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

MANAGEMENT SCIENCE
Vol. 64, No. 10, October 2018, pp. 4629–4647
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

The Value of Multidimensional Rating Systems: Evidence from a Natural Experiment and Randomized Experiments

Pei-Yu Chen,[a] **Yili Hong,**[a] **Ying Liu**[a]

[a] Department of Information Systems, W. P. Carey School of Business, Arizona State University, Tempe, Arizona 85287

Contact: peiyu.chen@asu.edu (PC); hong@asu.edu, http://orcid.org/0000-0002-0577-7877 (YH); yingliu_is@asu.edu (YL)

Received: April 13, 2015
Revised: May 21, 2016; November 8, 2016; March 19, 2017
Accepted: May 10, 2017
Published Online in Articles in Advance: October 13, 2017

https://doi.org/10.1287/mnsc.2017.2852

Abstract. Online product ratings offer information on product quality. Scholars have recently proposed the potential of designing multidimensional rating systems to better convey information on multiple dimensions of products. This study investigates whether and how multidimensional rating systems affect consumer satisfaction (measured by product ratings), based on both observational data and two randomized experiments. Our identification strategy of the observational study hinges on a natural experiment on TripAdvisor when the website started to allow consumers to rate multiple dimensions of the restaurants, as opposed to only providing an overall rating, in January 2009. We further obtain rating data on the same set of restaurants from Yelp, which controls for the unobserved restaurant quality over time and allows us to identify the causal effect using a difference-in-differences approach. Results from the econometric analyses show that ratings in a single-dimensional rating system have a downward trend and a higher dispersion, whereas ratings in a multidimensional rating system are significantly higher and convergent. Findings from two randomized experiments suggest that the multidimensional rating system helps consumers find products that better fit their preferences and increases the confidence of their choices. We also show that the observed results cannot be explained by the priming effect due to rating system interface or a list of other alternative explanations. The combined evidence from the natural experiment and randomized experiments support the view that the multidimensional rating system enhances rating informativeness and provide implications for designing online rating systems that help consumers match their preferences with product attributes.

History: Accepted by Anandhi Bharadwaj, information systems.
Supplemental Material: Data and the online appendix are available at https://doi.org/10.1287/mnsc.2017.2852.

Keywords: multidimensional rating systems • natural experiment • difference-in-differences • multimethod research • product quality • consumer preference • rating informativeness • priming effect

1. Introduction

The substantial increase in online word of mouth (WOM) in the form of online product ratings and reviews has transformed the way consumers acquire product information. According to an article in the *New York Times* (Streitfeld 2012), "Reviews by ordinary people have become an essential mechanism for selling almost anything online." And nearly 67% of consumers report that their purchasing decisions were influenced by online reviews, according to a recent survey (Hinckley 2015). While online reviews prove to drive product demand and predict sales (Chevalier and Mayzlin 2006, Duan et al. 2008), it is not conclusive how informative they are to consumers (Godes and Silva 2012), especially for single-dimensional rating systems (herein referred to as "SD system"), wherein review platforms only collect and aggregate overall product ratings. Recent research suggests that SD ratings cannot comprehensively reflect information on different product attributes (Archak et al. 2011). Because consumers tend to focus on the average valence and generally ignore other important information (De Langhe et al. 2016), the effectiveness of SD ratings in improving consumers' decision making may be limited. This leads to our contention that because product quality is often comprised of multiple dimensions (Garvin 1984), a multidimensional rating system (herein referred to as "MD system") may outperform an SD system in conveying product information.

Given the theoretical importance and practical significance of online WOM systems, a rigorous examination of the design of rating systems is of great importance, particularly the informational value of the MD system. The MD system allows consumers to provide and obtain not only the overall ratings, but also ratings on different product dimensions. Using restaurant rating systems as an example, consumers could rate on food, service, and ambiance of restaurants; similarly, prospective consumers can obtain information on these different dimensions. This suggests two potential

Management Science, 2018, vol. 64, no. 10, pp. 4629–4647, © 2017 INFORMS

effects of the MD system on product ratings: higher ratings (satisfaction) because of more informative MD ratings (rating consumption), and higher ratings due to the rating system interface that may prime reviewers (rating generation).

In the rating-consumption stage, the MD system provides more information because it allows previous consumers to share their experiences in terms of ratings in different dimensions, which can be more informative than a single overall rating to subsequent consumers, particularly when consumers derive the utility of a product from different attributes (dimensions). Essentially, products may exhibit vertical differentiation in terms of quality (Cremer and Thisse 1991), and horizontal differentiation in terms of fit or preference. Therefore, when making a decision, consumers face two types of uncertainty: product quality uncertainty on the vertical quality dimension and product fit uncertainty on the horizontal quality/preference dimension (Hong and Pavlou 2014). Consumers rely on online ratings for information to resolve such uncertainties (Kwark et al. 2014). However, given the potential multidimensional nature of product attributes and consumer preferences thereof, matching the idiosyncratic preferences of consumers with an overall rating, as in the SD system, is difficult. For example, when it comes to the decision to have dinner at a restaurant, different consumers have different preferences in terms of food quality, service, and restaurant ambiance. Furthermore, the same consumer may have different preferences toward these dimensions on different occasions. For instance, a consumer on a romantic date is more likely to emphasize services and ambiance than when he is simply looking to get a delicious meal. The MD system provides systematic information on both vertical (quality) and horizontal (preference) dimensions, and consumers who access information from the MD system could obtain a more accurate estimate of the utility from consuming a product. Note that it is also possible that reviewers could obtain dimensional information from text reviews in the SD system. However, it takes consumers more time, effort, and cognitive resources to obtain useful information from text reviews. Taken together, we would expect the MD system to have higher "rating informativeness," as it facilitates the matching between consumers and products by providing information on different attributes of products and mitigates consumers' product uncertainty. Such better matching enables consumers to make a more informed decision regarding whether they should purchase a product or not, reducing the likelihood of buying a product that would not fit their preferences.

In the rating-generation stage, the SD system and the MD system may systematically lead to different overall ratings because of the rating system interface. The interface of the SD system only requires that the consumer provide an overall rating with the option to write a text review. It is likely that a consumer's rating reflects predominantly her experience in a certain dimension. For example, research has found that consumers have a tendency to recall negative aspects of their experiences (Sundaram et al. 1998, Hennig-Thurau et al. 2004). The interface of the MD system, however, may prime consumers about different dimensions of the product. Considering different aspects of the product comprehensively when consumers post the overall rating may mitigate this negativity recall bias toward a particular dimension. In summary, empirically examining and quantifying the extent to which the MD system affects product ratings has considerable value to both researchers and practitioners. Further, understanding the mechanism underlying the observed effects provide important theoretical and practical insights to this important problem. If we find evidence that MD indeed reduces consumer uncertainty and enhances decision, despite the possibility of the priming effect, it would suggest considerable value to embrace MD rating systems. Therefore, the goal of this research is to address the following research questions:

(1) *How does the implementation of an MD system affect product ratings*?

(2) *Was the effect driven by MD being more informative or/and the priming effect of the MD system*?

We use a multimethod approach to empirically evaluate our research questions. Specifically, we conducted studies based on both observational data and data from lab experiments. First, we collected observational data from two leading restaurant review websites: Yelp and TripAdvisor. Specifically, we constructed our panel data by sampling 1,209 restaurants in New York City and obtaining reviews for the same set of restaurants from the two websites. We examined how these same restaurants are rated in different rating systems. Our main econometric identification strategy hinges on a natural experiment that took place on TripAdvisor, which in January 2009 reengineered its rating system from SD to MD by allowing consumers to provide ratings on multiple aspects of the restaurants, and providing a summary of dimensional reviews after the restaurants start to receive MD ratings. By contrast, Yelp did not make such a change and maintains an SD system, and therefore forms our control group. Tracking identical restaurants on Yelp enables us to control for unobserved restaurant quality change over time. The system change at TripAdvisor allows us to specify our empirical model in a quasi-experimental difference-in-differences (DID) framework. The results show that, on average, the overall rating of a restaurant on TripAdvisor increases by 0.4 after implementing the MD system. Additionally, ratings on the MD system are convergent. These findings provide evidence of the MD system relative to the SD system on product ratings. These results can be driven by either rating informativeness or a possible priming effect due to system interface. To uncover the mechanism on how the MD system affects ratings, we conducted two additional analyses on the observational data and two randomized experiments. First, to assess the role of rating informativeness, we analyzed the variation in MD rating accumulation across restaurants by examining how the effect varies as more MD ratings are accumulated. We found the effect to be notably stronger for restaurants that accrued more MD ratings, lending support for the rating informativeness mechanism. To assess the presence of the priming effect, we compare SD and MD ratings within TripAdvisor and, to mitigate the threat of self-selection, we focus on reviewers who have provided both SD and MD ratings after the system change at TripAdvisor. The results do not lend support for the priming mechanism. Second, we conducted two randomized experiments to triangulate with the findings from the observational data to further explore the mechanisms. Results from the first experiment provide evidence that the MD system enables better matching of preferences and that it enhances decision confidence by reducing product uncertainty. The second experiment assessed the priming effect of the MD system interface that may have led consumers to systematically report higher ratings, and again, we did not find evidence to support this mechanism being the driver of our observed results. Lastly, we rule out the alternative explanations of the two websites being not comparable before the system change, consumer bases of TripAdvisor changing because of the system change, and the parallel pretreatment trend assumption. Overall, our study makes a pioneering effort in establishing the value of MD rating systems.

2. Related Literature

A mature body of scholarly research is available on online product reviews across different fields, such as information systems (IS), marketing, and economics. Much of the prior work has focused on the effect of online product reviews on sales (e.g., Godes and Mayzlin 2004, Chevalier and Mayzlin 2006, Liu 2006, Dellarocas et al. 2007, Duan et al. 2008, Forman et al. 2008) and antecedents to review characteristics (e.g., Goes et al. 2014, Hong et al. 2016, Huang et al. 2016).

Extant research has started to explore different dimensions of product attributes using text mining approaches. For example, Hu and Liu (2004) identified product features for which consumers expressed their opinions. Decker and Trusov (2010) estimated the relative effect of product attributes and brand names on the overall evaluation of products. Ghose and Ipeirotis (2011), Archak et al. (2011), and Ghose et al. (2012) explored aspects of text reviews to identify important text-based features and their impact on review helpfulness and product sales. In summary, consumers do consider information on different dimensions of a product prior to consumption. In an SD system, consumers may look for information on the different dimensions of a product from text reviews. In an MD system, ratings on multiple product dimensions are presented to consumers, which facilitates the matching of consumer preferences with product attributes, leading to potentially more efficient matching and more satisfied purchases. More recently, IS researchers have looked at how online product reviews may reduce product uncertainty (Kwark et al. 2014, Sahoo et al. 2015, Wang et al. 2016). Notably, no research has directly compared the MD system with the SD system in affecting product ratings. The present study addresses this void.

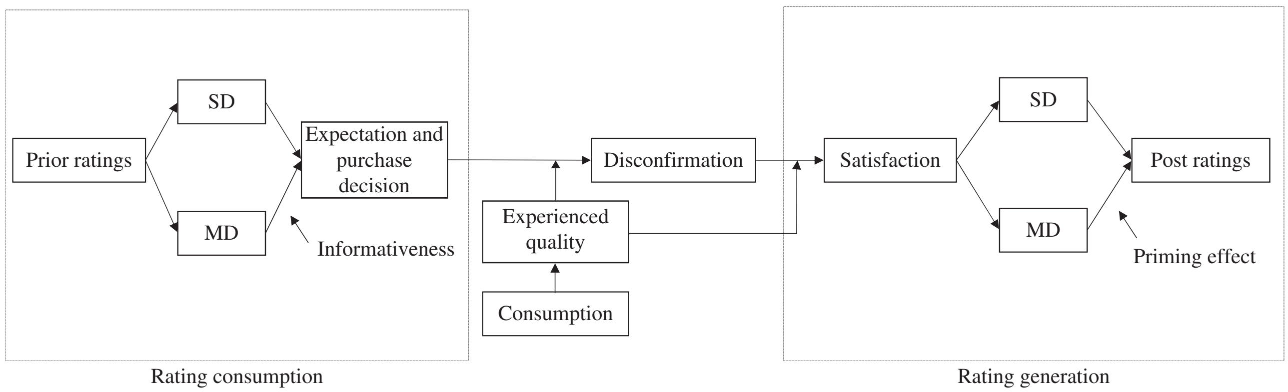
3. Theory and Hypothesis Development

The MD system can potentially affect ratings in two ways. First, MD ratings are likely more informative, either because information is presented in a way that reduces uncertainty or because it contains more information. Therefore, consumers who have access to the MD ratings may be more informed of whether a product will fit their preferences and make a better decision regarding whether to purchase the product or not. This facilitates matching of consumers to product attributes: consumers who purchase a product are likely consumers who indeed like the product, and consumers who would not like the product will not purchase it. As a result, the MD system could positively affect ratings because of lower decision error from consumers and a better matched pool of customers who indeed like the product. We draw on information transfer theory and expectation-confirmation theory (ECT) to propose the effect of the MD system (relative to the SD system) on ratings through improved rating informativeness. MD systems could also affect rating generation through a priming effect, beyond improving rating informativeness. While consumers may report lower ratings in the SD system because of the negativity bias (only recall the bad experience with a particular product dimension), the MD system may prime users about different product dimensions and lead consumers to consider all dimensions when they provide an overall rating, thus mitigating the negativity bias. By incorporating the above two potential effects of MD, we propose our conceptual model (Figure 1) and hypotheses below.

3.1. Rating Informativeness

Social scientists consider an information system as the dynamic interactions among three components: the

Figure 1. Conceptual Model



user (consumer), the knowledge resource, and the intermediary mechanism between the knowledge resource and the user (Belkin 1984). According to Belkin, the knowledge resource contains information that is represented and organized in certain ways. The user initiates the system to solve problems using the information obtained from the knowledge resource, and the intermediary mechanism mediates the user's requirement and knowledge resource. The system facilitates information transfer from the knowledge resource to the user via the intermediary. Based on this view, the rating system can be regarded as an information system, whereas existing ratings and reviews serve as the knowledge resource that a user may refer to when he or she needs to make a purchase decision that initiates the system. An information system is considered "effective" when the system design improves decision making and satisfaction. In the context of rating systems, consumers obtain product information that helps make purchase decisions from prior ratings using either an SD system or an MD system. The SD and MD systems are two systems that may differ in their effectiveness. As a knowledge resource, they may contain a different amount or quality of information. As an intermediary, they present information in different ways that could affect how information is consumed. Both could affect the effectiveness of the information system. When the information is more effective, the consumers are able to make better purchase decisions and they may provide a higher rating to the product. Note that we focus on the outcome "effectiveness" of the information system, but not solely on the efficiency of information transfer. Throughout the paper, we use the term "rating informativeness" to capture the effectiveness of the two rating systems as an information system.

3.2. Expectation and Disconfirmation

Expectation-confirmation theory (ECT) is widely used in the IS and marketing literature to understand system adoption (e.g., Bhattacherjee 2001, Brown et al. 2012, Lin et al. 2012, Brown et al. 2014, Diehl and Poynor 2010, Venkatesh and Goyal 2010) and consumer satisfaction (e.g., Anderson and Sullivan 1993, Churchill and Suprenant 1982, Kim et al. 2009, Oliver 1980). Drawing on adaptation level theory (Helson 1964), Oliver (1980) posited that one's expectation level of product performance is an adaptation level. The degree to which the product exceeds, meets, or falls short of one's expectation may cause satisfaction deviations from the adaptation level. Subsequent research found that perceived quality and disconfirmation of expectation have direct effects on satisfaction (Anderson and Sullivan 1993). They also reported an asymmetric (dis)confirmation effect, such that negative confirmation (disconfirmation) has a greater effect on satisfaction than positive confirmation.

We combine the expectation-confirmation theory (Anderson and Sullivan 1993) and the information transfer theory (Belkin 1984) to conceptualize how the MD system and the SD system affect ratings in the rating-consumption stage (Figure 1). Similar to extant studies on online product ratings, we assume that the reported ratings of consumers reflect their satisfaction after consumption. In other words, ceteris paribus, a satisfied consumer posts higher ratings than a dissatisfied consumer. According to ECT, satisfaction is a function of experienced quality and expectation (dis)confirmation. Disconfirmation is defined as the extent to which *experienced quality* falls short of the *expected quality*. Based on the information transfer theory, existing ratings serve as a knowledge resource for consumers. Consumers can form a priori expectation of utility through this knowledge. After consuming the product, consumers derive actual utility based on their experienced quality. Consumers then compare their experienced quality with their expectation, which may be the same as or different from the expected quality. The experienced quality, which is derived at the time of actual consumption, is independent of the rating system. However, the rating system will affect expectations because consumers may form different expectations based on the information they obtain from different rating systems. When the product ratings are informative, consumers should be able to estimate the utilities from the product and form reasonable expectations of utility from consuming that product. When expectations are formed more precisely, consumers are less likely to be disappointed because the consumption utility is more likely to confirm the expected utility. In other words, output ratings are likely to match input ratings.[1]

In the SD system, product quality is presented as a single number (i.e., the overall rating). It is difficult or would take consumers significantly more efforts (by going through text reviews) to form a precise expectation of product utility because the information in SD ratings is not sufficient to reflect the MD nature of product attributes. Consider the restaurant example: A two-star rating of a restaurant could be interpreted as "the restaurant has really bad food." However, the restaurant may actually serve good food; the consumer who posted the two-star rating simply did not like its service. In such cases, subsequent consumers who primarily care about food served by the restaurant may be misinformed. Even though different product attributes may be extracted from the text reviews (Archak et al. 2011, Ghose and Ipeirotis 2011, Ghose et al. 2012), the unstructured nature of text reviews makes it difficult for consumers to mentally parse and obtain the information they desire. Moreover, consumers may fail to mention one dimension or another in their text reviews.

In the MD system, besides the overall rating, separate ratings on multiple different dimensions (e.g., food quality, service, and ambiance) are also available, thereby conveying comprehensive yet nonredundant information on the product. When consumers make a purchase decision, they can effectively take different dimensions of product attributes into account. Thus, the MD system helps consumers choose the restaurant that best fits their preferences/occasions by better resolving product uncertainty, which has been found to hinder effective consumer decision making (Hong and Pavlou 2014). Even though the text reviews may also include information on different product attributes (Wang et al. 2016), the MD system at least alleviates consumers' cognitive burden to process all of the textual information, making the MD ratings more informative. In this way, an MD system acts as a better matching and screening system. For example, a consumer who places high weight on restaurant "ambiance" but low weight on "food" may find a restaurant, say A, with an overall rating of 4, an ambiance rating of 5, and a food rating of 3 more attractive than another restaurant, say B, with the same overall rating of 4, but with an ambiance rating of 3 and a food rating of 5. This consumer will be satisfied with her choice of restaurant A and may not choose restaurant B at all (and therefore will not give low ratings to restaurant B). This example suggests that better matching and screening of consumers to product attributes will improve the overall ratings.

3.3. Priming Effect

The above discussion suggests that MD systems likely increase ratings because it improves rating informativeness in the rating-consumption stage. It is also important to consider how consumers choose to share their overall consumption experiences of product with different attributes. In the rating-generation stage, the rating system interface (MD versus SD) may exert a "priming effect" on consumers who provide a review, possibly leading to higher ratings in the MD systems. Prior research has established the existence of a priming effect and its possible impact on consumer behaviors (Meyer and Schvaneveldt 1971, Bargh and Tanya 2000, Sela and Shiv 2009). The priming effect is incorporated to the rating-generation process on the right side of the conceptual model in Figure 1. In the SD system, consumers are only required to provide an overall rating. In such cases, consumers likely recall the most salient product attributes to themselves. Related to this point, several studies have established the negativity bias (Sundaram et al. 1998, Hennig-Thurau et al. 2004)—i.e., consumers may put more weight on negative attributes of the product and ignore positive attributes (Hu et al. 2006). In the MD system, because consumers are primed to consider all key aspects of the product instead of only focusing on the most salient dimension that they recall when posting a review, they may have a higher likelihood to account for all dimensions more objectively when they provide the overall evaluation of a product. In this way, the negativity bias is mitigated and consumers are likely to provide a higher overall rating in the MD system than the SD system. This priming effect mechanism will lead to the same prediction as the mechanism of increasing rating informativeness. Therefore, it presents a challenging empirical task to disentangle the mechanisms.

To summarize, there are two ways the MD system can affect ratings. First, in the rating-consumption stage, MD ratings act as input that can affect expectation formation and decision confidence. Second, in the rating-generation stage, MD system primes users with different dimensions and can affect their rating behaviors by mitigating the possible negativity bias. Based on the mechanism we discuss above, we propose several hypotheses on how rating systems affect product ratings. We document the findings in Section 5, and we seek to offer evidence on the most plausible underlying mechanisms in Section 6.

3.4. Hypotheses

Based on the above discussion, we propose the following hypotheses. First, prior literature has documented that early consumers of a product tend to be more enthusiastic about the product and tend to provide higher ratings than later consumers, and therefore downward trend of ratings is commonly observed because of such self-selection (Li and Hitt 2008, Godes and Silva 2012). We expect that both SD and MD systems are subject to such self-selection bias; however, the effect of self-selection bias should be attenuated when the rating system enhances consumers' decision making. Within our framework of information transfer, it is possible that later consumers will be misled by forming unreasonable expectations from high ratings provided by early reviewers. In the SD system, only a single overall rating is provided. Consumers are not able to match the rating to a specific product dimension that they care most about. Therefore, consumers are more likely to experience product uncertainty, leading to a higher likelihood for mismatch. For example, an early reviewer may rate a restaurant as 4.5 stars simply because of his or her enthusiasm about its great food, but a subsequent consumer who is looking for high-quality service may misinterpret the 4.5 stars as reflecting service and become disappointed. Therefore, the inability of the SD ratings to resolve product uncertainty may aggravate the downward trend. On the other hand, if ratings provided in MD system are indeed more informative, then they will improve consumers' decision making and attenuate the effect of self-selection bias. To understand if informativeness really plays a role in explaining rating trend beyond self-selection, we can compare the rating trends in the SD system and the MD system. Presumably, regardless of the rating system, reviewers are subject to the same self-selection bias over time, if it is present. Therefore, if self-selection is the only explanation, and the SD and MD systems do not differ in rating informativeness, we should expect ratings to exhibit similar downward trends in both systems. However, if rating informativeness is the dominant factor, then we expect to see a downward trend in the SD system but not in the MD system. We propose the following hypotheses:

Hypothesis 1A (H1A). *The SD system exhibits a downward trend of ratings.*

Hypothesis 1B (H1B). *The MD system exhibits a downward trend of ratings.*

Empirically, we would expect strong support for H1A yet weak or no support for H1B if ratings in MD system are more informative for consumers to make a better decision.

We proceed to discuss the effect of MD system on product ratings. The MD system may affect product ratings because it provides more informative ratings (i.e., rating consumption) or/and because of the priming effect (i.e., rating generation). As our earlier discussion suggests, when ratings are informative, we should expect the ratings to help resolve product uncertainty and enhance consumer satisfaction. The MD system organizes and presents information in a way that allows subsequent consumers to process the information more easily and help improve the formation of expectation and facilitate decision making. Therefore, compared with SD ratings, MD ratings are more informative because they help consumers to form reasonable expectations and choose a restaurant that better fits their preferences. Second, in the rating-generation stage, there is a potential for a priming effect in the MD system because consumers are primed to take all dimensions into account. In the SD system, it is likely that consumers focus on certain dimensions. If consumers focus more on positive attributes, then ratings in SD will be higher than in MD. On the other hand, if consumers put more emphasis on negative attributes (i.e., negativity bias is present), ratings in SD will be lower than in MD. Taken together, we expect that ratings in MD are more likely to be higher because of increased rating informativeness and/or because the MD system is likely to mitigate the likely presence of negativity bias in the SD system. Therefore, we propose the following:

Hypothesis 2 (H2). *Ceteris paribus, the overall ratings in the MD system are higher than those in the SD system.*

Similarly, when MD systems increase rating informativeness, it helps consumers form more reasonable expectations, and thus consumers' experienced quality is more likely to confirm the expected quality. Therefore, we expect less deviation in ratings for the MD system. In other words, we will observe ratings to converge over time (i.e., a consumer's overall rating is more likely to be similar to those reported by prior consumers).

Hypothesis 3 (H3). *Ceteris paribus, the overall ratings in the MD system are less likely to deviate from prior average ratings than those in the SD system.*

4. Data and Empirical Analyses

4.1. Observational Data

We first address our research questions by studying restaurant reviews in different rating systems. We choose restaurants as our context because restaurants are well known for having different dimensions of quality (e.g., food, service, and ambiance) and have attracted significant attention in recent academic literature (e.g., Anderson and Magruder 2012, Dai et al. 2012, Gu et al. 2012, Luca and Zervas 2016, Mangold et al. 1999, Huang et al. 2016). Our empirical analysis utilizes restaurant review data gathered from two leading consumer review websites: Yelp.com (*Yelp*) and TripAdvisor.com (TripAdvisor) covering the observation window between November 2004 and April 2013. Like most review websites, Yelp provides an SD system on a scale of one to five stars. TripAdvisor had been using an SD system until January 2009, when the website reengineered its system and implemented an MD system, which allows its users to submit overall ratings and optional ratings for different dimensions of the restaurants, including food, service, ambiance, and value, using the same five-star rating scale.

We used two customized web crawlers for data collection. We first collected all restaurants in New York City on TripAdvisor and Yelp. To eliminate restaurant differences and control for unobserved quality changes in the restaurants, we focused on data of identical restaurants across the two review websites. Therefore, the differences between the ratings in the two review systems for the same restaurant cannot be attributed to unobserved restaurant effects. Specifically, we matched the restaurants on Yelp and TripAdvisor according to restaurant names, addresses, and phone numbers. The two websites had a total of 1,209 restaurants in common in New York City during the time of data collection. We collected all available reviews for these common restaurants. Note that the two websites have always required both a numerical rating and textual review. For each review, we collected the time stamp of each review, the consumer ID, and the star rating (i.e., an integer between one and five).

After TripAdvisor changed from an SD to an MD system in 2009, some restaurants on TripAdvisor start to accumulate MD ratings. We collected the star ratings of each dimension for each restaurant on TripAdvisor whenever they are available. Even though the rating system of TripAdvisor has been reengineered from SD to MD, and aggregated dimensional ratings are provided to all consumers who come to TripAdvisor, consumers are not forced to rate all dimensions.[2] Consumers can either choose to provide only a single overall rating as in the SD system, or provide both the overall rating and the dimensional ratings. As a result, the number of total dimensional ratings of a restaurant on TripAdvisor is not equal to that of the total overall ratings of the restaurant. This variation warrants many robustness tests, as we describe in later sections. Particularly, it allows us to examine how informativeness improves over time as some restaurants accumulated more MD ratings. The details of this variation at the restaurant level and review level are shown in Figure 2.[3]

Table 1. Data Description

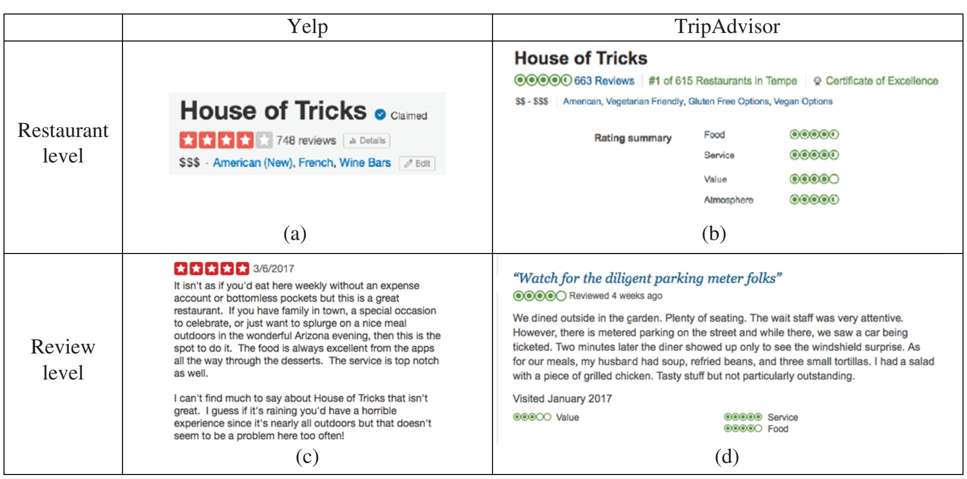
	Observations	Mean	Std. dev.
Yelp			
Before system change	18,388	3.77	1.02
After system change	166,670	3.67	1.14
TripAdvisor			
Before system change	8,917	3.91	1.11
After system change	69,024	4.16	1.03

Table 1 provides the summary statistics of the data. A notable difference based on the summary statistics is that TripAdvisor ratings increase after the system change, whereas at the same time, Yelp ratings decrease. This is just an aggregate view, and we describe our identification strategy and conduct formal econometric analyses in the next sections.

4.2. Research Design and Identification Strategy

To address our research question, we explore both cross-website and within-website variations by matching data from two websites (Yelp and TripAdvisor) on the same set of restaurants and leveraging a quasi-natural experiment on TripAdvisor. Our key econometric identification strategy hinges on the system change that occurred on TripAdvisor in January 2009 concerning the rating system design, which is exogenous to consumers. Summary statistics in Table 1 suggest that ratings of restaurants increase after the TripAdvisor

Figure 2. (Color online) Rating Systems Comparisons



system change. However, it is also possible that the rating change is due to restaurant quality change over time. Therefore, it is important to control for unobserved restaurant quality change over time. We track the same set of restaurants on Yelp as the "control group." In other words, the rating trend on Yelp for each of the restaurants in our sample serves as a proxy for any change in restaurant quality (e.g., change of chef, menu, or ownership). We therefore employ a difference-in-differences (DID) strategy to test how rating systems affect consumer ratings for the same set of restaurants. We employed two tests to ensure that DID is a valid strategy in our setting. First, we show that before the system change, the same restaurants on TripAdvisor and Yelp show similar rating trends. This suggests that we have a fair comparison to start with. Second, we employ a formal time effect test, and the results suggest that the observed rating trend after change did not occur prior to the system change. That is, it is not a pretreatment trend that caused the observed effects. The test further suggests that the observed effect starts to manifest in the second half of 2009. The details of these two tests are reported in Sections 7.1 and 7.2, which suggest that the common DID assumptions are satisfied. Therefore, after differencing out the rating changes at Yelp, the changes in rating trend of TripAdvisor can be attributed to the rating system change. In the following section, we present specific tests for each hypothesis.

4.3. Empirical Framework

4.3.1. Rating Trends: Test of Hypotheses 1A and 1B. We analyze how ratings change as more reviews are being accumulated. We estimate the following model:

$$Rating_{ij|k} = \beta_0 + \beta_1 * \log n_{ij|k} + \alpha_i + \epsilon_{ij|k}, \qquad (1)$$

where i indexes the restaurant and k denotes each website. We sort all reviews in order by the time of their postings, and j indexes the temporal order of the review for each restaurant. The dependent variable (DV), $Rating_{ij|k}$, is the jth rating submitted for restaurant i on website k. Here, $n_{ij|k}$ is the total number of previous ratings prior to rating j for restaurant i on website k. Therefore, β_1 captures the rating trend as a function of the number of reviews being accumulated. A negative coefficient indicates a downward trend. And α_i denotes a restaurant fixed effect (FE). As discussed in Section 3.3, self-selection may also lead to a downward trend in both the SD and MD systems. However, if informativeness dominates, we should expect β_1 to be negative in the SD system but nonnegative in the MD system.

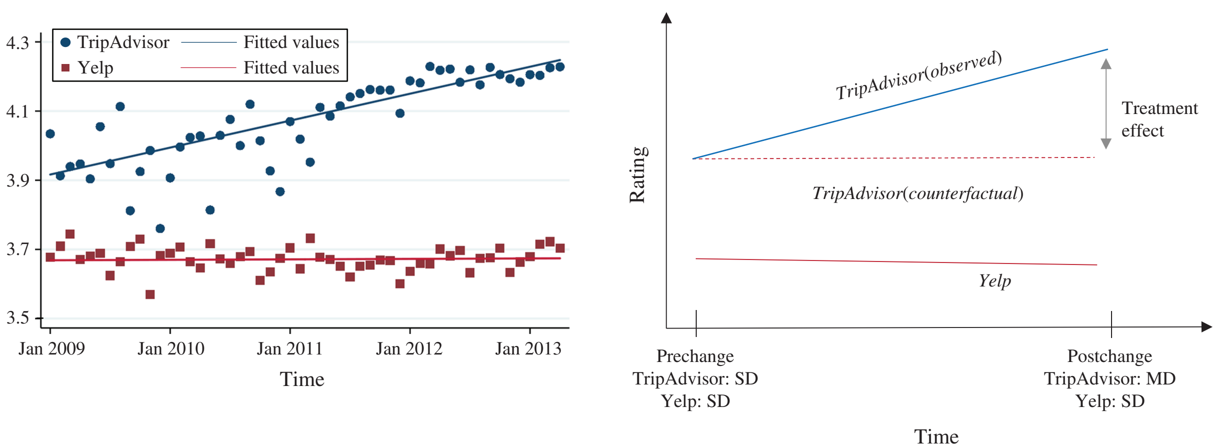
4.3.2. DID Rating Model: Test of Hypothesis 2. We utilize the DID approach to estimate the effect of the rating system change from SD to MD on the overall ratings. As discussed in Section 4.2, we choose identical restaurants on Yelp as the "control group," and the rating trend for each of these restaurants on Yelp serves as a proxy for any change in restaurant quality. The first difference takes away the rating changes for each restaurant at *Yelp* before we attribute the rating changes on TripAdvisor to the change of the rating system. Figure 3 demonstrates the relationship.

We estimate the following equation to test H2:

$$Rating_{ij|k} = \beta_0 + \beta_1 * Post + \beta_2 * Post * Treat + \beta_3 * Treat + \beta_4 * \log n_{ij|k} + \alpha_i + \epsilon_{ij|k}, \qquad (2)$$

where *Treat* is a dummy for the treatment group (i.e., TripAdvisor), which is set to one if the ratings are on TripAdvisor, and zero if on Yelp; and *Post* is a dummy variable set to one if the rating is made after the system change, and zero if the rating is provided prior to the system change. The coefficient, β_2, of the interaction term *Post* ∗ *Treat* measures the effect caused by the change of the rating system, after controlling for changes in restaurant quality over time, systematic website differences, and restaurant heterogeneity (fixed effects). We expect β_2 to be positive for H2.

Figure 3. (Color online) Difference-in-Differences Illustration



4.3.3. Deviation Model: Test of Hypothesis 3. To test H3, we investigate whether the introduction of the MD system reduces the distance between the current rating and prior average overall rating. The dependent variable, $Deviation_{ij|k}$, is measured as the absolute difference between the jth rating for restaurant i on website k (i.e., $Rating_{ij|k}$) and the average of all ratings till the $j-1$th rating of restaurant i on website k denoted as $\mu_{i(j-1)|k}$:

$$Deviation_{ij|k} = |Rating_{ij|k} - \mu_{i(j-1)|k}|. \quad (3)$$

The following deviation model tests if ratings converge over time in the MD system, by relating the deviation of ratings to the review sequence. As before, we include restaurant fixed effects α_i to control for any systematic differences due to restaurants. And β_1 measures how the deviation changes over time as ratings accumulate. A positive β_1 means that deviation increases as more ratings are accumulated, whereas a negative β_1 means that deviation decreases with more ratings, which indicates the convergence of ratings:

$$Deviation_{ij|\text{Yelp}} = \beta_0 + \beta_1 * \log n_{ij|\text{Yelp}} + \alpha_i + \epsilon_{ij}, \quad (4)$$

$$Deviation_{ij|\text{TripAdvisor}} = \beta_0 + \beta_1 * \log n_{ij|\text{TripAdvisor}} + \alpha_i + \epsilon_{ij}. \quad (5)$$

We also directly compare the deviation effect in the SD system and that in the MD system using Equation (6), where β_3 captures the difference of the rating deviation with the rating sequence between Yelp and TripAdvisor. We expect β_3 to be negative, indicating less dispersion in the MD system in comparison to the SD system:

$$Deviation_{ij|k} = \beta_0 + \beta_1 * \log n_{ij|k} + \beta_2 * Treat + \beta_3 * \log n_{ij|k} * Treat + \alpha_i + \epsilon_{ij|k}. \quad (6)$$

5. Empirical Results

5.1. Results: Rating Trends (H1)

Table 2 presents the results of how ratings change as more ratings are accumulated. Models 1 and 2 show the rating trends of Yelp and TripAdvisor, respectively, before the system change when they were both using SD systems. The coefficients of interest (i.e., the total number of ratings) have the expected negative signs, which suggests that both Yelp and TripAdvisor follow a downward trend when they were both on SD systems. Models 3 and 4 present the rating trends of Yelp and TripAdvisor after the rating system change, when *Yelp* is still using an SD system, and TripAdvisor adopted an MD system. The result of Yelp is consistent with that of Models 1 and 2, offering additional evidence that SD ratings are downward trending. Most interestingly, the coefficient in Model 4 is positive, which suggests that the MD system helps correct the downward trend observed in the SD system. Comparing Yelp (Model 3) and TripAdvisor (Model 4) during the same period after TripAdvisor adopts the MD system, the downward trend continues in Yelp, but the trend is reversed in TripAdvisor. Similarly, comparing the before (Model 2) and after the change (Model 4) at TripAdvisor, the downward trend before the system change is corrected and ratings start to exhibit an upward trend after TripAdvisor adopts the MD system. Overall, our results provide support of H1A but not for H1B. Notably, the different rating trends in SD and MD systems are consistent with the predictions based on the rating informativeness, beyond self-selection.

Table 2. Downward Trend in the SD System (DV = *Rating*)

	Before system change		After system change	
Sample	Yelp (SD)	TripAdvisor (SD)	Yelp (SD)	TripAdvisor (MD)
	Model 1	Model 2	Model 3	Model 4
log(# *of reviews*)	-0.115^{***}	-0.058^{**}	-0.021^{**}	0.084^{***}
	(0.011)	(0.018)	(0.007)	(0.010)
Constant	4.127^{***}	4.070^{***}	3.774^{***}	3.795^{***}
	(0.035)	(0.049)	(0.031)	(0.043)
Restaurant FE	Yes	Yes	Yes	Yes
Prob > *F*	0	0.0014	0.0011	0
Observations	18,388	8,917	166,670	69,024

Note. Cluster-robust standard errors in parentheses.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

5.2. Results: Rating Analysis (H2)

Table 3 presents the estimation results of DID analysis for Equation (2). First column presents the results using all data. The significant positive coefficient of *Post* ∗ *Treat* indicates that the change of the rating system from SD to MD significantly increased ratings by 0.4. In column (2), we narrow our time window to two years prior and two years after system change (i.e., from 2007–2010) to test if the effects manifest within a shorter time window around the rating system change.

Table 3. DID Analysis (DV = *Rating*)

Sample	(1) All data		(2) Two years before and after system change (2007–2010)	
Post	-0.124^{***}	(0.016)	-0.034	(0.018)
Treat	0.078^{***}	(0.022)	0.061^{*}	(0.024)
Post ∗ *Treat*	0.400^{***}	(0.021)	0.149^{***}	(0.027)
log(# *of reviews*)	0.001	(0.005)	-0.048^{***}	(0.009)
Constant	3.793^{***}	(0.021)	3.925^{***}	(0.034)
Restaurant FE	Yes		Yes	
Prob > *F*	0		0	
Observations	262,999		79,826	

Note. Cluster-robust standard errors in parentheses.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 4. Estimation of Rating Deviation (DV = *Rating Deviation*)

	Yelp		TripAdvisor		Yelp and TripAdvisor	
Sample	After system change		After system change		After system change	
	Model 1		Model 2		Model 3	
log(# *of reviews*)	0.039***	(0.004)	−0.059***	(0.006)	0.015***	(0.003)
Treat					−0.001	(0.021)
Treat ∗ log(# *of reviews*)					−0.014***	(0.005)
Constant	0.655***	(0.017)	1.047***	(0.025)	0.776***	(0.015)
Restaurant FE	Yes		Yes		Yes	
Prob > *F*	0		0		0	
Observations	166,670		69,024		235,694	

Note. Cluster-robust standard errors in parentheses.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Note that the effect of the change in the rating system does not instantly cause the ratings to increase; it takes time for consumers to realize this system change and generate adequate MD reviews and show observable effects. Therefore, we need to have sufficient observations of MD reviews across all restaurants to test the effect, and two years within system change appears to be a reasonable time frame to see if an effect takes place. Specifically, we observe a significant and positive coefficient of *Post* ∗ *Treat*, and the result suggests that the restaurant ratings increased by 0.149 on average within two years of the implementation of the MD system. Therefore, H2 is supported. Note that this result can be driven either by rating informativeness in the rating-consumption stage or a priming effect in the rating-generation stage. We conduct further analyses on the observational data and two randomized experiments in Section 6 to uncover the mechanisms.

5.3. Results: Rating Deviation Analysis (H3)

The results of the deviation analysis are presented in Table 4. Models 1 and 2 separately analyze the data of Yelp and TripAdvisor after the system change. Model 3 uses the pooled data of Yelp and TripAdvisor after the system change. Restaurant fixed effects are included. The coefficients of sequence (log(# *of reviews*)) are significantly positive for Yelp and negative for TripAdvisor. The results suggest that rating deviation, the absolute difference between the previous average rating and the following rating, increases in the SD system but decreases in the MD system as more ratings accrue on each platform. Similarly, the result of Model 3 shows that the difference between the SD system and MD system is significant. In other words, the deviation over time is significantly smaller for the MD system than for the SD system. The results also suggest that ratings converge in the MD system. Therefore, H3 is supported.

6. Uncovering the Mechanisms

As noted before, more insights can be obtained by understanding the main mechanism that underlies the rating increase after adopting an MD system. If the results are driven solely by priming effects, we could simply fix the SD system by reminding consumers to take into account different dimensions, without necessarily redesigning the rating system. If, however, results suggest that the MD system improves rating informativeness, then it supports the value of adopting an MD system to improve consumer decision making. In the following sections, we first conduct analyses with the observational data to understand the main driver of our results: rating informativeness, priming effect, or both. Then, we also report results from two randomized experiments to provide further evidence on the mechanisms.

6.1. Within-MD System Analysis: Rating Informativeness

Our earlier DID model tests the average treatment effect of the MD system relative to the SD system; however, the results can be driven by either rating informativeness or the priming effect due to the rating interface. System change by itself should not cause an instantaneous effect on rating informativeness as it takes time to accumulate MD ratings. If the rating informativeness argument holds, we expect the effect of MD system on product ratings to be stronger as more MD ratings are accumulated. We therefore explore variation in MD rating accumulation across restaurants by examining how the effect changes as some restaurants accumulate more MD ratings. We consider the following two models:

$$Rating_{ij|\text{TripAdvisor}} = \beta_1 + \beta_2 * fraction_{ij}^{d} + \beta_3 * \log n_{ij} + \gamma_r + \epsilon_{ijr}, \quad (7)$$

$$Rating_{ij|\text{TripAdvisor}} = \beta_1 + \beta_2 * \log multi_{ij}^{d} + \beta_3 * \log n_{ij} + \gamma_r + \epsilon_{ijr}, \quad (8)$$

Table 5. Within the MD System Analysis (DV = *Rating*)

Sample	After system change		Within two years after change (2009–2010)	
Fraction of dimensional reviews	0.323*** (0.026)		0.398*** (0.100)	
log(# *of dimensional reviews*)		0.218*** (0.017)		0.233*** (0.061)
log(# *of reviews*)	0.036*** (0.003)	−0.180*** (0.016)	0.0240* (0.011)	−0.209*** (0.061)
Constant	3.725*** (0.024)	4.026*** (0.013)	3.582*** (0.082)	3.939*** (0.054)
Reviewer FE	Yes	Yes	Yes	Yes
Prob > *F*		0.0001	0	0.0003
Observations	126,146	126,146	15,567	15,567

Notes. Cluster-robust standard errors in parentheses. In this analysis, because we explore cross-restaurant variation, we do not control for restaurant fixed effects, but a reviewer fixed effect.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

where $fraction_{ij}^{d}$ denotes the fraction of prior reviews that are MD to the total number of reviews till rating j for restaurant i; $\log multi_{ij}^{d}$ denotes the log transformation of the total number of MD ratings accumulated prior to rating j for restaurant i; and γ_r is reviewer fixed effects. We test the effect using data after the system change.

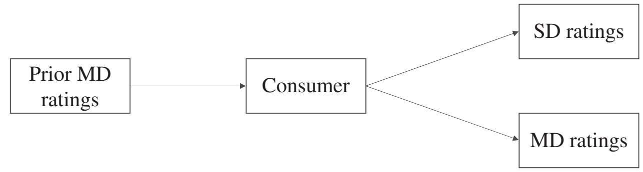
Given the focus on within-TripAdvisor analysis here, we take advantage of all data we had from TripAdvisor that included ratings and reviews of all restaurants in New York City, instead of just a subset of restaurants that matched to those covered at Yelp. Estimation results are shown in Table 5. The first two columns use all data after the system change, and the last two columns use data within two years after the system change (2009–2010). The signs of coefficients of fraction and absolute number are both positive, and the results are consistent using all data or data of 2009 and 2010. The significant positive coefficient indicates that the effect increases with the number of MD ratings generated by consumers. This analysis further strengthens the conclusion that the MD system enhances rating informativeness.

6.2. Testing the Priming Effect

To examine whether the priming effect could be solely responsible for the results in our observational data, we conduct an additional analysis by exploiting the fact that TripAdvisor does not force a consumer to provide ratings on multiple dimensions after the system change. While MD ratings are always visible to all consumers in the consumption stage, consumers have the options to either provide only single overall ratings or provide an overall rating plus dimensional ratings (as shown in Figure 4).

This analysis focuses on the rating-generation stage and tests whether the ratings from consumers when

Figure 4. Consumer's Report on Optional MD Ratings on TripAdvisor



they provide only single overall ratings differ significantly from cases when they also provide ratings on dimensions. Specifically, we compare SD and MD ratings written by the same reviewers to eliminate the concern of different reviewers' self-selection into reporting dimensional ratings. If consumers rate differently across the SD and MD systems, it suggests that the priming effect is likely present.

We estimate Equation (9). *Multi* is a dummy variable that is set to one when an overall rating is provided along with MD ratings, and zero when only an overall rating is provided without MD ratings. We control the reviewer fixed effect with γ_r. Here, $\log n_{jr}$ measures the total number of reviews made by reviewer r till the current review by this reviewer, which captures any review trend by a particular reviewer. We also add an interaction term of *Multi* and $\log n_{jr}$. For the priming effect to be the explanation for our main findings in Section 5, we should observe a positive coefficient for *Multi*:

$$Rating_{ij\,|\,\mathrm{TripAdvisor}} = \beta_0 + \beta_1 * Multi + \beta_2 * \log n_{jr} + \beta_3 * Multi * \log n_{jr} + \beta_4 * \log n_{ij} + \gamma_r + \epsilon_{ijr}. \quad (9)$$

Similar to the previous analysis, we utilize all of the New York City restaurant data on TripAdvisor that we had for this analysis. This allows us to increase observations for each reviewer and have better coverage of their review activities of all New York City restaurants. As attested in Table 6, the coefficient of *Multi* is insignificant, suggesting that there is no significant difference between overall ratings with dimensional ratings and ratings without dimensional ratings from the

Table 6. Within the MD Ratings Analysis (DV = *Rating*)

Sample	After system change	
Multi	0.036	(0.021)
log(# *of reviews by the reviewer*)	−0.028*	(0.012)
Multi ∗ log(# *of reviews by the reviewer*)	0.026	(0.013)
log(# *of reviews*)	0.022***	(0.004)
Constant	4.004***	(0.021)
Reviewer FE	Yes	
Prob > *F*	0	
Observations	58,291	

Note. Cluster-robust standard errors in parentheses.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

same reviewers. Because we do not find evidence that would suggest that reviewers behave significantly different when providing SD or MD ratings, the priming effect does not appear to be the primary driver of our observed results.

6.3. Randomized Experiments

As noted in the theory section, the MD system could affect product ratings via two mechanisms. First, in the rating-consumption stage, MD ratings serve as input that affects expectation formation, resolves product uncertainty, and enhances decision confidence. Second, in the rating-generation stage, the MD system primes users about different dimensions and can affect their rating reporting behavior (e.g., priming effect). To show that the effect due to rating informativeness is at play and to further explore the existence of a priming effect (or the lack thereof), we conduct two controlled lab experiments. First, we examine whether the MD system facilitates preference matching and reduction of product uncertainty. If consumers could find the restaurants that better match their preferences, and if they are more certain about their choices when they use the MD system, it would suggest that MD system improves rating informativeness and decision making. Second, we explore whether the MD system interface indeed exerts a significant effect on rating reporting. If consumers with same set of information report different ratings in different systems, it would suggest the existence of a priming effect.

For the two experiments, participants from a large public university in the United States were recruited as experimental subjects.[4] We describe the procedures of these two randomized lab experiments and report the experimental findings below.

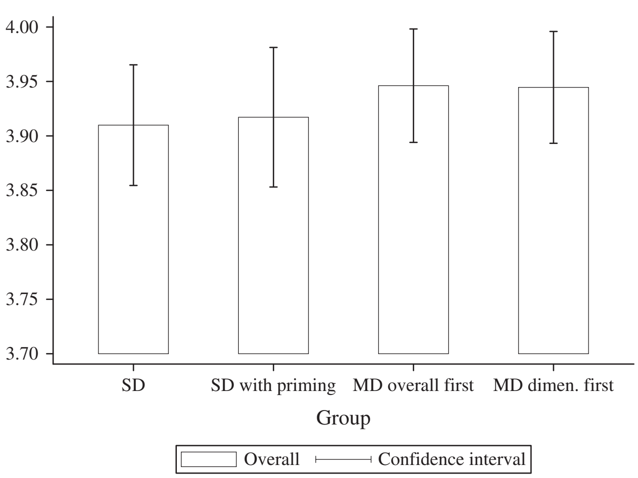
6.3.1. Experiment 1: Preference Matching. The purpose of this experiment is to examine whether the MD system facilitates preference matching and uncertainty reduction using a between-subject design.[5] Specifically, participants were shown information of one certain restaurant and asked to answer questions about how the information on the restaurant helped them determine if the restaurant matches their preference and if the provided information contributes to reducing their uncertainty about the restaurant. Participants were randomly assigned to either the treatment group, in which MD ratings were provided, or the control group, in which SD rating was provided. Subjects were first primed about a scenario in which they would go for lunch near campus. Then, they were shown one restaurant review page and asked to answer questions about preference matching and uncertainty. Restaurant information, such as name, description, overall rating, price level, and a sample of text reviews, was provided on the review page. Subjects also had the option to read more text reviews, as it is in the real setting. Identical information on the restaurant was provided to subjects in different groups, while additional MD ratings were provided to subjects in the MD group. Respondents were then asked to answer whether they wanted to go to this restaurant, and based on the information given to them about the restaurant, whether the restaurant fit their preference and if the provided information helped reduce their uncertainty (or increase their confidence) about the preference match with the restaurant. The questions on preference match were adapted from the "product fit uncertainty" scale from Hong and Pavlou (2014). We use the five-point Likert-type scale to measure the degree of preference match (where 1 indicates the subject strongly agrees that the restaurant is a good match for her/him) and degree of confidence (inverse of uncertainty) about preference match (where 1 indicates the subject is certain about the degree of preference match).[6] A total of 449 observations passed the sanity checks (based on the time used to answer all of the questions) and manipulation checks (whether the subjects did notice the MD ratings in MD systems and only overall ratings in the SD system), 221 for the SD group and 228 for the MD group. t-Tests show significant differences for groups that received MD ratings versus groups that received SD ratings, such that the MD system helps consumers choose the restaurant that better fits their preferences ($t = 2.19$, $p < 0.05$) and helps consumers reduce uncertainty when they make their restaurant choices ($t = 2.21$, $p < 0.05$). Therefore, MD ratings are more informative than merely an overall rating in the SD system.

6.3.2. Experiment 2: Priming Effect. The purpose of this experiment is to understand whether the priming effect does affect rating generation in a significant way. The experiment simulates an environment where consumers report ratings for restaurants after having the same consumption experience. Specifically, subjects were first asked to read a set of reviews describing the dining experiences and, based on those experiences, rate the restaurant using either the SD or the MD system. Pretests were conducted to ensure that subjects properly understood this task. Subjects were first shown four pieces of reviews. Subjects were informed that all of these reviews are authentic and that they needed to read all four reviews. We choose four reviews—a relatively small number—to reduce subjects' cognitive burden and make sure that all reviews have been read and all subjects obtain the same set of information. Then, subjects randomly enter into any of the four experimental groups with different rating interfaces: SD, SD with textual priming of product dimensions (termed "SD with priming"), MD with overall rating first, and MD with dimensional rating first. Subjects in the first experimental condition were asked to rate the restaurant using an SD system.

Subjects in the second condition were asked to consider different attributes of the restaurant (for example, service, food, ambiance, etc.) and then rate the restaurant using the SD system. If the priming effect exists, we would expect ratings of group 2 to be higher than ratings of group 1. Subjects in the third condition were asked to first rate the restaurant on the overall rating, and then rate different dimensions of the restaurant (food, service, atmosphere, and value). Subjects in the fourth condition were asked to first rate different dimensions of the restaurant (food, service, atmosphere, and value), and then rate the restaurant with the overall rating. Filler questions were used, and the time spent to finish answering the questions was recorded to make sure the subjects paid attention to the experiment. In total, we received 1,615 responses who passed the sanity check. We also checked to make sure that the results were consistent when we use a different cutoff time for valid responses based on time spent on the experiment.

Figure 5 shows the average overall ratings across four groups. We could tell that the average overall ratings of group 3 and group 4 are slightly higher than those of group 1 and group 2. However, ANOVA and pairwise t-tests suggest that the differences across groups are not statistically significant (p-values for the pairwise t-tests between any two groups are greater than 0.35; for details, see Table A9 in Online Appendix D). Therefore, there are no significant differences in the overall ratings across four groups using different rating systems. While we cannot completely rule out that the priming effect is a possible mechanism, the experimental evidence suggests that priming does not lead to significant rating differences in the SD and MD systems. Besides, we also checked the web design of *TripAdvisor* to see if the interface is likely to induce a priming effect. In theory, priming effect is likely more salient when reviewers were prompted to review different dimensions before giving an overall rating. However, in the design of TripAdvisor, reviewers were asked to provide the overall rating first, and reviewers had to scroll down the web page to provide the dimensional ratings. This design likely alleviates the priming effect because consumers were not primed about the different dimensions before giving the overall rating, unless they made an effort to scroll back to the top and change the overall rating, when filling out the dimensional ratings.[7]

Figure 5. The Priming Effect on the Overall Rating



7. Additional Analyses

7.1. Robustness Checks: Addressing Alternative Explanations of User Differences

7.1.1. Website Difference. In our main analysis, we used DID as our key identification strategy in Section 5 and used identical restaurants on Yelp as a control group. One possible concern with this approach is that the two websites may have different users to start with; in particular, the results could be explained if TripAdvisor users are more positive than Yelp users. If this explanation holds, we should observe significant rating differences across the two websites before the TripAdvisor system change. We conduct the following test to determine if a systematic difference regarding user rating exists between the two websites before the system change of TripAdvisor. If there is already a significant difference in the ratings of Yelp and TripAdvisor before the system change, the observed results could be due to a continuation of the difference. On the other hand, if there is no significant difference in ratings across the two websites prior to the system change, we can conclude with a higher confidence that the observed effects after system change are due to the system change. We test the difference in Equation (10):

$$Rating_{ij|k} = \beta_0 + \beta_1 * Treat + \beta_2 * \log n_{ij|k} + \alpha_i + \epsilon_{ij|k}. \quad (10)$$

Here, the coefficient β_1 captures the average difference in ratings between Yelp and TripAdvisor for the same restaurants before the system change. This allows us to check if rating differences are due to systematic differences between the two websites without the "shock" of system change.

Table 7 shows the estimation results of Equation (10), which aims to examine whether Yelp and TripAdvisor have any systematic difference in terms of overall rating prior to the system change. The first column reports results using all data from the two websites before the system change (from 2004 to the end of 2008). In the second column, we only use the rating data in years 2007 and 2008 that are within two years prior to the system change that took place in TripAdvisor. This allows us to investigate if ratings are comparable just before the system change. Similar to the main analyses, we control for restaurant fixed effect and number of reviews. The negative coefficient of $\log n_{ij|k}$ indicates that the ratings follow a downward

Table 7. Websites Difference (DV = *Rating*)

Sample	Before system change		Two years before system change	
log(# *of reviews*)	-0.087^{***}	(0.010)	-0.090^{***}	(0.014)
Treat	0.007	(0.023)	0.013	(0.025)
Constant	4.074^{***}	(0.031)	4.082^{***}	(0.047)
Restaurant FE	Yes		Yes	
Prof > *F*	0		0	
Observations	27,305		24,057	

Note. Cluster-robust standard errors in parentheses.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

trend, which is consistent with the results in Table 2. The insignificant coefficient of *Treat* indicates that Yelp and TripAdvisor have no systematic difference before the system change, regardless of the horizon of the time window. Overall, these results provide evidence that Yelp and TripAdvisor are comparable prior to the TripAdvisor system change. We also present a formal time effect test in the next section to show that the observed results were not merely a pretreatment trend that continued.

7.1.2. Within-Reviewer Analysis. We have provided evidence that Yelp and TripAdvisor have comparable ratings for the same restaurants before the system change; however, some may argue that the consumer bases of TripAdvisor may change because of the system change. That is, it is possible that those who provide ratings before the system change and after the system change are not the same set of consumers. To rule out this confounding effect, we focus on repeat reviewers and examine whether their overall ratings change under the MD system. Given that the same set of reviewers from the same website are tracked, the difference in ratings cannot be due to different user base or website difference. We test whether or not the result holds for only repeat reviewers (Equation (11)). We control for the amount of information any consumer receives prior to the consumption by adding the number of existing reviews for the restaurant as a control variable, and we test if this same set of reviewers indeed provide higher ratings after system change:

$$Rating_{ijr\,|\,\text{TripAdvisor}} = \beta_0 + \beta_1 * Post + \beta_2 * \log n_{ij} + \gamma_r + \epsilon_{ijr}. \tag{11}$$

The coefficient of interest is β_1, which measures whether the same reviewer provides higher ratings after the system change. To have enough repeat reviewers to understand their behaviors, we again utilize the data of all restaurants in New York City on TripAdvisor that we sampled. In total, we have 216,986 observations, which consists of 39,470 repeat reviewers with 140,418 ratings and 76,568 one-time reviewers. Details of data are presented in Tables 8 and 9.

Table 8. Data Description: Valence of Ratings

	Number	Obs.	Mean rating	Std. dev.
Repeat reviewers	39,470	140,418	4.09	1.01
One-time reviewers	76,568	76,568	3.97	1.21

Table 9. Data Descriptions: Number of Ratings

	Obs.	Mean	Std. dev.	Min	Max
Repeat reviewers	39,470	3.56	3.51	2	191

One-time reviewers are approximate twice the number of repeat reviewers with approximately half the number of total ratings. On average, each repeat reviewer had four reviews on average, 97.4% had no more than 10 reviews, and 99.95% had no more than 50 reviews. Four people had more than 100 reviews and were considered outliers and are removed from our analyses.

We report results of Equation (11) under different selection rules of reviewers in Table 10. The first column ($N \leq 100$) presents results with four outlier reviewers removed, and the second column presents results with reviewers who have authored no more than 50 reviews while the third column shows results with reviewers who have authored no more than 10 reviews. Across all selection rules of repeat reviewers, the positive coefficient of *Post* suggests that, among repeat reviewers, the overall rating increases significantly after the rating system change. In other words, we observe that same reviewers on TripAdvisor provide higher ratings after the system change. This result provides further support for our main finding that consumers are more satisfied after using the MD system as reflected in their ratings. Overall, since we focus on repeat reviewers within TripAdvisor, we are able to rule out the alternative explanation that reviewers may be different in the two websites or that there might be a different set of reviewers before and after the system change.

Table 10. Within-Reviewer Analysis (DV = *Rating*)

Sample	$N \leq 100$	$N \leq 50$	$N \leq 10$
Post	0.095^{*}	0.088^{*}	0.145^{**}
	(0.044)	(0.043)	(0.053)
log(# *of reviews*)	0.019^{***}	0.019^{***}	0.017^{***}
	(0.002)	(0.002)	(0.003)
Constant	3.933^{***}	3.940^{***}	3.910^{***}
	(0.043)	(0.043)	(0.053)
Reviewer FE	Yes	Yes	Yes
Prob > *F*	0	0	0
Observations	139,843	138,794	121,885

Note. Cluster-robust standard errors in parentheses.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

7.2. Dynamics of the Rating System Change

We conduct a time effect test to examine when exactly the observed effect of the MD system starts to take place by investigating the dynamic effects of the MD system change. We employ the approach suggested by Angrist and Pischke (2008), a commonly used method in the economics literature. This test serves two purposes. First, this analysis allows us to check a critical assumption for using the DID analysis that the observed effect does not occur pretreatment. If the effect occurred before the system change, then it is likely that something else is responsible for the observed outcome. Second, it allows us to narrow down the time the effect starts to manifest, therefore offering evidence that the effect is caused by the system change.

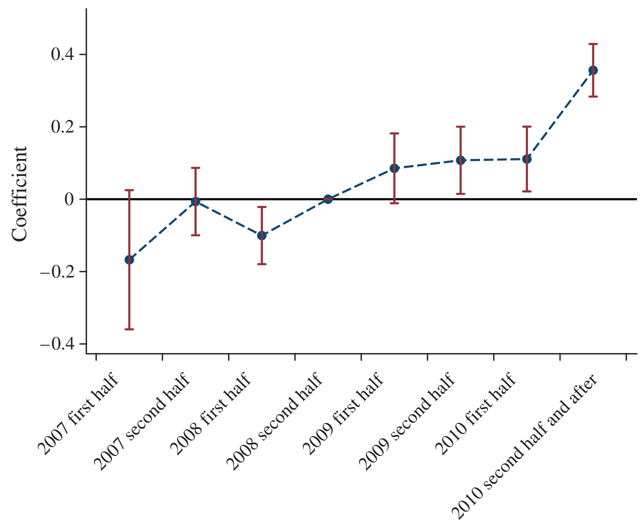
Specifically, we estimate the interactions of treatment (TripAdvisor) and the leads and lags of the rating system change. Specifically, we add three indicator variables for each half year before the system change, three indicators for each half year after the system change, one indicator for two years after the system change and afterward, and the interaction terms of indicators and treatment. The reference point is the ratings during the second half of 2008 (i.e., right before the system change). Table 11 presents the results. There is no specific trend on overall ratings within two years before the rating system change. Significant increase in ratings is observed in the second half of 2009; specifically, overall ratings increase by 0.107 in the second half of 2009 at TripAdvisor. No such positive effect is observed before the system change; therefore, what is observed is suggestive evidence that the observed effects occurred after system change and that the effects were not a pretreatment trend that continued. The results also show that the overall ratings continue to increase after the system change. We visualize this pattern in Figure 6. The results in Table 11 and Figure 6 corroborate with our findings in the previous section that the adoption of the MD system increases the overall ratings compared to those of the SD system.

Table 11. Results for Parallel Trend Analysis (DV = *Rating*)

Treat	0.146***	(0.037)
Treat ∗ *Year 2007 first half*	−0.168	(0.098)
Treat ∗ *Year 2007 second half*	−0.007	(0.047)
Treat ∗ *Year 2008 first half*	−0.100*	(0.040)
Treat ∗ *Year 2008 second half*	Omitted	(baseline)
Treat ∗ *Year 2009 first half*	0.085	(0.049)
Treat ∗ *Year 2009 second half*	0.108*	(0.047)
Treat ∗ *Year 2010 first half*	0.111*	(0.046)
Treat ∗ *Year 2010 second half and after*	0.356***	(0.037)
Time dummies		
Year 2007 first half	0.173***	(0.031)
Year 2007 second half	0.152***	(0.027)
Year 2008 first half	0.122***	(0.023)
Year 2008 second half	Omitted	(baseline)
Year 2009 first half	0.021	(0.021)
Year 2009 second half	−0.017	(0.022)
Year 2010 first half	0.001	(0.023)
Year 2010 second half and after	−0.019	(0.024)
log(*# of reviews*)	0.005	(0.007)
Constant	3.665***	(0.031)
Restaurant FE	Yes	
Prof > *F*	0	
Observations	259,751	

Note. Cluster-robust standard errors in parentheses.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Figure 6. (Color online) Time Passage Relative to Year of Adoption of the MD System



7.3. Other Robustness Checks

In this section, we summarize results from a number of additional robustness checks. First, it is possible that restaurant operations are affected by season. When we control for seasonality, we find that restaurant ratings are higher in winter. Yet the results of variables of interest still hold after we control for seasonality.[8]

Second, we used more than 50,000 observations in our main analysis. With such a large sample, small p-values are easier to obtain (Lin et al. 2013). To rule out large sample being the reason of the strong statistical significance, we randomly sampled 200 restaurants, which leaves us fewer than 10,000 observations and the significant effects persist.

Third, although our dynamic time dummy interaction model based on Angrist and Pischke (2008) provides evidence that the observed effect was due to system change, some may argue that other system changes introduced in TripAdvisor, which are unrelated to rating systems, may cause the rating difference. While it is difficult to track all of the system changes, we checked major changes, such as management response. Because both Yelp and TripAdvisor allow owners to respond to comments, we believe it is not a serious concern to the validity of the results since their effects will be differenced out in the DID model. But future research is warranted to study how management response would affect consumer ratings. Overall, our results, based on both observational data and randomized experiments, provide strong support showing that presenting product information on different dimensions increase rating informativeness, leading to better decision making and higher consumer satisfaction reflected in ratings.

8. Discussion

8.1. Key Findings

This study extends the limited understanding of different online rating system designs. Based on a unique dataset from two leading online review platforms, the results of this study first show that ratings trend down in the SD system but not in the MD system. Although downward trending of ratings has been observed in prior studies and explained by self-selection (Li and Hitt 2008, Godes and Silva 2012), the rating trend in the MD system is inconsistent with the self-selection explanation. We identified a plausible mechanism of rating informativeness that at least partially explains the different rating trends in SD and MD systems. We show that the overall ratings are convergent and increasing by 0.40 after the adoption of the MD system. These results suggest that the MD system enhances consumers' ability to make more informed decisions and that they are more satisfied with their decisions after using the MD system. We also want to note that the current system of TripAdvisor does not have MD ratings side by side with overall ratings, but dimension-specific numeric ratings are included in the "restaurant details" section, and when the individual review is lengthy, individual MD ratings are shown when a consumer clicks to expand and read more about the review. Any effects that arise from introducing the MD system must operate on those people who actually are exposed to restaurant details or the individual MD ratings. This suggests that stronger effects are likely to be observed if these dimensional ratings are displayed more prominently.

To offer further support for the theoretical mechanism of rating informativeness and/or priming due to the rating interface, we performed additional analyses and conducted two randomized experiments. The impact of the MD system on rating informativeness is further corroborated by the results that the effects are stronger when additional dimensional ratings are accumulated. The experimental results triangulate the findings from the observational study by showing that the MD system facilitates preference matching and uncertainty reduction, both of which allow consumers to process rating information more efficiently, leading to higher satisfaction. While we cannot fully rule out that the priming effect does not exist, our analyses and experimental results suggest that its effect, if present, is not strong enough to create a statistically significant difference in ratings and therefore is not a primary source of our observed results.

We also ruled out a number of alternative explanations. First, we show that TripAdvisor and Yelp are comparable to start with. These two websites have no significant difference regarding average ratings before the system change. Moreover, to rule out the possibility that reviewer bases may be different across the two websites and before and after the system change, we focus on repeat reviewers on TripAdvisor and show that higher overall ratings after system change still exist among those repeat reviewers. We further ensure that the adoption of the MD system increases the overall ratings compared to those of the SD system through investigating the dynamics of overall rating after the rating system change. This helps us rule out the alternative explanation that the observed effect is due to a pretreatment trend that continued.

In summary, this study provides robust evidence suggesting that switching from the SD to MD system provides substantial benefits to consumers, particularly for experience goods and services like restaurants, for which the quality is difficult to observe and fitness is difficult to infer prior to consumption. When multiple dimensions are rated, it is easier for consumers to obtain both quality and fit information, allowing consumers to form more rational expectations, leading to better decision making and higher satisfaction.

8.2. Implications

While previous research has investigated the effects of different product attributes on pricing power, hotel ranking, and review helpfulness (Decker and Trusov 2010, Archak et al. 2011, Ghose and Ipeirotis 2011, Ghose et al. 2012), as well as the effects of crowd and friends on consumer reviews (Lee et al. 2015, Wang et al. 2018), this study is, to our knowledge, the first to directly compare SD and MD systems and examine the effect of system design on product ratings. This study addresses whether the MD system enhances rating informativeness by examining whether the MD system leads to more informed purchase decisions and more satisfied consumers. We extend the limited understanding of online rating system designs endorsed by many IS scholars (Li and Hitt 2010, Archak et al. 2011, Ghose and Ipeirotis 2011). Our model relates the product information that consumers can gain from online rating system to product uncertainty (Dimoka et al. 2012, Kwark et al. 2014), which is further integrated to consumer expectation and satisfaction based on ECT (Anderson and Sullivan 1993), information transfer (Belkin 1984) and priming due to rating interface. We revisit prior work on the dynamic effects of ratings where the downward trend of SD ratings is observed (Li and Hitt 2008, Godes and Silva 2012, Moe and Schweidel 2012). In addition to providing a complementary explanation for a downward trend in the SD rating system based on rating informativeness, we also show that such limitations can possibly be addressed by using the MD system. This study also extends the extant research on how IT-enabled technologies can reduce various sources of consumer product uncertainty (Dimoka et al. 2012, De et al. 2013, Hong and Pavlou 2014).

This study also has important managerial implications. First, the results from this study inform practitioners about whether or not adopting the MD system can improve the performance of online product reviews and also provide insights on the effective design of informative rating systems. The MD system is particularly useful for products and services for which consumers have heterogeneous preferences while the SD system would suffice when the utility of products and services can be unambiguously summarized in one numerical value. Our results also indicate that the effect of the MD system design varies depending on the number of MD ratings accumulated. Therefore, review websites should incentivize consumers to provide dimensional ratings. Besides, for effective design of the MD system, it is important to identify orthogonal dimensions/attributes of products and services for which consumers show heterogeneous preferences. Altogether, the MD system increases rating informativeness through more informative reviews and by presenting information in a way that enables easier preference matching.

8.3. Limitation and Future Research

As with most empirical studies, this study is not free of limitations. First, while we have provided evidence supporting the value of adopting the MD system, the results may not be generalized to other types of products, particularly search goods with attributes that are easily conveyed with product descriptions. A potential future study is to look at whether different performance effects are observed for different types of products (i.e., search, experience, and credence goods) when the MD system is introduced. Second, although we expended significant efforts to tease out differences between Yelp, TripAdvisor, and their respective users, the available data limit us from fully ruling out the possibility that the changes are driven by some unobserved or confounding changes (on Yelp, on TripAdvisor, or offsite) that we are not aware of. We have attempted to address the limitations of observational data using experiments, and we were able to offer some suggestive evidence that the higher rating informativeness in the MD system is at least an important underlying mechanism for the observed effect of the MD system on product ratings. We also offered evidence that one important alternative explanation—the priming effect due to different rating interface—does not significantly explain the findings. Still, there could be other mechanisms at play. Future research could dig further on how different rating system interfaces affect consumers' rating behavior. Third, another limitation is that in this paper, we assume that consumers who write a review in a platform also read the reviews from the same platform before they dine at a restaurant. This is not likely to jeopardize our conclusion, because if some consumers consult both TripAdvisor and Yelp before dining at a restaurant, it is unlikely we will observe a significant effect of implementing the multidimensional rating system. Therefore, our estimates are likely conservative. Fourth, it would also be interesting to examine the dynamic effects of MD ratings in a future study. It is possible that the effect of the MD system is stronger when the number of prior reviews is small, the prior reviews are very short and uninformative, or when very recent reviews included MD information, in addition to the overall rating. Lastly, in one of our robustness tests, we utilized consumers who had rated on both SD and MD systems (for different occasions), it will be interesting to systematically examine why the same consumer would switch between MD and SD systems in TripAdvisor and when a consumer tends to provide MD ratings. For example, it is likely that consumers provide MD ratings when they perceive variation in quality in different dimensions.

8.4. Concluding Remark

Based on a quasi-experimental DID analysis in a natural experiment setting corroborated with evidence from two randomized experiments, this paper uses a multimethod approach to investigate the value of adopting the MD system relative to the SD system, from the perspective of rating informativeness and priming due to rating interface. Our results from both the observational data and lab experiments provide evidence that the MD system enhances rating informativeness because the way information is presented on different dimensions allows better preference matching and uncertainty reduction, leading to more informed purchasing decisions and more satisfied consumers. Given the societal importance of online product review systems, our study serves as an important first step toward establishing the value of rating system designs.

Acknowledgments

The authors thank the department editor, the associate editor, and the three anonymous reviewers for a constructive and developmental review process, and participants of the 2014 International Conference on Information Systems (ICIS) for helpful discussions and valuable feedback on earlier versions of the paper.

Endnotes

[1] Note that in this study, we assume consumers read and write reviews on the same website. If users read and write reviews on both

Yelp and TripAdvisor, we are less likely to see an effect from the MD system. See Section 8.3 for further discussions on this assumption.

[2] Note that the aggregated dimensional ratings are available on the "Restaurant Details" page and are also shown with any individual review a consumer may click on.

[3] We obtained the figures from the following two pages for the sample restaurant "House of Tricks": Yelp, https://www.yelp.com/biz/house-of-tricks-tempe (accessed March 10, 2017); TripAdvisor, https://www.tripadvisor.com/Restaurant_Review-g31377-d334370-Reviews-or10-House_of_Tricks-Tempe_Arizona.html (accessed March 10, 2017).

[4] Institution Review Board (IRB) approval from the authors' university was obtained before the experiments started. Participants were informed that they would receive extra credits and a random draw of a cash gift if they finished the study attentively by passing screening and manipulation check questions.

[5] Details and screenshots of the two experiments are available in Online Appendix B.

[6] Questions (1a) and (2a) are about preference match; Questions (1b) and (2b) are about the degree of (un)certainty of preference match: (1a) This restaurant would fit my preference. (1b) I am sure that this restaurant would fit my preference. (Note that this question asks how confident you are in answering the previous question.) (2a) This restaurant with these characteristics is what I am looking for. (2b) I am sure that this restaurant with these characteristics is what I am looking for. (Note that this question asks how confident you are in answering the previous question.)

[7] A screenshot (taken in March 2017) of how users rate on the overall rating and dimensional ratings on TripAdvisor can be found in Online Appendix C.

[8] Details for robustness checks are available in Online Appendix A.

References

Anderson M, Magruder J (2012) Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *Econom. J.* 122(563):957–989.

Anderson EW, Sullivan MW (1993) The antecedents and consequences of consumer satisfaction for firms. *Management Sci.* 12(2):125–143.

Angrist JD, Pischke JS (2008) *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, Princeton, NJ).

Archak N, Ghose A, Ipeirotis PG (2011) Deriving the pricing power of product features by mining consumer reviews. *Management Sci.* 57(8):1485–1509.

Bargh JA, Tanya LC (2000) The mind in the middle: A practical guide to priming and automaticity research. Reis HT, Judd CM, eds. *Handbook of Research Methods in Social and Personality Psychology* (Cambridge University Press, New York), 253–85.

Belkin NJ (1984) Cognitive models and information transfer. *Soc. Sci. Inform. Stud.* 4(2):111–129.

Bhattacherjee A (2001) Understanding information systems continuance: An expectation-confirmation model. *MIS Quart.* 25(3): 351–370.

Brown SA, Venkatesh V, Goyal S (2012) Expectation confirmation in technology use. *Inform. Systems Res.* 23(2):474–487.

Brown SA, Venkatesh V, Goyal S (2014) Expectation confirmation in Information Systems research: A test of six competing models. *MIS Quart.* 38(3):729–756.

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

Churchill GA, Suprenant C (1982) An investigation into the determinants of consumer satisfaction. *J. Marketing Res.* 19(4):491–504.

Cremer H, Thisse JF (1991) Location models of horizontal differentiation: A special case of vertical differentiation models. *J. Indust. Econom.* 39(4):383–390.

Dai W, Jin GZ, Lee J, Luca M (2012) Optimal aggregation of consumer ratings: An application to Yelp.com. National Bureau of Economic Research Working Paper 18567.

De P, Hu Y, Rahman MS (2013) Product-oriented web technologies and product returns: An exploratory study. *Inform. Systems Res.* 24(4):998–1010.

Decker R, Trusov M (2010) Estimating aggregate consumer preferences from online product reviews. *J. Marketing Res.* 27(4): 293–307.

De Langhe B, Fernbach PM, Lichtenstein DR (2016) Navigating by the stars: Investigating the actual and perceived validity of online user ratings. *J. Consumer Res.* 42(6):817–833.

Dellarocas C, Zhang X, Awad N (2007) Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *J. Interactive Marketing* 21(4):23–45.

Diehl K, Poynor C (2010) Great expectations?! Assortment size, expectations, and satisfaction. *J. Marketing Res.* 47(2):312–322.

Dimoka A, Hong Y, Pavlou PA (2012) On product uncertainty in online markets: Theory and evidence. *MIS Quart.* 32(3):395–426.

Duan W, Gu B, Whinston AB (2008) Do online reviews matter?—An empirical investigation of panel data. *Decision Support Systems* 45(4):1007–1016.

Forman C, Ghose A, Weisenfeld B (2008) Examining the relationship between reviews and sales: The role of consumer identity disclosure in electronic markets. *Inform. Systems Res.* 19(3):291–313.

Garvin DA (1984) What does "product quality" really mean? *Sloan Management Rev.* 26(1):25–43.

Ghose A, Ipeirotis PG (2011) Estimating the helpfulness and economic impact of product reviews: Mining text and consumer characteristics. *IEEE Trans. Knowledge Data Engrg.* 23(10): 1498–1512.

Ghose A, Ipeirotis PG, Li B (2012) Designing ranking systems for hotels on travel search engines by mining user-generated and crowd sourced content. *Marketing Sci.* 31(3):493–520.

Godes D, Mayzlin D (2004) Using online conversations to study word-of-mouth communication. *Marketing Sci.* 23(4):545–560.

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

Goes PB, Lin M, Yeung C-mA (2014) "Popularity effect" in user-generated content: Evidence from online product reviews. *Inform. Systems Res.* 25(2):222–238.

Gu B, Park J, Konana P (2012) The impact of external word-of-mouth sources on retailer sales of high-involvement products. *Inform. Systems Res.* 23(1):182–196.

Helson H (1964) *Adaptation-Level Theory* (Harper & Row, New York).

Hennig-Thurau T, Gwinner KP, Walsh G, Gremler DD (2004) Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *J. Interactive Marketing* 18(1):38–52.

Hinckley D (2015) New study: Data reveals 67% of consumers are influenced by online reviews. *Moz* (blog) (September 2), https://moz.com/blog/new-data-reveals-67-of-consumers-are-influenced-by-online-reviews.

Hong Y, Pavlou PA (2014) Product fit uncertainty in online markets: Nature, effects and antecedents. *Inform. Systems Res.* 25(2): 328–344.

Hong Y, Huang N, Burtch G, Li C (2016) Culture, conformity and emotional suppression in online reviews. *J. Assoc. Inform. Systems* 17(11):737–758.

Hu M, Liu B (2004) Mining and summarizing customer reviews. *Proc. 10th ACM SIGKDD Internet Conf. Knowledge Discovery and Data Mining* (*KDD*-2004) (ACM, New York), 168–177.

Hu N, Paul P, Jennifer Z (2006) Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word-of-mouth communication. *Proc. 7th ACM Conf. Electronic Commerce* (ACM, New York), 324–330.

Huang N, Burtch G, Hong Y, Polman E (2016) Effects of multiple psychological distances on construal and consumer evaluation: A field study of online reviews. *J. Consumer Psych.* 26(4):474–482.

Kim DJ, Ferrin DL, Rao HR (2009) Trust and satisfaction, two stepping stones for successful e-commerce relationships: A longitudinal exploration. *Inform. Systems Res.* 20(2):237–257.

Kwark Y, Chen J, Raghunathan S (2014) Online product reviews: Implications for retailers and competing manufacturers. *Inform. Systems Res.* 25(1):93–110.

Lee YJ, Tan Y, Hosanagar K (2015) Do I follow my friends or the crowd? Examining informational cascades in online movie reviews. *Management Sci.* 61(9):2241–2258.

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

Li X, Hitt LM (2010) Price effects in online product reviews: An analytical model and empirical analysis. *MIS Quart.* 34(4): 809–831.

Lin M, Lucas HC, Shmueli G (2013) Research Commentary—Too big to fail: Large samples and the *p*-value problem. *Inform. Systems Res.* 24(4):906–917.

Lin TC, Wu S, Hsu JSC, Chou YC (2012) The integration of value-based adoption and expectation–confirmation models: An example of IPTV continuance intention. *Decision Support Systems* 54(1):63–75.

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

Luca M, Zervas G (2016) Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Sci.* 62(12): 3412–3427.

Mangold W, Miller F, Brockway G (1999) Word-of-mouth communication in the service marketplace. *J. Services Marketing* 13(1): 73–89.

Meyer DE, Schvaneveldt RW (1971) Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *J. Experiment. Psych.* 90(2):227–234.

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

Oliver RL (1980) A cognitive model of the antecedents and consequences of satisfaction decisions. *J. Marketing Res.* 17(11): 460–469.

Sahoo N, Dellarocas C, Srinivasan S (2015) The impact of online product reviews on product returns. Boston University School of Management Research Paper 2491276, https://ssrn.com/abstract=2491276.

Sela A, Shiv B (2009) Unraveling priming: When does the same prime activate a goal versus a trait? *J. Consumer Res.* 36(3):418–433.

Sundaram DS, Mitra K, Webster C (1998) Word-of-mouth communications: A motivational analysis. *Adv. Consumer Res.* 25(1): 527–531.

Streitfeld D (2012) The best book reviews money can buy. *New York Times* (August 25), http://www.nytimes.com/2012/08/26/business/book-reviewers-for-hire-meet-a-demand-for-online-raves.html.

Venkatesh V, Goyal S (2010) Expectation disconfirmation and technology adoption: Polynomial modeling and response surface analysis. *MIS Quart.* 34(2):281–303.

Wang C, Zhang X, Hann I-H (2018) Socially nudged: A quasi-experimental study of friends' social influence in online product ratings. *Inform. Systems Res.*, ePub ahead of print May 10, https://doi.org/10.1287/isre.2017.0741.

Wang Y, Ramachandran V, Sheng O (2016) The causal impact of fit valence and fit reference on product returns. *Proc. Internat. Conf. Inform. Systems, Dublin, Ireland*.