The value of rating diversity within multidimensional rating system: Evidence from hotel booking platform

Abstract

An increasing number of hotel booking platforms and review websites deploy multidimensional rating systems to encourage users to provide additional evaluations of a product/service besides the overall rating. However, providing the itemized ratings of a product/service requires extra effort which makes users tend to post identical ratings (default setting). According to the information transfer theory, we postulate that rating diversity enhances review usefulness within the multidimensional rating system and; based on the loss aversion phenomenon, we also posit that the positive effect of rating diversity on review usefulness should be greater among negative reviews. Using a sample of 1,720,429 hotel reviews, this study reveals that while rating diversity enhances review usefulness, its effect varies across the valence of each review; specifically, the positive effect of rating diversity mainly functions among negative reviews. These findings yield direct implications for hotel booking platforms or review websites that deploy multidimensional rating systems.

Keywords: rating system; rating diversity; review valence; review usefulness; information transfer; loss aversion

1. Introduction

Hotel booking platforms (e.g., Expedia, Booking.com, and Ctrip) involves three parties: consumers, hotel managers, platform operators. From the perspective of consumers, the usefulness of online reviews is critical for potential consumers to evaluate a product/service in hospitality (Li et al., 2019; Li et al., 2020; Leung, 2021), the quality of which is very difficult to infer before consumption (Liu & Park, 2015). From the perspective of hotel managers, useful reviews facilitate information matching and thus promote online sales. From the perspective of hotel booking platforms, informative reviews can help potential consumers to make informed booking decisions, thus benefiting these platforms in the long run. Therefore, informative reviews are essential in the process of information matching and uncertainty reduction before making purchase decisions on a hotel booking platform.

Hotel reviews typically have two parts: review text and rating score. Compared with review texts, rating score is more straightforward and easy to be perceived by potential consumers who are browsing reviews to make booking decisions (Chen et al., 2018; Ghose et al. 2012). In current business practice, hotel booking platforms usually adopt a single-dimensional rating system or a multidimensional rating system. Related studies have shown that single-dimensional rating systems—wherein the platforms only collect and aggregate the overall rating—cannot comprehensively reflect the information of the different attributes of a product or service, thereby reducing the usefulness of reviews in helping potential consumers to make booking decisions (Archak et al., 2011; De Langhe et al., 2016). From this perspective, a multidimensional rating system (consumers are required to provide both the overall rating and itemized ratings on their orders) should perform better in conveying the comprehensive information of a product or service, because the multidimensional ratings (e.g., service, location, etc.) facilitate information screening and information matching (Chen et al., 2018). That is, potential consumers can evaluate whether a product or service fits their requirements and/or preferences through the itemized ratings on each attribute of the product or service.

Although many hotel booking platforms and hotel review websites deploy multidimensional rating systems for users to rate their orders after consumption; the fact is that providing itemized ratings requires additional effort, which may lead users to provide identical ratings (the itemized ratings are the same as the overall rating). If this is the case, the multidimensional rating system would lose effectiveness. Based on a sample of over a million hotel reviews collected from Qunar.com, a leading hotel booking platform in China, only 23% reviews provide diversified ratings. In other words, the itemized ratings are exactly the same as the overall rating for the rest 77% reviews. Although the multidimensional rating system offers users optional mechanisms to provide additional information and evaluations, up to 77% users provide identical ratings after their consumption.

However, most of prior studies mainly focused on the effect of review valence, review text, emotion, and reviewer characteristics on review usefulness (e.g., Chua and Banerjee, 2015; Fang et al., 2016; Huang et al., 2020; Kwok & Xie, 2016), how rating diversity affects review usefulness within the multidimensional rating system remains unexplored. Given that multidimensional rating systems are widely deployed by hotel booking platforms, whether reviews with diversified ratings facilitate information matching between consumers and hotels is very critical for the operation of a platform and involves the interests of consumers and hotel managers. Therefore, the research purpose of this study is to examine the relationship between rating diversity and review usefulness at the review level and explore the moderating role of review valence in this relationship. Leveraging the research context of the multidimensional rating system on Qunar, we aim to achieve two research objectives. First, measure the rating diversity of each review and evaluate its influence on review usefulness. Particularly, we use the coefficient of variation among the overall rating and itemized ratings of each review to measure rating diversity. On the basis of information transfer theory, reviews with diversified ratings should be more useful in facilitating information matching within the context of multidimensional rating systems. Second, explore whether review valence moderates the effect of rating diversity on review usefulness in the framework of loss aversion, because research has shown that negative reviews are usually perceived as more useful (Park & Nicolau, 2017).

2. Literature review

2.1. Review valence and review usefulness

Using attraction reviews, restaurant reviews, and hotel reviews collected from mainstream platforms such as TripAdvisor and Yelp, extensive studies have investigated the factors that can influence the usefulness (helpfulness) of online reviews (Kwok & Xie, 2016; Park & Nicolau, 2015, 2017). A large proportion of these studies have focused on the relationship between the valence of a review and its usefulness. For example, using hotel reviews obtained from TripAdvisor, Kwok and Xie (2016) showed that the rating of a review is negatively associated with its helpfulness and user experience (e.g., number of cities visited, years of membership, etc.) is positively correlated with review helpfulness. Moreover, using restaurant reviews collected from Yelp, Park and Nicolau (2015) revealed that extreme ratings (positive or negative) are perceived as more useful and enjoyable in comparison with moderate ratings. Using hotel reviews collected from TripAdvisor, Park and Nicolau (2017) further reported that negative reviews rather than positive reviews are perceived as more useful and the ratings of specific hotel attributes differently affect review usefulness.

2.2. Review text and review usefulness

Besides of numerical ratings, the textual content offers additional descriptions and explanations of consumer experience. Research has shown that review text is associated with the usefulness of a review. For example, Cao et al. (2011) found that the semantic characteristics of online reviews are more influential than other review features in affecting review usefulness, Willemsen et al. (2011) suggested that the content characteristics of a review are crucial in understanding its usefulness, and Li et al. (2019) observed that the temporal and explanatory cues of review texts exert an influence on review usefulness. Particularly the length of a review (also known as review depth) which indicates the amount of information provided is positively correlated with review usefulness, because longer reviews usually contain more product/service related information (Chua and Banerjee, 2015; Pan and Zhang, 2011). Potential consumers who read reviews before making booking decisions tend to perceive those with adequate length and depth as more practical and useful than short reviews (Cheng and Ho, 2015; Mudambi and Schuff, 2010). Moreover, Leung (2021) found that long reviews described in literal language are usually perceived as more useful than reviews presented in figurative language.

2.3. Emotion and review usefulness

Recently, an increasing number of studies investigated how emotion affects review usefulness. Through analysis of Yelp data, Hair and Ozcan (2018) observed that review usefulness is greater among positive reviews containing (vs. lacking) profanity. Using a sample of 727 participants, Li (2019) found that a review is perceived as more useful if certain positive emotion (e.g. happiness) rather than uncertain positive emotion (e.g. hope) is expressed for both low- and high-involvement products. Moreover, Huang et al. (2020) conducted one lab experiment with restaurants and one empirical study with hotel reviews and found that positive emoticons in narrative-based reviews enhance review helpfulness and negative emoticons in list-based reviews increase review helpfulness. Li et al. (2020) further showed that positive (negative) emotional intensity present a negative (positive) influence on review usefulness. Because negative reviews are helpful to avoid potential risks, consumers often perceive negative reviews more useful than positive ones (Eslami et al., 2018). For example, Chatterjee (2020) found that negative emotion with low arousal rather than high arousal makes a hotel review more helpful.

2.4. Reviewer characteristics and review usefulness

Besides studies focusing on the effect of review valence on review usefulness, researchers also examined the influence of reviewer characteristics on review usefulness (Schindler and Bickart, 2012; Fang et al., 2016). For instance, using restaurant reviews collected from Yelp, Liu and Park (2015) found that user online reputation is positively associated with review usefulness, and Chua and Banerjee (2015) also observed that Amazon users have a proclivity for reviews posted by those with a positive track record. Moreover, using the information of attraction reviews collected from TripAdvisor, Fang et al. (2016) found that the perceived value of online reviews (represented by the number of helpful votes each review received) is correlated with text readability and author characteristics. In addition, it was also found that reviewer rank, reviewer identity disclosure, and number of followers are positively associated with review usefulness (Cheng and Ho, 2015; Lee and Choeh, 2016).

Although existing studies have explored the potential factors that may influence review usefulness from many aspects (e,g., review valence, review text, emotion, and reviewer characteristics), rare research has examined whether and how rating diversity affects review usefulness. Therefore, this study is one of the first attempts to investigate the effect of rating diversity on review usefulness at the review level, highlighting the importance of rating diversity within the multidimensional rating system. Furthermore, this study also investigates the moderating effect of review valence on the relationship between rating diversity and review usefulness as existing studies have found that review valence is highly correlated with review usefulness (Kwok & Xie, 2016).

3. Theory and hypothesis development

3.1. Multidimensional rating system and rating diversity

In this research, two concepts are involved: multidimensional rating system and rating diversity. Multidimensional rating system is presented as opposing single-dimensional rating systems, and rating diversity refers to diversified vs identical ratings within the multidimensional rating system. Figure 1 (a-b) shows an example of multidimensional rating system (for example, TripAdvior and Qunar), wherein users need to provide both the overall rating and itemized ratings (e.g., service, location, etc.). Figure 2 presents an example of a single-dimensional rating system (for example, Yelp), wherein users can only post an overall rating on their orders. Within the multidimensional rating system, Figure 1a versus Figure 1b illustrate rating diversity (a new concept proposed in this research). According to our dataset, over 70% of the whole sample are in the case that the itemized ratings are the same as the overall rating of a review (the default setting when rating an order). We thus utilize the coefficient of variation among the overall rating and itemized ratings of a review to evaluate its rating diversity.

Previous studies have examined the difference between multidimensional rating system and single-dimensional rating system (Chen et al., 2018). This study aims to go a step further and examine how rating diversity affects review usefulness within the multidimensional rating system. Therefore, in this study, multidimensional rating system is the research context and rating diversity is the focus.

Figure 1a. Multidimensional rating system (diversified ratings).

Figure 1b. Multidimensional rating system (identical ratings).

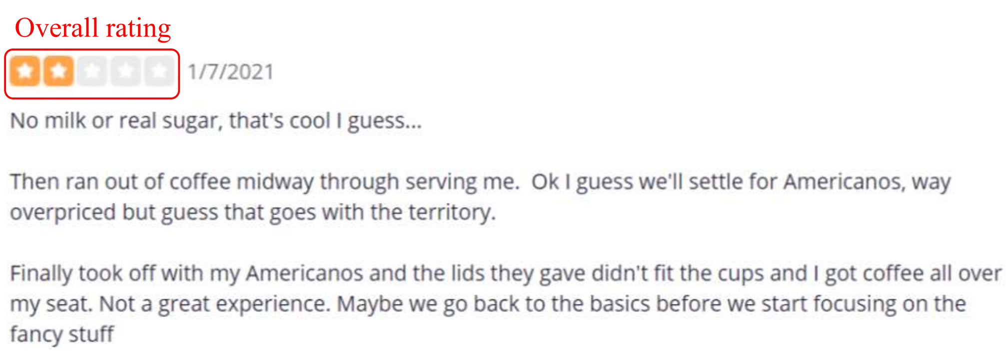

Figure 2. Single-dimensional rating system.

3.2. Information transfer theory: rating diversity and review usefulness

According to the information transfer theory (Belkin, 1984), the rating system on a platform (e.g., hotel booking platforms, review websites, etc.) functions as an information system, whereas existing ratings and reviews (especially for recently posted ones) serve as the knowledge resource. Research has shown that social networking sites (e.g., Twitter and Facebook) facilitate information seeking/exchange and personality also affects the usage of these sites (Hughes et al., 2012). Based on diverse information sources, Kim et al. (2016) compared the macro-level information transfer and found that news is the most active, SNS (social networking sites) is the most reactive, and blog is the most persistent. Moreover, Bartschat et al. (2022) investigated how potential consumers allocate their effort in searching WOM (e.g., face-to-face offline communications, product reviews on Internet opinion sites, and recommendations on social media platforms) before making booking decisions. The results suggest that offline communications and Internet opinion sites are usually consulted earlier than social media which serves as the final information sources. For the case of experience goods such as hospitality related products or services, which are difficult to evaluate before consumption (Liu & Park, 2015), potential consumers depend heavily on this knowledge resource to make informed booking decisions. Although product attributes can be described in review texts (Archak et al. 2011; Ghose et al. 2012), the multidimensional ratings are more straightforward and informative for potential consumers to evaluate whether a product/service will fit their requirements and/or preferences before making purchase decisions (Chen et al., 2018). Therefore, within the multidimensional rating system, reviews with diversified ratings are more informative for potential consumers to increase matching efficiency and decrease decision error when making booking decisions.

Moreover, multidimensional rating systems—characterized by showing not only overall ratings but also itemized ratings—enhance rating usefulness in comparison with single-dimensional rating system—which presents overall ratings only—because itemized ratings provide extra information for potential consumers to reduce uncertainty (consumers can make choices according to specific attribute ratings rather than just the overall rating of a product or service) before making booking decisions. Manes and Tchetchik (2018) found that user-generated content (e.g., rating scores and review texts) plays an important role in reducing information asymmetry of experience goods. From this perspective, diversified ratings enhance information dimension and facilitate matching of consumers to product attributes, thus reducing product/service uncertainty.

Therefore, although a platform may offer a multidimensional rating system, if reviewers post ratings for itemized attributes that have the same values as the overall rating, the potential positive effect of a multidimensional rating system in providing extra information disappears. In this case, reviewers do not make effective use of the multidimensional rating system by assessing each item independently. Diversified ratings facilitate potential consumers to look at the rating of each attribute and make their own decisions, which increase matching efficiency and reduce uncertainty; thus, reviews with diversified ratings should be perceived as more useful. Accordingly, we hypothesize that:

H1. Rating diversity enhances review usefulness.

3.3. Loss aversion phenomenon: asymmetric effects

The loss aversion characteristic of Kahneman and Tversky's (1979) Prospect Theory suggests that people weigh losses heavier than gains, thereby paying greater attention to the information that may suggest that a potential loss can happen derived from a possible action (i.e., making a reservation). Loss aversion (losses loom larger than gains) is widely accepted and applied in the research of social sciences (Gal and Rucker, 2018). Extensive studies have used the framework provided by the loss aversion phenomenon to analyze online reviews and user-generated content. For example, Kim et al. (2021) found a greater effect of negative green content on the attitude about environment-related issues as compared to environmental content that is positive; these asymmetrical effects are also found by Lee and Choi (2020) in the context of tourism attributes, by Mellinas and Nicolau (2020) in the "Internet speed – overall rating" relationship, by Tanford and Kim (2019) in the relationship between location and other attributes, by Nicolau et al. (2020) in their analysis of the halo effect, by Mehraliyev et al. (2020) in their examination of sensory experiences. Despite the evidence in support of loss aversion, inconsistent or contradictory phenomenon is not rare (e.g., Ert and Erev, 2013; Yechiam and Hochman, 2013). However, as Gal and Rucker (2018) discussed and stated "the acceptance of loss aversion as a general principle remains pervasive and persistent among social scientists".

More relevant to our case is the research that has shown that negative reviews are more useful in helping potential consumers to make booking decisions (Chatterjee, 2020; Eslami et al., 2018; Kwok & Xie, 2016; Park & Nicolau, 2017), because consumers are more concerned with reducing losses than increasing gains in the hospitality context (Chang and Cheng, 2021). For example, using hotel attribute ratings within the framework of prospect theory, Mellinas et al. (2019) found that consumers punish the hotel more harshly for dissatisfaction than praise it lavishly for satisfaction; Sharma et al. (2020) observed that negative deviations in ratings bring about a higher impact on review sentiment than positive deviations do. These empirical results are consistent with the loss aversion phenomenon. Based on the loss aversion property, as losses weigh heavier than gains, any attempt to reduce any potential uncertainty—and its asymmetric information thereof—should have a higher effect on the loss region than the gain region. If potential consumers observe positive ratings, processing that information is relatively free of concerns; however, if potential consumers see very negative ratings, it is evident that something is wrong. If the information allows potential consumers to exactly identify the specific attribute that is underperforming, the potential relief derived from the extra knowledge obtained from breaking down the assessment and showing different ratings for each attribute should increase the usefulness of the rating—regardless of whether these potential consumers eventually opt for this product/service or not. Therefore, we hypothesize that:

H2. The effect of rating diversity on review usefulness is more salient for negative reviews.

4. Data and methodology

4.1. Data collection

We collect data from Qunar, a leading hotel booking platform that covers over 730,000 hotels in 71,360 cities. To gain comprehensive and robust empirical results, we focus on the largest cities in China, i.e., Shanghai, Beijing, and Shenzhen. A Python-based crawler is developed to retrieve the information of each review posted on the hotels in these three cities. As Qunar.com only keeps the reviews posted for the latest three years, the sample period in this study is May 2017-April

2020 (the data is collected in May 2020). To avoid the sample being affected by the impact of COVID-19, the observations from 2020 are deleted.

4.2. Variables and econometric models

The dependent variable in this study is the usefulness of a review (*Useful*) and, accordingly, we use the number of usefulness votes a review received, which is a direct measure to evaluate whether a review is informative in helping consumers to make booking decisions (Fang et al., 2016; Liu & Park, 2015). The independent variable of interest is rating diversity (*Diversity*), which reflects whether the multidimensional rating system works. For example, as shown in Figure 1a, if all the itemized ratings are the same as the overall rating, the itemized ratings would not deliver extra information, thereby reducing the usefulness of a review. We thus use the coefficient of variation among the overall rating and itemized ratings of a review (CV_1) and the coefficient of variation among the itemized ratings of a review (CV_2) to comprehensively measure the rating diversity of a review. Model 1 specifies the econometric model to test H1.

$$Useful_i = \theta_0 + \theta_1 Diversity_i + Controls_i + \varepsilon_i \quad (1)$$

We also employ the valence of a review as a moderator to test its moderating effect on the relationship between review usefulness and rating diversity. In particular, we utilize the rating score (one to five) and the sentiment of a review to measure review valence. The sentiment of the review text of each review is evaluated by Baidu PaddleHub (Module: Bidirectional Long Short Term Memory), ranging from 0 (the most negative) to 1 (the most positive). And then introduce their interactions with rating diversity into Model 1 to develop Model 2 (*Rating*) and Model 3 (*Sentiment*), separately. Through the significance and sign of parameter θ_3 in Model 2 and Model 3, we can infer the moderating effect of review valence (H2).

$$Useful_i = \theta_0 + \theta_1 Diversity_i + \theta_2 Rating_i + \theta_3 Diversity_i \cdot Rating_i + Controls_i + \varepsilon_i \quad (2)$$

$$Useful_i = \theta_0 + \theta_1 Diversity_i + \theta_2 Sentiment_i + \theta_3 Diversity_i \cdot Sentiment_i + Controls_i + \varepsilon_i \quad (3)$$

To control for other factors that may influence review usefulness, we include the following variables in Models 1-3. First, the length of a review and the number of photos uploaded in a review, because research has shown that review length and the number of photos are positively associated with the usefulness of a review (Mudambi & Schuff, 2010). Second, we calculate the number of days each review has been posted to control the review date of a review. Third, we also include the hotel star fixed effect and travel type (e.g., business, family, couple, friends, etc.) fixed effect to control for differences at the business level. As the dependent variable is the number of usefulness votes each review received, a count variable with overdispersion (mean=0.126 and variance=0.678), the negative binomial regression with robust standard errors is appropriate for the estimation of Models 1-3 (Chen & Lurie, 2013). In addition, review usefulness (the number of usefulness votes each review received after published) should not be endogenous with the rating diversity (independent variable: *CV*) or review valence (moderator: rating or sentiment) which is generated before a review is published.

5. Results

5.1. Descriptive statistics

Table 1 presents the summary statistics of the main variables used in the following analysis. The dataset of Shanghai is utilized in the main analysis and the dataset of Beijing and Shenzhen are used for robustness analysis, the total number of reviews obtained from these three cities is 1,720,429. For brevity, we only summarize the dataset of Shanghai. The average number (0.126) of usefulness votes a review received is relatively small, indicating that many of the reviews have not received usefulness votes from potential consumers who read reviews for decision making. The average magnitude of rating diversity (CV_1 and CV_2) is small and the itemized ratings are exactly the same as the overall rating for over 70% of the whole sample, indicating that most users are not willing to provide diversified ratings which require extra effort and time. The average of the overall rating (sentiment) of a review is approximately 4.5 (0.8), suggesting a relatively high degree of satisfaction across these cities. In addition, the average length of the reviews collected from Shanghai is approximately 43 and the average number of photos uploaded in a review is 0.444.

Table 1. Summary statistics (mean).

City	Shanghai	Beijing	Shenzhen
Useful	0.126	0.102	0.079
Diversity (CV_1)	0.050	0.059	0.048
Diversity (CV_2)	0.048	0.057	0.046
Rating	4.506	4.393	4.453
Sentiment	0.798	0.762	0.779
Length	42.737	38.548	36.874
Photo	0.444	0.350	0.382
ReviewDate	565.881	553.249	548.695
Observations	631,356	652,346	436,727

5.2. Estimation results

The empirical results in this section are based on the dataset of Shanghai. Table 2 presents the negative binomial regression results with robust standard errors. Rating diversity is indexed by CV_1 (the coefficient of variation among the overall rating and itemized ratings of a review) in columns 1-3. The positive and significant coefficient (0.967***) of rating diversity in column 1 indicates that reviews with diversified ratings are perceived as more useful than reviews with identical ratings, thereby supporting H1 that reviews with diversified ratings are more informative and thus receive more usefulness votes from potential consumers who read reviews before making booking decisions. Columns 2-3 test the moderating effect of review valence indexed by rating score and text sentiment, respectively. The negative and significant coefficient (-0.857***) of *Diversity·Rating* in column 2 indicates that low-rating reviews with diversified ratings are perceived as more useful than high-rating reviews with diversified ratings. The significantly negative coefficient (-2.769***) of *Diversity·Sentiment* in column 3 further confirms H2. Figure 3 visualizes the moderating effect of review valence and shows that the positive effect of rating diversity on review usefulness mainly exists among negative (low-rating or low-sentiment) reviews.

Columns 4-6 report the estimation results using CV_2 (the coefficient of variation among the itemized ratings of a review) to measure rating diversity. We observe that the significance, sign, and magnitude of the coefficients of *Diversity* and its interaction with rating score or text sentiment are essentially the same as those in columns 1-3 of Table 2. These results support our hypotheses that within the context of multidimensional rating system, rating diversity enhances review usefulness and this positive effect is more salient for negative reviews with low rating score or text sentiment.

Table 2. Negative binomial regression results (Shanghai).

	Diversity: CV_1			Diversity: CV_2		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	*Useful*	*Useful*	*Useful*	*Useful*	*Useful*	*Useful*
Diversity	0.967***	1.881***	1.414***	0.879***	1.855***	1.290***
	(0.047)	(0.102)	(0.077)	(0.046)	(0.104)	(0.076)
Rating		-0.151***			-0.153***	
		(0.009)			(0.009)	
Diversity·Rating		-0.857***			-0.794***	
		(0.034)			(0.033)	
Sentiment			-0.029			-0.068**
			(0.031)			(0.030)
Diversity·Sentiment			-2.769***			-2.613***
			(0.158)			(0.144)
Length	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Photo	0.186***	0.196***	0.189***	0.185***	0.196***	0.189***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
ReviewDate	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-5.130***	-4.420***	-5.079***	-5.121***	-4.417***	-5.046***
	(0.028)	(0.052)	(0.039)	(0.028)	(0.050)	(0.038)
Hotel Star FE	YES	YES	YES	YES	YES	YES
Travel Type FE	YES	YES	YES	YES	YES	YES
Pseudo R^2	0.1319	0.1371	0.1336	0.1316	0.1369	0.1334
Observations	631,356	631,356	631,356	631,356	631,356	631,356

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

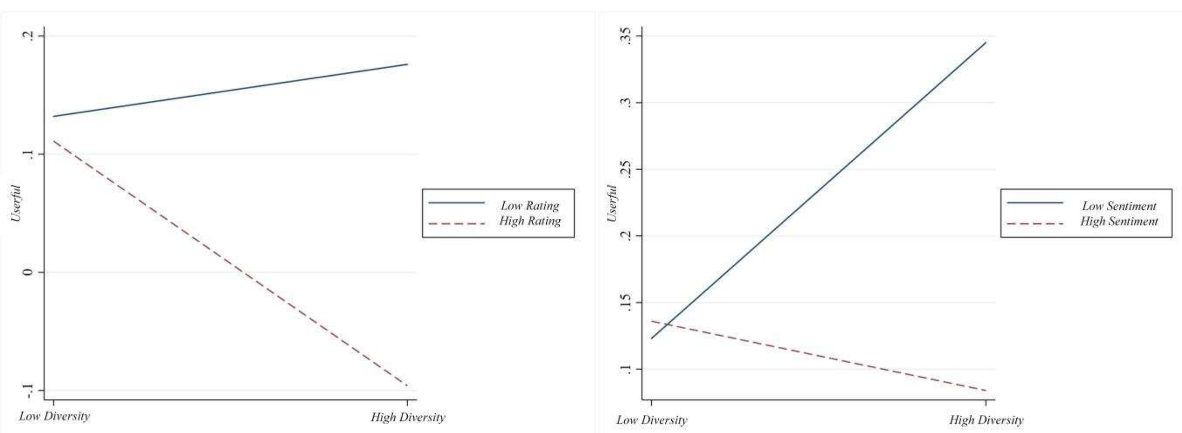

Figure 3. The moderating effect of review valence.

5.3. Robustness analysis

Using reviews posted on hotels in Shanghai, we find that reviews (especially for negative reviews with low rating score or text sentiment) with diversified ratings rather than identical ratings are more useful in the context of multidimensional rating system. To validate and generalize this finding, we collect data from Beijing and Shenzhen and replicate the results in Table 2. The estimation results are reported in Table 3 and Table 4, respectively. We observe that the coefficients of interest in Table 3 and Table 4 are essentially the same as those in Table 2, suggesting the robustness of our empirical results. That is, negative reviews with diversified ratings are perceived as more useful by potential consumers, which is in line with the argument that any action to diminish uncertainty should have a greater effect on the loss region than the gain region (the loss aversion phenomenon). In addition, we observe that review length, the number of photos uploaded in each review, and review date also exert certain effect on review usefulness. These results indicate that review usefulness can be affected by other review level factors (Fang et al., 2016; Liu et al., 2019).

Table 3. Negative binomial regression results (Beijing).

	Diversity: CV_1			*Diversity*: CV_2		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	*Useful*	*Useful*	*Useful*	*Useful*	*Useful*	*Useful*
Diversity	1.064***	1.268***	0.947***	0.948***	1.199***	0.810***
	(0.040)	(0.075)	(0.056)	(0.040)	(0.077)	(0.055)
Rating		-0.235***			-0.237***	
		(0.006)			(0.006)	
Diversity·Rating		-0.691***			-0.617***	
		(0.031)			(0.030)	
Sentiment			-0.349***			-0.385***
			(0.022)			(0.022)
Diversity·Sentiment			-2.007***			-1.872***
			(0.149)			(0.143)
Length	0.004***	0.003***	0.004***	0.004***	0.003***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Photo	0.166***	0.178***	0.174***	0.164***	0.178***	0.174***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
ReviewDate	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-5.035***	-4.000***	-4.757***	-5.023***	-3.998***	-4.726***
	(0.027)	(0.037)	(0.032)	(0.027)	(0.037)	(0.032)
Hotel Star FE	YES	YES	YES	YES	YES	YES
Travel Type FE	YES	YES	YES	YES	YES	YES
Pseudo R^2	0.1235	0.1335	0.1266	0.1230	0.1332	0.1264
Observations	652,346	652,346	652,346	652,346	652,346	652,346

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 4. Negative binomial regression results (Shenzhen).

	Diversity: CV_1			*Diversity*: CV_2		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	*Useful*	*Useful*	*Useful*	*Useful*	*Useful*	*Useful*
Diversity	1.499***	1.385***	1.344***	1.397***	1.323***	1.190***
	(0.059)	(0.119)	(0.077)	(0.061)	(0.129)	(0.077)
Rating		-0.236***			-0.240***	
		(0.008)			(0.008)	
Diversity·Rating		-0.552***			-0.495***	
		(0.067)			(0.069)	
Sentiment			-0.326***			-0.373***
			(0.030)			(0.030)
Diversity·Sentiment			-1.719***			-1.523***
			(0.276)			(0.286)
Length	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Photo	0.179***	0.192***	0.188***	0.177***	0.193***	0.188***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
ReviewDate	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-5.036***	-3.987***	-4.765***	-5.024***	-3.968***	-4.723***
	(0.041)	(0.051)	(0.046)	(0.041)	(0.050)	(0.045)
Hotel Star FE	YES	YES	YES	YES	YES	YES
Travel Type FE	YES	YES	YES	YES	YES	YES
Pseudo R^2	0.1231	0.1312	0.1254	0.1225	0.1310	0.1250
Observations	436,727	436,727	436,727	436,727	436,727	436,727

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6. Conclusion

Using 1,720,429 hotel reviews collected from a leading hotel booking platform in China, our empirical results reveal that reviews with diversified ratings are perceived to be more useful than those with identical ratings. These results are robust across different cities (Shanghai, Beijing, and Shenzhen). However, the positive effect of rating diversity on review usefulness is greater for negative (i.e., low-rating or low-sentiment) reviews; as negative elements weigh heavier than positive ones, in line with prospect theory's loss aversion, any informational clue that helps consumers make their decision is valued more when this clarifying clue refers to the negative element compared to the positive one. In addition, our results should not be influenced by endogeneity, because review usefulness (the number of usefulness votes each review received after published) cannot be endogenous with the rating diversity (independent variable: *CV*) or review valence (moderator: rating or sentiment) which is generated before a review is published.

Theoretically, this study is among the first to examine whether rating diversity enhances the usefulness of user-generated reviews and extends theories of information transfer into the analysis of users' decision making behavior, highlighting the value of rating diversity within the multidimensional rating system. Although information transfer theory is widely applied in the analysis of information seeking, information sharing, information exchange, and information diffusion, this study extends its application into information matching in the context of rating diversity and multidimensional rating systems. Specifically, while the "personalization" component claimed in information transfer theory, i.e. the one-to-one transfer of knowledge among customers, only happens through online platforms; the "codification" component, i.e. the conversion of knowledge into knowledge devices, is certainly extended through the use of multidimensional rating systems. Furthermore, the effect of rating diversity on review usefulness is not constant across the ratings, rather it exerts a greater effect on negative reviews with low-rating or low-sentiment, in line with the loss aversion phenomenon during decision making in the context of hotel booking platform. Our empirical results further confirms the loss aversion effect in the context of multidimensional rating systems on hotel booking platforms. These empirical results not only are in line with the idea that consumers punish the hotel more harshly for dissatisfaction than praise it lavishly for satisfaction (Mellinas et al., 2019) and that negative deviations in ratings bring about a higher impact on review sentiment than positive deviations do (Sharma et al., 2020), but our outcomes go well beyond these expected results and make a significant advancement from a theoretical viewpoint: if consumers are able to exactly identify the specific attributes that are underperforming (and remember that if potential consumers see very negative ratings, it is evident that something is wrong), the potential relief derived from the extra knowledge obtained from breaking down the assessment and showing different ratings for each attribute increases the usefulness of the rating.

The original intention of multidimensional rating systems is to encourage consumers to rate a product or service from multidimensional attributes, because ratings from multidimensional attributes of a product or service facilitate information matching and screening which are essential for potential consumers to make informed booking decisions. However, according to our statistics, the majority of consumers post identical ratings which makes multidimensional rating systems out of operation. The main reason leading to such a situation may be the default setting that the itemized ratings are automatically set to the value of the overall rating when consumers rate a hotel. Although this default setting makes it easier for consumers to provide itemized ratings, it also induces consumers to provide identical ratings. We thus suggest these online booking platforms and review sites cut away this default setting or forbid the submission of identical ratings when posting reviews. In this regard, platforms can provide incentives to encourage customers to post diversified ratings; for example, as there are platforms that reward the reviewers (for instance, via badges) based on the quality of the review they post, these platforms can include diversified ratings as a criterion to measure the quality of a review and then, grant the rewards accordingly. This is critical to prompt consumers to post diversified ratings which are essential for potential consumers to make informed booking decisions on a platform. From the perspective of hotel managers, they can timely monitor the performance of each attribute of their hotels to locate the weakness needing improvement. Moreover, managers may want to pay special attention to those reviews that have diversified ratings because these ratings have an increased impact on customers' decisions, particularly, when it comes to diversified ratings with low-rating or low-sentiment. From the perspective of consumer, if a customers is especially sensitive to an attribute (e.g., location), they can keep off hotels with low ratings on this attribute –that is against their preferences or requirements. Additionally, given that diversified ratings are found to be more useful, customers seeking information might first look at the diversity of the rating as it would be indicative of reliability with which the reviewer posted the ratings and wrote the review for that matter.

This study also has limitations which deserve future research effort. First, the usefulness votes are giving prior to consumption, thus, their effectiveness in forming booking decisions is limited or inconclusive. That is, although some reviews received many usefulness votes from potential consumers, whether these consumers made booking decisions after voting these reviews is uncertain. An ideal effectiveness measure of rating diversity will be whether the chosen hotel, which was based on certain ratings, indeed matched the expectation these ratings created. Second, the empirical results of this study are based on one popular hotel booking platform that deployed a multidimensional rating system, thus, results obtained from other popular platforms like

TripAdvivor would help to better reveal the general effect of rating diversity on review usefulness. Third, this study only uses the number of usefulness votes each review received to measure the usefulness of each review; consequently, future research can conduct topic analysis (e.g., LDA) to examine whether reviews with diversified rating contain more topics which are also important for potential consumers to make booking decisions. Fourth, review usefulness can also be affected by other factors, such as the readability of each review and the preference of readers. Future research may consider these issues via experimental designs.

References

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.

Baidu PaddleHub (2020). https://aistudio.baidu.com/aistudio/projectdetail/215814.

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.

Bartschat, M., Cziehso, G., & Hennig-Thurau, T. (2022). Searching for word of mouth in the digital age: Determinants of consumers' uses of face-to-face information, internet opinion sites, and social media. Journal of Business Research, 141, 393-409.

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

Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. Decision Support Systems, 50(2), 511-521.

Chang, K. C., & Cheng, Y. S. (2021). How online service recovery reviews influence behavioral intentions in the hospitality context: Regulatory focus and loss aversion perspectives. Journal of Hospitality and Tourism Management, 46, 440-455.

Chatterjee, S. (2020). Drivers of helpfulness of online hotel reviews: A sentiment and emotion mining approach. International Journal of Hospitality Management, 85, 102356.

Chen, P. Y., Hong, Y., & Liu, Y. (2018). The value of multidimensional rating systems: Evidence from a natural experiment and randomized experiments. Management Science, 64(10), 4629-4647.

Chen, Z., & Lurie, N. H. (2013). Temporal contiguity and negativity bias in the impact of online word of mouth. Journal of Marketing Research, 50(4), 463-476.

Cheng, Y. H., & Ho, H. Y. (2015). Social influence's impact on reader perceptions of online reviews. Journal of Business Research, 68(4), 883-887.

Chua, A. Y., & Banerjee, S. (2015). Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. Journal of the Association for Information Science and Technology, 66(2), 354-362.

De Langhe, B., Fernbach, P. M., & Lichtenstein, D. R. (2016). Navigating by the stars: Investigating the actual and perceived validity of online user ratings. Journal of Consumer Research, 42(6), 817-833.

Ert, E., & Erev, I. (2013). On the descriptive value of loss aversion in decisions under risk: Six clarifications. Judgment and Decision Making, 8(3), 214-235.

Eslami, S. P., Ghasemaghaei, M., & Hassanein, K. (2018). Which online reviews do consumers find most helpful? A multi-method investigation. Decision Support Systems, 113, 32-42.

Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. Tourism Management, 52, 498-506.

Gal, D., & Rucker, D. D. (2018). The loss of loss aversion: Will it loom larger than its gain?. Journal of Consumer Psychology, 28(3), 497-516.

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.

Hair, M., & Ozcan, T. (2018). How reviewers' use of profanity affects perceived usefulness of online reviews. Marketing Letters, 29(2), 151-163.

Huang, G. H., Chang, C. T., Bilgihan, A., & Okumus, F. (2020). Helpful or harmful? A double-edged sword of emoticons in online review helpfulness. Tourism Management, 81, 104135.

Hughes, D. J., Rowe, M., Batey, M., & Lee, A. (2012). A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. Computers in Human Behavior, 28(2), 561-569.

Li, Y. (2019). Consumers' perceived usefulness of online reviews: Effects of emotional certainty and product involvement. Social Behavior and Personality, 47(9), 1-16.

Liu, X., Law, R., & Xu, Y. (2019). Does identity disclosure affect review extremity? Annals of Tourism Research, 77, 171-174.

Kim, E. J., Tanford, S., & Book, L. A. (2021). The effect of priming and customer reviews on sustainable travel behaviors. Journal of Travel Research, 60(1), 86-101.

Kim, M., Newth, D., & Christen, P. (2016). Macro-level information transfer in social media: Reflections of crowd phenomena. Neurocomputing, 172, 84-99.

Kwok, L., & Xie, K. L. (2016). Factors contributing to the helpfulness of online hotel reviews: Does manager response play a role? International Journal of Contemporary Hospitality Management, 28(10), 2156-2177.

Lee, J. S., & Choi, M. (2020). Examining the asymmetric effect of multi-shopping tourism attributes on overall shopping destination satisfaction. Journal of Travel Research, 59(2), 295-314.

Lee, S., & Choeh, J. Y. (2016). The determinants of helpfulness of online reviews. Behaviour & Information Technology, 35(10), 853-863.

Leung, D. (2021). Unraveling the interplay of review depth, review breadth, and review language style on review usefulness and review adoption. International Journal of Hospitality Management, 97, 102989.

Li, H., Liu, H., & Zhang, Z. (2020). Online persuasion of review emotional intensity: A text mining analysis of restaurant reviews. International Journal of Hospitality Management, 89, 102558.

Li, H., Wang, C. R., Meng, F., & Zhang, Z. (2019). Making restaurant reviews useful and/or enjoyable? The impacts of temporal, explanatory, and sensory cues. International Journal of Hospitality Management, 83, 257-265.

Liu, X., Law, R., & Xu, Y. (2019). Does identity disclosure affect review extremity? Annals of Tourism Research, 77, 171-174.

Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. Tourism Management, 47, 140-151.

Manes, E., & Tchetchik, A. (2018). The role of electronic word of mouth in reducing information asymmetry: An empirical investigation of online hotel booking. Journal of Business Research, 85, 185-196.

Mehraliyev, F., Kirilenko, A. P., & Choi, Y. (2020). From measurement scale to sentiment scale: Examining the effect of sensory experiences on online review rating behavior. Tourism Management, 79, 104096.

Mellinas, J. P., & Nicolau, J. L. (2020). Let's hook up fast! Hotel reviews and Wi-Fi flaws. Annals of Tourism Research, 80, 102842.

Mellinas, J. P., Nicolau, J. L., & Park, S. (2019). Inconsistent behavior in online consumer reviews: The effects of hotel attribute ratings on location. Tourism Management, 71, 421-427.

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.

Nicolau, J. L., Mellinas, J. P., & Martín-Fuentes, E. (2020). The halo effect: A longitudinal approach. Annals of Tourism Research, 83, 102938.

Pan, Y., & Zhang, J. Q. (2011). Born unequal: A study of the helpfulness of user-generated product reviews. Journal of Retailing, 87(4), 598-612.

Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. Annals of Tourism Research, 50, 67-83.

Park, S., & Nicolau, J. L. (2017). Effects of general and particular online hotel ratings. Annals of Tourism Research, 62, 114-116.

Park, S., & Nicolau, J. L. (2019). Image effect on customer-centric measures of performance. Annals of Tourism Research, 76, 226-238.

Schindler, R. M., & Bickart, B. (2012). Perceived helpfulness of online consumer reviews: The role of message content and style. Journal of Consumer Behaviour, 11(3), 234-243.

Sharma, A., Park, S., & Nicolau, J. L. (2020). Testing loss aversion and diminishing sensitivity in review sentiment. Tourism Management, 77, 104020.

Tanford, S., & Kim, E. L. (2019). Risk versus reward: when will travelers go the distance? Journal of Travel Research, 58(5), 745-759.

Willemsen, L. M., Neijens, P. C., Bronner, F., & De Ridder, J. A. (2011). "Highly recommended!" The content characteristics and perceived usefulness of online consumer reviews. Journal of Computer-Mediated Communication, 17(1), 19-38.

Xu, Y., Zhang, Z., Nicolau, J. L., & Liu, X. (2020). How do hotel managers react to rating fluctuation? International Journal of Hospitality Management, 89, 102563.

Yechiam, E., & Hochman, G. (2013). Losses as modulators of attention: Review and analysis of the unique effects of losses over gains. Psychological Bulletin, 139(2), 497-518.