Exploring the Mechanisms of Information Sharing

by

Ying Liu

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2018 by the
Graduate Supervisory Committee:

Pei-yu Chen, Co-Chair
Yili Hong, Co-Chair
Bin Gu

ARIZONA STATE UNIVERSITY

December 2018

ABSTRACT

Online product ratings offer consumers information about products. In this dissertation, I explore how the design of the rating system impacts consumers' sharing behavior and how different players are affected by rating mechanisms. The first two chapters investigate how consumers choose to share their experiences of different attributes, how their preferences are reflected in numerical ratings and textual reviews, whether and how multi-dimensional rating systems affect consumer satisfaction through product ratings, and whether and how multi-dimensional rating systems affect the interplay between numerical ratings and textual reviews. The identification strategy of the observational study hinges on a natural experiment on TripAdvisor when the website reengineered its rating system from single-dimensional to multi-dimensional in January 2009. Rating data on the same set of restaurants from Yelp, were used to identify the causal effect using a difference-in-difference approach. Text mining skills were deployed to identify potential topics from textual reviews when consumers didn't provide dimensional ratings in both SD and MD systems. Results show that ratings in a single-dimensional rating system have a downward trend and a higher dispersion, whereas ratings in a multi-dimensional rating system are significantly higher and convergent. Textual reviews in MDR are in greater width and depth than textual reviews in SDR. The third chapter tries to uncover how the introduction of monetary incentives would influence different players in the online ecommerce market in the short term and in the long run. These three studies together contribute to the understanding of rating system/mechanism designs and different players in the online market.

ACKNOWLEDGMENTS

I am grateful to all of those with whom I have received support and encouragement. I would like to thank my supervisor, Prof. Pei-yu Chen, for the patient guidance, encouragement and advice she has provide throughout the past few years. I was so lucky to have her as my supervisor without whose guidance and help this dissertation would not have been possible. I would like to thank my committee co-chair, Prof. Yili Hong who has provided me extensive professional guidance and taught me a great deal about how to improve my research. I would also like to thank my dissertation committee member Prof. Bin Gu who has been super supportive during this long journey. In addition, a thank you to Prof. Raghu Santanam, who provided extraordinary support to me throughout the program. Furthermore, I must express my gratitude to all the members of the Department of Information Systems (faculty, staff and PhD students) for their continued support and encouragement. Finally, I would like to thank my husband, my parents and my sister, whose love are with me whatever I pursue.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

THE VALUE OF MULTI-DIMENSIONAL RATING SYSTEMS

1.1 Introduction

The substantial increase in online word of mouth (WOM) in the form of online product reviews and ratings has transformed the way consumers acquire product information. Online product reviews enable consumers to acquire product information and simultaneously share their experience of product usage. According to a recent article in The New York Times (2012), “Reviews by ordinary people have become an essential mechanism for selling almost anything online.” Given that most consumers refer to online reviews before they make a purchase decision, reviews are expected to have a significant effect on sales. However, mixed findings on the effects of ratings on consumer decision-making have been reported (Chevalier and Mayzlin 2006, Duan et al. 2008, Godes and Silva 2012). Such mixed findings question the assumption that either ratings or text reviews efficiently convey all the dimensions of product quality. Furthermore, these findings are based on single-dimensional rating systems, in which consumers report only their overall satisfaction. Therefore, this lack of consensus may be a result of the limitation of single-dimensional ratings in efficiently transferring product quality information. Some scholars endorse multi-dimensional rating systems as a relatively better means of conveying quality information because product quality is often comprised of multiple dimensions (Archak et al. 2011). This study directly investigates whether or not, and to what extent, multi-dimensional rating systems enhance information transfer efficiency among consumers compared to single-dimensional rating systems.

In practice, single-dimensional rating systems allow consumers to submit a numerical rating of the product (i.e., usually on a discrete interval scale of 1–5 stars) with an option to submit additional text reviews. These numerical ratings are then aggregated and presented as an average value (valence) or a rating distribution. However, products consist of multiple attributes, and people usually have heterogeneous preferences and place different weights on different attributes. Thus, such ignorance of consumer heterogeneity in single-dimensional rating systems can limit information transfer efficiency because consumers may have different interpretations of the ratings. Most current online product rating systems follow such single-dimensional systems (e.g., Amazon.com, Yelp.com, etc.), with a few exceptions that acquire and present ratings in multiple dimensions (e.g., TripAdvisor.com).

Given the theoretical importance and practical significance of online WOM systems, information system (IS) scholars (Li and Hitt 2010, Archak et al. 2011) have called for the rigorous examination of the design of rating systems, particularly the informational value of multi-dimensional rating systems. At first glance, multi-dimensional rating systems provide more information because they allow previous consumers to share their consumption experiences in terms of different dimensions, which can be more meaningful to future consumers than a single overall rating, particularly when consumers derive the utility of a product/service from different key attributes (dimensions). For example, when consumers plan to dine at a restaurant, different consumers have different preferences in terms of food quality, service, and restaurant ambience. Furthermore, these consumer preferences may also vary for different occasions. Consumers essentially face two types of uncertainty in such scenarios: product quality uncertainty (vertical quality dimension) and fit uncertainty (horizontal quality/preference dimension). Consumers rely on online ratings for information that help resolve such uncertainties (Kwark et al. 2014). Given the potential multi-dimensional nature of consumer preferences for most products, particularly experience products, matching the idiosyncratic preferences of consumers with a single numerical rating, as in single-dimensional rating systems, is difficult. Multi-dimensional ratings provide systematic information on both vertical (quality) and horizontal (preference) dimensions, and consumers who gain information from multi-dimensional rating systems presumably obtain a more accurate estimate of the utility from consuming a product. In other words, multi-dimensional rating systems should facilitate matching between consumers and products because they contain more information.

However, more information does not necessarily translate to high information transfer efficiency. First, excessive information can lead to the cognitive overload of consumers (Simon 1982). Given that the information contained in multi-dimensional rating systems possibly lead to higher evaluation costs for consumers, it is not clear if additional multi-dimensional ratings result in a net increase in decision performance relative to a single rating. Only when information transfer is efficient will consumers make more informed decisions and be more satisfied with their purchases. Moreover, both single-dimensional and multi-dimensional rating systems provide text reviews. The text reviews in single-dimensional rating systems can be informative because previous consumers express their evaluations of different dimensions in their detailed text reviews, which can also help future consumers to resolve their uncertainties (Archak et al. 2011). Therefore, the value of and need for multi-dimensional ratings decrease when consumers can effectively obtain quality information from texted reviews. In terms of system implementation, re-designing a single-dimensional rating system into a multi-dimensional system is also costly. Furthermore, consumers may also find rating different dimensions time consuming given the extra effort required, which may potentially reduce content generation quantity and quality. In summary, empirically examining whether or not a multi-dimensional system makes information transfer easier among consumers and quantifying its value have considerable value to both researchers and practitioners.

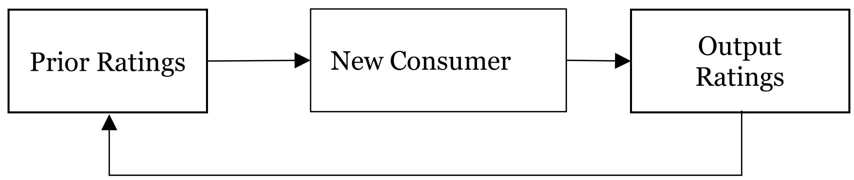

Figure 1. Dual Role of Ratings

Building on information transfer theory in social science, we define information transfer efficiency as the extent to which information from a knowledge resource help solve the "problem" of consumers. In the context of ratings, the "problem" is the purchase (or consumption) decision. When information is transferred efficiently from knowledge resource to consumers, consumers should be able to make more informed decisions and become more satisfied with their decisions. We examine the efficiency of information transfer by comparing the dynamics of ratings over time in a single-dimensional rating system re-engineered into a multi-dimensional rating system versus a constant single-dimensional rating system. Ratings serve dual roles in information transfer as shown in Figure 1. First, existing ratings serve as input, from which consumers can form expectations of consumption utility. Moreover, consumers can also rate the product according to their expectations and realized consumption experience, thereby generating the output ratings that reflect their satisfaction with their purchase decisions. The efficiency of the information transfer can be inferred by comparing the input and output ratings. The deviation of the output ratings from the input ratings suggests that the consumption experiences of consumers don't match their expectations, which are formed based on the reviews of prior consumers. Such deviation indicates that information is not efficiently transferred, so that inaccurate expectations are formed. By contrast, when information transfer is efficient, it is easier for consumers to distinguish between different products based on input ratings and form reasonable expectations of utility from consuming a product. As a result, consumers are less likely to be disappointed because the consumption utility is likely to confirm the expected utility that is formed based on prior ratings and reviews. In other words, output ratings are likely to match input ratings.

To address our research questions, we collected observational data from two leading restaurant review websites: Yelp and TripAdvisor. We sampled 1207 restaurants in New York City and obtained reviews for these same restaurants from the two websites to construct our panel data set. We then examined how these same restaurants are rated in these different rating systems. Our main econometric identification strategy hinges on a natural experiment that took place on TripAdvisor, which re-engineered its rating system from single-dimensional to multi-dimensional and implemented the multi-dimensional system in January 2009. By contrast, Yelp did not make such change and continues to maintain a single-dimensional rating system. Such system change allows us to specify our empirical model in a quasi-experimental difference-in-difference (DID) framework. Tracking identical restaurants on the two review sites essentially allow us to control for unobserved restaurant quality change over time.

Several interesting results emerge from our econometric analyses. First, we show that the overall ratings no longer follow a downward trend after a multi-dimensional rating system is adopted, in contrast to those in a single-dimensional rating system. On the average, the overall rating of a restaurant on TripAdvisor increases by 0.154. The increase in ratings becomes notably stronger as more dimensional ratings are accumulated. This result is consistent with the view that the multi-dimensional rating system enables and enhances information transfer efficiency among consumers, thereby leading to more effective purchase decisions and more satisfied customers. Second, we show that ratings on multi-dimension rating systems are convergent, which suggests that consumer consumptions meet their expectations. This result is consistent with the finding that multi-dimensional rating systems enhance information transfer efficiency. Overall, our study makes a pioneering effort in establishing the causal effect of adopting a multi-dimensional rating system using a real-world quasi-natural experiment.

1.2 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 has 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.

1.3 Theory and Hypotheses Development

The focus of this study is the examination of whether or not multi-dimensional rating systems increase information transfer efficiency. We first provide a theory on information transfer efficiency. We then leverage expectation-confirmation theory (ECT) to theorize the effect of multi-dimensional rating systems (relative to single-dimensional rating systems) on information transfer efficiency. In this section, we measure the effect by comparing the overall ratings for identical products on two websites that adopt single- and multi-dimensional rating systems, respectively.

Social scientists consider information system as the dynamic interaction among three components: the user (consumer), knowledge resource, and intermediary mechanism between the knowledge resource and user (Belkin 1984). According to Belkin (1984), the knowledge resource contains texts (i.e., in the semiotic sense) that are represented and organized in certain ways. The user initiates the system because of some problem, goals, or intentions, whose management or realization he or she believes may be enhanced using the information obtained from the knowledge resource. The intermediary mechanism mediates between the desires, requirements, knowledge, and so on of the user on the one hand, and the contents, representation, and organization of the knowledge resource on the other. The function of such system is information transfer, or the communication of useful information to the user from the knowledge resource via the intermediary. Based on this view, the rating system can be regarded as the intermediary mechanism, whereas existing ratings 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." Information transfer is considered effective or efficient when users are better able to understand and/or manage the problem that initiates the "system." In the context of purchase, this definition of information transfer efficiency suggests that consumers can make more informed purchase decisions and are satisfied with their purchase decisions, which are made based on the information obtained from the knowledge resource (i.e., the prior ratings). After the consumption of the information in the knowledge resource, consumer satisfaction regarding their purchase decisions is the key to understanding whether information transfer is efficient or not. This phenomenon is shown in Figure 2.

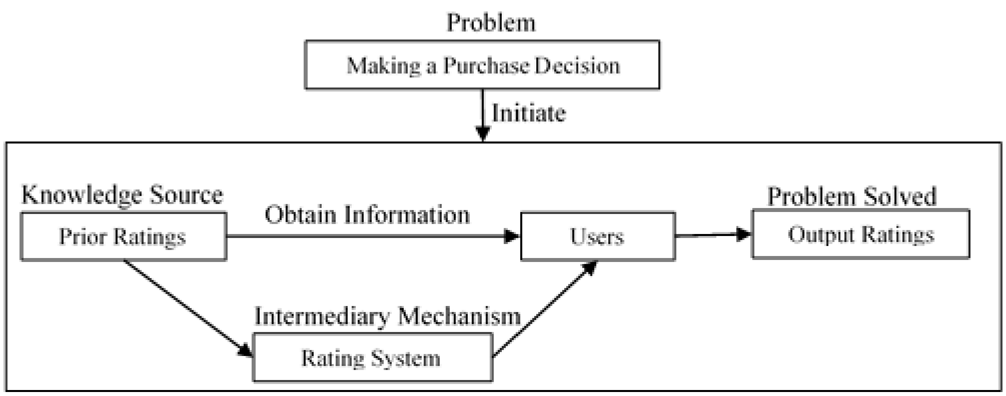

Figure 2. Information Transfer Model

ECT is widely used in the literature on information systems and marketing to understand system adoption (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 (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 level of expectation of product performance is an adaptation level. The degree to which the product exceeds, meets, or falls short of one's expectation may cause post-decision deviations from the adaptation level. Subsequent research (Anderson and Sullivan 1993) found that perceived quality and disconfirmation of expectation have a direct effect on satisfaction. They also reported an asymmetric (dis)confirmation effect, in which negative confirmation (disconfirmation) has a greater effect on satisfaction than positive confirmation.

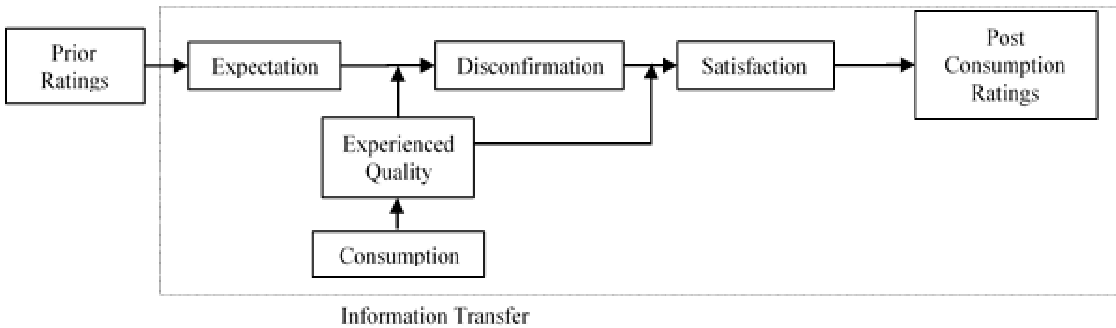

Figure 3. Information Transfer and ECT Model

We adopt the ECT model from Anderson and Sullivan (1993) and combine it with information transfer perspective (Belkin, 1984) to form the basis of our theoretical development. Ratings serve dual functions (i.e., as input and as output) that allow us to examine the efficiency of information transfer by observing the dynamics of ratings over time in single-dimensional versus multi-dimensional rating systems. As discussed earlier, existing ratings serve as knowledge resource or input for consumers in search of information. Consumers can form an expectation of consumption utility from using the products through this input. Alternatively, consumers can also rate the product (i.e., output rating) according to the information they obtain from a knowledge resource and from their own consumption experiences. Comparing the input ratings and output ratings gives us a signal of the information transfer efficiency. Inefficient information transfer occurs when the consumption experiences don't match the input ratings that consumers received from other consumers. Therefore, consumers are likely to be disappointed. By contrast, a match between the output ratings and the input ratings suggests that consumption experience confirms the product evaluation of previous customers. Overall, we can get a measure of information transfer efficiency by studying the dynamics of ratings over time.

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 due to 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. We expect to see a downward trend in the SD system but not in the MD system. We propose the following hypotheses:

Hypothesis 1 (H1): The SD system exhibits a downward trend of ratings.

We proceed to discuss the effect of MD system on product ratings. The MD system may affect product ratings because information is transferred more effectively. 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. We expect that ratings in MD are more likely to be higher. Therefore, we propose:

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 transfer efficiency, 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.

1.4 Research Methodology

1.4.1. Data

We draw on consumer review data to address our research questions by studying restaurant reviews in different rating systems. We choose restaurants as our context because restaurants have well-known different dimensions of services (e.g., food and location) and attract significant attention in academic literature. Our empirical analysis utilizes restaurant review data gathered from two leading consumer review websites: Yelp.com (Yelp) (i.e., covering Nov 2004 to April 2013) and TripAdvisor.com (TripAdvisor) (i.e., covering May 2004 to April 2013). Like most review websites, Yelp provides a single-dimensional rating system on a scale of five stars. TripAdvisor had been using a single-dimensional rating system until January 2009, when the website re-engineered its system and implemented a multi-dimensional rating system, which provides not only overall ratings but also ratings for the dimensional characteristics of restaurants, such as food, service, atmosphere, and value, using the same five-star rating scale.

We used two customized web crawlers for data collection. To eliminate restaurant differences and control for unobserved quality changes in the restaurants, we obtain data for exactly the same 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 effect. Specifically, we matched the restaurants on Yelp and TripAdvisor according to restaurant names, addresses, and phone numbers. The two websites have a total of 1,207 restaurants in common in New York City. We collected all available reviews for these common restaurants. For each review, we collected the time stamp of when the review was reported, the consumer ID, and the star rating (i.e., an integer between 1 and 5).

1.4.2 Research Design and Identification Strategy

At any given point of time, one specific rating system design (either single dimensional or multi-dimensional) is generally implemented across an entire website to maintain consistency. Therefore, randomizing system designs is practically impossible (e.g., implement a single dimensional system for some products while multi-dimensional system for others). Thus, the variation of rating system designs need to come from one of the two sources: cross two different websites, or within a website (before/after). In this paper, we explore both cross-websites and within-website variations by matching data from two websites (Yelp and TripAdvisor) and leveraging a quasi-natural experiment on TripAdvisor, respectively. Our key econometric identification strategy hinges on the system change that occurred on TripAdvisor with regard the rating system design, which is exogenous to consumers. Specifically, TripAdvisor re-engineered and implemented its rating system from single-dimensional to multi-dimensional in January 2009, which provides us a natural experiment setting to test effect of the rating system change with a difference-in-difference specification. Furthermore, we track the same set of restaurants on Yelp as the "control group" to control for any unobserved restaurant quality change (with restaurant level fixed effect). In other words, the rating trend on Yelp for each of these restaurants serves as a proxy for any change in restaurant quality. Therefore, such a research design controls for any factor related to restaurants (e.g., change of chef or menu).

To claim that the change of rating system causes the differences between the ratings of these two websites, we must eliminate multiple alternative explanations. One possible explanation is that ratings are different to start with, because these two websites may attract different crowds of users, and TripAdvisor users may have the tendency to be more positive than Yelp users. We can test this explanation by investigating if rating difference persists even before the system change. We conduct a rigorous test to determine if a systematic difference existed between the two websites before the system change of TripAdvisor.

1.4.3 Empirical Models

In this section, we introduce the empirical models that we use for parameter estimation, hypotheses testing and the elimination of the alternative explanations.

First, we analyze the effect of the total number of reviews by estimating models with the following specification:

$$Rating_{it} = \beta_0 + \beta_1 * logn_r + \alpha_i + \epsilon_{it} \tag{1.1}$$

where i indexes the restaurants, and t indexes the time when the rating is made. The dependent variable, $Rating_{it}$, is the consumer rating submitted for restaurant i at time t. $logn_r$ is the log transformation of the number of overall ratings. Therefore, β_1 captures the trend of the rating as a function of number of reviews. A negative coefficient indicates a downward trend. α_i denotes the fixed effect of the restaurant.

As noted earlier, we must control for website difference to estimate the effects of the multi-dimensional rating system. In particular, we must examine (1) if the two websites are comparable to begin with and (2) if Yelp has experienced any significant changes in ratings, which have nothing to do with the system change of TripAdvisor. The first test investigates whether or not the ratings of Yelp and TripAdvisor before the system change of TripAdvisor have any significant differences (Equation 1.2). The second test examines whether or not the ratings of Yelp before and after the system change of TripAdvisor are comparable. In other words, we test whether or not Yelp had significant differences before and after the system change (Equation 1.3).

$$Rating_{it|k} = \beta_0 + \beta_1 * Treat + \beta_2 * logn_r + \alpha_i + \epsilon_{it} \tag{1.2}$$

$$Rating_{it|Yelp} = \beta_0 + \beta_1 * Time + \beta_2 * logn_r + \alpha_i + \epsilon_{ikt} \tag{1.3}$$

where k indexes the website. The dependent variable, $Rating_{it|k}$, is the consumer rating submitted for restaurant i at time t on website k. $Time$ is a dummy that equals one if the time period is after the change of the rating system, and zero otherwise. Here, the coefficient β_1 in Equation 1.2 captures the average difference in ratings between Yelp and TripAdvisor before the system change. This allows us to check if ratings difference is due to systematic difference between the two websites without the "shock" of system change. $Treat$ is a dummy that equals one if the ratings are made on TripAdvisor, and zero if on Yelp. ; the β_1 in Equation 1.3 captures the average difference in ratings of Yelp before and after the system change.

Once we have evidence that the two websites are comparable to start with and that Yelp itself has not experience any significant rating changes, we can utilize the DID approach to estimate the effect of the rating system change from single- to multi-dimensional on the overall ratings. Recall that we choose the exact same restaurants on Yelp as "control." Therefore, the rating trend for each of these restaurants on Yelp serves as the proxy for any change in restaurant quality. Therefore, the additional rating changes on TripAdvisor are caused by the change of the rating system after controlling for the rating trend at Yelp. Figure 4 demonstrates the relationship.

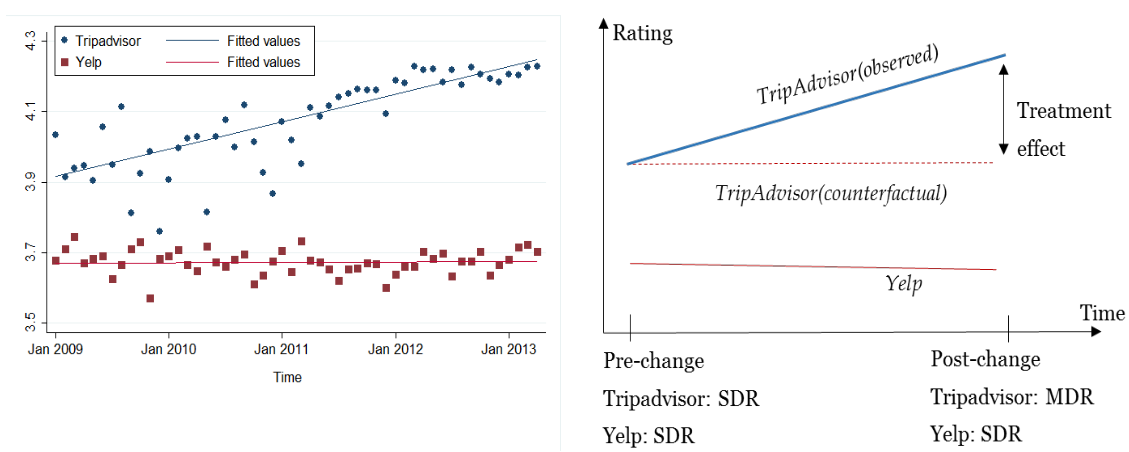

Figure 4. DID Analysis

We summarize the DID approach as follows:

$$Rating_{it|k} = \beta_0 + \beta_1 * Time + \beta_2 * Time * Treat + \beta_3 * Treat + \beta_4 * logn_r + \alpha_i + \epsilon_{ikt} \quad (1.4)$$

As an additional robustness test, we further look at how the effects change as more multi-dimensional ratings are accumulated. If multi-dimensional ratings indeed enable information transfer, then we should observe stronger effects as more dimensional ratings are aggregated. As previously mentioned, consumers are not forced to provide dimensional ratings on TripAdvisor. More multi-dimensional reviews provide stronger confirmation of quality information and further reduce the uncertainty related to consumption experience, which in turn makes information transfer easier.

We now extend the model (i.e., in Equation 1.4) to test whether or not the effect is stronger for those restaurants with more dimensional ratings. Let n_d denote the number of existing dimensional ratings on TripAdvisor at time t, and $logn_d$ the log transformation of n_d. We use the following difference in difference in difference (DDD) formulation:

$$Rating_{it|k} = \beta_1 + \beta_2 * Treat + \beta_3 * Time * Treat * logn_d + \beta_4 * logn_r + \alpha_i + \epsilon_{ikt} \quad (1.5)$$

β_3 measures the effects of each additional multi-dimensional rating on the overall rating.

To test hypothesis H3, we present the deviation model to determine if ratings indeed converge over time in multi-dimensional rating systems. We relate the deviation of ratings to the nominal sequence value of the rating at time t. The rating deviation of restaurant i at time t is measured as the absolute difference between the rating of a consumer at time t and a previously observed (overall) rating. We compute previously observed ratings as the average of all the ratings made before time t. We include restaurant fixed effects α_i as controls in the analysis to control for any systematic differences due to restaurants. β_1 measures the relationship between the deviation of the current rating from previously observed ratings and sequences. A positive β_1 means that a deviation from previously observed ratings increases with the rating sequence, whereas a negative β_1 means that a deviation (from previously observed ratings) decreases with the rating sequence, which indicates the convergence of ratings.

$$Deviation_{it} = |r_{it} - \mu_{it-1}| \quad (1.6)$$

$$Deviation_{it|Yelp} = \beta_0 + \beta_1 * Sequence + \alpha_i + \epsilon_{it} \quad (1.7)$$

$$Deviation_{it|TripAdvisor} = \beta_0 + \beta_1 * Sequence + \alpha_i + \epsilon_{it} \quad (1.8)$$

We also directly compare the deviation effect in the single-dimensional system and that in the multi-dimensional system using Equation 1.9, where β_3 captures the difference of the rating deviation with the rating sequence between Yelp and TripAdvisor.

$$Deviation_{it|k} = \beta_0 + \beta_1 * Sequence + \beta_2 * Treat + \beta_3 * Sequence * Treat + \alpha_i + \epsilon_{it} \quad (1.9)$$

1.5 Results

Table 1

Downward Trend in Single Dimensional Rating Systems

Sample:	Before System Change		After System Change		All Data
	Yelp	*TripAdvisor*	*Yelp*	*TripAdvisor*	*Yelp*
	Model 1	Model 2	Model 3	Model 4	Model 5
$\log n_r$	−0.115***	−0.058**	−0.022**	0.084***	−0.042***
	(0.012)	(0.018)	(0.007)	(0.010)	(0.005)
Constant	4.127***	4.070***	3.774***	3.795***	3.876***
	(0.035)	(0.049)	(0.031)	(0.043)	(0.024)
Restaurant FE	Yes	Yes	Yes	Yes	Yes
Observations	18,388	8,917	166,670	68,024	185,058

Table 1 presents the regression results of the effect of the total number of ratings on ratings (Equation 1.1). Models 1 and 2 use the data of Yelp and TripAdvisor respectively before the change in the rating system when they were all still using a single-dimensional rating system. The parameter of interest (i.e., total number of ratings) has the expected signs, which suggests that a single-dimensional rating system follows a downward trend. Models 3 and 4 use the data of Yelp and TripAdvisor after the rating system change, when Yelp was still using a single-dimensional rating system, and TripAdvisor adopted a multi-dimensional rating system. The result of Yelp is consistent with that of Models 1 and 2, offering additional confirmation that single-dimensional ratings are downward trending. Most interestingly, the coefficient in Model 4 is positive, which suggests that the ratings in multi-dimensional rating system are upward trending instead of downward trending. In Model 5, we use the entire data of Yelp to check the stability, which shows that the estimates are robust. Overall, we provide very strong support of H1. Particularly, we offer an additional explanation as to why single dimensional rating systems likely exhibit a downward trend based on information transfer efficiency. In particular, when information is not transferred efficiently from prior consumers to future consumers, ratings would follow a downward trend.

Table 2

Websites Difference

Variables	Model 1	Model 2
$\log n_r$	−0.099*** (0.024)	−0.078*** (0.022)
Treat (Comparison between *Yelp* and *TripAdvisor* before the system change)	−0.006 (0.031)	
Time (Comparison of *Yelp* before and after the system change)		0.036 (0.028)
Constant	4.114*** (0.088)	3.985*** (0.077)
Restaurant FE	Yes	Yes
Observations	16,194	39,912

Table 2 shows the results of Equations 1.2 and 1.3, which aim to investigate whether or not Yelp and TripAdvisor have any systematic differences. Equation 1.2 uses all the ratings before the system change on the two websites, and the results are shown in Model 1. The negative coefficient of logn_r shows that the ratings on both websites follow a decreasing trend, which is consistent with the results in Table 2. The insignificant coefficient of $Time$ shows that Yelp and TripAdvisor have no systematic difference before the system change after controlling for the downward trend of the ratings. Equation 1.3 uses the data on the ratings of Yelp before and after the system change, and the results are shown in the second column. A downward trend is still observable. Moreover, the insignificant coefficient of $Treat$ shows that Yelp had no significant difference before and after the rating system change. The results show that Yelp and TripAdvisor are comparable prior to TripAdvisor's system change, and that Yelp itself also exhibits consistent rating trends before and after TripAdvisor's system change.

Table 3

DID Analysis (DV = Rating)

Variables	Model 1 (DID)	Model 2 (DDD)
Time	−0.016 (0.026)	
Treat	0.069*** (0.028)	0.084*** (0.026)
Time*Treat	0.154*** (0.030)	
Time*Treat*logn		0.055*** (0.011)
logn	−0.047*** (0.011)	−0.048*** (0.011)
Constant	3.901*** (0.046)	3.879*** (0.056)
Restaurant FE	Yes	Yes
Observations	50,153	50,153

Table 3 presents the estimation results of a DID analysis for Equation 1.4. Model 1 presents the results for the regression, including restaurant fixed effects. The significant positive coefficient of $Time*Treat$ indicates that the change of the rating system from single-dimensional to multi-dimensional significantly increased ratings by 0.154. In other words, the restaurant ratings increased by 0.154 on the average as a result of the implementation of the multi-dimensional rating system. The increase in the ratings suggests that consumers are "happier" in the sense that they are able to form a rational expectation based on the information gathered from multi-dimensional rating systems, which match their preference well. Therefore, H2 is supported.

The results of Equation 1.5 are reported in Model 2 of Table 3. In this estimation we use a difference-in-difference-in-difference (DDD) approach. As expected, the significant positive coefficient of $\mathbf{Time} * \mathbf{Treat} * \mathbf{logn_d}$ supports the assumption that the effect of the system change is dependent on the number of multi-dimensional ratings generated by the consumers. In contrast to the downward trend in single-dimensional rating systems, this finding provides additional evidence that an upward trend is likely observable in multi-dimensional rating systems as more multi-dimensional ratings are accumulated. Notably, previous literature has shown that the downward trending of ratings is widely observed (Li and Hitt, 2008) because of self-selection, in which consumers who are enthusiastic about a product rate earlier. Such high ratings "trick" later consumers to try the product, and these customers are likely to be more disappointed than the enthusiastic consumers. Therefore, ratings exhibit a downward trend over time. Our findings suggest that multi-dimensional rating systems can alleviate this effect because multi-dimensional ratings give consumers more reasonable expectations of how much they will like the product. Therefore, later consumers are less likely to be "tricked." The DDD analysis further strengthens the conclusion that multi-dimensional ratings enhance information transfer efficiency.

The results for the deviation are presented in Table 4. Models 1 and 2 separately use the data of Yelp and TripAdvisor. Model 3 uses the combined data of Yelp and TripAdvisor. We also add restaurant fixed effects that control for any unobserved restaurant effects. The coefficients of sequence are significantly positive for Yelp and negative for TripAdvisor. The results suggest that the absolute difference between the previous average rating and the next rating increases in single-dimensional rating systems with an increase in the number of ratings but decreases in multi-dimensional rating systems. Model 3 shows that the difference between single-dimensional rating systems and multi-dimensional rating systems is significant. In other words, the deviation from previous ratings is smaller for multi-dimensional ratings than for single-dimensional ratings. The results also suggest that ratings converge in multi-dimensional systems. H3 is supported.

Table 4

Estimation of Rating Deviation (DV=rating deviation)

	Yelp	*TripAdvisor*	*Yelp* and *TripAdvisor*
	Model 1	Model 2	Model 3
Sequence	0.0001*** (0.0000)	−0.0003*** (0.0000)	0.0001*** (0.000)
Treat			−0.0250*** (0.007)
Sequence*Treat			−0.0002*** (0.000)
Constant	0.816*** (0.005)	0.831*** (0.007)	0.843*** (0.004)
Restaurant FE	Yes	Yes	Yes
Observations	166,670	69,024	235,694

1.6 Discussion

This study extends the limited understanding of information transfer efficiency of different online rating system designs. Based on unique data from two leading online review platforms, the results of this study first show that ratings trend down in a single-dimensional rating system but not in a multi-dimensional rating system. Although this phenomenon has been observed in many prior studies (Li and Hitt 2008, Godes and Silva 2012), we provide an additional theory based on information transfer efficiency to explicate the reasons behind this phenomenon. We then show that consumers are generally more satisfied when consuming information in a multi-dimensional rating system. Therefore, high and convergent ratings are reported after using a multi-dimensional rating system because they are better able to form realistic expectations and make more informed decisions using information derived from multi-dimensional ratings. This finding provides evidence that multi-dimensional rating systems enable a more efficient information transfer.

We eliminated many possible alternative explanations. First, we ensure that TripAdvisor and Yelp are comparable. In other words, these two websites have no significant difference in terms of average ratings before the system change. Second, we exclude the possibility of unobserved change in Yelp after the system change. In other words, the ratings on Yelp before and after the system change have no significant difference when the number of ratings is controlled for. In addition, some may argue that TripAdvisor may have introduced other strategies unrelated to rating systems that may lead to a difference in ratings. For example, TripAdvisor allows owners to respond to comments. We do not observe owner comments in our data set. We also eliminate the possibility that the higher ratings in multi-dimensional rating systems are due to self-selection (i.e., consumers that provide multi-dimensional ratings may be more enthusiastic consumers).

In summary, this study provide significant and robust findings that suggest that switching from single-dimensional rating systems to multi-dimensional rating systems provides benefits, especially for experience goods which product attributes are difficult to observe before consumption. Information is effectively transferred to consumers, and consumers form more rational expectations after adopting multi-dimensional rating systems.

Although previous research has investigated the effects of different product attributes on pricing power, hotel ranking, and review helpfulness (Archak et al. 2011, Decker and Trusov 2010, Ghose et al. 2009, Ghose and Ipeirotis 2011, Ghose et al. 2012), as well as the effects of crowd and friends on consumer reviews (Wang et al. 2015, Lee et al. 2015), this study is to our knowledge the first to directly compare single-dimensional and multi-dimensional rating systems. This study addresses whether or not multi-dimensional ratings facilitate information transfer efficiency and whether or not multi-dimensional ratings lead to more informed purchase decisions and more satisfied consumers. We extend the limited understanding of the importance of the different designs of online rating systems 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), which is further integrated to consumer expectation and satisfaction based on ECT (Anderson and Sullivan, 1993) and the perspective of information transfer (Belkin 1984). We revisit prior work on the dynamic effects of ratings where the downward trend of single-dimensional 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 trending in single dimensional rating system based on information transfer theory, we also show that such biases can be reduced using multi-dimensional rating systems. Our results show an upward and convergent trend of multi-dimensional ratings, which indicate that, after adopting multi-dimensional rating systems, consumer preferences are better matched with the attributes of the restaurants because information from multi-dimensional rating systems is efficiently transferred to them. This study also extends the extant research on how IT-enabled technologies can reduce different sources of consumer product uncertainty (Dimoka et al. 2012, Kwark et al. 2014, Hong and Pavlou 2014).

This study also has two important managerial implications. First, the results from this study inform practitioners about whether or not adopting multi-dimensional rating systems can improve the performance of online product reviews and also provide insights on the effective design of informative rating systems. For products or services with higher inherent quality and fit uncertainty, we suggest that multi-dimensional rating systems be adopted, especially for experience products with attributes that provide idiosyncratic utilities to consumers (Nelson 1981). Our results also suggest that the effect of multi-dimensional rating system design change depends on the number of multi-dimensional ratings accumulated. Therefore, review websites should incentivize consumers to provide dimension ratings.

As with most empirical studies, this study is not free of limitations. First, while we have provided evidence supporting the value of adopting multi-dimensional rating systems, we should acknowledge that the results may not be generalized to other types of products, particularly search goods with product attributes that are easily observable by consumers. A potential interesting future study is to look at whether or not different performance effects are observed for different types of products when multi-dimensional rating systems are introduced, such as search, experience, and credence goods.

Second, in the present study, we focus on information transfer efficiency, which cannot be identified by estimating the effect of multi-dimensional ratings on sales because it is difficult to distinguish a happy purchase and a regretful purchase. For example, higher sales with most consumers who feel “dissatisfied” after purchases is actually a signal of low information transfer efficiency because information does not help consumers to match their preferences with their ideal products. The post-purchase satisfaction of consumers is the correct metric to understand information transfer efficiency. Given that we’ve provided significant evidence supporting that multi-dimensional ratings enable more efficient information transfer, a natural extension of our work is the examination of relationships between multi-dimensional ratings and sales or other performance data. Previous research has tried to link single-dimensional ratings to firm revenue and the stock market (Godes and Mayzlin 2004, Chevalier and Mayzlin 2006, Liu 2006, Dellarocas et al. 2007, Duan et al. 2008, Forman et al. 2008). Similar empirical analyses can be performed to examine the effect of multi-dimensional rating systems.

Third, we do not consider the possibility of fake reviews because we are unable to track down fake reviews. However, fake reviews do not pose a serious concern for this study because, first, the long-term effect of fake reviews are likely to be negligible (Dellarocas, 2006), and second, these review websites spend a huge amount in warding off fake reviews, including legislatures (CNET 2013). Moreover, it is possible that additional information can be obtained from text reviews (Pavlou and Dimoka 2006) or different product attributes be extracted from a review of the texts (Ghose and Ipeirotis 2011, Archak et al. 2011, Ghose et al. 2012), which we do not control in this study. However, it involves extra time and effort to read texts and obtain useful information. To the contrary, consumers can very quickly obtain a basic understanding of the different dimensions of the restaurant through the average dimension ratings displayed on the restaurant's home page. Obtaining information on single-dimensional rating systems by reading text reviews may take several times longer. Future research is warranted in examining the interplay between multi-dimensional ratings and text reviews.

CHAPTER 2

THE EFFECT OF RATING SYSTEM DESIGN ON OPINION SHARING

2.1 Introduction

Online review platforms allow consumers to share their opinions about products. Ubiquitous and accessible online reviews provide a wealth of information about goods and services to consumers in their search, evaluation, and choice of products. Literature has endorsed that both numerical ratings (Godes and Mayzlin 2004, Chintagunta et al. 2010, Rosario et al. 2016) and textual reviews (Archak et al. 2011, Ghose et al. 2012) have an impact on consumers' decision making, but they also play different roles. Hu et al 2014 suggests that consumers may use ratings to reduce the decision sets and use textual reviews to do further evaluation to arrive at a decision. Because numerical ratings require less cognitive effort and consumers resort to simplifying strategies and heuristics to arrive at a decision due to cognitive limitation. However, one potential issue of numerical ratings is they may not be representative of the information embedded in textual reviews. Ratings cannot comprehensively reflect information on different product attributes (Archak et al 2011), primarily because a product usually comprises of multiple attributes and consumers, who are heterogeneous (Li and Hitt 2008, Godes and Silva 2012), may form different levels of preferences towards different product attributes. For example, a consumer may prefer high image quality than other attributes when he evaluates a camera. It is possible that he may provide a 5-star rating when the camera performs well in image quality, ignoring other attributes. On the other hand, he may elaborate his opinions on image quality in textual reviews or share his experience on all aspects of the camera. Although product attributes could be identified from textual reviews (Hu and Liu 2004, Archak et al 2011), there is limited understanding of the distribution of number of attributes covered in each review. It is also not clear how consumers choose to provide ratings and reviews, and the relationship between ratings and reviews: consumers may have taken into account all dimensions when giving a rating, or the rating may reflect only a particular dimension which matters the most to a consumer. Similarly, when providing textual reviews, consumers may focus on the dimension that drives the rating, or they may choose to provide additional information not necessarily reflected in their ratings; they may focus on only positive attributes or only negative attributes, or they may provide comprehensive reviews covering all dimensions. Given consumers use ratings and textual reviews differently and potentially at different stages in their decision making, ideally, it will be great to have comprehensive ratings and comprehensive reviews. The goal of this paper is to take a deeper look at how consumers choose to provide ratings and reviews, and how ratings and reviews are related, and most importantly, if multi-dimensional rating system (MD system) may help achieve the goal of having more comprehensive ratings and more comprehensive reviews. MD system allows a user to rate different dimensions/attributes of their product experiences. A multi-dimensional rating system is found to be more informative to users, reducing user uncertainty and leading to higher consumer satisfaction (Liu et al. 2014). Yet, it is not clear how the introduction of a multi-dimensional rating system affects the content of reviews. On one hand, reviewers may not find the need to write comprehensive and long reviews because they may think they already adequately expressed their opinions through the multi-dimensional ratings (substitution effect). On the other hand, reviewers may attempt to justify their ratings on different dimensions (justification effect), leading to a review that is comprehensive and covering all dimensions. Taken together, the reviews in a multi-dimensional rating system may become either longer or shorter in length, and either broader (cover more dimensions) or narrower (cover fewer dimensions) in terms of number of topics. In addition, the introduction of MD system may also affect linguistic features on each product dimension. Reviews could either be deeper (longer reviews) or more superficial (shorter reviews) on each product dimension. Consumers could also focus more on positive aspects or more on negative aspects. Given MD primes consumers of different aspects of their consumption experiences, it is also likely that MD reviews become more objective (or neutral). Bearing the above in mind, in this study, we are interested in answering the following questions:

RQ1: How ratings and reviews reflect consumers' heterogeneous preference?

RQ2: Do ratings complement or substitute textual reviews?

RQ3: How does rating system moderate the interplay?

We collected data on the same set of restaurants from Yelp and TripAdvisor and adopted the DID method to control for restaurant quality change. Our results suggest that MD ratings do not substitute text reviews. To the contrary, consumers tend to share more information in textual reviews in a more objective way using the MD system. MD reviews have greater breadth (more dimensions) and depth (longer). An experimental study is corroborated with the observational study to further uncover the mechanism. MD system primes consumers to generate a more comprehensive numerical overall rating of all dimensions. Our study makes a pioneering effort in establishing the value of rating system design on opinion sharing.

2.2 Theory and Hypotheses Development

While a single numerical rating may not reflect consumers' overall experiences across multiple dimensions in an SD system, consumers could provide more information in textual reviews. Previous research has found that textual reviews contain information of different dimensions of product attributes using text mining approaches (e.g., natural language processing). Decker and Trusov (2010) considered rating heterogeneity and estimated the relative effect of product attributes and brand names on the overall evaluation of products. Ghose et al. (2012) estimated consumer demand and various product attributes using hotel reservation data and consumer-generated reviews and proposed a new ranking system that reflects the multidimensional preferences of consumers for products. Ghose et al. (2009) demonstrated that different dimensions indeed differentially affect the pricing power of sellers. However, these product attributes are extracted from the whole corpus of textual reviews. And it is still not clear how much information referred to product attributes is covered for each piece of textual review. Textual reviews do provide more details of product information, but it is possible that each piece of textual review only expand what consumers want to express in the numerical rating.

The introduction of MD system may lead to a change in the content generation in textual reviews in a few different ways. First, a substitution effect may exist. Given consumers' opinions have already been incorporated into numerical ratings on different dimensions, it is likely that consumers do not find the need to write a long review and elaborate on different dimensions. . In an SD system, since consumers can't express their opinions in one numerical rating, they may try to provide more details in their textual reviews, to make up for the deficiency in the single rating. For example, a consumer may feel bad about the service but good on other dimensions when he goes to a restaurant, and he may rate a 3-star in SD system to release bad emotions and explain in the textual review why he rates a 3-star and how he hates the service. And in MD system, he may just rate 3 on service and 5 on other dimensions and feel no need to explain in the textual review. Hence, we propose that:

H1a: Textual reviews substitute numerical ratings in MD system. Consumers tend to write shorter textual reviews in MD system than in SD system.

Alternative to the substitution effect proposed above, a justification effect could exist. MD system may lead to more content generation of textual reviews through the priming mechanism, since consumers are primed with "multiple dimensions" in an MD system. In this case, we would expect consumers to write reviews which cover more dimensions in MD system compared to reviews in SD system. Much as dimensional ratings contain more information compared to a single rating in SD system, they also leave more information to be explained. Because consumers now provide both overall rating and dimensional ratings in MD system, it is possible that their overall ratings are not consistent with dimensional ratings. For example, in SD system, consumers only need to explain why they provide a 3-star overall rating. However, in MD system, consumers may attempt to explain why a 3-star on one attribute and a 4-star on the other attribute. According to the attribution theory, people tend to attach meaning to their behavior. In another word, they may tend to explain every dimensional rating they provide. Cognitive dissonance theory (Festinger 1957) suggests that people tend to seek consistency among their cognitions. When there is an inconsistency between attitudes or behaviors, something must change to eliminate the dissonance. When the inconsistency happens, consumers may feel like he needs to achieve consistency by rationalization and excuses, and he would explain in the textual reviews why he gives these dimensional ratings and the overall rating. And again, we would expect textual reviews in MD systems to be longer and cover more dimensions. Further, the number of product attributes listed itself could affect consumers' behavior. Sela and Berger (2012) argue that attribute numerosity is a heuristic cue for usefulness (Thompson et al. 2005), and according to the principle of multi-attribute diminishing sensitivity (Nowlis and Simonson 1996), increasing perceived usefulness through attribute numerosity should benefit more on hedonic than utilitarian options. That is, when choosing from different options, the number of attributes listed could imply more useful, and it benefits more on hedonic options. Hedonic options may be perceived more useful with more attributes listed. In our study, it is possible that consumers try to provide perceived useful information in textual reviews in SD system. And in MD system, the existence of dimensional ratings itself increases perceived usefulness which may lead to more information shared on attributes that are not "useful" in textual reviews. In this case, we would again expect more dimensions are covered in textual reviews in MD system than in SD system. Thus, we propose that:

H1b: Textual reviews complement numerical ratings in MD system. Consumers tend to write longer textual reviews in MD system than in SD system.

H2: On average, textual reviews in MD system are in greater breadth and depth than textual reviews in SD system.

While it is possible that consumers' preferences for product attributes be reflected in both numerical ratings and textual reviews in a single-dimensional rating system (SD system), several theories and prior findings suggest that numerical ratings may not fully reflect consumer experience on all product dimensions because consumers tend to place more weight on certain product attributes/dimensions toward which consumers have extreme feelings, either positive or negative. First, consumers are motivated to share positive or negative WOM for impressions. Previous research (Chung and Darke 2006; Hennig-Thurau et al. 2004; Sundaram et al. 1998) find that people are more likely to share positive things because they want to be perceived as being positive. At the same time, people are also motivated to share negative things to show discriminating tastes because reviewers were seen as more intelligent, competent, and expert when they wrote negative as opposed to positive reviews (Amabile 1983). Second, consumers are motivated to share positive or negative WOM for emotion control. According to the Balance Theory (Heider 1946, 1958, Newcomb 1953), people have a basic desire for balance in their lives (Zajonc

1971). Thus, when experiencing a strong unbalance from either a strong positive or negative consumption experience, consumers may attempt to restore the equilibrium by expressing related positive emotions and negative feelings in reviews. This motive is referred as Homeostase Utility (Hennig-Thurau et al. 2004). For example, angry consumers (Wetzer et al. 2007) or dissatisfied customers (Anderson 1998) are more likely to share negative word of mouth to vent or to punish the company. Taken together, a consumer's overall satisfaction is likely to skew towards dimensions with the extreme sentiment, leading to ratings that are not comprehensive and are biased.

On the other hand, in an MD system, consumers may not only rate the restaurant overall but they also have an option to rate on different dimensions. MD system may cause consumers to report a more comprehensive overall rating. Compared to SD system, consumers are still motivated to share extreme feelings, however, their motivations of self-impression and emotion regulation could now be captured in dimensional ratings instead of the overall rating. And for the overall rating, the MD system may exert a priming effect (Neely 1977, Tipper 1985, Tulving and Schacter 1990). Consumers are primed with "multiple dimensions", which could remind consumers to take into account different dimensions, either positive or negative, and report a rating more representative of overall consumption experience. Therefore, we propose that:

H3: The overall ratings in MD system tend to reflect more dimensions of consumers' consumption experience compared to the overall ratings in SD system.

2.3 Data

We address our research questions by studying restaurant ratings and reviews in different rating systems. We choose restaurants as our context because restaurants have well-known different dimensions of services (e.g., food and location) and attract significant attention in academic literature. We gathered data from two leading consumer review websites: Yelp.com (Yelp) and TripAdvisor.com (TripAdvisor). Like most review websites, Yelp provides a single-dimensional rating system on a scale of five stars. TripAdvisor, on the other hand, provides a multi-dimensional rating system, which allows not only overall ratings but also ratings for the dimensional characteristics of restaurants, such as food, service, and ambiance, using the same five-star rating scale. Figure 5 shows ratings of an identical restaurant, The Eddy in New York, on these two websites.

We used two customized web crawlers and collected data from these two websites. We obtained data for the identical restaurants of two review sites to eliminate restaurant differences and control for unobserved quality changes in the restaurants. Therefore, the differences between the ratings in the two review systems for the identical restaurants cannot be attributed to the unobserved restaurant effect. We specifically match the restaurants according to restaurant names, addresses, and phone numbers in New York City. Finally, we obtained a sample of 698 restaurants. For each restaurant, we extracted the overall rating, dimensional ratings and reviews.

Figure 5. Ratings of an Identical Restaurant on the Two Rating Websites

For each piece of textual review, we measured word count (WC), positive affect (PA) and negative affect (NA). We follow Golder and Macy (2011)'s approach to measure PA and NA by the proportion of positive emotion words, and the proportion of negative emotion words respectively. A higher value of PA denotes a higher portion of positive emotion words are used in the review. Besides, we move one step further to dig deeply what consumers are writing in text reviews.

We use a sophisticated machine learning method denoted as AIRS (Li et al. 2015) to automatically discover and measure the topic and sentiment for each review. The intuition of this method is a topic is a cluster of frequently co-occurred words and the sentiment for a topic is reflected by a mixture of negative and positive terms about the topic. Different topic mining methods such as LDA (Blei et al. 2003) have been applied to many business contexts, such as the analysis of blog content (Singh et al. 2014), the measure of business proximity (Shi et al. 2016) and the impact of keyword ambiguity on search advertising (Gong et al. 2017). Compared with these methods, the advantage of AIRS model is that it could generate not only the dimension (i.e., topic) probability but also the sentiment score for each dimension for each review. Empirical studies with online review data have demonstrated that the AIRS method could generate robust results (Li et al. 2015). While the dimension probability reflects the presence of dimension in each review, the sentiment score for a dimension reflects the user's sentiment (i.e., positive, negative, neutral) on the dimension expressed in his textual review. These twofold results allow us to study how the review system design (MD versus SD review system) impact the review breath/depth and dimensional sentiment.

The AIRS model (Li et al. 2015) takes not only review text but also overall rating as input to infer the probability and sentiment for each dimension of each review. As there are four predefined dimensions of restaurant in our review data, we set the number of dimensions (i.e., topics) as four in this study. Consequently, we obtain four dimension probability and four dimensional sentiment score for each review as the output of AIRS model. The sentiment score will be scaled to the range of (1, 5), which is the same as that of overall rating in our restaurant review data. For instance, we may get (0.1, 0.2, 0.4, 0.3) as dimension probability and (2.2, 3.1, 4.0, 4.8) as dimensional sentiment score for one review. Furthermore, to improve the robustness of AIRS model, we select a set of seed words for each of four restaurant dimensions (e.g., meat, soup, salad for food dimension) and use them as a prior topic words to guide the machine learning process. With this assistant, we are able to assign the mined dimension probability and sentiment score with each of four predefined restaurant dimensions. For each piece of review, we take the log transformation of the product of word count and dimension probability loading to estimate the depth of each dimension. For example, if a review has 100 words, and the loadings of four dimensions are 0.25 respectively, then the depth of each dimension is log25. A higher number denotes higher depth that is more words are used to express opinions on this specific dimension. We use this measure instead of loading to control for word count. And then we compute the breadth of each review, basically, we use the number of dimensions mentioned in each review. Our method extracts four topics as well as their probabilities for each review. However, the probability loadings of some dimensions could be extremely low, we try to tease out these dimensions when measuring breadth of each review. We only count dimensions whose Z-scores of probability loadings are greater than -2, in another word, we don’t consider those dimensions outside two standard deviations from the mean loading of this dimension across all reviews. To sum it up, a larger number suggests more dimensions are covered and a higher breadth of a review.

2.4 Research Setting and Methodology

Yelp adopts a single-dimension rating system, while TripAdvisor changed its rating system from single-dimension to multi-dimension in January 2009. To identify whether there is any effect of multi-dimensional rating system on emotion sharing, we compare variables of interest of TripAdvisor before and after the system change. However, there might be other reasons causing the change in variables of interest. For example, the quality of the restaurant might increase or decrease. In this case, we can't tell which factor causes the change. Here we take the difference in difference (DID) approach. We choose the exact same restaurants on Yelp as 'control group', therefore any trend on Yelp for each of these restaurants will serve as a proxy of change in restaurant quality. Besides, emotion sharing change at TripAdvisor, after controlling for the trend at Yelp, will be due to the change of the rating system.

We summarize this difference in difference approach below:

$$Y_{ij|k} = \beta_0 + \beta_1 * Time + \beta_2 * Time * Treat + \beta_3 * Treat + \beta_4 * \boldsymbol{X_{ijk}} + \alpha_i + \epsilon_{ijk} \quad (2.1)$$

where i indexes the restaurant, j indexes the position of the review in the review sequence for each restaurant and k denotes the website. Dependent variables are a list of variables including the overall rating, word count, positive and negative emotions, and breadth and depth from text mining. Treat is a dummy that equals one if the ratings are made on TripAdvisor, and zero if on Yelp. Time is a dummy that equals one if ratings are made after the system change, and zero if before the system change. The coefficient of the interaction term measures the difference caused by the change of the rating system, after controlling for changes in restaurant quality over time and systematic website differences. X_ijk is a vector of control variables. For example, we control ratings for word count and emotions. And we also control word count for emotions as emotions here are calculated as a portion of emotional words out of all words.

2.5 Results

2.5.1 Substitute vs. Complement

Table 5 shows the results of the DID analysis. The dependent variable of the first column is the overall rating. And then in the following columns, dependent variables are word count, positive affect, and negative affect of each piece of textual review separately. The significant positive coefficient of the interaction term indicates that the change of the rating system from SD system and MD system significantly increase ratings by 0.409. The result is consistent with Liu et al (2014) where they find that the overall rating may increase due to increased information transfer efficiency when adopting the MD system. And the results from the second column show that word count increases by almost 60 which indicates that consumers tend to write more when they use MD system. Our conjecture is that consumers tend to explain more on why they provide these dimensional ratings. They try to make their ratings more reasonable and credible. H1b is supported. Textual reviews complement numerical ratings.

Table 5

DID Analysis

	(1)	(2)
	Rating	WC
Time	-0.155***	-10.80***
	(0.0147)	(1.316)
Treat	0.0837**	-99.41***
	(0.0266)	(2.526)
Time*treat	0.409***	58.48***
	(0.0279)	(2.355)
Rating		-11.25***
		(0.290)
N	143885	143885
Restaurant FE	Yes	Yes

We also investigates whether the use of MD system reduces consumers' likelihood to read text reviews. We conduct an experiment in which respondents were traced if they clicked to read reviews of the restaurant after the rating information was provided to them (See Appendix A.1). Subjects were first primed about a scenario that they will go for lunch near campus. Subjects were then shown four restaurants, two of which came with SD ratings (Restaurant 1 and 2), while the other two with MD ratings (Restaurant 3 and 4). Restaurant 1 and 2 were provided with only an overall rating while Restaurant 3 and 4 were presented with MD ratings. The overall ratings of Restaurant 2 and 4 are higher than those of Restaurant 1 and 3. Besides, for each restaurant, respondents were asked whether they want more information about the restaurant. If respondents answered yes, information concerning price level, restaurant description would be provided, and a further question of whether they want to read more text reviews would be asked. And the text reviews were shown in random order if respondents chose to read the text reviews. The display of information is similar to what one would see on the website. The respondents were then asked if they would choose to have lunch at each of the four restaurants. After the choice had been made, respondents were asked to answer a list of questions related to demographics, etc. The results show that on average, less than 30% of participants chose to read text reviews. The results also show that consumers' decisions to read reviews are not affected by the rating system. That is, the likelihood to read reviews is comparable in SD and MD systems. Results combined from the DID analysis and the experiment show that consumers would read textual reviews in the MD system and they would write more textual reviews. These results provide support that MD ratings do not substitute text reviews.

2.5.2 Breadth and Depth

Table 6 shows the results of how breadth and depth change after the adoption of MD system. The dependent variable of the columns (1) to (4) are the depth of four dimensions. A larger coefficient of the interaction term suggests consumers write more about this dimension. The significant positive coefficients of the interaction terms indicate that consumers tend to write more about all dimensions in MD system. MD system leads to greater depth of each dimension in text reviews. Consumers are not expanding their opinions on one or two specific dimensions, instead, they try to talk deeper in each dimension. The dependent variable of column (5) is the breadth of each review which is the number of dimensions covered in each review. The positive and significant coefficient of the interaction term suggests that on average MD reviews cover more dimensions. Results from Table 6 suggest that MD reviews have greater breadth and depth than SD reviews. H2 is supported.

Table 6

DID Analysis of Breadth and Depth

	(1)	(2)	(3)	(4)	(5)
	Depth_food	Depth_service	Depth_value	Depth_atmo	Breadth
Time	-0.148*** (0.0129)	-0.163*** (0.0156)	-0.134*** (0.0150)	-0.164*** (0.0138)	-0.0529*** (0.00354)
Treat	-1.896*** (0.0434)	-1.816*** (0.0586)	-1.653*** (0.0418)	-1.361*** (0.0399)	-0.938*** (0.0325)
Time*treat	1.390*** (0.0450)	1.788*** (0.0535)	1.189*** (0.0374)	1.197*** (0.0410)	0.968*** (0.0327)
Rating	-0.0791*** (0.00348)	-0.154*** (0.00389)	-0.133*** (0.00386)	-0.00545 (0.00352)	-0.0131*** (0.00115)
Restaurant FE	Yes	Yes	Yes	Yes	Yes

2.5.3 Within TripAdvisor Analysis

An alternative explanation to our findings may be that the differences are not from the system change but from website differences. That is, Yelp and TripAdvisor may attract a different set of audience who may have different writing styles. Therefore, in the following sections, we focus on within TripAdvisor analysis. And in case the user base of TripAdvisor itself may change due to the system change, we only focus on TripAdvisor data after the system change. Consumers of TripAdvisor are not forced to use MD system, instead, they could provide overall ratings with or without multi-dimensional ratings. This setting allows us to compare ratings and reviews from consumers who provide only SD ratings and who provide MD ratings.

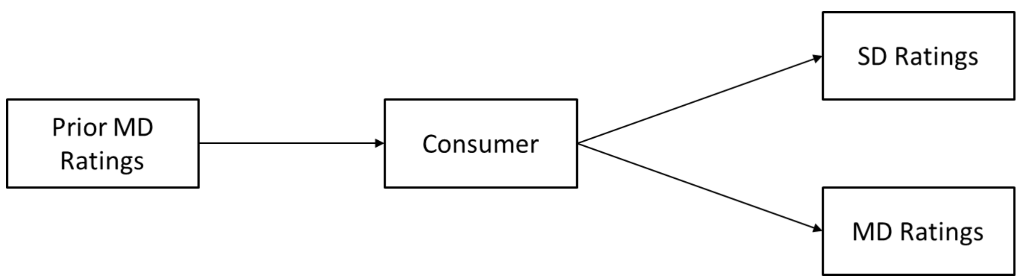

Figure 6. Within *Tripadvisor* after the System Change

We estimate the following equation.

$$Y_{ijr} = \beta_0 + \beta_1 * Multi + \beta_2 * X_{ijr} + \alpha_i + \gamma_r + \epsilon_{ijr} \qquad (2.2)$$

Multi is a dummy variable that when it is set to one, it indicates when an overall rating is provided along with multi-dimensional ratings, and when it is zero, an overall rating is provided without multi-dimensional ratings. We also control for rating for word count and emotions. And we control for word count for emotions. And since we only use data within TripAdvisor, we are able to control both restaurant and reviewer fixed effect. The results shown in table 7 and table 8 are quite consistent with what we have in table 5 and 6. There is no difference of ratings within TripAdvisor as consumers obtained the same set of information according to Liu et al 2014. We could still see an increase in word count, and increase in both depth and breadth.

Table 7

Within TripAdvisor after the System Change

	Rating		WC	
Multi	-0.02	(0.037)	32.5***	(2)
Rating			-10.7***	(0.72)
Restaurant FE	Yes		Yes	
Reviewer FE	Yes		Yes	

Table 8

Within Tripadvisor after the System Change_Breadth and Depth

	Depth_food	Depth_service	Depth_value	Depth_atmo	Breadth
Multi	0.503*** (0.0326)	0.580*** (0.0368)	0.360*** (0.0345)	0.358*** (0.0339)	0.0800*** (0.0105)
Rating	-0.0592*** (0.0117)	-0.160*** (0.0132)	-0.166*** (0.0123)	0.0718*** (0.0121)	-0.00410 (0.00377)
Restaurant FE	Yes	Yes	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes	Yes	Yes

Some may also argue that self-selection issue may exist. Consumers self-select to use either SD or MD. It is possible that consumers who tend to write longer reviews would tend to use MD ratings. In this section, we not only focus on within Tripadvisor data but also only consider reviewers who provide both SD and MD reviews as depicted in Figure 7. We didn't observe any time trend that consumers would use SD first and then stick to MD. That is, empirically, users just switch between MD and SD randomly. Results are shown in Table 9 and 10. Again, we see similar results as in previous sections.

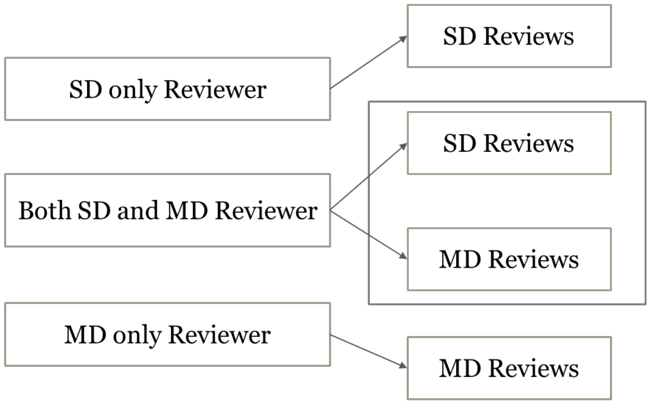

Figure 7. Within *Tripadvisor* after the System Change SD and MD Reviewers

Table 9

Within TripAdvisor after the System Change_SDMD

	Rating		WC	
Multi	-0.00776	(0.0392)	31.63***	(2.168)
Rating			-10.75***	(1.153)
Restaurant FE	Yes		Yes	
Reviewer FE	Yes		Yes	

Table 10

Within TripAdvisor after the System Change_SDMD_Breadth and Depth

	Depth_food	Depth_service	Depth_value	Depth_atmo	Breadth
Multi	0.507***	0.566***	0.376***	0.335***	0.077***
	(0.0370)	(0.0414)	(0.0381)	(0.0378)	(0.0121)
Rating	-0.0601**	-0.190***	-0.152***	0.0913***	-0.00569
	(0.0197)	(0.0220)	(0.0202)	(0.0201)	(0.00645)
Restaurant FE	Yes	Yes	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes	Yes	Yes

2.5.4 Emotions

Next we consider how emotions are shared in the MD system. Results in Table 11 show that positive affect decreases significantly, suggesting fewer positive words are being used, while there is no significant change of negative affect. Overall, these results indicate that more neutral words are being used. The results are interesting because consumers are providing higher ratings which may suggest they are more satisfied, at the same time, consumers use fewer positive emotional words and more neutral words, suggesting that consumers are more objective.

Table 11

DID Analysis_Emotions

	(1)		(2)	
	PA		NA	
Time	0.644***	(0.0404)	-0.0233	(0.0146)
Treat	5.662***	(0.302)	0.158	(0.0822)
Time*treat	-6.031***	(0.299)	-0.154	(0.0804)
Rating	1.061***	(0.0142)	-0.492***	(0.00816)
WC	-0.0169***	(0.000218)	-0.000567***	(0.0000628)
N	143885		143885	
Restaurant FE	Yes		Yes	

Table 12

DID Analysis_Emotions_Rating Valence

	(1)		(2)	
	PA		NA	
Time*treat Rating<3	-2.31***	(0.349)	-1.14**	(0.435)
Time*treat Rating=3	-7.63***	(0.557)	0.013	(0.120)
Time*treat Rating>3	-6.26***	(0.344)	-0.043	(0.052)
Restaurant FE	Yes		Yes	

Results in Table 12 analyzes whether the effect is consistent across rating valence. Due to space constraint, we only report the coefficients of the interaction terms. Results from column (1) are consistent with Table 10. Consumers tend to write fewer positive words are used in all conditions, which suggest longer and more objective reviews in MD system. Results from column (2) show that fewer negative words are used when ratings are low. Again, the results suggest more objective reviews.

2.5.5 Discrepancy between Overall Ratings and Dimensional Ratings

We are also able to analyze whether the discrepancies between the overall rating and the dimensional ratings lead reviewers to write longer reviews. The dependent variable is the word count. Discrepancy is the vector of the absolute differences between the overall rating and the dimensional estimated dimensional sentiment scores. We control for restaurant fixed effects. Results in Table 13 show that consumers tend to write longer reviews when the dimensional ratings are not consistent with the overall ratings, and the impact increase after the system change.

$$WC_{ij|k} = \beta_0 + \beta_1 * Time * Treat * Discrepancy + \beta_2 * \boldsymbol{X_{ijk}} + \alpha_i + \epsilon_{ijk} \quad (2.3)$$

Table 13

Review Length and Rating Discrepancy

D_food*Time*Treat	2.609	(1.467)
D_service*Time*Treat	13.76***	(1.464)
D_atmo*Time*Treat	10.07***	(1.197)
D_value*Time*Treat	6.271***	(1.470)
Rating	-13.29***	(0.311)
Restaurant FE	Yes	

2.5.6 Comprehensive MD Ratings

Next we examine how consumers refer their preferences into ratings in SD and MD using Equation 2.4. The dependent variable is the overall rating. Dimensentiment is the estimated sentiment score. The coefficient of the interaction term captures how dimensional ratings affect the overall rating before and after the system change. As can be seen in Table 14, the coefficients of food and value are negative and those of service and atmosphere are positive, which implies that after the system change from SD to MD, the impact of food and value on the overall rating decrease and impact of service and atmosphere increase. The results suggest that consumers put different weight on different dimensions and after the system change, the weights are more balanced. MD system primes consumers to generate a more comprehensive evaluation of all dimensions. H3 is supported.

$$Rating_{ij|k} = \beta_0 + \beta_1 * Time * Treat * DimenSentiment + \alpha_i + \epsilon_{ijk} \quad (2.4)$$

Table 14

Overall Rating and Dimensional Sentiments

Senti_food*Time*Treat	-0.0272*	(0.0124)
Senti_service*Time*Treat	0.213***	(0.0143)
Senti_atmo*Time*Treat	0.117***	(0.0157)
Senti_value*Time*Treat	-0.0983***	(0.0130)
Restaurant FE	Yes	

We also conduct an experiments to better understand how consumers refer their preferences into ratings in SD and MD. We try to mimic an environment where consumers have the same consumption experience while using different rating systems. Specifically, subjects were asked to read someone else's dining experiences and then rate the restaurant using the two rating systems. Subjects were first shown four pieces of reviews (See Appendix A.2). We choose a relatively small number to reduce cognitive burdens and make sure that all reviews have been read and all subjects obtain the same set of information. Subjects were told that all these reviews are authentic and they need to read all and consider these reviews as their own consumption experiences. And then subjects were randomly divided into four groups, SD, SD with priming, MD with overall rating first, and MD with dimensional rating first. Subjects in the first group were asked to rate the restaurant using SD system. Subjects in the second group were asked to think about different attributes of the restaurant (for example, service, food, ambiance, etc) and then rate the restaurant using SD system. Subjects in the third group 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 group were asked to first rate different dimensions of the restaurant (food, service, atmosphere, and value), and then rate the restaurant on the overall rating. Time used to read and rate reviews are recorded and subjects who didn't spend enough time will be teased out of the sample. Besides, on the same page, subjects will be asked to indicate the month of the day when they take the experiment, and those who didn't provide the correct answer will be teased out. And then, all subjects were asked to indicate what attributes (food, service, atmosphere, value, and other) have been considered when generating the overall rating and the importance of each of these attributes by allocating 100 points among the attributes. In total, we receive 2682 responses passing manipulation test out of 2745 responses. Results are shown in Figure 8 using 45 seconds as the cutoff which leads to 1615 valid responses. And we use different cutoffs of time spent on reading and rating reviews and the results are consistent.

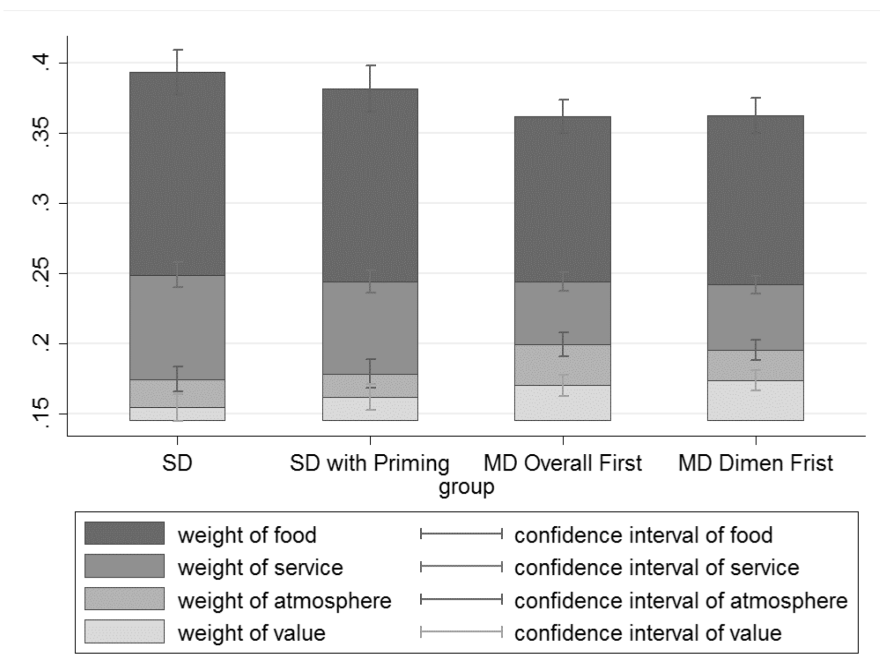

Figure 8. The Priming Effect on Dimensional Ratings

Figure 8 shows the average percentages of different dimensions used to generate the overall rating. Column SD shows that the average weights of food, service, atmosphere and value are approximately 0.40, 0.25, 0.18, and 0.16 respectively. And the last column MD Dimen First show that the weights change to approximately 0.18, 0.2, 0.25, 0.37. The weights of atmosphere and value increase significantly when using the MD system. The results suggest a higher diversity of weights of different dimensions in the SD system than in the MD system. On average, consumers would put more dimensions on food in the SD system, however, in the MD system they significantly consider more other dimensions. MD system primes consumers to generate a more comprehensive evaluation of all dimensions.

2.6 Discussion

We corroborated an observational study with an experimental study to examine how consumers reflect their overall consumption experience in ratings and reviews in different rating systems. Our results suggest that MD ratings do not substitute text reviews. Consumers in an MD system tend to share more information and cover more dimensions in textual reviews in a more objective way. A natural question following is that are higher depth and breadth reviews really helpful? Consumers read textual reviews rather than relying simply on summary statistics (Chevalier and Mayzlin 2006) to resolve their uncertainty about product attributes (Pavlou and Dimoka 2006). Review depth has a positive effect on the helpfulness of the review (Mudambi and Schuff 2010). There is limited understanding of how review breadth impact review helpfulness which could be a potentially interesting topic for a future study. Results from randomized experiments corroborate that MD ratings do not substitute text reviews. Consumers' decisions to read reviews are not affected by the rating system. MD system primes consumers to generate a more comprehensive numerical overall rating of all dimensions as well as more comprehensive textual reviews. In addition, consumers are also found to use more neutral words in their textual reviews. Future research will dig further impact on other linguistic features and robustness checks of within-reviewer and between-reviewer variation. Our study contributes to rating system design and provides a better understanding of how ratings and reviews reflect consumers' experiences, and our findings also increase online retailers' understanding of the role rating system play in opinion sharing.

CHAPTER 3

THE IMPACT OF ECONOMIC INCENTIVES ON DIFFERENT PLAYERS IN ONLINE MARKETPLACE

3.1 Introduction

Ratings and reviews are considered to play an important role in the online marketplace. However, not everyone who transacts online may review. Various strategies have been employed to encourage customers to post reviews. Using financial incentives to attract users has become a common practice in recent years. Reviewers could get rewards, such as reward points or a small amount of money after consumption based on the number of reviews written or based on the quality of the review written. Vendors or platforms could also offer free or discounted products before consumption in exchange for reviews. There are other differences between before and after consumption incentives except for when the financial incentives are received. For example, after-consumption financial incentives are usually offered by the platform (Khernamnuai et al. 2018), while before-consumption financial incentives could be offered by both the platform and the vendor. Most research focuses on incentives offered by the platform with conflicting findings (Stephen et al. 2012, Wang et al. 2012, Yu et al. 2018). It is not clear how the impact differ when offered by the platform and by the vendors. For example, Amazon provides highly-ranked reviewers with free products and expects for high quality Vine reviews. However, Amazon also banned incentivized reviews of free or discounted products offered by vendors instead of the platform in Oct 2016. It is ambiguous how financial incentives affect different players and the interaction in the online marketplace, such as the vendor, the incentivized and non-incentivized reviewers, and the platform. On one hand, financial incentives from the platform may attract new reviewers (Khernamnuai et al. 2018), and enjoy more positive reviews from incentivized reviewers (ReviewMeta.com). On the other hand, financial incentives from product vendors may lower satisfaction of more non-incentivized reviewers who consume previous incentivized reviews (Stephen et al. 2014). It could also reduce consumers' trust in the reviews platform as a whole. The competing arguments from the few prior studies and the lack of direct evidence motivate the need for additional studies. The main interest of our study is to provide a comprehensive understanding of how before-consumption financial incentives affect different players and the interaction in the online marketplace, such as the vendor, the incentivized and non-incentivized reviewers, and the platform. This study contributes to the IS literature in a few significant ways. For instance, we empirically demonstrate how financial incentives differ when offered before and after consumption by different players in the online marketplace. And we explored how economic incentives offered by vendors affect reviewer behavior in a real-world setting. Our findings help academic better understand the role of financial incentives in online product reviews and offer practical implications on the design of online product review systems.

3.2 Related Literature

IS literature has explored the impact of financial incentives on customer referral (Ryu and Feick, 2007; Kornish and Li, 2010; Reimer and Benkenstein, 2016; Yili et al., 2017) and on reviews regarding quantity, quality and valence, however yielding inconclusive results (Stephen et al., 2012; Wang, Pavlou and Gong, 2016; Burch et al., 2018; Yu, Khern-am-nuai and Pinsonneault, 2018).

Two mechanisms of after consumption financial rewarding, completion-contingent and performance-contingent, have been investigated. Completion-contingent rewards are given based on the quantity of reviews written and is frequently used by online retail platforms, while performance-based mechanism incentivizes on quality of the reviews but is difficult to implement in practices (Khern-am-nuai et al., 2018). While the majority of current studies assumed completion-contingent incentives and yielded mixed results. (Stephen et al., 2012; Wang et al., 2016, Burtch et al., 2018) on review valance and quality, performance-contingent incentives were found to have a positive effect on quality (Want et al., 2012) and quality (Yu et al., 2018). In both mechanisms, consumers are given the financial incentives after they complete the consumption. And the incentives are provided by the platform across all retails. However, there is limited understanding of the impact of financial incentives using before consumption mechanisms. Qiao et al. (2017) found that financial incentive would lower review quality, however, it is not clear how other review characteristics and other players are affected. This research extends prior research by empirically examining the impacts of before consumption financial incentive on opinion sharing in terms of review quantity, quality, and valance.

3.3 Theory and Hypotheses Development

In this section, we develop the research hypotheses examining how finical incentives affect firms, reviewers and the platform. A summarized framework is presented in Figure 9. Both firms and the platform could provide financial incentives, and consumers decide whether to participate in the campaign. Reviewers are eligible to receive free or discounted sample products in exchange for writing honest and unbiased product reviews. Reviewers who ever received before-consumption financial incentives are identified as incentivized reviewers. Those incentivized reviewers could write both incentivized and non-incentivized reviews. Products whichever offered before-consumption financial incentives are identified as incentivized products. And incentivized products could receive reviews from both incentivized reviewers and non-incentivized reviewers.

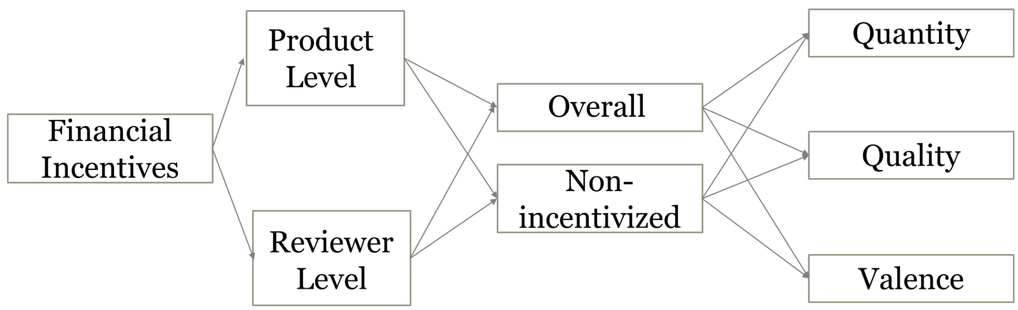

Figure 9. Research Framework

Quantity

Extrinsically motivated behaviors could be understood as actions that take place to earn external rewards and benefits (Lepper and Green, 1978), while intrinsically motivated behaviors are rewarded with the satisfaction from the actions themselves (Deci, 1971). Accordingly, besides relying on the customers to intrinsically share their satisfaction with the products or services, another strategy to generate more eWOM is providing monetary incentives to promote review writing through external benefits. The presence of monetary incentives could increase the intention to write a review by enhancing the extrinsic motivation of potential reviewers. Even though it might also threaten to hinder intrinsic motivation – the self-determination – of the reviewers (Sun et al. 2016), the increase of extrinsic motivation can be larger than the decrease of intrinsic motivation. Therefore, we propose that

Hypothesis 1: Financial incentives lead to an increase in review quantity for incentivized reviews of incentivized reviewers.

Hypothesis 2: Financial incentives lead to an increase in review quantity for incentivized reviews of incentivized products.

Incentivized reviewers could also write non-incentivized reviews. A spillover effect might exist when consumers continue to write positive reviews for other products which don't provide monetary incentives.

Hypothesis 3: Financial incentives lead to an increase in review quantity for non-incentivized reviews of incentivized reviewers.

Incentivized products could also receive non-incentivized reviews. The increase in review quantity brought by incentivized reviews could positively impact sales of the focal product leading to more non-incentivized reviews.

Hypothesis 4: Financial incentives lead to an increase in review quantity for non-incentivized reviews of incentivized products.

Quality

Regarding quality, on one hand, offering explicit incentives may encourage reviewers to take a more professional approach to the review-writing task, thus lead to higher review quality. However, financial incentives might also have negative impacts on review quality. Reviewers receive the incentives before the consumption and they decrease the initial motivations including altruism and some intrinsic interests (Verlegh et al. 2004 and Martin 2015). As a result, they would be less likely to spend time and effort in writing the reviews. Thus, the review quality might decrease when receiving before consumption incentives.

Hypothesis 5: Financial incentives lead to a decrease in review quality for incentivized reviews of incentivized reviewers.

Hypothesis 6: Financial incentives lead to a decrease in review quality for incentivized reviews of incentivized products.

Hypothesis 7: Financial incentives lead to an increase in review quality for incentivized reviews of incentivized reviewers.

Hypothesis 8: Financial incentives lead to an increase in review quality for incentivized reviews of incentivized products.

Reviewers have limited time and efforts. Even if they could put more efforts on the incentivized reviews, it is likely that they won't have enough time to write higher quality non-incentivized reviews.

Hypothesis 9: Financial incentives lead to a decrease in review quality for non-incentivized reviews of incentivized reviewers.

Valance

Review valance of incentivized reviewers and of incentivized products could be affected in different directions. The concept of valence captures the attitude of the reviewers towards the products or services mentioned in the reviews (Khern-am-nuai et al., 2018; Yu et al., 2018). First, reciprocity could drive incentivized reviewers to reward the product by providing more positive reviews. Kim et al. (2016) found that producing incentivized positive eWOM improved the review writers' attitude towards the product and company through the mechanism of "saying is believing". A spillover effect might exist when consumers continue to write positive reviews for other products which don't provide monetary incentives. Therefore, we propose that

Hypothesis 10: Financial incentives lead to an increase in review valance for incentivized reviews of incentivized reviewers.

Hypothesis 11: Financial incentives lead to an increase in review valance for non-incentivized reviews of incentivized reviewers.

Different from after-consumption incentives, incentivized reviewers are required to release the incentive information in the reviews. On one hand, non-incentivized reviewers who read the incentivized reviews would impose skepticism and doubt on the credibility and the quality of the reviews and the product (Stephen et al., 2012; Godes et al., 2005; Martin, 2014), thus lead to negative reviews towards the incentivized products.

Hypothesis 12: Financial incentives lead to a decrease in review valance for non-incentivized reviews of incentivized products.

3.4 Data and Empirical Model

In order to perform our empirical analysis, we collect reviews from Amazon.com for one category cell phone and accessories. We started from the product codes and then collected all their reviews. In 2007, Amazon launched the Amazon Vine program, through which consumers could receive free product in exchange for unbiased reviews. On October 3rd, 2016, Amazon announced to eliminate any incentivized reviews which were offered to customers by product vendors. That allows us to identify how incentivized reviews affect consumer behavior. Reviewers are required to release incentivized information in their reviews. For example, “I received a free/discounted product in exchange for an honest/unbiased review.” We use four key words to match disclosed information in text reviews and then manually evaluate the performance. Reviews are then identified into incentivized reviews and non-incentivized reviews. And then we are able to identify incentivized reviewers, non-incentivized reviewers, incentivized products and non-incentivized products. Quantity is measured by the total number of reviews per month. Valance is the average rating per month. And Quality is measured by the average word count per month. These three variables are all measure on both product and reviewer level.

We are able to observe when a certain product/reviewer adopted incentivized reviews and when the incentivized reviews were prohibited by the platform. We explore the impacts of adopting and banning financial incentives separately. In the adoption process, we use data before 2014 and consider a DID model combined with matching techniques, such as the propensity score matching (PSM), to eliminate unobserved variable bias. The treatment groups are incentivized products and incentivized reviewers. We create “proper” control groups for treated products and treated reviewers separately by using PSM. We ensure that the control and treated groups are comparable in terms of observable characteristics. Monthly Average price, average sales rank and sub-category are used to match products exhibiting similar patterns. Monthly number of reviews, average word count, average rating and standard deviation of ratings are used to match reviewers in the control group. Then, we run the DID regression question. We estimate the Equation 3.1 and 3.2. Dependent variables include log transformation of total number of reviews of product i and of reviewer j in month t (LnNReviewsit, LnNReviewsjt), log transformation of total number of non-incentivized reviews of product i and of reviewer j in month t (LnNReviewsNIit, LnNReviewsNIjt), average rating of product i and of reviewer j in month t(AvgRatingit, AvgRatingjt), average non-incentivized rating of product i and of reviewer j in month t (AvgRatingNIit, AvgRatingNIjt), log transformation of average review word count of product i and of reviewer j in month t (LnWCit, LnWCjt), log transformation of average non-incentivized review word count of product i and of reviewer j in month t (LnWCNIit, LnWCNIjt). Treat denotes the treatment group. And the coefficient of the interaction term captures the effect of incentivized reviews. AdoptIR indicates the dummy for a certain product i or a certain reviewer j start to receive/provide incentivized reviews. We also control for product fixed effects and time fixed effects.

$$Y_{it} = \beta_0 + \beta_1 * AdoptIR_i + \beta_2 * AdoptIR_i * Treat + \beta_3 * Treat + \alpha_i + \delta_t + \epsilon_{it} \tag{3.1}$$

$$Y_{jt} = \beta_0 + \beta_1 * AdoptIR_j + \beta_2 * AdoptIR_j * Treat + \beta_3 * Treat + \alpha_i + \delta_t + \epsilon_{it} \tag{3.2}$$

In the prohibition process, we estimate the following equation focusing on product level. BanIR is the dummy for Amazon banning incentivized reviews (BanIR) in Equation 3.3. We control for the review sequence, which is the distance between current month and the first review month for each product, as well as product fixed effects and time fixed effects.

$$Y_{it} = \beta_0 + \beta_1 * BanIR + \beta_2 * LnSequence_i + \alpha_i + \delta_t + \epsilon_{it} \quad (3.3)$$

3.5 Results

The results of DID-PSM analysis for the total number of reviews, average rating and average word count are shown below. Table 15 shows the results of incentivized products. The positive coefficients of the interaction term in column 1 and 2 suggest that the introduction of financial incentives lead to an increase in the total number of both incentivized and non-incentivized reviews. The negative coefficients in column 3 and 4 suggest that the existence of incentivized reviews may cause reviewers to doubt the quality of the product and give lower ratings. In addition, financial incentives appear to have a negative impact on review quality. Reviewers spend less time and effort writing reviews.

Table 15

With Economic Incentives_Product Level (DID-PSM)

	(1)	(2)	(3)	(4)	(5)	(6)
	LnNReviews	LnNReviewsNI	AvgRating	AvgRatingNI	LnWC	LnWCNI
AdoptIR *Treat	3.420***	3.971***	-0.245***	-0.272***	-25.09***	-16.95***
	(0.209)	(0.224)	(0.0072)	(0.007)	(0.498)	(0.492)
N	414634	437894	414608	437868	414634	437894
Product FE	Yes	Yes	Yes	Yes	Yes	Yes
Yearmonth FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 16 reports the impact of financial incentives on incentive reviewers. Results from column 1 and 2 are consistent with results in Table 15. Financial incentives lead to more reviews. Incentivized reviewers write more reviews not only for the incentivized products, but also for other products which don't provide financial incentives. Average rating increase as indicated in column 3 and 4. Reviewers reciprocate to the product by providing positive reviews. The positive spillover effect exists. Results don't support a significant impact on review quality.

Table 16

With Economic Incentives_Reviewer_Level (DID-PSM)

	(1)	(2)	(3)	(4)	(5)	(6)
	LnNReviews	LnNReviewsNI	AvgRating	AvgRatingNI	LnWC	LnWCNI
AdoptIR *Treat	0.512***	1.141***	0.126***	0.107***	-3.457	6.840
	(0.116)	(0.138)	(0.028)	(0.0278)	(5.005)	(4.686)
N	18970	18970	18970	18970	18970	18970
Product FE	Yes	Yes	Yes	Yes	Yes	Yes
Yearmonth FE	Yes	Yes	Yes	Yes	Yes	Yes

Results from Table 17 show that the Amazon's decision to prohibit incentivized reviews decrease the number of total reviews and non-incentivized reviews, and no significant quality change. The overall ratings increase for both incentivized and non-incentivized reviews. The coefficients are in the opposite direction of those in Table 15 which suggest consistent results.

Table 17

Without Economic Incentives_Product Level

	(1)	(2)	(3)	(4)	(5)	(6)
	LnNReviews	LnNReviewsNI	AvgRating	AvgRatingNI	LnWC	LnWCNI
BanIR	-1.755**	-1.965**	2.593**	2.576**	0.0839	-1.164
	(-2.72)	(-3.11)	(3.03)	(2.75)	(0.11)	(-0.77)
LnSequence	0.183***	0.228***	-0.138***	-0.137***	-0.26***	0.069*
	(12.45)	(13.80)	(-7.08)	(-5.59)	(-14.31)	(2.03)
N	13269	10482	13269	10482	13268	13268
Product FE	Yes	Yes	Yes	Yes	Yes	Yes
Yearmonth FE	Yes	Yes	Yes	Yes	Yes	Yes

3.6 Discussion

In this study, we examine the impact of the before-consumption financial incentives on different players in the online marketplace. Using a natural experimental design combining with propensity score matching, we test the influence of adopting and prohibiting financial incentives on the quantity, quality, and valance of reviews. We found that similar to after-consumption incentives, before-consumption incentives encourage people to provide more reviews, however, low quality. Additionally, we find that the release of incentivized information may cause reviewers to doubt the credibility and quality of the reviews and of the product and lead to lower average ratings. This study advances our understanding of financial incentives by investigating the before-consumption incentive mechanism and providing the empirical evidence of how the disclose of incentive information affect consumers' behavior on information sharing.

REFERENCES

Anderson, E. W., & Sullivan, M. W. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing science*, 12(2), 125-143.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management science*, 57(8), 1485-1509.

Amabile, T. M. (1983). Brilliant but cruel: Perceptions of negative evaluators. *Journal of Experimental Social Psychology*, 19(2), 146-156.

Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, 53(3), 297-318.

Belkin, N. J. (1984). Cognitive models and information transfer. *Social Science Information Studies*, 4(2-3), 111-129.

Benbasat, I., & Zmud, R. W. (2003). The identity crisis within the IS discipline: Defining and communicating the discipline's core properties. *MIS quarterly*, 183-194.

Bhattacherjee, A. (2001). Understanding information systems continuance: an expectation-confirmation model. *MIS quarterly*, 351-370.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research* (3), pp. 993-1022.

Brown, S. A., Venkatesh, V., & Goyal, S. (2012). Expectation confirmation in technology use. *Information Systems Research*, 23(2), 474-487.

Brown, S. A., Venkatesh, V., & Goyal, S. (2014). Expectation confirmation in information systems research: A test of six competing models. *MIS Quarterly*, 38(3).

Burtch, G., Hong, Y., Bapna, R., & Griskevicius, V. (2016). Stimulating Online Reviews by Combining Financial Incentives and Social Norms.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3), 345-354.

Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science*, 29(5), 944-957.

Chung, C. M., & Darke, P. R. (2006). The consumer as advocate: Self-relevance, culture, and word-of-mouth. *Marketing Letters*, 17(4), 269-279.

Churchill, G. A., & Carol, S. (1982). An investigation into the detenninants of consumer satisfaction. *Journal of Marketing Research*, 19(4), 491-504.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. Journal of *Personality and Social Psychology*, 18(1), 105–115.

Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, *27*(4), 293-307.

Dellarocas, C. (2006). Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management science*, *52*(10), 1577-1593.

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

Diehl, K., & Poynor, C. (2010). Great expectations?! Assortment size, expectations, and satisfaction. *Journal of Marketing Research*, *47*(2), 312-322.

Dimoka, A., Hong, Y., & Pavlou, P. A. (2012). On product uncertainty in online markets: Theory and evidence. *MIS quarterly*, 395-426.

Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter?—An empirical investigation of panel data. *Decision support systems*, *45*(4), 1007-1016.

Engel, J. F., Blackwell, R. D., & Miniard, P. W. (1995). Consumer behavior (8th eds.). *Forth Worth-Tokyo: The Dryden Press,-992 pp*.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, *19*(3), 291-313.

Festinger, L. (1957). A theory of cognitive dissonance: Stanford Univ Pr. *Fornell, C., & Larcker, DF (1981). Evaluating structural equation models with*.

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

Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., ... Verlegh, P. (2005). The Firm's Management of Social Interactions. *Marketing Letters*, 16(3/4), 415–428.

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

Golder, S. A., & Macy, M. W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, *333*(6051), 1878-1881.

Ghose, A., Ipeirotis, P., & Sundararajan, A. (2009). The dimensions of reputation in electronic markets.

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, *23*(10), 1498-1512.

Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, *31*(3), 493-520.

Goh, K. Y., Heng, C. S., & Lin, Z. (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. *Information Systems Research*, *24*(1), 88-107.

Gong, J., Abhishek, V., Li, B. (2017) Examining the Impact of Contextual Ambiguity on Search Advertising Keyword Performance: A Topic Model Approach. MIS Quarterly Forthcoming.

Grewal, D. (1995). Product quality expectations: Towards an understanding of their antecedents and consequences. *Journal of business and Psychology*, *9*(3), 225-240.

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of psychology*, *21*(1), 107-112.

Heider, F. (2013). *The psychology of interpersonal relations*. Psychology Press.

Helson, H. (1964). Adaptation-level theory: an experimental and systematic approach to behavior.

Hennig-Thurau, T., Gwinner, K.P., Walsh, G., Gremler, D.D. (2004). Electronic word- of-mouth via consumer-opinion platforms: what motivates consumers to ar- ticulate themselves on the Internet? *J. Interact. Mark.* 18 (1), 38–52.

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

Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decision support systems*, *57*, 42-53.

Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). ACM.

Hong, Y., Chen, P. Y., & Hitt, L. M. (2012). Measuring product type with dynamics of online product review variance.

Hong, Y., & Pavlou, P. A. (2014). Product fit uncertainty in online markets: Nature, effects, and antecedents. *Information Systems Research*, *25*(2), 328-344.

Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I* (pp. 99-127).

Khern-am-nuai, W., Kannan, K., & Ghasemkhani, H. (2017). Extrinsic Versus Intrinsic Rewards for Contributing Reviews in an Online Platform.

Kim, D. J., Ferrin, D. L., & Rao, H. R. (2009). Trust and satisfaction, two stepping stones for successful e-commerce relationships: A longitudinal exploration. *Information systems research*, *20*(2), 237-257.

Kim, J., Naylor, G., Sivadas, E., & Sugumaran, V. (2016). The unrealized value of incentivized eWOM recommendations. *Marketing Letters*, 27(3), 411–421.

Kirmani, A., & Rao, A. R. (2000). No pain, no gain: A critical review of the literature on signaling unobservable product quality. *Journal of marketing*, *64*(2), 66-79.

Kivetz, R., Urminsky, O. & Zheng, Y. (2006). The Goal-Gradient Hypothesis Resurrected: Purchase Acceleration, Illusionary Goal Progress, and Customer Retention. Journal of *Marketing Research*, (1), 39.

Kwark, Y., Chen, J., & Raghunathan, S. (2014). Online product reviews: Implications for retailers and competing manufacturers. *Information systems research*, *25*(1), 93-110.

Lee, M., Liang, M.-C., & Kang, E. (2009). Getting Paid to Blog about Products: The Effects of Disclosure and Amount of Monetary Compensation on Brand Attitude. *American Academy of Advertising Conference Proceedings*, 91.

Lee, Y. J., Tan, Y., & Hosanager, K. (2009). Do I follow my friends or the crowds? Examining informational cascades in online movie reviews. *Proceedings of WITS 2009*.

Li ,H., Lin, R., Hong, R., Ge, Y. (2015) Generative Models for Mining Latent Aspects and Their Ratings from Short Reviews. *IEEE International Conference on Data Mining*, pp. 241-250.

Li, X. (2016). Could deal promotion improve merchants' online reputations? The moderating role of prior reviews. *Journal of Management Information Systems*, *33*(1), 171-201.

Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, *19*(4), 456-474.

Li, X., & Hitt, L. M. (2010). Price effects in online product reviews: An analytical model and empirical analysis. *MIS quarterly*, 809-831.

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

Liu, Y., Chen, P. Y., & Hong, Y. (2014). Value of multi-dimensional rating systems: An information transfer view. *Proceedings of the 35th International Conference on Information Systems (ICIS)*, Auckland, New Zealand.

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

Marinescu, I., Klein, N., Chamberlain, A., & Smart, M. (2018). Incentives can reduce bias in online employer reviews. *Academy of Management Proceedings* (Vol. 2018).

Martin, W. C. (2015). Independent versus incentivized word-of-mouth: effects on listeners. *Academy of Marketing Studies Journal*, (1), 1.

Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: an empirical investigation of online review manipulation. *American Economic Review*, 104(8), 2421–2455.

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

Mudambi, S. M., & Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. *MIS quarterly*, 185-200.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of experimental psychology: general*, *106*(3), 226.

Nelson, P. J. (1981). Consumer information and advertising. In *Economics of information* (pp. 42-77). Springer, Dordrecht.

Newcomb, T. M. (1953). An approach to the study of communicative acts. *Psychological review*, *60*(6), 393.

Nowlis, S. M., & Simonson, I. (1996). The effect of new product features on brand choice. *Journal of marketing research*, 36-46.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of marketing research*, 460-469.

Packard, G., & Wooten, D. B. (2013). Compensatory knowledge signaling in consumer word-of-mouth. *Journal of Consumer Psychology*, *23*(4), 434-450.

Pavlou, P. A., & Dimoka, A. (2006). The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research*, *17*(4), 392-414.

Reimer, T., & Benkenstein, M. (2016). Altruistic eWOM marketing: More than an alternative to monetary incentives. *Journal of Retailing and Consumer Services*, 31, 323–333.

Reimer, T., & Benkenstein, M. (2018). Not just for the recommender: How eWOM incentives influence the recommendation audience. *Journal of Business Research*, 86, 11–21.

Sela, A., & Berger, J. (2012). How attribute quantity influences option choice. *Journal of Marketing Research*, *49*(6), 942-953.

Singh, P. V., Sahoo, N., and Mukhopadhyay, T. (2014) How to Attract and Retain Readers in Enterprise Blogging? *Information Systems Research* (25:1), pp. 35-52.

Shi, Z., Lee, G., Whinston, A. (2016) Toward a Better Measure of Business Proximity: Topic Modeling for Industry Intelligence. *MIS Quarterly* (40:4):1035–1056.

Stephen, A., Bart, Y., Du Plessis, C., & Goncalves, D. (2012). Does paying for online product reviews pay off? The effects of monetary incentives on content creators and consumers. *ACR North American Advances*.

Sun, Y., Dong, X., & McIntyre, S. (2017). Motivation of user-generated content: Social connectedness moderates the effects of monetary rewards. *Marketing Science*, 36(3), 329-337.

Sundaram, D. S., Mitra, K., & Webster, C. (1998). Word-of-mouth communications: A motivational analysis. *ACR North American Advances*.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *The quarterly journal of experimental psychology*, *37*(4), 571-590.

Thompson, D. V., Hamilton, R. W., & Rust, R. T. (2005). Feature fatigue: When product capabilities become too much of a good thing. *Journal of marketing research*, *42*(4), 431-442.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*(4940), 301-306.

Venkatesh, V., & Goyal, S. (2010). Expectation disconfirmation and technology adoption: polynomial modeling and response surface analysis. *MIS quarterly*, 281-303.

Wang, S., Pavlou, P. A., & Gong, J. (2016). On Monetary Incentives, Online Product Reviews, and Sales. *ICIS 2016 Proceedings*, 1–20.

Wang, C., Zhang, X., & Hann, I. H. (2018). Socially nudged: A quasi-experimental study of friends' social influence in online product ratings. *Information Systems Research*.

Wetzer, I. M., Zeelenberg, M., & Pieters, R. (2007). "Never eat in that restaurant, I did!": Exploring why people engage in negative word-of-mouth communication. *Psychology & Marketing*, *24*(8), 661-680.

Yili, H., Pavlou, P. A., Nan, S., & Kanliang, W. (2017). on the Role of Fairness and Social Distance in Designing Effective Social Referral Systems. *MIS Quarterly*, 41(3), 787-A13.

Yu, Y., Khern-am-nuai, W., & Pinsonneault, A. (2018). When Paying for Reviews Pays Off: The Case of Performance-Contingent Monetary Incentives.

Zajonc, R. B. (2017). The concepts of balance, congruity, and dissonance. In *Attitude Change* (pp. 63-85). Routledge.

APPENDIX A

HOW CONSUMERS REPSOND TO DIFFERENT RATING SYSTEMS

A.1

In this experiment, subjects are randomly divided into two groups, SD and MD. If subjects are in the SD group, he/she would be shown the following page.

Café Tempe

55 Reviews

$
Downtown Tempe
American, International, Contemporary, Healthy

Dining options: Breakfast,Lunch,Dinner, Late Night, Delivery, Reservations, Takeout
Description: Cafe Tempe Bistro is a charming art district eatery nestled among the galleries of downtown Scottsdale. Featuring New American Comfort Food, a fresh & healthy style of cooking delicately prepared and dedicated to the use of all natural locally grown produce & organic ingredients whenever possible. Established in 1996, open for lunch & dinner, reservations welcomed.

Consumer reviews:

"Good food, good prices, pleasant staff"
The food was good, the service was great, and the prices were good.

"Convenient but that's about all"
Food was mediocre and access was extremely limited by strip mall construction. Service was adequate. Price was inexpensive.

"Good Quality and Service"
The staff is always friendly and helpful. The food is fresh, tastes great, and is made the way I like it. If not, they'll make it right. I usually have a difficult time deciding- everything looks so good.
I got the slab of ribs with fries and coleslaw. The wife got fried shrimp (and a lot of them!). The ribs are excellent; smoky, fall of the bone, fork tender and moist. The sauce is tasty with mild heat. The shrimp were cooked well, not overdone or greasy.

Want to read more consumer reviews:

Yes No

Figure 10. Restaurant Page Provided to Respondents

After that, subjects were asked to answer a list of same demographic questions, and then they were asked to answer a few manipulation questions.

Gender

Male	Female

Are you a native English speaker?

Yes	No

Age

How often do you check restaurant ratings online before you decide to go to a restaurant?

0%	25%	50%	75%	100%

I am familiar with restaurant ratings.

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

Generally speaking, people can not be trusted.

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

Generally speaking, it is difficult for me to make a decision in my daily life.

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

All things considered, I am satisfied with my life as a whole these days.

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

Do you like American foods?

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

How often do you go to restaurants?

- Less than once a month
- Once a month
- 2-3 Times a Month
- Once a Week
- 2-3 Times a Week
- Daily

Figure 11. Demographic Questions

A.2

Respondents were asked to read the reviews:

Please read all the following reviews carefully:

"Very Nice Restaurant"

This is the second visit to this family owned restaurant. I wasn't disappointed the first time; and I was once again pleased. They have a knowledgeable and friendly staff. The menu choices were more than adequate. The food was properly prepared and served fresh out of the kitchen. I would be more than pleased to visit this wonderful, little restaurant again.

"It doesn't get much better than this!"

What a beautiful setting in their patio. Very romantic. The food is very good, but not exceptional. But the setting and the staff really make it memorable. This is a special occasion place!

"Lovely ambiance/average food"

My husband and I took my 86-year-old mother for a special occasion. The food is overpriced. The service also was not on par with the price point. We were in a room with only one other seated table, and rarely saw our waiter.

"Truly one of the best meals I have had"

I came here with my boyfriend for his birthday. WOW the service was cunning, the atmosphere is the most incredible part and the food paired with wine is so good it makes you want to scream. Gorgeous place inside and especially out. Everything has a place to go and every detail is thought out!

Figure 12. Reviews Provided to Respondents

And then they were randomly divided into four groups: SD, SD with Priming, MD Overall First, and MD Dimen First, and then they were asked to indicate the month of the day as a screening test.

Please rate the restaurant (1 denotes extremely bad and 5 denotes extremely good)

	1	2	3	4	5
Overall	○	○	○	○	○

Please indicate the month of today

August	Sepetember	July	October
○	○	○	○

Figure 13. Respondents in SD Group

Please think about different attributes of the restaurant (for example, service, food, ambiance, etc) and then rate the restaurant(1 denotes extremely bad and 5 denotes extremely good)

	1	2	3	4	5
Overall					

Please indicate the month of today

August	Sepetember	July	October

Figure 14. Respondents in SD with Priming Group

Please rate the restaurant (1 denotes extremely bad and 5 denotes extremely good)

	1	2	3	4	5
Overall					

Please rate different attributes of the restaurant(1 denotes extremely bad and 5 denotes extremely good)

	1	2	3	4	5
Food					
Service					
Atomosphere					
Value					

Please indicate the month of today

August	Sepetember	July	October

Figure 15. Respondents in MD Group

Please rate different attributes of the restaurant(1 denotes extremely bad and 5 denotes extremely good)

	1	2	3	4	5
Food					
Service					
Atomosphere					
Value					

Please rate the restaurant (1 denotes extremely bad and 5 denotes extremely good)

	1	2	3	4	5
Overall					

Please indicate the month of today

August	Sepetember	July	October

Figure 16. Respondents in MD with Priming Group