NBER WORKING PAPER SERIES

CHANNELS OF IMPACT:
USER REVIEWS WHEN QUALITY IS DYNAMIC AND MANAGERS RESPOND

Judith A. Chevalier
Yaniv Dover
Dina Mayzlin

Working Paper 23299
http://www.nber.org/papers/w23299

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2017

The authors acknowledge support from the National Science Foundation (Chevalier, Award 1128322) and the Israel Science Foundation (Dover, Grant 1124/16). We thank numerous seminar participants for helpful comments and thank Revinate and STR for providing data. The authors have no material financial relationships related to this study. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Channels of Impact: User Reviews when Quality is Dynamic and Managers Respond
Judith A. Chevalier, Yaniv Dover, and Dina Mayzlin
NBER Working Paper No. 23299
March 2017
JEL No. L15,L86,M31

ABSTRACT

We examine the effect of managerial response on consumer voice in a dynamic quality environment. We argue that, in this environment, the consumer is motivated to write reviews by, in addition to altruism, the possibility that the reviews will impact the quality of the service directly. We examine this empirically in a scenario in which reviewers receive a credible signal that the service provider is listening. Specifically, we examine the managerial response feature allowed by many review platforms. We hypothesize that managerial responses will stimulate reviewing activity and that, because managers respond more and in more detail to negative reviews, we hypothesize that managerial responses will particularly stimulate negative reviewing activity. Using a multiple-differences specification, we show that reviewing activity and particularly negative reviewing is indeed stimulated by managerial response. Our specification exploits comparison of the same hotel immediately before and after response initiation and compares a given hotels reviewing activity on sites with review response initiation to sites that do not allow managerial response.

Judith A. Chevalier
Yale School of Management
135 Prospect Street
New Haven, CT 06520
and NBER
judith.chevalier@yale.edu

Dina Mayzlin
USC Marshall School of Business
701 Exposition Boulevard
Los Angeles, CA 90089-0808
mayzlin@marshall.usc.edu

Yaniv Dover
The Jerusalem School of Business Administration,
Mt. Scopus,
The Hebrew University,
Jerusalem.
Israel
yaniv.dover@mail.huji.ac.il

1 Introduction

Many Internet sites rely extensively on user contributions. User-generated online reviews have become an important resource for consumers making purchase decisions; retailers such as Amazon rely on reviews to help match consumers with products, and information portals like TripAdvisor and Yelp have user-generated reviews at the core of their business models. While many types of platforms rely on user reviews, there is a substantial amount of variation in the details of platform design. For example, some sites verify that reviewers purchased the product, while others do not. Some platforms allow reviewers to edit reviews after posting them; some do not. In this paper, we are interested in the role played by a feature adopted by some platforms that allows managers to respond publicly to consumer reviews. We examine the effect of managerial response on consumer expression of voice.

To understand the effect of managerial response on reviewing behavior, we consider what motivates consumers to expend time and energy to post reviews. Here we focus on two motivations behind consumer voice: altruism (Sundaram et al. (1998) and Hennig-Thurau et al. (2004)) and impact (Hirschman (1970), Wu and Huberman (2008)). That is, we model reviewers as driven by the needs and concerns of the audience. We specifically consider the audience for the review as including not only other consumers, but also the service provider.

The previous literature on online reviews has primarily focused on an environment where product quality is time-invariant, studying physical products such as books, movies, digital cameras etc. Here we study a product category for which product quality is dynamic, hotels. For static-quality products, customer reviews are most naturally thought of as an avenue for customers to share information with each other. In contrast, for dynamic-quality goods and services, managerial investments may alter product quality over time. In such cases, the consumer may reasonably view both the management and other consumers as an audience for the review. Reviewing could be motivated by an intent to impact the manager, not just other consumers.[1] Furthermore, a number of platforms that deal with dynamic-quality categories have increased the salience of the management as an au-

[1] There is a sense in which the quality of physical products such as books or beauty products could change over time. A book could become dated, or a product's attributes could become dominated by a newly-introduced product. However, these changes are not a result of managerial investments as with hotels and restaurants and thus, the impact issues we discuss are not completely relevant.

dience for reviews by allowing managers to directly respond to customer reviews on those platforms. (See Figure 1 for an example of a hotel review

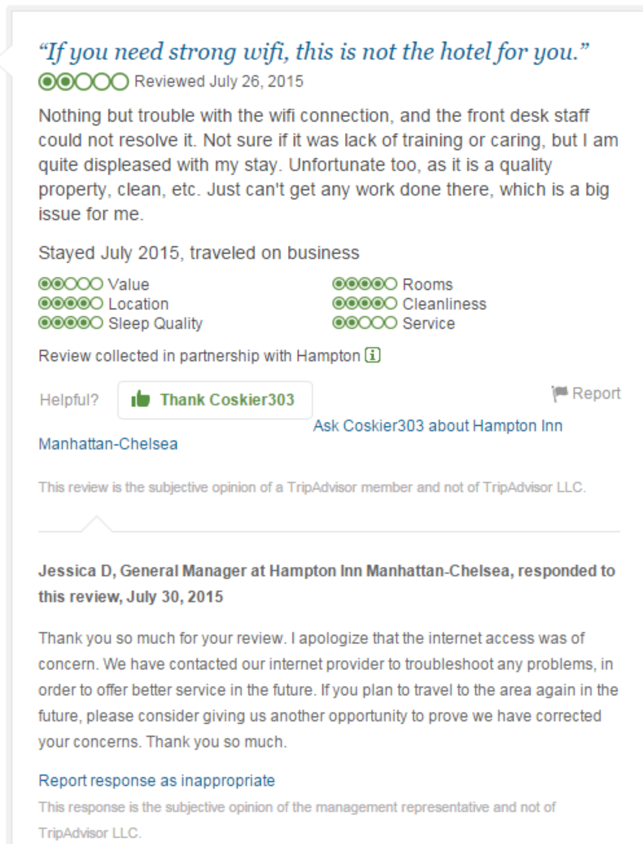

Figure 1: A TripAdvisor hotel review with managerial response

on TripAdvisor and the managerial response that followed).

The entry of the firm into the conversation potentially changes the nature of the discourse, which in turn may impact the customers' incentives to post reviews. The response functionality transforms a peer-to-peer review system into a hybrid peer review/customer feedback system. We study whether managerial response stimulates consumer reviewing and changes the nature of the reviews posted. Understanding the effect of managerial actions on consumer voice sheds further light on what motivates consumers to post feedback, which in turn has implications for all three actors (platforms, managers and consumers). That is, understanding these incentives can aid in optimal platform design, help firms improve their response strategy, and help consumers of reviews make better decisions.

We examine managerial response in the empirical setting of midtier to luxury hotels. The primary hypothesis that we examine is whether public managerial response communication stimulates reviewing activity. The basic idea is the following: because the manager is listening, consumers' feedback is more likely to impact the product's future quality. This appeals to both the altruism and the impact motivations of reviewers. Specifically, we investigate whether the introduction of managerial responses leads consumers to write more reviews, and whether it leads them to put more effort into reviewing, as measured by writing longer reviews. In addition, as we discuss at length below, we hypothesize that managerial response activity disproportionately stimulates negative review production since negative feedback may be seen as particularly impactful by reviewers.

In particular, we study whether reviewing activity changes for a hotel following the first day of posting of managerial responses on three websites that allow managers to respond to reviews: TripAdvisor, Expedia, and Hotels.com. Of course, in testing our hypotheses, we are faced with an identification challenge. Clearly, hotels that post managerial responses are different from hotels that do not. The decision to commence posting responses is an endogenous decision of the manager. It is possible, for example, that managers begin responding to reviews when something is going on at the hotel that leads the manager to anticipate more reviews, and more negative reviews being posted in the future.

To handle this identification challenge, we employ an extension of the techniques that were used to handle similar identification challenges in Chevalier and Mayzlin (2006) and in Mayzlin et al. (2014). Specifically, our identification strategy has four components.

First, we undertake an "event study" technique in which we examine reviewing activity for a given hotel for very short (6 week) time windows before and after the posting of the first managerial response. Therefore, we are not identifying the impact of review responses by straightforwardly comparing hotels that do and do not post responses; we are examining the change for a given hotel over time. Furthermore, by examining only the before/after change surrounding a discrete event in a very short time window, we are discarding changes that may derive from longer run investments in facilities, position, or quality of the hotel.

Second, we undertake specifications in which the reviewing changes on each of the focal sites on which managers may post responses are measured controlling for reviewing changes on other sites where managerial responses are not allowed. Specifically, managerial responses are not allowed on the popular booking sites Priceline and Orbitz, and our reviewing activity changes are measured from specifications which control for the changes in reviewing activity on Priceline and Orbitz. Thus, if a manager undertakes his/her first response on TripAdvisor in anticipation of some physical change in the hotel that will lead to negative reviews, the controls for changes in Priceline and Orbitz reviews should remove that source of review changes. This is similar to the approach used in Chevalier and Mayzlin (2006) and Mayzlin et al. (2014), though the appropriate "treatment" and "control" sites differ across the papers.

Third, all of our specifications are conducted measuring the change in the hotel's reviews relative to changes in average reviews in the geographic area on the same reviewing site. For example, when we are measuring changes in TripAdvisor reviews around the commencement of managerial response, we are measuring the change in reviewing activity for the hotel relative to other hotels in the local area. This differencing strategy will prevent us from attributing the changes in reviewing activity on TripAdvisor to the managerial response by the hotel if, in fact, the change in reviewing activity was caused by, for example, increased TripAdvisor advertising in the local area. This is a variant of a strategy that was employed in Mayzlin et al. (2014).

Finally, given the prominence of TripAdvisor as a review site, one may be concerned that a hotel that starts to respond on TripAdvisor may in fact also make physical TripAdvisor-specific investments. For example, TripAdvisor reviewers may especially value faster Wi-Fi, and the hotel's first online review response (a virtual TripAdvisor-specific investment) may be accompanied by Wi-Fi upgrades (a physical TripAdvisor-specific investment). Due to this concern we obtain data from from two other sites that allow responses (Expedia and Hotels.com) in addition to TripAdvisor, as we discuss further in Section 5.

In sum, this "multiple-differences" strategy controls for many of the underlying sources of the identification challenge. In Section 5 we discuss in more detail the extent to which remaining confounding effects can be eliminated and our results can be interpreted as causal. We find that reviewing activity for a given hotel increases on TripAdvisor in the six-week window following a managerial response relative to the six weeks prior and relative to the same hotel on Priceline and Orbitz, and relative to the TripAdvisor reviews of other hotels in the geographic area in the same six-week window.

We also find, relative to the controls, a statistically significant decrease in review valence and a statistically significant increase in the length of reviews. We have many fewer managerial response episodes for Expedia and Hotels.com. However, we conduct the same exercise for these platforms. For these sites, as with TripAdvisor, we find an increase in reviewing activity following the posting of a managerial response. For Expedia, we also find a significant decrease in review valence, and for Hotels.com, a significant increase in review length. We also explore further the decrease in review valence result using our large TripAdvisor sample. While we are cautious about interpreting this particular set of specifications causally, it suggests that responding only to negative reviews increases negative reviewing effort.

Our paper contributes to the recent literature on how firm intervention impacts consumer voice (see Ma et al. (2015) and Proserpio and Zervas (2016)). As we elaborate below in more detail, our setting and identification strategy are different from Ma et al. (2015), which gives rise to important differences in consumer incentives and overall results. While our setting is the same as that of Proserpio and Zervas (2016), our sampling and our identification strategy differs from theirs in important ways that we believe contributes to more robust results.

Our paper proceeds as follows. Section 2 reviews the literature. Section 3 presents the theoretical development in more detail. Section 4 describes our data and provides summary statistics about our sample. Section 5 describes our methodology. Section 6 provides our basic results. Section 7 discusses robustness issues. Section 8 concludes.

2 Related Literature

What drives a consumer to engage in word of mouth in the first place? Berger (2014) differentiates between two types of drivers of word of mouth: those that are self-driven (such as impression management for example) versus those that are audience-driven. Here we focus on the latter, and in particular, altruism and impact.

We first turn to the evidence that consumers are motivated by altruism when engaging in word of mouth. Sundaram et al. (1998) conduct interviews where respondents are asked to recall recent episodes of word of mouth. For both positive and negative word of mouth, respondents listed "altruism" as one of the major motivators of word of mouth. Hennig-Thurau et al. (2004) survey an online panel of German opinion platform users to investigate the motivations behind the generation of electronic word of mouth.

They find that one of the major factors that drives posting is "concern for other consumers."

In addition to altruism, previous research suggests that impact, the ability to influence the audience's actions, also motivates the poster's actions. Wu and Huberman (2008) provide evidence that individuals are more likely to post reviews on Amazon when their reviews will have a greater impact on the overall average review. That is, reviewers are more likely to post reviews when their opinion differs more from the consensus and when there are fewer reviews. Given the overall positive valence of reviews on review sites, the impact hypothesis is consistent with the findings of Moe and Trusov (2011). They examine data from a beauty products site and demonstrate that reviews become increasingly negative as ratings environments mature. Godes and Silva (2012) test the hypotheses of Wu and Huberman (2008) using book data from Amazon.com and show that low-impact reviews are more likely to be posted when reviewing costs are low. Specifically, they demonstrate that the sequential decline in ratings is more pronounced for high- than for low-cost reviews.[2] Notably, these papers, in using books and beauty products as their empirical settings, have focused on environments in which later reviews necessarily have more limited impact because the underlying true quality of the product has gradually become known. However, products with a substantial service component such as hotels and restaurants provide potentially quite different reviewing environments, as the quality of service can evolve over time. In dynamic-quality settings, consumers can influence the audience by alerting the audience to changes in product quality. Dynamic quality settings also create the possibility, as we discuss in the theory section, of reviews having an impact on managerial quality decisions.

Our paper is also related to recent literature on platform design and mechanisms for eliciting consumer feedback. Klein et al. (2016) shows that an increase in transparency in the review solicitation mechanism on eBay leads to a decrease in strategic bias in buyer ratings and an increase in welfare. Several papers document that many consumers who use an internet site do not post reviews. For example, Brandes et al. (2015) show that only thirteen percent of buyers posted a review on a European hotel booking portal. Nosko and Tadelis (2014) show that sixty percent of consumers leave

[2] In a context distinct from product reviews, Zhang and Zhu (2011) provide support for the impact hypothesis using data from Chinese Wikipedia postings. Zhang and Zhu (2011) show that contributions to Chinese Wikipedia by contributors outside mainland China declined dramatically when the site was blocked to readers in mainland China. This suggests that contributors receive social benefits from posting their contributions; reducing the size of potential readership reduces the benefits of contributing.

feedback on Ebay while Fradkin and Pearson (2015) show that about two thirds of consumers leave reviews on Airbnb. The results of both Nosko and Tadelis (2014) and Fradkin and Pearson (2015) suggest that eliciting negative feedback from consumers may be particularly difficult. For example, Nosko and Tadelis (2014) suggests that consumers that have had negative experiences are less likely to leave feedback. Similarly, Horton and Golden (2015) demonstrates a positive skew in reviewing on the temporary-labor site ODesk.

Next, we consider the nascent literature on the impact of managerial response on consumer voice. Ma et al. (2015) examine the effect of a firm's service intervention in response to a compliment or a complaint on Twitter on the consumer's subsequent Twitter comments. The authors find that redress-seeking is a major driver of complaints, and hence an intervention may actually encourage future complaints. Note that there are important differences between our studies. First, the identification strategies are very different – the earlier study builds a dynamic choice model of voice behavior and intervention on one platform only. Second, in our setting the managerial response is not accompanied by a service intervention. In this sense our results apply to settings where a response may be a form of communication and not involve an actual service intervention.

Several papers in this area have examined the effect of managerial response to hotel reviews. The papers of which we are aware are Park and Allen (2013); Ye and Chen (2009); Proserpio and Zervas (2016) and Kim and Brymer (2015). Park and Allen (2013) use a case study of four luxury hotels; the authors examine why management choose to be active or not in review responses. Kim and Brymer (2015) use proprietary data from an international hotel chain and show a correlation between responses to negative comments in online reviews and hotel performance. Ye and Chen (2009) conduct an experiment similar to ours and Proserpio and Zervas (2016). Using data from two Chinese travel agents, Ye and Chen (2009) show that reviewing activity and valence increases for hotels that post manager responses on the travel site that allows responses relative to the travel site that does not. However, the number of manager responses is quite low.

The closest paper to ours is Proserpio and Zervas (2016). This paper examines Texas hotels, comparing the reviews in the time period before and after a hotel's initiation of managerial response on TripAdvisor. The identification scheme is similar to ours; however, these authors use Expedia as a control for TripAdvisor, constraining the sample to hotels that do not use Expedia's response function. Interestingly, the authors find that valence of reviews increases following managerial response, which they explain is due to a lowered cost of leaving a positive review. We elaborate on the differences between the two studies in Section 7.

Finally, since online reviews in the presence of managerial response can be thought of as a hybrid complaint/word of mouth system, our paper relates to the literature on customer complaint management (see Hirschman (1970), Fornell and Wernerfelt (1988), Fornell and Wernerfelt (1987)). We discuss these papers in more detail in the Theory Section below.

3 Theoretical Relationship between Managerial Response and Subsequent User Reviews

We examine the effect of initial managerial response on the incentive to post assuming that posters of reviews are motivated by altruism and impact. The primary audience for a review in the "before" period are potential customers who are considering the product. A review may benefit a potential customer by impacting her purchase decision. Hence both the altruism and the impact motivations are present in this period and form the potential reviewer's baseline incentive to post.

The first managerial response credibly signals to potential reviewers that the manager is listening to consumer feedback. Hence, in the "after" period, in addition to potential customers, the audience for a review also consists of the manager. That is, while the firm always has the ability to collect and analyze customer reviews,[3] the ability to respond allows the firm to credibly signal that it is reading reviews and responding to suggestions therein. Since the review may now also help the manager and impact product quality (which is variant in this setting), we expect to see an *increase* in both the altruism and the impact motivations of potential reviewers over the baseline. We expect this increase to lead to both an increase in the incidence and the quality of reviews.

Hypothesis 1 *Managerial response increases the probability that a review will be posted.*

Hypothesis 2 *Managerial response increases the effort put into posting reviews, operationalized here as the length of reviews posted.*

[3] The fact that consumer reviews contain managerially-useful information is illustrated by White (2014) which discusses the trend of hotels using customer reviews as "blueprints" for renovation.

The dynamic nature of the product is an essential condition for the results above to hold since in this setting reviews can impact the manager's actions. Firms may, of course, be an audience for reviews in a static setting as well in that they may consider reviews in designing future products. However, given that the products being reviewed will not change as a function of the review, the owner or manager of the product is likely not considered by consumers to be the primary audience for the review. In contrast, in a dynamic setting the manager can take direct actions to improve quality.

Our next hypothesis is based on another application of the "impact" principle. In an environment with managerial response, there are two reasons why reviewers may perceive that negative reviews will have more potential impact than positive reviews and thus be differentially incented to post negative versus positive reviews.

First, let's consider the potential review's impact on product quality. It seems a priori clear that reviewers who point out a product's flaw expect to have a bigger impact on subsequent product quality than reviewers who bring up positive points. Negative reviews of managerially-dependent hotel attributes implicitly suggest a potential change in the the manager's behavior. For example, a negative review may result in a manager fixing the bathrooms or training front desk staff. Positive reviews do not. This reasoning is very consistent with Hirschman (1970) who proposed that an unhappy loyal customer may exercise voice in order to improve the product.

Second, let's consider the potential review's impact on managerial response behavior. One of the primary motivations for a manager to respond to a review may be to encourage potential consumers to view the experience described in a negative review as unlikely to be repeated. Indeed, much of the online discussion about managerial response strategies emphasizes the importance of responding to negative reviews and advice about how to respond to them.[4] We will demonstrate that overall, more negative reviews are more likely to receive managerial responses on reviewing platforms than are more positive reviews, and that responses to negative reviews tend to be much longer (and therefore, presumably more substantive) than the responses to more positive reviews. However, this general pattern in response behavior magnifies the differential impact of negative reviews.

This effect is also very much related to the idea of "audience tuning" (Berger (2014)). That is, the sender of word of mouth may change the

[4]See for example the many articles that TripAdvisor has written in its "TripAdvisor Insights" article series on the topic of managerial responses to negative reviews, http://www.tripadvisor.com/TripAdvisorInsights/t16/topic/management-responses.

information that she shares based on the needs and desires of her audience. In this sense, when the manager publicly becomes an audience member, reviewers are more likely to share more negative reviews which may be more helpful and are more likely to receive a response.

Hypothesis 3 *Managerial response decreases the valence of reviews posted.*

There are alternative hypotheses that one may hold about the motivations for reviewing. For example, one possibility, also related to Hirschman (1970), is that reviewers are motivated to write negative reviews in order to solicit individual compensation for their negative experiences. For example, Ma et al. (2015) finds that redress-seeking is a major force in driving complaints on Twitter. In fact, Ma et al. (2015) finds that a service intervention on the part of the firm may result in more complaints being posted on Twitter. There are indeed many situations in which this motivation may be operative, but we believe it is muted in this context since the platforms that we study use public reviews and public managerial responses. Offers of specific individual remuneration in these responses appear to be rare. However, this may be operable elsewhere. For example, Yelp is a prominent example of a reviewing platform that allows private communication. Yelp also (unusually) allows reviewers to change their reviews, and the private communication channel is often used to encourage reviewers to change a review.[5]

4 Data

The starting point of our data collection efforts is the identification of 50 focal cities. Similarly to Mayzlin et al. (2014), we identified the 25th to 75th largest US cities to include in our sample. Our goal was to use cities that were large enough to have many hotels, but not so large and dense that competition patterns amongst hotels would be difficult to determine. Using data available on the TripAdvisor website in mid-2014, we identified all hotels that TripAdvisor identifies as operating in these focal cities. We also obtained data from Smith Travel Research, a market research firm that provides data to the hotel industry (www.str.com). STR attempts to cover the universe of hotels in the US. We use name and address matching to hand-match the TripAdvisor data to STR.

[5]See, for example, the Yelp Official Blog, http://officialblog.yelp.com/2011/03/tactics-for-responding-to-online-critics-new-york-times-boss-blog.html.

From STR, we obtain many characteristics of hotels including their class gradings of hotels. STR grades hotels by brand and objective characteristics into six quality tiers. For this paper, we exclude the bottom two quality tiers. The excluded chains consist largely of roadside budget motel chains such as Red Roof Inn, Super 8, Motel 6, Quality Inn, and La Quinta Inns. We include hotels in STR's "Upper Midtier" range and higher. "Upper Midtier" includes Holiday Inn, Hampton Inn, Fairfield Inn, and Comfort Inn. We focus on "Upper Midtier" hotels and higher because "Economy Class" and "Midscale Class" hotels have significantly fewer reviews per hotel, and hotel shoppers can perhaps be expected to be less quality-sensitive. Our sample of hotels is more homogeneous. Furthermore, the Revinate data that we will use for reviews has much better coverage for "Upper Midtier" hotels and higher. In total, we obtain a sample of 2104 hotels that match between STR and TripAdvisor in the "Upper Midtier" or higher categories.

Finally, our main source of review data comes from Revinate, a guest feedback and reputation management solutions provider. Among other services, Revinate provides client hotels a Review Reporting Dashboard. Client hotels can view daily social media feed from all major reviewing sites on one page, view the equivalent feed for their competitors, and respond to reviews from multiple sites from the single interface. In order to provide this service, Revinate has invested in creating robust matching of the same hotel across multiple reviewing platforms.

Revinate has excellent coverage of the 2104 hotels in our target sample. Revinate has substantial market share; they serve more than 28000 client hotels worldwide.[6] Many large chains subscribe to Revinate services for all of their hotels. Crucially for us, they track not only their clients, but a large group of client competitors. Overall, of the 2104 hotels in our target sample, we are able to obtain Revinate data for 88 percent of them, for a total of 1843 hotels.

Even with this excellent coverage, the imperfect coverage presents a selection bias. However, note that the hotels that we track contain both Revinate clients and non-clients. The 261 hotels that we could not track through Revinate are clearly not Revinate clients. In the analysis that follows, we will undertake weighted specifications in which the non-clients receive more weight to equate the weight on non-clients in the sample to the weight of non-clients in the overall population. Note that because Revinate's coverage is extensive, the extra weight assigned to non-clients is relatively small. All

[6]This number was obtained from a conversation with Revinate executives in January of 2016.

our results are substantially the same with or without the weighting scheme.

The Revinate data contains full text review information with date and time stamps for every review for the sites that Revinate tracks for our 1843 hotels. We examine the flow of reviews for a six year period; the earliest review in our sample is posted on January 1st, 2009, and the the latest review is posted on Dec. 8th, 2014. Revinate also collects full text of all managerial responses, again with date and time stamps. The time period we study is well-suited to our question. Managerial response increased in popularity over the time period of our data. In early 2010, USA Today reported that fewer than 4 percent of negative reviews on TripAdvisor get a response (according to TripAdvisor), but that utilization of the function increased 203 percent in 2009.[7] There are a few data issues with the data that we receive from this provider. Due to a peculiarity of the way that Booking.com displays review data, our review data for Booking.com does not contain older reviews for the site. We include some summary information about Booking.com in this section, but do not use Booking data at all in our analyses. Furthermore, due to a reporting format change at Hotels.com, we were unable to obtain the date of managerial response after sometime in the 2011 period. Our summary data here includes all Hotels.com data, but our analysis requiring the response date uses only the earlier period data.

Our primary analysis compares pre-managerial response reviewing to post-managerial response reviewing. However, since our data has a starting point and an end point, there will be some truncation of the time period. That is, there are 31 hotels where the initial managerial response occurred less than 6 weeks from the beginning of our data set and 6 hotels where the initial managerial response occurred less than 6 weeks before the end of our data set. For these hotels, either the "before" or the "after" period is less than 6 weeks. Because all of our analysis uses a difference-in-difference, comparing the difference between pre- and post- across platforms, and given that the exact same truncation occurs for all the platforms, we do not exclude the hotels with truncated pre- or post- periods. However, our results are robust to excluding these hotels from our sample.

Table 1 contains summary statistics that describe the reviewing and response data for our sample of hotels for six of the most popular booking and reviewing sites: Booking.com, Expedia, Hotels.com, Orbitz, Priceline, and TripAdvisor.

Table 1 reveals several interesting stylized facts about the data. First, note that there are no review responses on Booking, Priceline, and Orbitz

[7] See Yu (2010).

Table 1: Summary Statistics - Reviews

	Booking	Expedia	Hotels.com	Orbitz	Priceline	TripAdvisor
Number of reviews						
noncustomer	248065	252781	223450	94321	256143	624555
customer	80596	106207	81122	36973	93841	265944
Response share noncustomer	0	0.056	0.015	0	0	0.457
customer	0	0.132	0.041	0	0	0.517
weighted	0	0.078	0.022	0	0	0.474
Mean Rating						
noncustomer no response	4.142	4.165	4.237	3.892	3.988	4.139
customer noresponse	4.146	4.158	4.241	3.891	4.011	4.194
weighted no response	4.143	4.163	4.238	3.892	3.994	4.155
noncustomer response		3.963	4.131			4.0259
customer response		3.950	4.015			4.102
weighted response	0.000	3.959	4.101	0.000	0.000	4.048
Weighted: response vs no response						
Percentage difference		-4.9%	-3.2%			-2.6%
For the universe of hotels that have responded 5 times before the data of this review:						
Mean Rating						
noncustomer no response	4.142	4.163	4.234	3.887	3.989	4.135
customer no response	4.143	4.156	4.238	3.888	4.012	4.190
weighted no response	4.143	4.161	4.235	3.887	3.995	4.151
noncustomer response		3.964	4.131			4.026
customer response		3.949	4.015			4.102
weighted response		3.960	4.101			4.048
Weighted: response vs no response						
Percentage difference		-4.8%	-3.2%			-2.5%

All weighted means are weighted to overrepresent Revinate non-customers.

because they do not allow responses. TripAdvisor, Expedia, and Hotels.com allow responses. In our sample, roughly half of TripAdvisor reviews receive responses while only about 8 percent of Expedia reviews receive responses and only 2 percent of Hotels.com reviews receive responses.[8]

Revinate customers are much more likely to respond to reviews than non-customers. This is likely partly due to selection; they have demonstrated their interest in social media management by becoming customers. However, this is also likely due to the causal impact of the Revinate platform. The platform makes it very easy to respond to reviews. Notice that the customer vs. noncustomer review response rates are most disparate for non-TripAdvisor sites. This makes sense because the Revinate interface makes it equally easy to view all of the sites and post all of the responses; non-Revinate customers may have more tendency to monitor only the perceived most influential site, TripAdvisor. Again, our reweighting of customers vs. non-customers is important to achieve representativeness.

Important for our purposes is the disparity between the valence of reviews that are responded to and reviews that are not responded to. Reviews that are responded to average 4.0 for TripAdvisor versus 4.2 for reviews that do not receive responses. This disparity is greater for the other sites.

The differences in the valence of reviews responded to versus not responded to is largely due to the reviews selected for response rather than the characteristics of the hotels that respond to reviews. Limiting the sample of hotels to only hotels that have responded to at least five prior reviews (frequent responders) produces similar summary statistics. In unreported specifications, for each of the three sites that allow responses, we take the sample of all reviews as observations, and regress the indicator variable for "manager responded" on the review rating and hotel fixed effect. The coefficient for the review rating is strongly negative and significant. This suggests that hotels are more likely to respond to their more negative reviews.

[8]Toward the end of our data period, in summer of 2013, Hotels.com and Expedia, who have a common owner, initiated some review sharing on the two sites. This data-sharing was not reflected in Revinate's data collection. While we found instances when some reviews from Expedia (which were marked as "Expedia-verified") appeared on Hotels.com, we did not find examples of the reverse occurring. We also found that the policy was not consistent across all hotels - while in one instance only a small subset of all available Expedia reviews was displayed on Hotels.com, in another instance it was a bigger subset. There are several reasons why we think that this merger was not a major issue for us. First, for unrelated reasons, as discussed above, we do not use post-merger data for Hotels.com, so this is not an issue for our Hotels.com estimation. Second, we did not see examples of Hotels.com reviews displayed on Expedia, which suggests that this is not a major issue for our Expedia.com estimation. Finally, this occurred relatively late in our data set.

Table 2 provides summary data on review length of each of the sites. Clearly, there are differences across sites in typical review length. For example, reviews on Booking.com are particularly short and reviews on TripAdvisor.com are particularly long. More negative reviews tend to be significantly more detailed across all sites.

Table 2: Review length and managerial response length

	Booking	Expedia	Hotels.com	Orbitz	Priceline	TripAdvisor
Review stars ≤ 3 review length	221.291	473.814	359.919	464.867	324.862	964.677
Review stars>3 review length	138.711	362.464	229.833	347.809	231.359	683.175
Percentage difference	-37.3%	-23.5%	-36.1%	-25.2%	-28.8%	-29.2%
Review stars ≤ 3 response length		518.891	514.179			703.007
Review stars>3 response length		387.266	361.037			516.219
Percentage difference		-25.4%	-29.8%			-26.6%

All means are weighted to overrepresent Revinate non-customers.

Table 2 also provides summary data on review response length for the three sites that allow review responses. Managers tend to provide longer review responses on TripAdvisor versus the other two sites. Across all sites, managers appear to put more effort into responding to negative reviews. The summary table shows that, for all three sites that allow responses, responses to more positive reviews (greater than three stars) are 25 to 30 percent shorter than responses to more negative reviews.

5 Methodology

As discussed above, we consider the hotels' adoption of managerial response on a particular site to potentially present a discrete change in reviewer incentives. Thus, our methodology focuses on changes in reviewing activity in a very tight time window around the day that responses are first posted by the hotel.

Table 3: Characteristics of first day of responses

	Expedia	Hotels.com	TripAdvisor
Reviews responded to the first day			
Average star Noncustomer	3.54	3.78	3.12
Average star Customer	3.65	3.56	3.20
Average star weighted	3.57	3.72	3.14
Total number responded Noncustomer	1.91	1.69	1.94
Total number responded Customer	1.99	1.32	2.04
Total Number responded weighted	1.93	1.60	1.97

In Table 3, we examine summary statistics on the behavior of hotels on the day of first managerial response posting. The first thing to notice about Table 3 is that managers frequently respond to more than one review the first time that review responses are posted. The average star of reviews responded to, unsurprisingly, are more negative than the overall population of reviews. For example for TripAdvisor, of the 1807 hotels that post responses, 641 of the first day responses are to reviews with an overall average star rating of strictly less than three.

Our identification scheme relies on a differences methodology. Our primary specifications examine the 6 week window before and after the posting of first managerial response. We undertake robustness specifications below, but describe our primary measurement strategy here.

We will describe the platforms in which the manager initially posts a response as the "treatment" platforms. These are TripAdvisor, Expedia, and Hotels.com which we will consider separately in separate specifications (since the response posting time windows are different for the three sites). The "control" platforms are Orbitz and Priceline.

First, in order to control for contemporaneous factors that may cause within-platform changes in reviews in a geographic area, we straightforwardly difference the before vs. after review measurements from the before vs. after review measurements for the geographic area. We construct the geographic mean for each TripAdvisor geocode imposing no restrictons on the hotels included in the mean calculation. Our measure of review changes measures the difference over time from the geographic mean for both treatment and control platforms. Thus, for example, in all specifications that use the number of TripAdvisor reviews in a time window, TripAdvisor reviews for the observation hotel is calculated as the difference between TripAdvisor reviews in the time window minus the average number of reviews for all other hotels in the same city for the same time window.

We have two platforms that we use as controls, Orbitz and Priceline. These controls are meant to capture other physical investments in quality that the hotel makes concurrently with response adoption. An important issue in using one platform as a control for another is that the platforms certainly may cater to different populations. This might be particularly true for TripAdvisor vs. the other platforms, as TripAdvisor is primarily a review platform while Expedia, Orbitz, and Priceline are booking platforms. In addition, the scales of the platforms are very different. In including two controls, we allow the data to "choose" the combination of platforms that are the best control. Also, calculating a control platform-specific coefficient captures differences in scale across platforms. Thus, for any review measurement constructed for our treatment sites TripAdvisor, Expedia, and Hotels, we construct the corresponding measure for both Orbitz and Priceline and use these as control variables in our regression specifications.

Table 4 summarizes the cross-sectional correlation in number of reviews, review valence, and review length across sites but within hotels. Our identification strategy is predicated on the idea that hotel characteristics will be similarly measured across sites. Table 4 suggests high correlation among the sites for all of the measures. The correlation across sites is largest for review valence and smallest for review length. Hotels that inspire longer reviews on one site tend to receive longer reviews on other sites, but the effect is modest in magnitude. Our two control sites, Orbitz and Priceline are less correlated with each other across each of the measures than are any other pair of sites. This suggests that there is plausibly different information about the hotel captured by using each of them as a control.

Recall that our review measures of interest are: change in the number of reviews between the six week windows (to test Hypothesis 1), change in the valence of reviews (to test Hypothesis 3), and changes in measures of the length of reviews (to test Hypothesis 2). For each measure, M, for hotel i in city j, our simple estimating equation is:

Table 4: Correlations within hotel and across sites

Review count correlations					
	TripAdvisor	Expedia	Hotels.com	Priceline	Orbitz
TripAdvisor	1.00				
Expedia	0.78	1.00			
Hotels.com	0.63	0.58	1.00		
Priceline	0.38	0.34	0.31	1.00	
Orbitz	0.67	0.41	0.29	0.24	1.00

Average star correlations					
	TripAdvisor	Expedia	Hotels.com	Priceline	Orbitz
TripAdvisor	1.00				
Expedia	0.81	1.00			
Hotels.com	0.71	0.80	1.00		
Priceline	0.72	0.77	0.72	1.00	
Orbitz	0.63	0.66	0.60	0.59	1.00

Length correlations					
	TripAdvisor	Expedia	Hotels.com	Priceline	Orbitz
TripAdvisor	1.00				
Expedia	0.55	1.00			
Hotels.com	0.45	0.58	1.00		
Priceline	0.30	0.34	0.31	1.00	
Orbitz	0.34	0.41	0.29	0.24	1.00

$$
\begin{aligned}
&(M_{ij}^{Treat} - \overline{M}_j^{Treat})_{Post} - (M_{ij}^{Treat} - \overline{M}_j^{Treat})_{Pre} \quad = \\
\alpha + [&(M_{ij}^{Price} - \overline{M}_j^{Price})_{Post} - (M_{ij}^{Price} - \overline{M}_j^{Price})_{Pre}]\beta_1 + \\
[&(M_{ij}^{Orbitz} - \overline{M}_j^{Orbitz})_{Post} - (M_{ij}^{Orbitz} - \overline{M}_j^{Orbitz})_{Pre}]\beta_2 + \epsilon_{ij}
\end{aligned} \tag{1}
$$

In this equation, "Pre" denotes the six week period prior to first managerial response on the treatment site for hotel i, "Post" denotes the six week period following the first managerial response on the treatment site. The treatment site, Treat∈(TripAdvisor,Expedia,Hotels.com) and $\overline{M}_j$ denotes the city-average variable. Note that, for each specification, α is our variable of interest, as it is the change in the variable net of the changes in the controls. For example, consider the change in number of reviews for TripAdvisor following the first response. α measures the extent to which the number of reviews on posted on TripAdvisor increases above and beyond corresponding changes in the number of reviews on Priceline and Orbitz, as well as corresponding geographic averages.

Our identification strategy is helpful in overcoming several potential endogeneity challenges. (See Figure 2 for the summary of endogeneity concerns and our solutions to these concerns). When will our strategy fail? The primary weakness of our strategy is that it is possible that a hotel-specific time-specific platform-specific factor is correlated with both the initiation of managerial response for hotel i and with the future review process for hotel i on that specific platform. We find the possibility of one category of such confounds slightly more plausible for Expedia and Hotels.com than for TripAdvisor.com. This is because Expedia and Hotels.com are both booking sites, rather than purely review sites.[9] It is possible, for example that hotels systematically simultaneously commence participation in hotel-specific promotions on Expedia or Hotels.com and initiate managerial response on that site. Such a promotion might plausibly lead, through an increase in platform-specific bookings, to an increase in reviewing activity on those sites. In contrast, even if a hotel undertook some kind of promotion on

[9]Note that that the booking versus information only site distinction also leads to differences in how the platform can elicit reviews. For example, Expedia, Orbitz and Priceline send reminders to their customers to review their recent stays. In contrast, many of the reviews posted on TripAdvisor describe trips that were booked on other platforms. Hence, the majority of the reviews on TripAdvisor were posted spontaneously and not due to a reminder. We also expect these difference to affect the timing of review posting across sites. Since there are many observable and unobservable differences across platforms, it is important for us to take differences over time since they keep the platform constant.

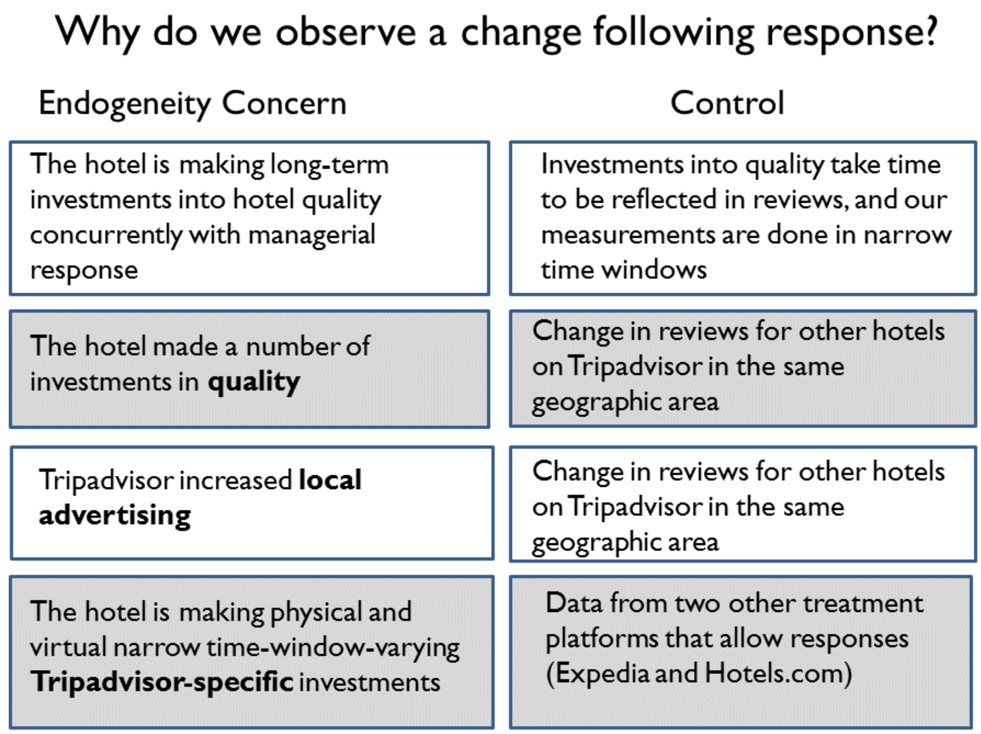

Figure 2: Our Methodology and Endogeneity Concerns

TripAdvisor, the consumer would click out to a different site to book the reservation. Hence, there is a weaker connection between promotion, the number of bookings and the number of reviews. While this is a concern, we note three things about this concern. First, many of the plausible confounds that we can think of (such as promotions) might lead people to book more rooms in the short run, but the booking activity should somewhat more slowly work its way into staying and reviewing activity (since many people book for a stay occurring somewhat in the future). Second, it is not entirely clear why such a promotion would also lead people to be systematically differentially satisfied or unsatisfied with their experience; that is, there is not a natural prediction for review valence. Third, there is also not a natural prediction for review length.

A second scenario that would undermine our strategy is the possibility that hotels that commence review response on a site are "getting organized" about catering to the users of the site more generally. This might be a particular concern for TripAdvisor, since TripAdvisor has become so important in the industry. Several factors, we believe, mitigate this concern in our setting. First, using the same logic that TripAdvisor reviews are particularly important and salient in the industry, this should be less of an issue for Expedia and Hotels.com. Second, the natural bias stories of this type seem to cut in the opposite direction of our hypothesis and findings. If a hotel were launching a coherent TripAdvisor management system with immediate implications, it is hard to see how that would systematically lead to a decrease in review valence. Third, there is not a clear natural prediction of this possibility for review length.

6 Results

Table 5 presents summary statistics of the variables used to estimate Equation 1.

Table 6 provides estimates of Equation 1 where TripAdvisor is the treatment site. Column 1 examines the variable of primary interest to us, the change in the number of reviews for the sample of 1807 first responders on TripAdvisor. This provides our primary test of Hypothesis 1. Note that both the Orbitz control and the Priceline control coefficients are positive and significant at least at the ten percent level. The number of reviews positively comoves across the sites. This is not surprising, as presumably all of the sites get more reviews in periods of particularly high occupancy for the hotel relative to the local area. Note that both Orbitz reviews and Priceline reviews also experience review increases over the period. The constant term is positive and statistically significant at the one percent level. This suggests that, net of the controls, the number of reviews increases by 1.2 reviews. The average number of reviews in the pre-review period is 4.5 (from Table 5), so this represents a substantial sudden increase in reviewing activity relative to the controls.

Column 2 examines average stars, which pertains to Hypothesis 3. When considering average stars, we restrict the sample to hotels that had reviews on Tripadvisor in both the pre- and post-six week windows. We see that average stars decrease significantly for hotels that have commenced managerial response. This is consistent with our hypothesis that reviewers respond with more substantive reviews when they receive feedback that managers are listening. Interestingly, our results contrast Proserpio and Zervas (2016) in this regard. The point estimate is -0.06. This may seem small, but recall that

Table 5: Summary statistics for six week difference variables

Summary statistics for first TripAdvisor response	Number of Reviews	Average Stars	Average Length
TripAdvisor average	4.50	4.01	762.60
(sum six weeks before)	(8.29)	(0.82)	(412.21)
Diff in Diff TA	1.27	-0.06	50.61
	(5.64)	(0.94)	(514.00)
Diff in Diff Orbitz	0.14	0.00	6.54
	(2.96)	(2.35)	(344.55)
Diff in Diff Priceline	0.12	0.05	3.56
	(2.78)	(2.28)	(174.78)

Summary statistics for first Expedia response	Number of Reviews	Average Stars	Average Length
Expedia average	4.97	4.19	351.13
(sum six weeks before)	(6.55)	(0.66)	(191.49)
Diff in Diff Expedia	0.33	-0.07	9.94
	(4.59)	(0.81)	(239.75)
Diff in Diff Orbitz	-0.04	-0.13	-5.76
	(2.20)	-(3.34)	(364.92)
Diff in Diff Priceline	0.12	0.07	16.59
	(3.64)	(2.20)	(255.96)

Summary statistics for first Hotels.com response	Number of Reviews	Average Stars	Average Length
Hotels.com	9.49	4.25	236.77
(sum six weeks before)	(13.19)	(0.56)	(132.19)
Diff in Diff Hotels.com	1.06	0.00	24.23
	(8.57)	(0.61)	(207.88)
Diff in Diff Orbitz	-0.04	0.30	-1.19
	(3.61)	(2.10)	(337.62)
Diff in Diff Priceline	-0.26	0.19	8.04
	(4.24)	(2.16)	(192.40)

Standard deviations are in parentheses. Observations are weighted using the probability weight for Revinate customers vs noncustomers.

Table 6: TripAdvisor changes in reviewing activity

VARIABLES	(1) Num Reviews	(2) Avg Star	(3) Avg Length	(4) Length 1/2/3 star reviews	(5) Length 4/5 star reviews
Orbitz controls	0.117**	0.001	-0.015	0.039	0.007
	(0.059)	(0.011)	(0.048)	(0.079)	(0.048)
Priceline controls	0.110*	-0.003	-0.020	0.104	0.089
	(0.061)	(0.013)	(0.076)	(0.224)	(0.074)
Constant	1.240***	-0.064**	50.783***	19.492	31.235**
	(0.130)	(0.027)	(14.781)	(32.402)	(15.548)
Observations	1,807	1,210	1,210	516	1,037

Robust standard errors in parentheses
Regressions weighted to overrepresent Revinate non-customers
*** p<0.01, ** p<0.05, * p<0.1

the average stars obtained in the pre-period are 4.01 with a cross-sectional standard deviation of only 0.8. Thus, small movements in the average stars can significantly move a hotel up or down the rank ordering of hotels in a local area. An important issue to consider is that the Orbitz and Priceline controls are not statistically significant in the regression. The hotels also have essentially no average star movement of Priceline or Orbitz over the two six week windows. This is to be expected if underlying quality at the hotel is, indeed, not significantly changing.

Column 3 displays results for the average length of review, with Columns 4-5 demonstrating review length for reviews of different star levels. These results pertain to Hypothesis 2. Again, we restrict the sample to hotels that had reviews in both the pre- and post- period on Tripadvisor. Review length has not changed significantly over the 6 week windows for Priceline and Orbitz, suggesting that perhaps nothing has changed at the hotel that fundamentally leads consumers to want to leave longer reviews. However, reviews on TripAdvisor increase significantly. The point estimate of the increase is 51 characters. This is a large change given the average length of reviews on TripAdvisor for the period prior to the managerial response is 763 characters. We examine the change in review length for negative valence (1,2, and 3 star) reviews and positive valence (4 and 5 star) reviews separately. In each case, we must restrict the sample to hotels that have reviews in that category in both the pre and post windows on TripAdvisor. This increase in reviewing effort is estimated to be positive both for negative and positive valence reviews, although is is statistically different from zero only for the 4 and 5 star reviews.

Table 7 provides estimates of Equation 1 where Expedia is the treatment site. Of course, the sample of review responses is much smaller. Interestingly, the number of reviews that the hotel has in the six weeks prior to the first response is about the same for Expedia as for TripAdvisor, suggesting that responders on Expedia tend to be hotels that garner a lot of Expedia reviews. Despite the smaller number of observations, we find somewhat similar, albeit weaker, results. The increase in the number of reviews net of the controls is 0.3, about one-quarter of the magnitude of the effect estimated for TripAdvisor, but still significantly different from zero at standard confidence level. The decrease in review valence of 0.07 is very similar to the TripAdvisor results. The increase in review length is small and insignificant, however.

Table 8 provides estimates of Equation 1 where Hotels.com is the treatment site. Here, the sample of responding hotels is again quite small, as we have only 332 hotels that provide responses on Hotels.com or for which the date of the first managerial response date is available. Nonetheless, we do find a significant increase in the number of reviews posted, no measurable change in average star, a small (and significant at the ten percent level) increase in review length overall.

In sum, we interpret the results in Tables 6, 7, and 8 as demonstrating that reviewing activity increases following managerial response, that valence decreases (at least modestly), and that the length of reviews (a measure of reviewing effort) increases.

Should we conclude based on the valence result that a hotel is better off not responding to reviews? Given the popularity of response among hotels (including the hotel's competitors), not responding at all may not be feasible. Furthermore, it is quite likely (and untestable given current data) that consumer purchasing behavior responds differently to a given review depending on whether it has received a thoughtful response. Instead, we explore an aspect of the manager's response strategy that is under her control – the extent to which managers respond to positive versus negative reviews. That is, so far we have hypothesized that perceived higher impact of negative reviews combined with the empirical regularity that negative reviews

Table 7: Expedia changes in reviewing activity

VARIABLES	(1) Num Reviews	(2) Avg Star	(3) Avg Length	(4) Length 1/2/3 Star reviews	(5) Length 4/5 Star reviews
Orbitz controls	0.135	0.006	0.026	-0.051	-0.005
	(0.086)	(0.012)	(0.024)	(0.054)	(0.032)
Priceline controls	0.086*	0.017	-0.029	0.007	-0.017
	(0.052)	(0.013)	(0.046)	(0.077)	(0.053)
Constant	0.329**	-0.072**	10.568	-30.860	9.557
	(0.145)	(0.030)	(9.055)	(24.113)	(9.056)
Observations	984	709	709	275	652

Robust standard errors in parentheses
Regressions weighted to overrepresent Revinate non-customers
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Hotels.com changes in reviewing activity

VARIABLES	(1) Num Reviews	(2) Avg Star	(3) Avg Length	(4) Length 1/2/3 Star Reviews	(5) Length 4/5 Star Reviews
Orbitz controls	0.412***	0.004	0.028	-0.036	-0.012
	(0.132)	(0.019)	(0.038)	(0.077)	(0.038)
Priceline controls	0.159	-0.022	-0.032	0.128	-0.026
	(0.135)	(0.019)	(0.055)	(0.096)	(0.065)
					(0.065)
Constant	1.116**	0.001	24.515*	12.047	19.924
	(0.460)	(0.038)	(12.835)	(30.852)	(13.860)
Observations	332	265	265	136	256

Robust standard errors in parentheses
Regressions weighted to overrepresent Revinate non-customers
*** p<0.01, ** p<0.05, * p<0.1

are more likely to receive a response by the managers imply that managerial response results in a decrease in subsequent review valence. Another implication of our reasoning is that a manager who is more likely to favor responses to negative over positive reviews will amplify this effect further. That is, by disproportionately responding to reviews that contain criticisms, an individual manager signals to consumers that criticisms will be read by management, possibly responded to, and possibly acted upon.

Below we explore whether we observe this regularity in the data. While overall, more negative reviews are more likely to receive managerial responses and more likely to receive more substantive managerial responses, the extent that this is true will vary across managers and hotels. We explore whether the hotels that are more likely to respond to negative reviews are also more likely to experience a larger fall in subsequent review valence.

Using our large sample of TripAdvisor first responses, we consider hotels first-day reviewing strategy. Note that this analysis is purely correlational and not causal since we do not know why hotels pursue different first-day reviewing strategies. Nonetheless, this exploratory analysis is suggestive of a pattern consistent with our main hypotheses, and in fact serves as a robustness check on our proposed mechanism.[10]

In Tables 9 and 10, we reestimate Equation 1 for TripAdvisor but separate the data into two separate groups. First, in Table 9, we consider the set of TripAdvisor respondees whose first day responses are to reviews with an average star value of less than three. Second, in Table 10, we consider the disjoint set of TripAdvisor respondees whose first day responses are to reviews with an average star value of three or more. For the 641 hotels that respond to low average star reviews in Table 9, we see that, in the six weeks after the first response, the number of reviews increases significantly, review valence decreases significantly and average length increases. The decrease in review valence is roughly double the magnitude of our estimate of the review valence differences using the overall sample. On closer inspection of the pattern of review length increases, review length increases are positive but insignificant for the two valence categories. Overall, the results in this table are suggestive that managerial responses that concentrate on responding to negative reviews encourages more and more detailed negative reviews from consumers.

[10] As another check on our proposed mechanism that posters are motivated by the impact of their reviews, we conduct an exploratory LDA analysis to investigate whether the topics of reviews change after managerial response. We find that the largest increase after the initiation of managerial response was in the propensity of reviews to discuss issues that are related to staff and service quality.

Table 9: Changes in reviewing activity for TripAdvisor hotels who respond to reviews with an average star value of less than three

VARIABLES	(1) Num Reviews	(2) Avg Star	(3) Avg Length	(4) Length 1/2/3 Star Reviews	(5) Length 4/5 Star Reviews
Orbitz controls	0.093	-0.004	-0.046	0.085	0.002
	(0.071)	(0.016)	(0.075)	(0.112)	(0.077)
Priceline Controls	-0.035	-0.000	-0.041	-0.091	0.107
	(0.055)	(0.017)	(0.108)	(0.225)	(0.107)
Constant	0.724***	-0.143***	62.186***	42.942	25.324
	(0.157)	(0.037)	(20.537)	(41.472)	(20.336)
Observations	968	673	673	305	564

Robust standard errors in parentheses
Regressions weighted to overrepresent Revinate non-customers
*** p<0.01, ** p<0.05, * p<0.1

For the 839 hotels that respond to higher average star reviews in Table 10, we see that, in the six weeks after the first response, the number of reviews increases significantly, review valences are unchanged (the point estimate is positive) and average review length increases modestly. The contrast between Table 9 and Table 10 is suggestive that negative reviewing activity is differentially stimulated when managers use the response function exclusively to respond to negative reviews. Again, this makes sense in an environment in which consumers post reviews in order to have an impact on hotel quality.

7 Relationship to Previous Literature

In this Section we highlight the contribution of our paper and reconcile the findings with the existing literature: Ma et al. (2015) and Proserpio and Zervas (2016).

Ma et al. (2015) finds that a service intervention by the firm may actually encourage negative redress-seeking tweets. This is in spirit related to our finding that a managerial response results in lower-valenced reviews. Despite the seeming similarity, there are also important differences between the two

Table 10: Changes in reviewing activity for TripAdvisor hotels who respond to reviews with an average star value of three or more

VARIABLES	(1) Num Reviews	(2) Avg Star	(3) Avg Length	(4) Length 1/2/3 Star Reviews	(5) Length 4/5 Star Reviews
Orbitz controls	0.161	0.007	0.012	-0.008	0.012
	(0.104)	(0.016)	(0.061)	(0.120)	(0.061)
Priceline controls	0.269**	-0.009	-0.005	0.348	0.071
	(0.113)	(0.018)	(0.108)	(0.402)	(0.105)
Constant	1.806***	0.034	37.490*	-9.838	38.204
	(0.210)	(0.040)	(20.990)	(50.474)	(24.044)
Observations	839	537	537	211	473

Robust standard errors in parentheses
Regressions weighted to overrepresent Revinate non-customers
*** p<0.01, ** p<0.05, * p<0.1

settings and the resulting mechanisms. That is, while Ma et al. (2015) examines concrete service interventions, here we examine communication only. Also, as we discuss above, while it is possible that redress-seeking is present in our setting, it is clearly not as central a motive as it is in Ma et al. (2015). There are many online platforms where word of mouth is exchanged, and where a service intervention is either not possible (due to the anonymous nature of the forum) or is not advisable (perhaps because the firm does not want to encourage excessive redress-seeking), but a firm may still be able to respond to feedback. In this sense our setting is broad.

Proserpio and Zervas (2016) find an increase in review valence on TripAdvisor relative to Expedia following a managerial response on Tripadvisor in a sample of Texas hotels. We discuss how differences in our methodologies could have contributed to different results. First, we use Priceline and Orbitz, both of which do not allow managerial responses, as controls rather than Expedia as is done in Proserpio and Zervas (2016). In fact, we use Expedia in one of our treatment analyses since it allows managerial response. While managerial response is less frequent on Expedia than TripAdvisor, it is not negligible. In Proserpio and Zervas's sample of Texas hotels, 17.5 percent of the hotels that receive any Expedia reviews eventually post a reply and have to be discarded; in our sample the majority of hotels with Expedia reviews eventually post at least one reply. More fundamentally, since both sites allow managerial response, in the sample in Proserpio and Zervas (2016), a hotel chooses to post a reply on TripAdvisor and not on Expedia. This raises endogeneity concerns. For example, it could be the case that the hotel favors TripAdvisor over Expedia customers in other ways, which would result in more and more positive reviews arriving on TripAdvisor compared to Expedia. In our paper this concern is lessened since 1) our control sites do not allow managerial reviews, and 2) we investigate several treatment sites besides TripAdvisor. That is, as we argue above, the "getting organized on TripAdvisor" effect becomes less plausible when applied to three different platforms. In contrast, the "getting organized on TripAdvisor" effect is more plausible in Proserpio and Zervas's environment where hotels have chosen to respond on TripAdvisor and have specifically chosen not to respond on Expedia.

Another important difference between our research design and that of Proserpio and Zervas (2016) is that they use all hotel types while we restrict our sample to upper Midtier and above hotels. The categories that we exclude are the "motel" categories Economy and Midscale. While both samples are valid, they differ in interesting ways. In our market areas, there are a total of 3757 hotels for which we have data from both TripAdvisor and data from STR on hotel characteristics. Of these, 2104 are in our target sample of Upper Midtier and above. These hotels represent 56 percent of the total hotel units but 74 percent of the total rooms in the sample, reflecting the smaller size of most motels. The motel sample that we exclude also tends to have less reviewing activity; this may reflect a lower sensitivity to quality and may also reflect the fact that motel nights are more likely to be booked the day of stay rather than in advance through search platform such as those we study. In our data, Upper Midtier and above hotels account for 56 percent of the hotel units in our data, but 81 percent of the total reviews. Hence we would argue that our sample is more representative of the online hotel marketplace.

In order to explore more thoroughly the issue of the role of the hotel tier, we examine potential heterogeneity of our results across hotel types. It is only feasible to do this for our TripAdvisor sample given the lower number of review responses at Expedia and Hotels.com. Recall that STR classifies all hotels into tiers using largely time-invariant characteristics. The tiers that we use in this paper are Luxury, Upper Upscale, Upscale, and Upper Midscale. Even among the tiers we use, there is considerable heterogeneity in the type of hotel and the characteristics of the customers. The lowest tier we use, Upper Midscale, includes ordinary traveler hotels such as Hampton Inns or Fairfield Inns. The Luxury category includes hotels such as the Four Seasons and the Ritz-Carlton. It is reasonable to expect that reviewing dynamics and the role of managerial responses will differ across those hotels.

We investigate this in Table 11 where we replace the constant term in each specification of Equation 1 with indicator variables for hotel type. Of the 1807 responding hotels, 560 are Upper Midscale, 635 are Upscale, 435 are Upper Upscale, and 149 are Luxury. Eventual responders represent 70 percent of the Upper Midscale hotels in our sample and over 90 percent of the other classes of hotels.

The results suggest that the change in reviewing activity is monotonically increasing in the ex ante hotel quality. That is, the increment to reviews is greatest for luxury hotels. However, the increases in review length are monotonically decreasing in the quality tier of hotels. That is, review length is particularly stimulated for Upper Midscale Hotels. The results for valence are mixed as large significant decreases in review valence are only found for the Luxury and Upscale hotels (although we do not find evidence for valence increases for any hotel type).

In summary, while our paper builds on the existing literature, we believe that it also makes concrete contributions above and beyond current work.

8 Robustness

We undertake a number of robustness specifications; for the robustness specifications, we focus on TripAdvisor, as we have a larger number of responders on TripAdvisor.

First, we investigate using a longer window both before and after the managerial response. While the narrow window is desirable in order to separate our hypotheses from longer-run changes in hotel quality, it raises the possibility that the reviewers have not had time to fully "react" to the initiation of managerial response by the end of the six week window. In Table 12, we undertake the basic specifications of Equation 1 for TripAdvisor, allowing a 10 week window before and after the managerial response. The cost of a longer window is that it is more plausible that long-run investments in hotel quality that could be systematically coincident with the advent of managerial response are being experienced by consumers. The benefit of a longer window is that, of course, over short time periods reviews and the correlation of the reviews across sites will be noisier.

Table 11: Heterogeneity in review responses across hotel classes

VARIABLES	(1) Num Reviews	(2) Avg Star	(3) Avg length	(4) Length 1/2/3 Star reviews	(5) Length 4/5 Star reviews
Priceline controls	0.097	-0.004	-0.816	0.037	0.006
	(0.060)	(0.013)	(5.880)	(0.079)	(0.048)
Orbitz controls	0.109*	0.001	-0.014	0.107	0.101
	(0.057)	(0.011)	(0.075)	(0.225)	(0.074)
Luxury class	2.143***	-0.129**	3.767	32.848	-63.935
	(0.556)	(0.057)	(43.256)	(102.777)	(40.269)
Upper Upscale Class	1.987***	0.013	31.283	-7.630	15.282
	(0.347)	(0.044)	(27.520)	(46.301)	(31.625)
Upscale Class	0.965***	-0.140***	56.468**	81.052	29.319
	(0.186)	(0.050)	(25.789)	(70.591)	(26.284)
Upper Midscale Class	0.788***	-0.040	78.070***	-7.604	89.169***
	(0.184)	(0.057)	(28.078)	(62.779)	(28.776)
Observations	1,807	1,210	1,210	516	1,037

Robust standard errors in parentheses
Regressions weighted to overrepresent Revinate non-customers
*** p<0.01, ** p<0.05, * p<0.1

The results in Table 12 look very similar to our base results in Table 6. The constant term in the number of reviews specification has increased from 1.24 to 2.27. Recall that we are measuring changes in reviews in levels; if the treatment effect of managerial response is permanent, we would anticipate roughly a 67 percent increase in the coefficient given the 67 percent increase in the time period. Our estimates represent an 83 percent increase in the number of reviews and we certainly cannot reject a 67 percent increase in the number of reviews. The average star measure is nearly identical in Table 11 and Table 6 as are the review length results.

Table 12: TripAdvisor specifications- longer time window

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Num Reviews	Avg Star	Avg length	Length 1/2/3 Star reviews	Length 4/5 Star reviews
Orbitz controls	0.108*	0.008	-0.034	0.059	0.039
	(0.055)	(0.010)	(0.047)	(0.061)	(0.043)
Priceline controls	0.205*	0.005	0.015	0.052	0.024
	(0.105)	(0.011)	(0.063)	(0.121)	(0.061)
Constant	2.270***	-0.059**	35.761***	14.992	29.455**
	(0.209)	(0.023)	(13.188)	(26.845)	(12.769)
Observations	1,807	1,385	1,385	738	1,248

Robust standard errors in parentheses
Regressions weighted to overrepresent Revinate non-customers
*** p<0.01, ** p<0.05, * p<0.1

We conduct a few other tests to investigate possible sources of misspecification. First, our empirical design uses an "event study" framework in which we look at a narrow window before and after the posting of managerial response. This design envisions the posting as a discrete shift in reviewer behavior. However, a plausible alternative explanation is that our empirical observations represent a continuation of a trend that commenced prior to the review response date. That is, for some reason, reviews were getting more numerous, longer, and more negative over time for the hotels that responded relative to those that did not (and on TripAdvisor relative to the other platforms). Therefore, a shift that we attribute to managerial response may instead arise simply due to a time trend.

In order to address this, we devise an alternative specification that specifically takes out the possible time trend by differencing out all the variables with respect to their lags. If the changes that we observe are due to the time trend alone, the de-trended data should no longer yield significant results. We refer to the (6- or 10-week) period after managerial response as "post", the (6- or 10-week) period before managerial response as "pre," and the (6- or 10-week) period prior to that as "pre-pre." Hence, the left-hand side of the specification in Equation 1 becomes:

$$\begin{aligned}&[(M_{ij}^{Treat} - \overline{M}_j^{Treat})_{Post} - (M_{ij}^{Treat} - \overline{M}_j^{Treat})_{Pre}] \quad - \\ &[(M_{ij}^{Treat} - \overline{M}_j^{Treat})_{Pre} - (M_{ij}^{Treat} - \overline{M}_j^{Treat})_{Pre-Pre}]\end{aligned} \tag{2}$$

The independent variables are also differenced accordingly. We perform this specification for both the 6 week window and the 10 week windows described above. The basic results for number of reviews, average star and length are presented in Table 13. The results suggest that our specifications are robust to this new specification, although the average star specification for the six week horizon is not statistically different from zero. Note that the average star specification for the 10 week window is still negative and significant.

Finally, in the original specifications we difference changes with respect to geographic averages. We also re-estimated our main Tripadvisor specification where we difference with respect to geographic-tier averages. The results remain qualitatively the same.

9 Conclusion

Allowing management to respond to user reviews has become a common feature of reviewing platforms, especially platforms geared to services such as hotels. We argue that allowing managerial response can fundamentally change the nature of the reviewing platform if users view themselves to be in a dialogue with management rather than only leaving information for future customers.

Casual empiricism and the advice proffered to hotel managers on the web suggests that managers attempt to use the response function to mitigate the impact of criticism, often by promising change. However, the prior literature suggests that consumers post reviews to have an "impact" on others and that they are more motivated to post when they perceive that they can have more impact. These observations lead to our hypothesis that the managerial response function promotes reviewing generally and

Table 13: Robustness: Time-shifted control specifications

Variables	(1) 6 week Num Rev	(2) 6 week Avg Star	(3) 6 week Avg Length	(4) 10 week Num Rev	(5) 10 week Avg Star	(6) 10 week Avg Length
Orbitz Controls	0.092*	-0.005	0.042	0.056	-0.001	-0.011
	(0.051)	(0.011)	(0.045)	(0.052)	(0.010)	(0.045)
Priceline Controls	-0.013	-0.001	-0.004	0.154**	0.002	0.002
	(0.058)	(0.011)	(0.070)	(0.066)	(0.010)	(0.059)
Constant	1.171***	-0.058	83.544***	1.940***	-0.120***	55.137**
	(0.175)	(0.045)	(25.319)	(0.301)	(0.040)	(21.673)
Observations	1,807	1,025	1,025	1,807	1,165	1,165

Robust standard errors in parentheses
Regressions weighted to overrepresent Revinate non-customers
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

promotes the production of critical reviews specifically. Our empirical results rely on a multiple-difference strategy to address endogeneity issues. Our results are generally supportive of the hypothesis that managerial response encourages critical reviewing.

While stimulating negative reviews is not a favorable outcome from the point of view of the firm, we do not claim that responding to reviews is a poor managerial strategy. Note that we do not study the impact of managerial response on actual bookings. It is possible that managerial responses do lead potential customers to view a hotel's negative reviews in a more favorable light and thus lead potential customers to be more likely to book the hotel conditional on the reviews. This is an interesting area for future research.

These results also have implications for the growing literature on online collaborative information creation. Our results suggest that seemingly small tweaks to the platform design can have measurable implications for consumer reviewing behavior and potentially the utility of reviews for future consumers.

Finally, our results contribute to the literature on the effect of managerial response on customer voice. While ours is not the first paper to address this issue, we believe that our setting, sampling strategy and methodology contribute to robustness and more general applicability of our results.

References

Berger, Jonah (2014) "Word-of-Mouth and Interpersonal Communication: An Organizing Framework and Directions for Future Research," *Journal of Consumer Psychology*.

Brandes, L., D. Godes, and D. Mayzlin (2015) "Controlling for Self-Selection Bias in Customer Reviews," University of Southern California Working Paper.

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

Fornell, C. and B. Wernerfelt (1987) "Defensive Marketing Strategy by Customer Complaint Management: A Theoretical Analysis," *Journal of Marketing Research*, Vol. 24, pp. 337–346.

——— (1988) "A Model for Customer Complaint Management," *Marketing Science*, Vol. 7, pp. 287–298.

Fradkin, Grewal E. Holtz D., A. and M. Pearson (2015) "Bias and Reciprocity in Online Reviews: Evidence from Field Experiments on Airbnb," MIT Sloan School of Management and Airbnb working paper.

Godes, David and Jose C. Silva (2012) "Sequential and Temporal Dynamics of Online Opinion," *Marketing Science*.

Hennig-Thurau, Thorsten, Kevin Gwinner, and Gianfranco Walsh (2004) "Electronic Word-of-Mouth via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themsleves on the Internet?" *Journal of Interactive Marketing*.

Hirschman, A. (1970) *Exit, Voice, and Loyalty*: Harvard University Press.

Horton, John J. and Joseph M. Golden (2015) "Reputation Inlfation in an Online Marketplace," NYU Stern working paper.

Kim, Lim Hyunjung, Woo Gon and Robert A. Brymer (2015) "The effectiveness of managing social media on hotel performance," *International Journal of Hospitality Management*.

Klein, Tobias, Christian Lambertz, and Conrad O. Stahl (2016) "Market Transparency, Adverse Selection, and Moral Hazard," *Journal of Political Economy*.

Ma, L., B. Sun, and S. Kekre (2015) "The Squeaky Wheel Gets the GreaseAn Empirical Analysis of Customer Voice and Firm Intervention on Twitter," *Marketing Science*, Vol. 34, pp. 627–645.

Mayzlin, Dina, Judith Chevalier, and Yaniv Dover (2014) "Promotional Reviews: An Empirical Investigation of Online Review Manipulation," *American Economic Review*.

Moe, W. W. and M. Trusov (2011) "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research*.

Nosko, C. and S. Tadelis (2014) "The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment," University of California at Berkeley Working Paper.

Park, Sun-Young and Jonathan P. Allen (2013) "Responding to Online Reviews: Problem Solving and Engagement in Hotels," *Cornell Hospitality Quarterly*.

Proserpio, Davide and Georgios Zervas (2016) "Online reputation management: Estimating the impact of management responses on consumer reviews," Boston U. School of Management Research Paper.

Sundaram, D.S., Kaushik Mitra, and Cynthia Webster (1998) "Word-Of-Mouth Communications: a Motivational Analysis," in *Advances in Consumer Research*, Vol. 25: Association for Consumer Research, pp. 527–531.

White, Martha C. (2014) "Hotels Use Online Reviews as Blueprint for Renovations," *New York Times*, Vol. Sept 22.

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

Ye, Gu Bin, Qiang and Wei Chen (2009) "Measuring the Influence of Managerial Responses on Subsequent Online Customer Reviews- A Natural Experiment of Two Online Travel Agencies," Working Paper.

Yu, Roger (2010) "Hotel managers monitor online critiques to improve service," *USA Today*, URL: `http://usatoday30.usatoday.com/travel/hotels/2010-03-23-businesstravel23_ST_N.htm`.

Zhang, Xiaoquan (Michael) and Feng Zhu (2011) “Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia,” *American Economic Review.*