions

	Condition 1		**Condition 2**		**Condition 5**		**Condition 9**		**Kruskal-Wallis**
Variable	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	
Travel Duration	8.16	4.20	8.18	5.36	8.16	4.53	8.18	5.39	0.304
Price	1,671	1,138	1,674	1,142	1,666	1,122	1,672	1,132	0.699
Travellers per Booking	2.34	0.98	2.35	0.98	2.34	0.98	2.35	0.99	7.813**
Bookings per Customer	1.11	0.63	1.12	0.80	1.12	0.69	1.11	0.62	1.568
Customer Age	41.12	13.29	41.16	13.32	41.09	13.27	40.99	13.24	3.317
	$N = 47,494$		$N = 47,513$		$N = 47,354$		$N = 47,481$		

Notes: Number of observations are for Travel Duration, Price, Travellers per Booking, and Bookings per Customer. For Customer Age, the associated values are $N = 42,397$, $42,419$, $42,411$, and $42,476$. **p=0.05

Table 4: Share of Extreme Reviews Across Conditions

	Condition 1 vs. Conditions 2, 5, and 9			Condition 2 vs. Conditions 5 and 9			Condition 5 vs. Condition 9		
Set of Days	1	2,5 and 9	z-stat.	2	5 and 9	z-stat.	5	9	z-stat.
Test 1: Day 1	0.55	0.61	3.41***						
Test 2: Days 1 to 4	0.55	0.61†	5.41***						
Test 3: Day 2				0.56	0.62	2.80***			
Test 4: Days 2 to 4				0.55	0.60	4.01***			
Test 5: Day 5							0.56	0.64	1.75*
Test 6: Days 5 to 8							0.55	0.61	2.51**

Notes: Displayed are proportions of extreme reviews across sets of days after end of travel and conditions. z-Stat. denotes z-statistic for tests of proportion equality across conditions. † This value is based only on Conditions 5 and 9. *** p<0.01, ** p<0.05, *p<0.10

Table 5: Logit Estimations for Likelihood of Extreme Reviews Across Conditions

	Condition 1 vs. Conditions 2, 5, and 9		Condition 2 vs. Conditions 5 and 9		Condition 5 vs. Condition 9	
Variable	Test 1: Day 1	Test 2: Days 1-4†	Test 3: Day 2	Test 4: Days 2-4	Test 5: Day 5	Test 6: Days 5-8
Condition 1 (Treatment)	-0.057***	-0.063***				
	(0.016)	(0.012)				
Condition 2 (Treatment)			-0.058***	-0.055***		
			(0.020)	(0.014)		
Condition 5 (Treatment)					-0.075*	-0.061**
					(0.041)	(0.024)
Travellers per Booking	-0.011	-0.008	0.016*	0.001	0.027**	0.012
	(0.008)	(0.006)	(0.009)	(0.007)	(0.011)	(0.008)
N	4,002	8,016	3,131	6,163	2,062	4,256
Wald	13.46***	31.34***	10.80***	16.06***	8.90**	8.55**
- LL	-2,729.06	-5,467.28	-2,128.27	-4,213.65	-1,405.89	-2,919.74

Notes: Displayed are marginal effects for Logit specifications. Robust standard errors are displayed in parentheses. † This effect is measured relative to Conditions 5 and 9. *** p<0.01, ** p<0.05, *p<0.10

Table 6: Share of Negative Extreme Reviews Across Conditions

	Condition 1 vs. Conditions 2, 5, and 9			Condition 2 vs. Conditions 5 and 9			Condition 5 vs. Condition 9		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Set of Days	1	2,5 and 9	z-stat.	2	5 and 9	z-stat.	5	9	z-stat.
Test 1: Day 1	0.19	0.28	4.55***						
Test 2: Days 1 to 4	0.19	0.26†	5.23***						
Test 3: Day 2				0.21	0.27	2.67***			
Test 4: Days 2 to 4				0.19	0.25	3.76***			
Test 5: Day 5							0.22	0.26	0.86
Test 6: Days 5 to 8							0.20	0.23	1.40

Notes: Displayed are proportions of negative extreme reviews across sets of days after end of travel and conditions. A review was classified as 'negative extreme' if it involved a rating score of 1,2 or 3. Positive extreme reviews with a rating score of 6 are excluded from this analysis. z-Stat. denotes z-statistic for tests of proportion equality across conditions. † This value is based only on Conditions 5 and 9. *** p<0.01, ** p<0.05, *p<0.10

Table 7: Share of Positive Extreme Reviews Across Conditions

	Condition 1 vs. Conditions 2, 5, and 9			Condition 2 vs. Conditions 5 and 9			Condition 5 vs. Condition 9		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Set of Days	1	2,5 and 9	z-stat.	2	5 and 9	z-stat.	5	9	z-stat.
Test 1: Day 1	0.50	0.54	2.24**						
Test 2: Days 1 to 4	0.49	0.55†	4.41***						
Test 3: Day 2				0.51	0.56	2.32**			
Test 4: Days 2 to 4				0.49	0.54	3.32***			
Test 5: Day 5							0.50	0.58	1.78*
Test 6: Days 5 to 8							0.49	0.56	2.45**

Notes: Displayed are proportions of positive extreme reviews across sets of days after end of travel and conditions. In these robustness checks, a review was classified as positive 'extreme' if it involved a rating score of 6. Negative extreme reviews with a rating score of 1,2, or 3 are excluded from this analysis. z-Stat. denotes z-statistic for tests of proportion equality across conditions. † This value is based only on Conditions 5 and 9. *** p<0.01, ** p<0.05, *p<0.10

Table 8: Logit Estimations for Likelihood of Extreme Reviews Across Conditions (with Log(Word Count) as Control)

	Condition 1 vs. Conditions 2, 5, and 9		Condition 2 vs. Conditions 5 and 9		Condition 5 vs. Condition 9	
Variable	Test 1: Day 1	Test 2: Days 1-4†	Test 3: Day 2	Test 4: Days 2-4	Test 5: Day 5	Test 6: Days 5-8
Condition 1 (Treatment)	-0.080***	-0.077***				
	(0.017)	(0.012)				
Condition 2 (Treatment)			-0.079***	-0.070***		
			(0.021)	(0.014)		
Condition 5 (Treatment)					-0.086**	-0.070***
					(0.042)	(0.024)
Travellers per Booking	-0.011	-0.008	0.016*	0.0001	0.028**	0.012
	(0.008)	(0.006)	(0.009)	(0.007)	(0.011)	(0.008)
Log(Word Count)	-0.037***	-0.032***	-0.035***	-0.030***	-0.019	-0.016**
	(0.008)	(0.005)	(0.009)	(0.006)	(0.011)	(0.008)
N	4,000	8,012	3,129	6,159	2,060	4,254
Wald	37.29***	66.55***	27.15***	40.17***	11.71***	13.04***
- LL	-2,715.60	-5,446.37	-2,118.66	-4,198.79	-1,403.04	-2,916.01

Notes: Displayed are marginal effects for Logit specifications. Robust standard errors are displayed in parentheses. † This effect is measured relative to Conditions 5 and 9. *** p<0.01, ** p<0.05, *p<0.10

Table 9: Robustness Check: Share of Extreme Reviews Across Conditions (Restrictive Extremity Measure)

	Condition 1 vs. Conditions 2, 5, and 9			Condition 2 vs. Conditions 5 and 9			Condition 5 vs. Condition 9		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Set of Days	1	2,5 and 9	z-stat.	2	5 and 9	z-stat.	5	9	z-stat.
Test 1: Day 1	0.48	0.53	2.73***						
Test 2: Days 1 to 4	0.48	0.53†	4.34***						
Test 3: Day 2				0.50	0.54	1.85*			
Test 4: Days 2 to 4				0.49	0.52	2.55**			
Test 5: Day 5							0.49	0.57	1.79*
Test 6: Days 5 to 8							0.48	0.54	2.06**

Notes: Displayed are proportions of extreme reviews across sets of days after end of travel and conditions. In these robustness checks, a review was classified as 'extreme' if it involved a rating score of 1,2 or 6. z-Stat. denotes z-statistic for tests of proportion equality across conditions. † This value is based only on Conditions 5 and 9. *** p<0.01, ** p<0.05, *p<0.10

Table 10: Robustness Check: Share of Extreme Reviews Across Conditions (Extensive Extremity Measure)

	Condition 1 vs. Conditions 2, 5, and 9			Condition 2 vs. Conditions 5 and 9			Condition 5 vs. Condition 9		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Set of Days	1	2,5 and 9	z-stat.	2	5 and 9	z-stat.	5	9	z-stat.
Test 1: Day 1	0.68	0.73	3.31***						
Test 2: Days 1 to 4	0.67	0.73†	4.90***						
Test 3: Day 2				0.69	0.74	2.65***			
Test 4: Days 2 to 4				0.67	0.72	3.48***			
Test 5: Day 5							0.69	0.75	1.43
Test 6: Days 5 to 8							0.68	0.74	2.70***

Notes: Displayed are proportions of extreme reviews across sets of days after end of travel and conditions. In these robustness checks, a review was classified as 'extreme' if it involved a rating score of 1,2,3,4 or 6. z-Stat. denotes z-statistic for tests of proportion equality across conditions. † This value is based only on Conditions 5 and 9. *** p<0.01, ** p<0.05, *p<0.10

Table 11: Robustness Check: Share of Extreme Reviews Across Conditions (One-Time Customers)

	Condition 1 vs. Conditions 2, 5, and 9			Condition 2 vs. Conditions 5 and 9			Condition 5 vs. Condition 9		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Set of Days	1	2,5 and 9	z-stat.	2	5 and 9	z-stat.	5	9	z-stat.
Test 1: Day 1	0.55	0.61	3.73***						
Test 2: Days 1 to 4	0.55	0.61†	5.45***						
Test 3: Day 2				0.56	0.62	2.75***			
Test 4: Days 2 to 4				0.55	0.60	3.76***			
Test 5: Day 5							0.57	0.65	1.84*
Test 6: Days 5 to 8							0.55	0.62	2.58***

Notes: Displayed are proportions of extreme reviews across sets of days after end of travel and conditions. In these robustness checks, only customers with a single, ending trip during our sample period were included. z-Stat. denotes z-statistic for tests of proportion equality across conditions. † This value is based only on Conditions 5 and 9. *** p<0.01, ** p<0.05, *p<0.10

APPENDIX

Mathematical Proofs

Proof of Theorem 1. 1. For the utility-based explanation, we have

$$E\left[P_x^t\right] = Nr_x\left[(1-r_x)(1-\phi)\right]^{(t-1)} \tag{A.1}$$

$$E\left[P_m^t\right] = Nr_m\left[(1-r_m)(1-\phi)\right]^{(t-1)} \tag{A.2}$$

Note that $E\left[P_x^t\right] > \left[P_m^t\right]$ iff

$$r_x(1-r_x)^{(t-1)} > r_m(1-r_m)^{(t-1)} \tag{A.3}$$

$$\iff \tfrac{r_x}{r_m} > \left[\tfrac{1-r_m}{1-r_x}\right]^{t-1} \tag{A.4}$$

At $t = 1$, Equation (A.4) holds since $r_x > r_m$. The right hand side of Equation (A.4) is monotonically increasing in t if $r_x > r_m$. Hence, there exists a t' such that Equation (A.4) holds iff $t < t'$. Thus, the expected number of extreme reviews is greater than the expected number of moderate reviews posted for low t only. □

2. For the attrition-based explanation, we have

$$E\left[P_x^t\right] = Nr\left[(1-r)(1-\phi_x)\right]^{(t-1)} \tag{A.5}$$

$$E\left[P_m^t\right] = Nr\left[(1-r)(1-\phi_m)\right]^{(t-1)} \tag{A.6}$$

Given $\phi_x < \phi_m$, $E\left[P_x^t\right] > \left[P_m^t\right]$ for all $t > 1$, and $E\left[P_x^t\right] = \left[P_m^t\right]$ for $t = 1$. □

3. For the base rate explanation, we have

$$E\left[P_x^t\right] = N_x r\left[(1-r)(1-\phi)\right]^{(t-1)} \tag{A.7}$$

$$E\left[P_m^t\right] = N_m r\left[(1-r)(1-\phi)\right]^{(t-1)} \tag{A.8}$$

Clearly, given $N_x > N_m$, the expected number of posted extreme reviews is strictly greater than the expected number of moderate reviews for all t. ■

Proof of Theorem 2. 1. To prove the claim for the attrition-based model, we will show that a reminder will increase the expected number of reviews of type m more than type x. Mathematically, this is equivalent to:

$$\frac{\partial \Delta_i^T}{\partial \phi_i} > 0$$

We begin by differentiating Δ_i^T with respect to ϕ_i:

$$\begin{aligned}
\frac{\partial \Delta}{\partial \phi_i} =& Nr(1-r) \times \left[\frac{1-(1-r)^T(1-\phi_i)^T}{1-(1-\phi_i)(1-r)} \right. \\
& \left. + \phi_i \frac{[1-(1-\phi_i)(1-r)]\, T(1-r)^T(1-\phi_i)^{T-1} - \left[1-(1-r)^T(1-\phi_i)^T\right](1-r)}{[1-(1-\phi_i)(1-r)]^2} \right] \\
=& \frac{Nr(1-r)}{[1-(1-\phi_i)(1-r)]^2} \times \left[[1-(1-r)^T(1-\phi_i)^T][1-(1-\phi_i)(1-r)] \right. \\
& \left. + \phi_i [1-(1-\phi_i)(1-r)]\, T(1-r)^T(1-\phi_i)^{T-1} - \phi_i(1-r)\left[1-(1-r)^T(1-\phi_i)^T\right] \right] \\
=& \frac{Nr(1-r)}{[1-(1-\phi_i)(1-r)]^2} \times \\
& \left[[1-(1-r)^T(1-\phi_i)^T]r + \phi_i [1-(1-\phi_i)(1-r)]\, T(1-r)^T(1-\phi_i)^{T-1} \right] > 0
\end{aligned}$$

2. For the utility-based theory, we demonstrate that, when the stated condition is met:

$$\frac{\partial \Delta_i^T}{\partial r_i} > 0$$

We begin by differentiating Δ_i^T with respect to r_i:

$$\begin{aligned}
\frac{\partial \Delta}{\partial r_i} = \frac{N\phi}{[1-(1-\phi)(1-r_i)]^2} \times \Big[& (1-2r_i)\left[1-(1-r_i)^T(1-\phi)^T\right][1-(1-\phi)(1-r_i)] \\
& + r_i T(1-r_i)^T(1-\phi)^T [1-(1-\phi)(1-r_i)] \\
& - r_i(1-r_i)(1-\phi)\left[1-(1-r_i)^T(1-\phi)^T\right] \Big]
\end{aligned}$$

We'll now define $\theta \equiv (1 - r_i)(1 - \phi)$ and rewrite the above as follows:

$$\frac{\partial \Delta}{\partial r_i} = \frac{N\phi}{[1-\theta]^2} \times \left[(1 - 2r_i) \left[1 - \theta^T\right] [1 - \theta] + r_i T \theta^T [1 - \theta] - r_i \theta \left[1 - \theta^T\right] \right] \quad \text{(A.9)}$$

In order to identify the sign of this expression, we need only focus on the sign of the quantity inside the brackets, which we define as Ψ. We first show that this term is increasing in ϕ by showing that it's decreasing in θ.

$$\begin{aligned}
\frac{\partial \Psi}{\partial \theta} &= (2r_i - 1) \left[T\theta^{T-1} + 1 - \theta^T(T+1)\right] + r_i T \left[T\theta^{T-1} - \theta^T(T+1)\right] - r_i \left[1 - \theta^T(1+T)\right] \\
&= \theta^T(T+1) [1 - r_i - r_i T] + \theta^{T-1} T [2r_i - 1 + r_i T] - (1 - r_i) \\
&= T \left[\theta^T - \theta^{T-1}\right] [1 - r_i(T+1)] + \theta^T [1 - r_i(T+1)] + r_i T \theta^{T-1} - (1 - r_i) \\
&< T \left[\theta^T - \theta^{T-1}\right] [1 - r_i(T+1)] + \theta^T [1 - r_i(T+1)] + \theta^{T-1} [r_i T - (1 - r_i)] \\
&= \left[\theta^T - \theta^{T-1}\right] [1 - r_i(T+1)] (T+1) \\
&< 0
\end{aligned}$$

where the final inequality follows from the premise of the theorem. To complete the proof, we return to (A.9) and note that the leading term is weakly positive. Since we've shown that $\frac{\partial \Psi}{\partial \phi} > 0$ in this region of the parameter space, this implies that, for $\phi \in (0, 1)$:

$$\begin{aligned}
\Psi \geq \Psi|_{\phi=0} &= (1 - 2r_i) r_i \left[1 - (1 - r_i)^T\right] + r_i^2 T (1 - r_i)^T - r_i (1 - r_i) \left[1 - (1 - r_i)^T\right] \\
&= r_i^2 \left[T(1 - r_i)^T - \left[1 - (1 - r_i)^T\right] \right]
\end{aligned}$$

which is positive iff:

$$(1 - r_i)^T > \frac{1}{T+1}$$

By the premise of the theorem, we know that:

$$(1-r_i)^T > \left[\frac{T}{T+1}\right]^T$$

Hence, $\Psi_{\phi=0} > 0$ if

$$\begin{aligned}
\left[\frac{T}{T+1}\right]^T &> \frac{1}{T+1} \\
\Leftrightarrow T^T &> (T+1)^{T-1} \\
\Leftrightarrow T\ln T &> (T-1)\ln(T+1)
\end{aligned} \tag{A.10}$$

We can rewrite the right-hand side as follows:

$$(T-1)\ln(T+1) = (T-1)\ln T + (T-1)\left[\ln(T+1) - \ln T\right]$$

and now incorporate this into (A.10):

$$\begin{aligned}
T\ln T &> (T-1)\ln(T+1) \\
\Leftrightarrow \ln T &> (T-1)\left[\ln(T+1) - \ln T\right] \\
\Leftrightarrow \frac{\ln T}{T-1} &> \left[\ln(T+1) - \ln T\right]
\end{aligned} \tag{A.11}$$

Now, we rewrite the left-hand side of (A.11) as a sum of differences, where $\ln(t+1) - \ln t$ is decreasing in t due to the concavity of logs:

$$\frac{\ln T}{T-1} = \frac{1}{T-1}\sum_{t=1}^{T-1}\left[\ln(t+1) - \ln t\right] > \left[\ln(T+1) - \ln T\right]$$

The last inequality is due to the fact that the average of decreasing terms is greater than the smallest term of the series. Thus, we've shown that $\Psi > 0$ as long as $r_i < \frac{1}{T+1}$. This, in turn, implies that, in this region of the parameter space, $\frac{\partial \Delta}{\partial r_i} > 0$.

3. The base rates result is obvious from inspection of (6). ■

Dataset Construction

The platform implemented the field experiment through a third-party that manages "after-travel" email correspondence. In particular, the third party performed the randomization across experimental conditions. While the platform provided us with bookings and review data, the third party provided the email information which, when merged with the platform data, identifies the traveler's experimental condition. Here we specify how we merged these data sets (bookings, reviews, and emails).

Each traveler who did not submit a review before their travel ended was assigned to one of the four conditions on the first day after their travel ended. Since our experiment ran from June 1, 2017 to September 26, 2017, the relevant travel end dates were May 31, 2017 to September 25, 2017. As noted on p. 19, we discard all observations with end dates after September 17, 2017 to ensure that all observations have enough time to receive their review-solicitation emails and that conditions are balanced and comparable. Again, the platform provided us with information on bookings and reviews. We obtained the entire set of the platform's 209,489 bookings with travel end dates between May 31, 2017 and September 17, 2017, as well as complete booking information, hotel details, and traveler information (e.g., gender, age). In addition, we received from the platform all 516,244 reviews that were provided between June 1, 2017 and September 26, 2017. Out of these reviews, 49,474 were posted by user IDs that appear at least once in our booking set. Recall that the platform, like TripAdvisor, accepts reviews both from its own customers and from those who booked through other channels. For each review, we observe the user ID, hotel ID, rating score, review text, entry date, and travel month.

We restricted our sample to include cases in which the same user ID had a maximum of one stay at the same hotel during the period of our study. We did this because the platform informed us that repeat bookings for the same hotel often come from bulk purchasers and travel agents that book through the platform on behalf of their customers. Similarly, we excluded all reviews in which the same user ID reviewed the same hotel more than once over the study period. A total of

199,397 (95.2%) of the bookings and 46,535 (94.1%) of the reviews met these two restrictions. Note that, even though they come from customers who traveled during our study period, a significant portion of the reviews observed at this stage come from bookings that occurred beforehand. These will be removed when we merge the reviews and bookings, as discussed below.

Next, we explain the merging of the bookings data and the email data. From the third party, we obtained details on 208,309 emails sent during the period June 1, 2017 to September 26, 2017. Each email observation included the corresponding booking ID, date of email, time of email, and an identifier. By comparing the date and type of the review-solicitation email with the customer's travel-end date, we can infer the experimental condition to which the customer was randomly assigned.[22] This process yielded 190,446 observations (95.5%) of travelers for whom we could match bookings to emails and an experimental condition.

Finally, we matched reviews to bookings based on user ID and hotel ID information in both data sets, and confirmed that the self-reported travel month in the review corresponded to either the travel start or end month of the booking. We deleted 601 booking-review observations for which the self-reported travel month differed from the platform-reported booking, and three observations for which the time to the first email did not correspond to the assigned experimental condition (e.g., eight days in Condition 9). Overall, this procedure yielded 35,238 matched reviews for 189,842 bookings associated with our four experimental conditions.

[22] There is slightly more complexity to this matching process for the interested reader. Specifically, there are two "types" of first review-solicitation emails consumers received, to which we'll refer here as Type A and Type B. The former were sent to all travelers who had not yet posted a review as of the end of the day prior to the day on which their experimental condition would dictate that they receive a review solicitation. For example, in condition 5, all travelers who hadn't posted a review as of the end of the fourth day following their return home would receive this email. In the email data file we received, each of these Type A email observations was coded specifically with the experimental condition. This was not the case with Type B emails which were sent to all consumers who had already posted a review as of the day of their scheduled email. Rather than asking them to write a review, this email thanked them for the review and asked them to post pictures of their trip. Since Type B identifiers did not specify the treatment condition, we accomplished this by comparing the date they returned with the date of the email. So, for example, if a customer received a Type B email on the fifth day after returning home, we know they are assigned to Condition 5.