Decision Support Systems 95 (2017) 1–11

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews

Baojun Gao, Nan Hu, Indranil Bose *

ARTICLE INFO

Article history:
Received 23 April 2016
Received in revised form 18 November 2016
Accepted 20 November 2016
Available online 25 November 2016

Keywords:
Consistency in rating
Helpful votes
Online hotel reviews
Rating bias
Rating difference
Tripadvisor

ABSTRACT

This study investigates if reviewers' pattern of rating is consistent over time and predictable. Two interesting results emerge from the econometric analyses using publicly available data from TripAdvisor.com. First, reviewers' rating behavior is consistent over time and across products. Furthermore, most of the variation in their future rating behavior can be explained by their rating behavior in the past rather than by the observed average rating. Second, reviews by reviewers with higher absolute bias in rating in the past receive more helpful votes in future. We further divide the bias in rating into intrinsic bias (driven by intrinsic reviewer characteristics) and extrinsic bias (driven by influences beyond intrinsic bias) and document that intrinsic bias plays a more significant role in influencing helpful votes for reviews than extrinsic bias. Our results are robust to different product categories and different definition of bias. Overall our results indicate that in the online review context, the observed average rating or an attention grabbing strategy may not be as important as believed in the past. This study provides insights into reviewers' rating behavior and prescribes actionable items for online vendors so that they can proactively influence online opinion instead of passively responding to them.

1. Introduction

During the past few years, online reviews generated by consumers have attracted much attention from both academia and industry. Online reviews play an important role in helping consumers make purchasing decisions [13,14,23,39], driving product sales [1,5–12,17,25,26,35,42], influencing firm equity value [28–30,43,50] and competition in the market [24].

Given the importance of online reviews, academic scholars and industrial practitioners are looking for the determinants of online ratings. Till now literature shows two schools of thought on this and these are the observed average rating (i.e., other consumers' average rating the focal reviewer observed when he/she posted a review) and attention grabbing. The former view based on social conformity theory suggests that the subsequent reviews will be affected by the average rating of prior reviews and conform to the social norm [31,40] because people build their own opinion on the basis of the groups' consensus and other consumers' average rating is a source of social influence on the focal reviewer's online product rating [40]. On the other hand, the attention grabbing point of view suggests that online reviewers strategically decide whether to differentiate from the current average rating based on their online reputation with the goal of gaining attention. The reviewers are more likely to deviate from social norm (the average rating) for popular products as well products with a higher number of preexisting reviews [38] because by doing so they are more likely to gain attention for the reviews they write. For both theories, it implies that an online reviewer's behavior may be largely driven by external factors and may not be consistent over time or across products.

Common wisdom states that human behavior is consistent over time. However, it seems that empirical results documented in the Word of Mouth (WOM) literature so far show inconsistency. Reviewers' intrinsic characteristics, both observable and unobservable, such as cultural values, personality, prior experience, geographic mobility, social connectedness and gender, are usually time-invariant and have significant impact on their ratings [19,31]. We believe that consumers' rating behavior, instead of being determined by the observed average rating or attention grabbing as documented in extant research, may be consistent over time and across products, and mainly driven by a reviewer's past behavior. Hence, we predict that on an average, a reviewer's future behavior is predictable and mainly determined by their past behavior and only partially driven by the observed average rating and attention grabbing.

To achieve the above goal, we retrieve the full rating history of reviewer *i* as well as the observed average ratings for product *j* when reviewer *i* posted the rating. Based on historical data, we estimate the average bias (formal definition is given below) for reviewer *i*. If reviewer behavior is consistent, then his/her bias in rating in the past can be used to predict his/her future review behavior.

* Corresponding author.
E-mail addresses: gaobj@whu.edu.cn (B. Gao), Nhu4@stevens.edu (N. Hu), bose@iimcal.ac.in (I. Bose).

http://dx.doi.org/10.1016/j.dss.2016.11.005
0167-9236/

Following previous literature [19], we define bias in rating (*Abs_Bias*) in terms of absolute value of difference in rating as follows:

$$\mathrm{Abs\_Bias_i} = \frac{\sum_{j=1}^{N_i} \mathrm{abs}(\mathrm{Rating\_difference_{ij}})}{N_i} = \frac{\sum_{j=1}^{N_i} \mathrm{abs}\left(\mathrm{Rating_{ij}} - \mathrm{Obs\_avg\_rating_j}\right)}{N_i} \quad (1)$$

where Abs_Bias_i (absolute bias in rating) is a proxy for the i^{th} reviewer's behavioral bias in rating. It measures, on an average, to what extent a reviewer's rating is different from public opinion, regardless of whether it is positive or negative. $Rating_{ij}$ is the rating given by reviewer i for product j while $Obs\_avg\_rating_j$ is the average rating of product j observed by reviewer i when he/she posted the review. $Rating\_difference_{ij}$ is the difference between $Rating_{ij}$ and $Obs\_avg\_rating_j$ and this can be either positive or negative. N_i is the number of ratings provided by reviewer i. For a robustness check, we also define bias based on the signed rating difference (as shown in Eq. 2).

$$\mathrm{Bias_i} = \frac{\sum_{j=1}^{N_i} \mathrm{Rating\_difference_{ij}}}{N_i} = \frac{\sum_{j=1}^{N_i} \left(\mathrm{Rating_{ij}} - \mathrm{Obs\_avg\_rating_j}\right)}{N_i} \quad (2)$$

Different from *Abs_Bias*, *Bias* can be either positive or negative. It measures, on an average, whether a reviewer tends to rate products higher or lower than public opinion and the extent of that difference.

Our first research question is to investigate whether bias in reviewer rating is consistent over time and across products, and to what an extent online rating behavior is determined by other reviewers' average rating versus the focal reviewer's intrinsic characteristics. If reviewers' rating behavior is mainly determined by their intrinsic characteristics and are consistent over time and across products, then reviewers with positive (negative) *Bias*/or high *Abs_Bias* in the past will probably rate products higher (lower) than the average rating or will deviate more from public opinion in future. However, if the bias in reviewer rating is just random, then there will be no significant relationship between the bias in rating in past and the rating in future.

Because review helpfulness has important managerial implications for online vendors and consumers, following the online review literature [19,38], our next research question is to study the impact of past rating bias on helpfulness of future reviews. Literature has shown that extreme reviews imply that a reviewer is trying to differentiate himself/herself from others [19] and is generally more able to attract attention [38]. As such, extreme reviews can get more helpful votes, [3,19]. However, those studies cannot prescribe actionable items to online vendors beforehand because online vendors have no idea whether one customer will write an extreme review until he or she finishes writing and posting it. At that point in time, it is too late. But, if we can prove that behavioral bias is consistent, then reviewers with greater absolute rating bias in the past will be more likely to give extreme reviews in the future, which will receive more helpful votes.

To answer the above research questions, we use a unique data set composed of hotel review data as well as the related reviewer information for New York City hotels in 2013[1]. Then for each reviewer, we retrieve his/her full rating history (hotels, restaurants, and attractions) before 2013, as well as the ratings by all the other customers before 2013 for the same hotel, restaurant or attraction. The data sets are used to construct our behavioral bias of each reviewer. We use OLS and ordered logistic regression models to identify the effect of reviewers' rating bias on their future ratings, and use a negative binominal model for the effect of reviewers' rating bias on perceived review value (helpful votes) of their future reviews.

Two interesting results emerge from the econometric analyses. First, the results indicate that reviewers' rating behavior is consistent over time and across products and their (absolute) bias in rating in the past is positively correlated with the deviation or difference in their future product ratings from public opinion. Furthermore, the future rating behavior of the reviewers can be explained better by their past rating behavior than by the observed average rating of products as in [31,40]. For example, reviewers' bias in rating in the past can capture 16.93% of the variation in the difference of their future rating from public opinion, whereas the observed average rating can explain only 0.38% of the variation.

Second, we find reviews by reviewers with greater absolute bias in rating in the past receive more helpful votes in the future. Furthermore, we divide the future bias in rating into intrinsic bias (driven by intrinsic reviewer characteristics) and extrinsic bias (driven by influences beyond intrinsic bias) and document that intrinsic bias plays a more significant role in influencing helpfulness of reviews than extrinsic bias. Our results indicate that in the context of online reviews, the observed average rating may not be as influential as believed in the past, and different reviews written by the same reviewers but for different products follow the same pattern over time. We prescribe actionable items to online vendors so that they can proactively influence online opinions instead of passively responding to online reviews.

The remainder of the paper is structured as follows. In section 2, we survey the extant literature and propose testable hypotheses for subsequent empirical investigation. Section 3 provides the description of the data and the proposed methodology. We report results of our empirical analyses and discuss the implications of the results in section 4. Finally, section 5 concludes the paper.

2. Theory and hypotheses

2.1. Behavioral bias of rating and its effects on future ratings

Previous studies have investigated whether a consumer's online product rating is prone to the observed average rating [31,40] and attention grabbling [38]. For example, people experience conformity pressures from their peers in a social group [40] and future reviewers are likely to be swayed positively by prior reviewers' rating [31,40]. However, Shen et al. [38] propose that attention is a scarce resource in the information-rich context, and to compete for attention, a reviewer will tend to post a differentiated rating to distinguish it from the average rating. Based on the data of Amazon books, results suggest that the more differentiated the rating, the more attention (measured in terms of total number of votes) a review can gain.

However, as learnt through common wisdom, a human being's behavior is generally consistent over a short period of time. Hence, there is a disconnect between the traditional literature and what have been documented in the literature on online reviews. We argue that customers' intrinsic characteristics, such as cultural background or country of origin, demographic, and personality traits, are normally stable and act as fundamental factors influencing their ratings or evaluations for products or services. In fact, both social influence theory and attention grabbing theory do not rule out the influence of reviewer's intrinsic characteristics on online reviews. For example, Sridhar and Srinivasan [40] have found that other consumers' online ratings (social influence) moderate the effects of both positive and negative features of product experience, product failure, and product recovery on a reviewer's product rating. Ma et al. [31] have revealed that reviewer characteristics (i.e., prior experience, geographic mobility, social connectedness, and gender) and review characteristics (length and time between reviews) are significant moderators of the relationship between prior reviews and subsequent ones. It is very important to understand the ultimate driving force of online reviews

[1] To make our results more likely to be generalizable, we choose New York City because it is the biggest city in the US with the highest number of hotels, restaurants, reviewers, reviews, and variety of travelers.

since some of these intrinsic characteristics (e.g., personality traits) are unobservable and hard to control for researchers. We elaborate on these traits in more detail below.

2.1.1. Culture

Many studies have concentrated on the impact of national culture on customers' perceptions of service quality and have found customers from different culture evaluate the service [4,19,32,33,44,45]. Most studies along this line have used Hofstede's six dimensions of culture [21] to measure cultural differences. The explanation of this effect has been that people born and raised with different cultural values inherit such values as their inherent traits, which in turn govern their attitudes and behavior [21].

Mattila [32,33] has noted that the most important cultural dimension affecting customer evaluations of hotel service is power distance. Mattila [33] has discovered that the mean evaluations of the service provider's performance were significantly higher for Western customers than for their Asian counterparts. Hong and Li [19] have examined the effect of one salient dimension of cultural values (i.e., individualism) on consumers' online WOM. They have found that compared to consumers from collectivistic cultures, consumers from individualistic cultures are more likely to give a rating that deviates from the average of prior consumers' ratings.

2.1.2. Gender

The impact of gender on evaluation of service or product has been investigated. Based on empirical evidence suggesting that men and women differ in their information processing styles [34], Mattila [33] has hypothesized that women provide more negative evaluation of a service encounter than men. However, no significant gender differences are found in that study. Ma et al. [31] have found that women are more likely to rate a product or service lower.

2.1.3. Personality

Consumer personality is another important factor affecting the level of consumer satisfaction and evaluation [15,22,46]. A consumer's personality reflects their distinctive way of perceiving the world and, therefore, consumers may have different preferences and attitudes towards the same services or products. This includes different evaluative criteria that each type of personality may use, expressed as different individual values, motives, emotions, and appraisals of experiences [15]. Gountas and Gountas [15] have found a direct relationship between consumer personality orientation and self-reported satisfaction of the service experience. Among the Big Five Personality Factors, Jani and Han [22] have indicated that extraversion, agreeableness, and neuroticism significantly affect customer satisfaction.

The literature above has indicated that consumers with certain cultural values, gender, and personality traits tend to rate products or services higher (lower) than others or tend to deviate from public opinion. As these characteristics are inborn and not prone to change, hence, we expect consumers' rating behavior to be consistent over time. For example, consumers who on an average deviate more from the social norm by rating higher (or lower) in the past will also rate higher (or lower) in the future. Such behavioral bias is influenced by consumers' intrinsic characteristics and, thus may not change significantly in the future. Hence, we propose hypothesis 1.

Hypothesis 1. *(H1): Reviewers' behavioral bias in rating which reflects his/her rating behavior in the past is positively related to his/her rating behavior in future.*

2.2. The effects of behavioral bias in rating on helpful votes

To draw policy implication following prior literature, we also study the impact of the past rating behavior on the helpfulness of future reviews. Up to now, papers related to determinants of online rating study the determining factors as well as the consequence (the helpfulness) of online reviews. However, till now these studies focus on linking the contemporary rating behavior to the contemporary consequence, such as rating helpfulness; while we are interested in using the reviewer's rating behavior in the past for predicting the helpfulness of the new reviews that he/she will post.

Table 1 gives a summary of this line of research. The characteristics of the review are the most frequently used determinants of helpfulness of review or helpful votes. This includes both numeric variables, such as valence, volume and variance of ratings, length of review, and textual features, such as readability [27,49], emotions [2,19,47], and stylistic and semantic characteristics [3]. Second, variables related to the reviewers have also appeared in some studies and these have included identity disclosure [11,12], experience, expertise, and reputation of reviewers [27,37], among others.

We notice that the dependent variables of these studies can be grouped into two categories. Some studies [36,47,48] have considered the ratio of the number of helpful votes to the total number of votes as the dependent variable and have defined that as *helpfulness*. These studies have used data from Amazon or other websites where both the options of *helpful* and *unhelpful* are provided. Unfortunately, for many other websites, such as Tripadvisor and Yelp, only the *helpful* option is provided. Thus, one cannot use this ratio to measure the value or quality of reviews. Instead, the number of helpful votes is used to measure the value of review [37,38].

Many studies conducted at the review level have studied whether reviews with extreme ratings or neutral ratings are more helpful. For example, Mudambi and Schuff [36] have found that for search products, extreme ratings are more helpful. Using the number of helpful votes as the dependent variable, most studies have supported that extreme reviews get more helpful votes [3,19]. The reason for the above results can be explained by the following logic. A review that deviates more from the average rating may reflect the reviewer's particular experience about a product/service, deliver obvious and decisive opinion, and hence is assessed as more informative by readers of the review [11,48]. Furthermore, for experiential service like stay in a hotel, consumers generally want to hear different voices towards the same hotel, so that they can assess whether their preferences can be matched with what the hotel has to offer. A rating with a deviation (either positive or negative) presents a different voice and is therefore perceived as having a higher value [19].

Extant research uncovers another important observation. For example, Shen et al. [38] have found that when a strategic reviewer posts a differentiated rating to distinguish it from the social normal, that review will gain more attention (measured by total votes) but at the same time, the same review will receive more unhelpful votes [38]. Yin et al. [48] have shown that deviation in rating has a negative effect on the perceived helpfulness of the review (the ratio of helpful votes to total votes) due to the confirmation bias, which says that humans prefer information that confirms their initial beliefs [41,51].

The inconsistent results presented in past research may be explained in terms of differences in data sources as well as the differences in the use of dependent variables. Our data source is based on TripAdvisor, which provides only *helpfulness* as the option. Literature using this data source and using the number of helpful votes as the dependent variable have tended to conclude that reviews with extreme ratings receive more helpful votes. Table 1 provides additional details. As mentioned above, a reviewer's behavioral bias is a measurement of the difference between their past ratings and the products' average ratings. If this behavioral bias is consistent, reviewers with a greater absolute bias in rating in the past will be more likely to give extreme ratings in future. Thus, reviews by these reviewers will receive more helpful votes in future. Hence, we propose the following hypothesis.

Hypothesis 2. *(H2): Future reviews posted by a reviewer with a greater absolute rating bias in the past will receive more helpful votes.*

Table 1
Prior studies on consumer perceptions about helpfulness of reviews.

References	Data source	Methods	Dependant variables	Review characteristics			Reviewer characteristics	
				(1)	(2)	(3)	(4)	(5)
Mudambi and Schuff [36]	Amazon	Tobit	Ratio	Yes/no	Yes			
Ghose et al. [12]	Amazon	Two stage least-squares	Ratio		Yes	Yes	Yes	Yes
Baek et al. [2]	Amazon Electronics	Hierarchical regression	Ratio		Yes	Yes		
Yin, Bond and Zhang [47]	Yahoo! Merchant Reviews	Tobit	Ratio	Yes	Yes	Yes		
Yin et al. [48]	Apple's App Store	Mixed effects logistic	Ratio	No	Yes	Yes		
Cao et al. [3]	CNET Download	Ordinal logistic	Votes	Yes	Yes	Yes		
Shen et al. [38]	Amazon	Negative binomial	Votes				Yes	
Hong and Li [19]	TripAdvisor restaurant	OLS	Votes	Yes	Yes	Yes		
Yin and Wei et al. [49]	Yelp restaurants	Negative binomial	Votes	Yes	Yes			Yes
Racherla and Friske [37]	Yelp	OLS	Votes	Yes				Yes
Liu and Park [27]	Yelp restaurants	OLS	Votes	Yes	Yes	Yes	Yes	Yes

Notes:
*Ratio represents the ratio of the number of helpful votes to the total votes.
*Column (1) refers to whether the results support that reviews with extreme rating receive more helpful votes. All studies use rating as a dependent variable. Column (2) indicates if the model includes length of review (number of words) as a dependent variable. Column (3) indicates if the model includes text-related aspects of reviews, such as positive or negative words, emotion or sentiment, readability, stylistic and semantic characteristics. Column (4) indicates if the model includes identity disclosure of the reviewers. Column (5) indicates if the model includes reviewers' experience, expertise, or reputation.

3. Research methodology

In this section we describe how we retrieve the data, select the samples and conduct econometric analysis. Moreover, we provide detailed information about the variables involved in the models, including the meaning of these variables and the logic behind their inclusion in the analysis.

3.1. Data

We combined several data sets from TripAdvisor.com for the empirical study. As shown in Fig. 1, the data preparation included five main steps.

As the first step we downloaded all New York City hotel reviews, including review level data, reviewer level data and hotel level data. The review level data included the numeric rating (1–5 points), text comments, date of posting for the review, and number of helpful votes received by the review. The reviewer level data contained the reviewer's ID and name, age and gender (if provided), location (described in open text, if disclosed), number of cities he or she has visited (reflecting the reviewer's offline travel experience), and the reviewer's contribution in TripAdvisor.com measured by the total number of reviews, ratings, and helpful votes (reflecting the reviewer's online experience). We also downloaded the basic hotel level information, including the hotel ID and name, star classification, average rating, and number of reviews. This comprises data set 1.

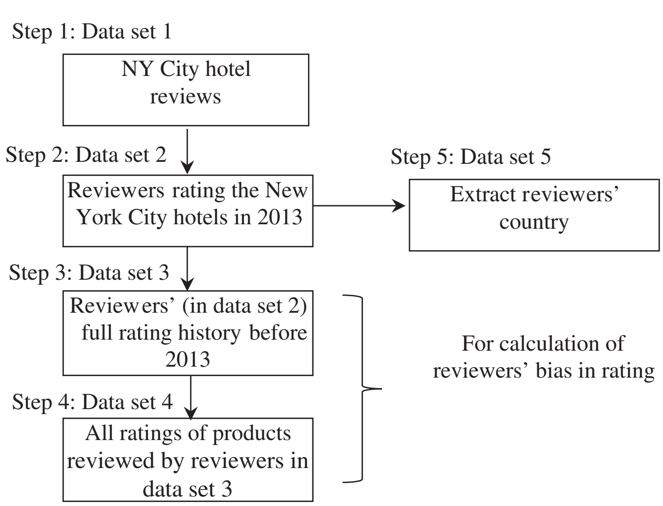


Fig. 1. Process for data preparation.

To estimate the reviewers' bias in rating, we first split the review data into two parts: the review data before 2013 and the review data from 2013. The data before 2013 is used only for calculating the reviewers' past bias in rating; while the data from 2013 onwards is used to validate the hypotheses.

As the second step we extracted the distinct reviewers who posted reviews on New York City hotels in 2013 and this resulted in 39,941 distinct reviewers. To avoid the problems produced by small samples and to get more reliable statistics for reviewers' bias in rating, we excluded reviewers whose reviews and ratings were less than 30 and only 8676 reviewers are finally considered.

The third and the fourth steps aimed to build the data set to calculate the bias in rating of each reviewer picked up in the second step. In order to calculate these reviewers' bias in rating, we traced the full rating history of each reviewer (third step) and the full rating history of each product that has been rated by them (fourth step). To be more specific, we retrieved the full rating history for each of the 8676 reviewers. For each review, we retrieved the name of reviewer, rating, date of review, and the link to the reviewed product that was a hotel, restaurant, park, or destination. This resulted in data set 3 which included 957,000 reviews.

As the fourth step, we downloaded all ratings for all products that appeared in data set 3. This is used to calculate the observed average rating of the product. This is data set 4 and it included 72 million ratings for 301,000 products.

Based on data sets 3 and 4, we calculated the reviewers' behavioral bias according to Eqs. (1) and (2). Since extant literature has shown that reviewers' cultural values [16] affect their ratings [19], and the cultural values are associated with the country, we also extracted the reviewers' country of origin to control for this effect[2], as the last step of data collection.

[2] On TripAdvisor, some reviewers disclosed their location in open text rather than through standard choices. Therefore, we first built a database for the country name and their abbreviations, the main city and districts in each country. For non-English countries, we listed both the English names and names in their native language. Then, we developed a program to extract reviewers' country of origin from the location text by matching the location text with the database. To validate that our method worked well, we randomly chose 1000 reviewers, and requested two graduate students to identify their country, and then compared that with the derived country names. The comparison showed that more than 94% of the country names derived using the automated algorithm are same as those identified by the students. This implied that the automated algorithm suitably identified the country of the reviewer.

3.2. Research variables

3.2.1. Dependent variables and variables of interest

The aim of this paper is to study whether reviewers' rating behavior is consistent over time and to what extent such behavior can be used to predict the helpfulness of their future reviews. As such, the research variables of interest in this study are reviewers' behavioral bias in rating as defined in Eqs. (1) and (2) through the variables *Abs_Bias* and *Bias*.

H1**.** states that reviewers' past rating behavior is positively correlated with their future ratings. To validate this hypothesis, we used different models and the dependent variables in these models are rating deviation, rating difference and rating, respectively. Rating deviation is measured as the absolute difference between a reviewer's rating and the average rating of the product that the reviewer observed when he/she posted the review [19].

H2**.** states that the future reviews by reviewers with greater absolute bias in rating in the past will receive more helpful votes. The dependent variable for H2 is the number of helpful votes that a review received. Unlike Amazon where customers can vote whether a review is helpful or unhelpful, TripAdvisor provides only the *helpful* option. Following previous literature [3,19], we used the total number of helpful votes as a measure of review value.

3.2.2. Control variables

Control variables are included to account for the reviewer specific, review specific, and hotel specific effects. The reviewer specific effect is captured by the variables age, gender, identity disclosure, country of origin, contributions, and number of cities visited. Review level characteristics are represented by the following: rating, observed average rating, length of review, readability of review, and the age of review. We control the hotel fixed effect to capture the impact of hotel level variation in the empirical model. More details about the control variables are provided below.

3.2.3. Identity disclosure, gender, age and reviewers' country

On TripAdvisor reviewers can choose to disclose their gender, age, and location. *No_Identity_Disc* is a dummy variable which equals one if a reviewer does not disclose both their age and gender, and zero otherwise. We coded the age of reviewers into three groups: *Young_Age* for reviewers younger than 35, *Mid_Age* for those between 35 to 49 and *Old_Age* for those older than 50. Gender effect is captured by the dummy viable *Women* which equals one if the reviewer is female. Hong and Li [19] have indicated that the cultural values of reviewers affect their rating behavior and different countries have different cultural values. To capture the effect of cultural values and other country level variations, we controlled the fixed effect of the reviewers' country of origin in the empirical model.

3.2.4. Reviewers' experience: cities visited and contribution

We captured two kinds of reviewer experience: online and offline. The offline experience of a reviewer is defined as the number of cities he/she has visited. Offline travel experience may affect the reviewer's rating behavior. If a person has a lot experience of travelling, he/she will be more adaptable to different cultures, which means he/she may be receptive of those things that exceed his/her expectations. Less experienced travelers, however, may not be able to tolerate all of the potential problems he/she may face during travel [19]. The online experience of a reviewer is measured by *Contributions* on TripAdvisor, which equals the total number of reviews, ratings, forum posts, and photos a reviewer has posted on TripAdvisor. This variable reflects the participation of a reviewer on TripAdvisor.com.

3.3. Review level control variables

3.3.1. Observed average rating

Observed average rating is the hotel's average rating at the time just before a reviewer submits his/her own rating. As such observed average rating is not constant. TripAdvisor adopts a half-star rating system and therefore the observed average rating on the website is a value within the set (1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, and 5). To calculate this variable, we first calculated the average rating according to the full rating history of the hotel when a reviewer posted his/her review and then rounded it to the nearest half star. Observed average rating is used when calculating the rating difference, rating deviation, and reviewers' bias in rating. Observed number of reviews is the review volume of a hotel that the reviewer observed when he/she posted a review. According to the attention grabbing theory [38], reviewers are more likely to deviate from the average rating for products with a higher level of crowdedness, i.e., products with a large number of reviews. To control the attention grabbing factors, we use the observed number of reviews as the proxy for the incentive grabbing.

3.3.2. Age, length and readability of review

Age of review is determined by the number of days that have elapsed since the review is posted. Length of review is measured by the word count of the review. Readability is measured by the Flesch-Kincaid index (FKI), which has been used in extant literature [12,40].

3.3.3. Fixed effect of hotels (Hotel level)

Admittedly, to study the effect of behavioral bias in ratings, choice of hotel is the most important fixed effect that should be taken into consideration. We controlled the fixed effect of hotels by incorporating the variable *Hotel_ID* in the empirical model. Table 2 summarizes the key variables used in the empirical model. Tables 3 and 4 present the descriptive statistics and correlations between variables, respectively.

3.4. Empirical model

In order to test H1, we first regressed rating deviation (in 2013) on reviewers' absolute bias in rating (before 2013) controlling for the

Table 2
Description of variables.

Variable	Description
Abs_Bias	Reviewers' absolute rating bias based on the data before 2013, as defined in Eq. (1)
Bias	Reviewers' (signed) rating bias based on the data before 2013, as defined in Eq. (2).
Rating	Online rating posted for a hotel by a reviewer.
Obs_Avg_Rating	A hotels' average rating at the time when a reviewer posted the review.
Obs_Rev_Number	The review volume of a hotel the reviewer observed when he/she post a review.
Rating_Difference	Difference between a rating and the observed average rating.
Rating_Deviation	Absolute value of *Rating_Difference*.
Helpful_Votes	Number of helpful votes a review received.
No_Identity_Disc	Equal to one if a reviewer didn't disclose gender and age, and zero otherwise.
Women	Equals to one if a reviewer's gender is female, and zero otherwise.
Age	Represents reviewers' age as three dummy variables, *Young_Age* for those younger than 35, *Mid_Age* for those between 35 to 49 and *Old_Age* for those older than 50.
Reviewer_Country	The country fixed effect.
Cities_Visited	Number of cities a reviewer has visited.
Contributions	Sum of the total number of reviews, ratings, forum posts, and photos a reviewer has posted on TripAdvisor.
Review_Age	Days elapsed since the review was posted.
Review_Length	Word count of the review text.
Flesch_Kincaid	Flesch Kincaid index for measuring the readability of reviews.
Hotel_ID	The hotel fixed effect

Table 3
Descriptive statistics of variables.

	Mean	St. dev.	Min	Median	Max
Bias	−0.05	0.31	−1.61	−0.04	0.98
Abs_Bias	0.24	Statistic	0	0.2	1.61
Rating	4.06	Bias	1	4	5
Obs_Avg_Rating	4.12	0.36	2.2	4.15	5
Obs_Rev_Number	1296.37	1009	11	929	4350
Rating_Difference	−0.06	0.83	−4	0.02	2.06
Rating_Deviation	0.65	0.52	0	0.53	4
Helpful_Votes	0.7	1.08	0	0	14
No_Identity_Disc	0.62	0.49	0	1	1
Woman	0.27	0.45	0	0	1
Mid_Age	0.39	0.49	0	0	1
Old_Age	0.21	0.41	0	0	1
Cities_Visited	108.58	138.76	2	65	4688
Contributions	129.97	308.77	3	72	13,725
Review_Age	143.5	76.11	2	146	276
Review_Length	164.09	134.65	24	126	3616
Flesch_Kincaid	7.38	3.83	0.44	6.9	123.9

reviewer level, review level, and hotel level characteristics, as shown in Eq. (3). This equation aims to test if the absolute bias in rating reflecting reviewers' past behavior is positively related to the deviation in their future rating.

$$\begin{aligned}&Rating\_Deviation_{ij}\\&=\ \alpha\ +\beta_1 Abs\_Bias_i+\beta_2 Obs\_Avg\_Rating_{ij}\\&+\beta_3 No\_Identity\_Disc_i+\beta_4 Women_i+\beta_5 Mid\_Age_i\\&+\beta_6 Old\_Age_i+\beta_7\log(Cities\_Visited_i)\\&+\beta_8\log(Contributions_i)+\boldsymbol{\delta}' Reviewer\_Country_i\\&+\boldsymbol{\theta}' Hotel\_ID_j+\ \ \varepsilon_{ij}\end{aligned}\tag{3}$$

In Eq. (3), α, β and ε_{ij} denote the intercept, coefficients, and disturbance term. The subscripts *i* and *j* indicate reviewer *i* and hotel *j*, respectively. Abs_Bias_i is the variable of interest and the expected sign of β_1 is positive. $Obs\_Avg\_Rating_{ij}$ is a review level control variable for capturing the effect of previous reviews. $No\_Identity\_Disc_i$, $Women_i$, Mid_Age_i, Old_Age_i, $Cities\_Visted_i$ and $Contributions_i$ are reviewer level control variables. The term $Reviewer\_Country_i$ and $Hotel\_ID_j$ refer to the vector of reviewers' country dummies and Hotel ID dummies which aim to control the fixed effects of reviewers' country and hotel, and their effects on rating difference are captured through the vectors $\boldsymbol{\sigma}'$and $\boldsymbol{\theta}'$ respectively.

To analyze the effect of absolute bias in rating of reviewers on the helpfulness of their future reviews (H2), the number of helpful votes for future reviews (in 2013) is regressed on the absolute value of reviewers' bias in rating (*Abs_Bias*) while controlling the reviewer level, review level, and hotel level characteristics. As the dependent variable in this regression is a count which ranges from zero to a certain positive number, it becomes necessary to use count data regressions instead of standard linear regression [18]. Count data can be modeled as a Poisson or negative binomial regression model. The Poisson regression model assumes that the mean and variance are equal. Since the variance of *Helpful_Votes* is higher than its mean the negative binomial regression model is more appropriate than the Poisson regression [18]. The usual functional form for the negative binomial regression is the log of the outcome predicted with a linear combination of the predictors. The model for testing hypothesis 2 is therefore specified as follows:

$$\begin{aligned}&\log(Helpful\_Votes_{ij})\\&=\alpha+\beta_1 Abs\_Bias_i+\beta_2\log\left(Review\_Age_{ij}\right)\\&+\beta_3\log\left(\boldsymbol{Review\_Length_{ij}}\right)+\beta_4 Flesch\_Kincaid_{ij}\\&+\beta_5 No\_Identity\_Disc_i+\beta_6 Women_i+\beta_7 Mid\_Age_i\\&+\beta_8 Old\_Age_i+\beta_9\log(Cities\_Visited_i)\\&+\beta_{10}\log(Contributions_i)+\boldsymbol{\delta}' Reviewer\_Country_i\\&+\boldsymbol{\theta}' Hotel\_ID_j\end{aligned}\tag{4}$$

In Eq. (4), Abs_Bias_{ij} is the variable of interest; $Review\_Age_{ij}$, $Review\_Length_{ij}$, and $Flesch\_Kincaid_{ij}$ are control variables for reviews; $No\_Identity\_disclosure_i$, $Women_i$, Mid_Age_i, Old_Age_i, $Cities\_Visted_i$ and $Contribution_i s$ are reviewer level control variables. *HotelID* is used to control the fixed effect of hotels. The term $Reviewer\_Country_i$ and $Hotel\_ID_j$ refer to the vector of reviewers' country dummies and hotel ID dummies, and their effects on the helpful votes are captured through the vectors $\boldsymbol{\sigma}'$and $\boldsymbol{\theta}'$respectively.

4. Results

Table 5 shows the results of the empirical analysis for H1.

4.1. Consistency in reviewer's ratings

In Table 5, Column (1) represents a null model that contains all control variables but excludes the research variable *Abs_Bias*. Even when we controlled the hotel fixed effects and the reviewers' country fixed effects with 325 and 93 degree of freedom, respectively, the adjusted R^2 is only 0.028, indicating that these variables together can interpret only a very small fraction of variation of the dependent variable. In column (2) when we add *Abs_Bias* the adjusted R^2 rises significantly to 0.098. Column (3) shows the results of the full model which contains *Abs_Bias* and all other control variables. The coefficients of *Abs_Bias* are positive and significant at the 0.001 level of significance in both Columns (2) and (3), indicating that reviewers with greater *Abs_Bias* in the past

Table 4
Correlations between variables.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	*Bias*	1.00																
2	*Abs_Bias*	−0.34	1.00															
3	*Rating*	0.33	−0.11	1.00														
4	*Obs_Avg_Rating*	−0.06	0.01	0.35	1.00													
5	*Rating_Difference*	0.38	−0.12	0.91	−0.07	1.00												
6	*Rating_Deviation*	−0.10	0.18	−0.44	−0.10	−0.43	1.00											
7	*Helpful_Votes*	−0.03	0.03	-0.04	0.05	−0.07	0.12	1.00										
8	*No_Identity_Disc*	0.00	−0.03	0.00	0.02	−0.01	−0.02	0.04	1.00									
9	*Woman*	0.07	−0.05	0.03	0.00	0.03	−0.01	0.02	0.48	1.00								
10	*Mid_Age*	−0.06	0.00	−0.04	0.00	−0.04	−0.01	0.03	0.63	0.33	1.00							
11	*Old_Age*	0.06	−0.03	0.04	0.02	0.03	−0.01	0.01	0.41	0.16	−0.42	1.00						
12	*Cities_Visited*	−0.02	−0.03	−0.01	−0.01	−0.01	−0.02	0.01	0.14	0.01	0.02	0.15	1.00					
13	*Contributions*	0.00	−0.03	0.00	0.00	0.00	−0.01	0.02	0.09	0.06	0.04	0.06	0.20	1.00				
14	*Review_Age*	0.02	0.00	−0.02	−0.04	0.00	0.00	0.05	0.01	0.00	0.01	−0.01	0.01	0.01	1.00			
15	*Review_Length*	−0.08	0.03	−0.10	0.00	−0.11	0.11	0.22	0.08	0.08	0.07	0.00	0.01	0.07	−0.01	1.00		
16	*Flesch_Kincaid*	0.00	−0.02	0.03	0.02	0.02	−0.03	0.00	0.00	0.00	0.00	0.00	0.01	0.01	−0.01	0.00	1.00	
17	*Obs_Rev_Number*	0.04	−0.02	−0.03	−0.07	0.00	0.02	−0.05	0.00	0.00	−0.01	0.01	0.00	0.00	−0.08	−0.01	0.00	1.00

Table 5
The effects of the past absolute rating bias on deviation in future rating.

	Dependent variable: Rating_Deviation		
	(1)	(2)	(3)
Abs_Bias		0.879***	0.876***
		(0.040)	(0.040)
Obs_Avg_Rating	0.357$^{\cdot}$		0.312$^{\cdot}$
	(0.194)		(0.184)
Obs_Rev_Number	−0.070*	−0.061*	−0.063*
	(0.032)	(0.029)	(0.029)
No_Identity_Disc	−0.039		−0.030
	(0.045)		(0.045)
Woman	−0.005		0.003
	(0.013)		(0.012)
MidAge	−0.052		−0.047
	(0.045)		(0.044)
OldAge	−0.051		−0.037
	(0.046)		(0.045)
log(*Contributions*)	0.002		0.002
	(0.007)		(0.007)
log(*CitiesVisited*)	−0.011*		−0.007
	(0.005)		(0.005)
Constant	−0.255	0.286$^{\cdot}$	−0.719
	(0.727)	(0.151)	(0.687)
Hotel fixed effects	Yes	Yes	Yes
Reviewers' country	Yes	Yes	Yes
Observations	10,220	10,220	10,220
R^2	0.068	0.135	0.136
Adjusted R^2	0.028	0.098	0.099
F Statistic	1.679***	3.649***	3.623***

Notes:
(1) Cluster-robust standard errors at the individual hotel level are reported in parenthesis. The null hypothesis that each coefficient is equal to zero is tested using robust standard errors.
(2) ***, ** and * indicate significance at the 0.001, 0.01 and 0.05 level, respectively.

are more likely to post a differentiated rating in the future. Therefore, reviewers' rating behavior is consistent over time.

Furthermore, by comparing models (2) and (3), we find that the R^2 are pretty much the same. This indicates that the behavioral biases have higher interpretive power over other predictors. We implemented ANOVA analysis to further reveal the percentage of variation of the dependent variable that can be explained by different independent variables.

Using Column (3) as an example, the untabulated analysis shows that reviewers' bias in past rating, which represent the reviewers' intrinsic characteristics, capture the largest fraction of variation (8.22%) of their deviation in future ratings. The observed average rating has a trivial impact on reviewers' rating behavior in future (0.07%), whereas the number of reviews, serving as a proxy for attention grabbing [38], also has trivial impact (0.08%). The other reviewer related variables, including *Gender*, *Age*, *Contribution*, *Cities_Visited* and reviewers' country fixed effects are not significant and can hardly account for any variation in the dependent variable. One possible explanation is that *Abs_Bias* already captures these intrinsic characteristics. As such, our results indicate that reviewers' rating behavior is consistent across products and over time since rating behavior in future is mainly determined by reviewer's rating bias in the past rather than by the observed average rating or attention grabbing. This supports hypothesis 1.

4.2. Reviewer's bias in rating in the past and helpfulness of future reviews

The negative binominal regression results for testing H2 are shown in Table 6. The appropriateness of our chosen approach is confirmed with the likelihood ratio (LR) test in which the base hypothesis is tested for the dispersion parameter alpha equals zero. If this hypothesis is rejected, then negative binominal regression is reduced to a Poisson regression model. The results of the LR test indicated alpha to be non-zero for all models, and the negative binominal model is the preferred model for testing the hypothesis.

Table 6
Negative binominal regression for the helpful votes.

	Dependent variable			
	Helpful_Votes			
	(1)	(2)	(3)	(4)
Rating_Deviation	0.321***		0.306***	
	(0.024)		(0.025)	
Abs_Bias		0.474***	0.180*	0.485***
		(0.084)	(0.088)	(0.084)
Extrinsic_Bias				0.306***
				(0.025)
log(*Review_Age*)	0.081***	0.083***	0.081***	0.081***
	(0.018)	(0.018)	(0.018)	(0.018)
log(*Review_Length*)	0.354***	0.381***	0.354***	0.354***
	(0.021)	(0.021)	(0.021)	(0.021)
Flesch_Kincaid	0.002	0.001	0.002	0.002
	(0.003)	(0.004)	(0.003)	(0.003)
Identity_Disclosure	−0.019	−0.027	−0.019	−0.019
	(0.120)	(0.121)	(0.120)	(0.120)
Woman	−0.065	−0.064	−0.063	−0.063
	(0.035)	(0.035)	(0.035)	(0.035)
Mid_Age	0.096	0.074	0.095	0.095
	(0.118)	(0.119)	(0.118)	(0.118)
Old_Age	0.067	0.053	0.069	0.069
	(0.120)	(0.121)	(0.120)	(0.120)
log(Contributions)	0.043*	0.041*	0.043*	0.043*
	(0.019)	(0.019)	(0.019)	(0.019)
log(Cities_Visited)	−0.007	−0.004	−0.005	−0.005
	(0.015)	(0.015)	(0.015)	(0.015)
Constant	−3.197***	−3.332***	−3.307***	−3.307***
	(0.393)	(0.398)	(0.396)	(0.396)
Hotel fixed effect	Yes	Yes	Yes	Yes
Country fixed effect	Yes	Yes	Yes	Yes
No. of observations	10,220	10,229	10,220	10,220
Log Likelihood	−10,734	−10,814	−10,731	−10,732
Theta	3.528***	3.192***	3.538***	3.538***
	(0.298)	(0.253)	(0.300)	(0.300)
Akaike Inference Criterion	22,325	22,488	22,323	22,323

Notes:
(1) Cluster-robust standard errors at the individual hotel level are reported in parenthesis. The null hypothesis that each coefficient is equal to zero is tested using robust standard errors.
(2) ***, ** and * indicate significance at the 0.001, 0.01 and 0.05 level of significance respectively.
(3) *Extrinsic_Bias* is defined as the difference between *Rating_Deviation* and *Abs_Bias*.

Since extant research has shown that reviews with extreme ratings will receive more helpful votes [3,19,49], before formally testing hypothesis 2, we examine the effect of extremeness in rating (based on the current rating) on helpful votes. Consistent with [19], we use *Rating_Deviation* (the rating deviation of the current review from the current product's average rating) to measure the extremeness in rating and present the results in Column (1).

It can be seen that the coefficients of *Rating_Deviation* are positive and significant at a 1% level of significance, which means reviews with extreme ratings will receive more helpful votes. This result is consistent with the literature [3,19,49]. As to the review level control variables, the coefficients of *Review_Age* and *Review_Length* are significantly positive which are consistent with the literature, while the coefficient of the readability measurement, *Flesch_Kincaid*, is not significant. For the reviewer level control variables, only the coefficient of Contributions is significant, indicating reviews by reviewers with more online experience tend to receive more helpful votes.

Column (2) shows the estimation results for Eq. (6) aiming to test Hypothesis 2, in which we replace the *Rating_Deviation*, a measure of the extremeness in current reviews with *Abs_Bias*, a measure of the reviewers' bias in rating in the past. We find that, the coefficients of *Abs_Bias* are positive and significant at a 0.001 level of significance. This means that reviews by reviewers with higher absolute bias in rating in the past tend to receive more helpful votes in the future. Therefore, H2 is supported. Furthermore, to quantify the effects of the absolute bias in rating of a reviewer in the past on the helpful votes of his/her future reviews, we calculate the incidence rate ratio (IRR), which is the exponential value of the coefficients [18]. The IRR of *Abs_Bias* is 1.61, indicating that holding the other variables in the model constant, for every one unit increase in *Abs_Bias* of a reviewer in the past, the count of helpful votes of his/her future reviews will increase by 61%.

Yin et al. [48] empirically demonstrated that the deviation of a review rating from the product's average rating (i.e., rating deviation) has a negative effect on the perceived helpfulness of the review. They termed this *confirmation bias*. To control for this potential confirmation bias, we include both *Abs_bias* (based on historical data) as well as *Rating_Deviation* (based on current data) as our independent variables and repeat our analysis and present the results in Column (3). The coefficient of *Abs_Bias* is still significant and positive (Coeff = 0.180 and S.D. = 0.088), indicating that even with the current potential confirmation bias controlled, future reviews written by reviewers with a larger past absolute rating deviation will be more likely to receive more helpful votes.

To further reveal the impact of past rating bias on future review helpfulness, we decompose *Rating_Deviation* into two components: the intrinsic rating bias (*Abs_Bias*)[3], which reflects a reviewer's historical intrinsic rating bias behavior, and the extrinsic rating bias *Extrinsic_Bias* (the difference between *Rating_Deviation* and *Abs_bias*), which represents bias above or below the historical intrinsic rating bias. The *Extrinsic_Bias* is captured on a specific review level and it goes beyond a reviewer's historical absolute bias and may be driven by a reviewer's social pressure, mood, or something else.

Column (4) shows the results with both the intrinsic rating bias and the extrinsic rating bias. The coefficient of intrinsic rating bias (*Abs_bias*) is 0.485 (with S.D. = 0.084), while the coefficient of extrinsic rating bias (*Extrinsic_Bias*) is 0.306 (with S.D. = 0.025). A formal Wald test shows that the null hypothesis of the coefficient of intrinsic rating bias equal to that of the extrinsic rating bias is rejected (*P*-value = 0.04). This is also confirmed by the significant coefficient of *Abs_Bias* (0.180) in column (3), which in fact captures the difference between intrinsic and extrinsic bias, revealing that intrinsic bias plays a more significant role in influencing helpfulness of future reviews than extrinsic bias.

4.3. Check for robustness

For a robustness check, we replace the dependent variable rating deviation ($Rating\_Deviation_{ij}$) with the signed rating difference ($Rating\_Difference_{ij}$), and replace the independent variable absolute bias (Abs_Bias_i) with signed bias ($Bias_i$ defined in Eq. (2)) to check if the past signed rating bias is consistent with the future rating difference, as shown in Eq. 5.

$$\begin{aligned}
&Rating_Difference_{ij}\\
&= \alpha + \beta_1 Bias_i\\
&+\beta_2 Obs_Avg_Rating_{ij} + \beta_3 Obs_Avg_Rating_{ij}\\
&+\beta_4 No_Identity_Disc_i + \beta_5 Women_i + \beta_6 Mid_Age_i\\
&+\beta_7 Old_Age_i + \beta_8 \log(Cities_Visited_i)\\
&+\beta_9 \log(Contributions_i) + \boldsymbol{\delta}' Reviewer_Country_i\\
&+\boldsymbol{\theta}' Hotel_{IDj} + \varepsilon_{ij}
\end{aligned} \tag{5}$$

Furthermore, we also regress reviewers' future ratings (in 2013) on their past rating bias (estimated on the basis of ratings before 2013) with other control variables in the same way as done in Eq. (3). Because online ratings (integer values from 1 to 5) are ordered and censored data variables that are not distributed normally, we employ the ordered logistic model according to the literature [20,31,40]. Let U_{ij} be the underlying latent variable that captures reviewer *i*'s evaluation on hotel *j*, and *k* denote a realized value of a rating with $k \in [1, 5]$; λ_1 through λ_5 are cutoffs and parameters to help identify intervals for each rank of the ratings. The ordered logistic model is therefore specified as shown below:

$$\begin{aligned}
&Pr\left(Rating_{ij} = k\right) = Pr(\lambda_{k-1} < U_{ij} \le \lambda_k)\\
&U_{ij} = \beta_1 Bias_i\\
&+\beta_2 Obs_Avg_Rating_{ij} + \beta_3 Obs_Avg_Rating_{ij} + \beta_4 No_Identity_Disc_i\\
&+\beta_5 Women_i + \beta_6 Mid_Age_i + \beta_7 Old_Age_i + \beta_8 \log(Cities_Visited_i)\\
&+\beta_9 \log(Contributions_i) + \boldsymbol{\delta}' Reviewer_Country_i + \boldsymbol{\theta}' Hotel_ID_j
\end{aligned} \tag{6}$$

The model accommodates nonlinear effects of the independent variables on U_{ij}. Except for estimating the coefficients of U_{ij}, we also estimate cutoffs λ_k ($k = 2$–5).

Columns (1)–(3) of Table 7 present the estimation results for Eq. (5), which aim to investigate how the past rating *Bias* of reviewers affect their future *Rating_Difference*. In Column (2) and Column (3), *Bias* is significant at the 0.001 level of significance and its coefficient is positive. This means a reviewer's rating bias in the past is positively related to his/her future *Rating_Difference*. Again, reviewers' rating behavior is consistent in terms of rating difference.

Interestingly, the coefficients of *Bias* are close to unity in Columns (2) and (3). We further test if the coefficient of *Bias* in Column (6) is significantly different from one. The Wald test indicates that the null hypothesis of $\beta_{Bias} = 1$ cannot be rejected (*P*-value = 0.94). This implies that one unit change in *Bias* will result in one unit change in the *Rating_Difference*. In other words, on an average, if reviewers' past ratings are lower (higher) than their observed average ratings, their future ratings will also be lower (higher) than their observed average ratings for exactly the same amount. This is very strong evidence that the reviewer's bias in rating is consistent. Therefore, Hypothesis 1 is supported.

Furthermore, based on the untabulated ANOVA analysis, in Column (3), *Bias* alone can interpret 16.93% of the variation of *Rating_Difference*; Hotel fixed effects (with 325 degree of freedom), *Obs_Avg_Rating* and *Obs_Rev_Number* can interpret only 6.65%, 0.38% and 0.04% of the variation, respectively. Again, our results indicate that reviewers' rating is consistent across products and over time, and future rating behavior is mainly determined by their past rating bias rather than by the observed average rating or attention grabbing, thus supporting hypothesis 1.

Columns (4)–(6) present the results of another robustness check for H1 in which we investigate the effects of reviewers' past rating bias on their future ratings. Again, Column (4) is the null model with only the control variables and Column (5) contains only *Bias* and hotel fixed effects. Column (6) is the full model that includes *Bias*, all control variables and the hotel fixed effect. From column (6) we observe that the coefficient of *Bias* (the signed bias in this case) is significant at the 0.001 level of significance while all other control variables are not significant. This indicates that reviewers with positive (negative) *Bias* will be more likely to give higher (lower) ratings in future. Therefore, the results of the ordered logistic model are consistent with the results of the OLS model and *Bias* is able to serve as predictor for reviewers' ratings in future.

Last, when estimating the bias of the reviewer in the past, we have used all reviews ever written by the same reviewer before year 2013, which includes his/her reviews under hotel, restaurant, and attraction related categories. To validate that reviewer rating behavior is consistent across product categories, we re-estimate the reviewer's rating bias based on his/her reviews in the above three categories respectively[4]. Then we link this product category specific bias in the past to the bias as well as helpfulness of his/her future review for hotels only. We present the results in Table 8 and Table 9. Due to space limitation, we only show the coefficients of the main variable of interest. As

[3] We acknowledge that Tripadvisor does not calculate and present this information on its website. However, since Tripadvisor does disclose all the ratings ever written by every reviewer, we believe consumers may use some heuristics to derive this piece of information.

[4] We thank an anonymous reviewer for his/her useful suggestions on checking robustness of the results across different types of products.

Table 7
Robustness check: the effects of rating bias on future rating.

	Dependent variable *Rating_Difference* *OLS* (1)	(2)	(3)	*Rating* *Ordered logistic* (4)	(5)	(6)
Bias		1.008^{***}	1.002^{***}		2.781^{***}	2.768^{***}
		(0.027)	(0.027)		(0.072)	(0.073)
Obs_Avg_Rating	-1.178^{***}		-1.067^{***}	-0.172		0.216
	(0.244)		(0.218)	(0.463)		(0.463)
Obs_Rev_Number	$0.078^{\cdot}$	0.066	$0.073^{\cdot}$	0.142	$0.160^{\cdot}$	$0.158^{\cdot}$
	(0.042)	(0.041)	(0.038)	(0.091)	(0.092)	(0.092)
No_Identity_Disc	-0.030		0.019	-0.102		0.050
	(0.077)		(0.074)	(0.178)		(0.185)
Woman	0.084^{***}		0.030	0.207^{***}		$0.091^{\cdot}$
	(0.021)		(0.019)	(0.050)		(0.052)
MidAge	-0.100		-0.014	-0.281		-0.059
	(0.076)		(0.073)	(0.176)		(0.182)
OldAge	-0.017		0.012	-0.059		0.034
	(0.077)		(0.074)	(0.179)		(0.185)
log(*Contributions*)	0.007		0.001	0.019		0.008
	(0.011)		(0.010)	(0.027)		(0.028)
log(*CitiesVisited*)	-0.013		0.003	-0.043^{*}		-0.005
	(0.009)		(0.008)	(0.021)		(0.021)
Constant	3.939^{***}	-0.220	3.477^{***}			
	(0.943)	(0.274)	(0.855)			
Intercept-2				$3.131^{\cdot}$	2.553^{***}	1.756
				(1.773)	(0.772)	(1.764)
Intercept-3				1.569	0.924	0.128
				(1.771)	(0.768)	(1.763)
Intercept-4				-0.287	-1.089	-1.886
				(1.770)	(0.768)	(1.762)
Intercept-5				-2.487	-3.557^{***}	-4.356^{*}
				(1.771)	(0.768)	(1.763)
Hotel fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10,220	10,220	10,220	10,220	10,220	10,220
R^2	0.075	0.198	0.202	0.214	0.341	0.341
Adjusted R^2	0.035	0.164	0.167			
F Statistic	1.871^{***}	5.766^{***}	5.792^{***}			
Chi^2				$2{,}221^{***}$	$3{,}805^{***}$	$3{,}812^{***}$

Notes:
(1) Cluster-robust standard errors at the individual hotel level are reported in parenthesis. The null hypothesis that each coefficient is equal to zero is tested using robust standard errors.
(2) ***, ** and * indicate significance at the 0.001, 0.01 and 0.05 level, respectively.

shown in Tables 8 and 9, regardless of whether the absolute bias or the extrinsic bias is estimated on the basis of reviews for hotels, restaurants, or attractions, the rating bias in the past is significantly and positively associated with the rating bias and helpfulness of future reviews for hotels. This validates that reviewer behavior is consistent across product categories.

Table 8
The effects of product category specific bias on ratings of hotels in future.

	Abs_Bias/Bias based on Hotel ratings			Restaurant ratings			Attraction ratings		
	OLS (1)	*OLS* (2)	*Logistic* (3)	*OLS* (4)	*OLS* (5)	*Logistic* (6)	*OLS* (7)	*OLS* (8)	*Logistic* (9)
Abs_Bias	0.782^{***}			0.491^{***}			0.702^{***}		
	(0.095)			(0.086)			(0.176)		
Bias		0.992^{***}	2.558^{***}		0.696^{***}	1.961^{***}		0.731^{***}	2.207^{***}
		(0.080)	(0.198)		(0.073)	(0.169)		(0.136)	(0.265)
Obs_Avg_Rating	0.364	-0.352	2.585^{***}	0.406	-0.352	2.306^{***}	-0.211	-3.258^{*}	2.190^{***}
	(0.288)	(0.446)	(0.182)	(0.273)	(0.435)	(0.153)	(0.959)	(1.546)	(0.212)
Hotel fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1176	1176	1176	1510	1510	1510	784	784	784
R^2	0.356	0.412	0.297	0.279	0.329	0.242	0.371	0.431	0.238
Chi^2			372.39^{***}			377.45^{***}			191.87^{***}

Notes:
(1) The dependent variables of columns (1), (4) and (7) are *Rating_Deviation*; the dependent variables of column (2), (5) and (8) are *Rating_Difference*; The dependent variables of column (1), (4) and (7) are *Rating*.
(2) Abs_Bias and Bias in columns (1)–(3), (4)–(6) and (7)–(9) are calculated on the basis of reviewers' hotel ratings, restaurant ratings and attraction ratings, respectively.
(3) Cluster-robust standard errors at the individual hotel level are reported in parenthesis. The null hypothesis that each coefficient is equal to zero is tested using robust standard errors.
(4) ***, ** and * indicate significance at the 0.001, 0.01 and 0.05 level, respectively.

Table 9
The effects of product category related bias on helpfulness of hotel reviews in future.

	Dependent variable: Helpful_Votes *Abs_Bias/Bias based on* Hotel ratings		Restaurant ratings		Attraction ratings	
	(1)	(2)	(3)	(4)	(5)	(6)
Abs_Bias	0.483*	0.539*	0.629**	0.814***	1.023**	1.152***
	(0.234)	(0.235)	(0.207)	(0.209)	(0.337)	(0.342)
Extrinsic_Bias		0.258***		0.336***		0.356***
		(0.078)		(0.068)		(0.102)
Hotel fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1176	1176	1510	1510	784	784
Log likelihood	−1143.343	−1137.956	−1510.751	−1498.771	−723.938	−717.969

Notes: (1) *Abs_Bias* and *Bias* in columns (1)–(2), (3)–(4) and (5)–(6) are calculated based on reviewers' hotel ratings, restaurant ratings and attraction ratings, respectively.
(2) Cluster-robust standard errors at the individual hotel level are reported in parenthesis. The null hypothesis that each coefficient is equal to zero is tested using robust standard errors.
(3) ***, ** and * indicate significance at the 0.001, 0.01 and 0.05 level, respectively.

Another robustness check we did is that we also constructed a new bias measures based on the future ratings in 2014[5], and then use the bias based on future ratings in 2014 to predict rating deviation as well as helpful votes of the past reviews in 2013. The results indicate that the future rating bias is also positively associated with the rating bias and helpfulness of votes of past reviews. Again, this indicates that the reviewer's rating behavior is consistent over time.

5. Conclusion

Online product reviews are a major informational source for consumers. Practitioners always look for a way to influence consumers' purchase decisions. Based on the hotel reviews of New York City hotels from TripAdvisor.com, our results indicate that consumer review patterns are consistent over time and across products, and those who gave a higher rating in the past (or deviated more from the social norm) will be more likely to leave a higher rating in the future (or deviate more from the social norm) when they rate new products. Reviewers' review pattern is governed mainly by their intrinsic characteristics rather than by the observed average rating. Given a reviewer's review pattern is consistent over time and predictable, online vendors can already be reasonable sure what kind of rating that customer will leave after consuming the product. Hence, online vendors may be able to proactively do something to influence the rating that the customer will leave on the review site.

This study has practical implications for individual consumers, online retailers, and information aggregators. For individual consumers, since they generally use the rating of online product reviews to infer product quality, our results indicate that consumers should take into account the existence of such reviewer rating biases since some reviewer may always post extreme reviews. For online retailers, we recommend that online vendors spend more time and resources in understanding the intrinsic needs of individual customers. First, given that reviewers are consistent over time and across product categories, when they receive an order from a customer, online retailers can already be reasonable sure about what kind of rating that customer will post after consumption. Hence, online retailers should proactively provide services to that customer if they want to influence that consumer's product experience. Second, online retailers should change the way they present the review information online if they believe that not disclosing such a reviewer's rating bias information causes a customer to make the wrong product choice. Last, if an online platform cares more about the usefulness of online reviews, then based on the review history of the reviewers, online vendors can selectively pick those who have tended to leave extreme reviews in the past and offer them free products to try. Online vendors can be reasonably sure that such reviewers will be more likely to write reviews that are considered useful. However, if an online platform cares about both the rating magnitude as well as rating helpfulness, then they can selectively pick those who tend to leave extremely positive reviews in the past since such reviewers will be more likely to post future reviews with ratings higher than average and, at the same time, which will be considered more useful. Finally, many companies (e.g., Epinions.com and BizRate.com) specialize on the collection, synthesis, and dissemination of online product reviews, and for them having a more accurate method to inform consumers about product quality can be a differentiating factor.

This study has certain limitations that create several interesting opportunities for future research. First, the 5-point scale of star ratings may not be sensitive enough to fully capture the bias in behavior. Thus, using a website whose review has a scale with more anchors (such as 10 points) may be more appropriate. Second, our results are based on data of New York City from TripAdvisor only. Hence, whether such a result can be generalized to other cities or other websites still remains unanswered. Third, future researchers can conduct a randomized field experiment with a random sample of both online consumers as well as online vendors and verify whether they are aware of such consistency and how they deal with it. This may lead online vendors to decide whether to pre-calculate and disclose such information on their websites.

[5] We thank an anonymous reviewer for his/her useful suggestions for this robustness check.

Acknowledgement

The work described in this paper was partially supported by the Key Program of National Natural Science Foundation of China (Grant No. 71532011), the National Natural Science Foundation of China (Grant No. 71272226), the Key Projects of Philosophy and Social Sciences Research, Ministry of Education (Grant No. 14JZD017), and Research Fund for Academic Team of Young Scholars at Wuhan University (Big Data and Business Analytics).

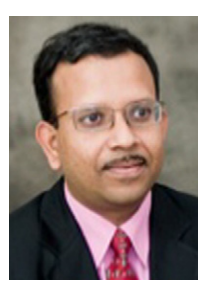
References

[1] N. Archak, A. Ghose, P.G. Ipeirotis, Deriving the pricing power of product features by mining consumer reviews, Management Science 57 (8) (2011) 1485–1509.
[2] H. Baek, J.H. Ahn, Y. Choi, Helpfulness of online consumer reviews: readers' objectives and review cues, International Journal of Electronic Commerce 17 (2) (2013) 99–126.
[3] Q. Cao, W. Duan, Q. Gan, Exploring determinants of voting for the “helpfulness” of online user reviews: a text mining approach, Decision Support Systems 50 (2) (2011) 511–521.
[4] O. Furrer, B.S.C. Liu, D. Sudharshan, The relationships between culture and service quality perceptions basis for cross-cultural market segmentation and resource allocation, Journal of Service Research 2 (4) (2000) 355–371.
[5] J.A. Chevalier, D. Mayzlin, The effect of word of mouth on sales: online book reviews, Journal of Marketing Research 43 (3) (2006) 345–354.
[6] P.K. Chintagunta, S. Gopinath, S. Venkataraman, The effects of online user reviews on movie box office performance: accounting for sequential rollout and aggregation across local markets, Marketing Science 29 (5) (2010) 944–957.

[7] E.K. Clemons, G.G. Gao, L.M. Hitt, When online reviews meet hyper differentiation: a study of the craft beer industry, Journal of Management Information Systems 23 (2) (2006) 149–171.
[8] C. Dellarocas, X.M. Zhang, N.F. Awad, Exploring the value of online product reviews in forecasting sales: the case of motion pictures, Journal of Interactive Marketing 21 (4) (2007) 23–45.
[9] V. Dhar, E.A. Chang, Does chatter matter? The impact of user-generated content on music sales, Journal of Interactive Marketing 23 (4) (2009) 300–307.
[10] W. Duan, B. Gu, A.B. Whinston, Do online reviews matter? An empirical investigation of panel data, Decision Support Systems 45 (4) (2008) 1007–1016.
[11] C. Forman, A. Ghose, B. Wiesenfeld, Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets, Information Systems Research 19 (3) (2008) 291–313.
[12] A. Ghose, P. Ipeirotis, Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics, IEEE Transactions on Knowledge and Data Engineering 23 (10) (2011) 1498–1512.
[13] A. Ghose, P.G. Ipeirotis, B. Li, Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content, Marketing Science 31 (3) (2012) 493–520.
[14] K.Y. Goh, C.S. Heng, Z. Lin, Social media brand community and consumer behavior: quantifying the relative impact of user-and marketer-generated content, Information Systems Research 24 (1) (2013) 88–107.
[15] J. Gountas, S. Gountas, Personality orientations, emotional states, customer satisfaction, and intention to repurchase, Journal of Business Research 60 (1) (2007) 72–75.
[16] G.H. Hofstede, G. Hofstede, Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations, Sage, 2001.
[17] B. Gu, J. Park, P. Konana, The impact of external word-of-mouth sources on retailer sales of high-involvement products, Information Systems Research 23 (1) (2012) 182–196.
[18] J.M. Hilbe, Negative Binomial Regression, Cambridge University Press, 2011.
[19] Y. Hong, C. Li, The "assertive" consumers: effects of culture in online word of mouthAvailable at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=24284072014 (last accessed April 22, 2016).
[20] Y. Hu, X. Li, Context-dependent product evaluations: an empirical analysis of internet book reviews, Journal of Interactive Marketing 25 (3) (2011) 123–133.
[21] N. Hu, N.S. Koh, S.K. Reddy, 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 (2014) 42–53.
[22] D. Jani, H. Han, Personality, satisfaction, image, ambience, and loyalty: testing their relationships in the hotel industry, International Journal of Hospitality Management 37 (2014) 11–20.
[23] Online reviews most helpful in purchase decisions Available at: http://www.marketingcharts.com/direct/online-product-reviews-most-helpful-in-purchasedecisions-offline-too-3437/2008 (last accessed April 22, 2016).
[24] Y. Kwark, J. Chen, S. Raghunathan, Online product reviews: implications for retailers and competing manufacturers, Information Systems Research 25 (1) (2014) 93–110.
[25] X. Li, L.M. Hitt, Self-selection and information role of online product reviews, Information Systems Research 19 (4) (2008) 456–474.
[26] Y. Liu, Word of mouth for movies: its dynamics and impact on box office revenue, Journal of Marketing 70 (3) (2006) 74–89.
[27] Z. Liu, S. Park, What makes a useful online review? Implication for travel product websites, Tourism Management 47 (2015) 140–151.
[28] X. Luo, J. Zhang, W. Duan, Social media and firm equity value, Information Systems Research 24 (1) (2013) 146–163.
[29] X. Luo, J. Zhang, How do consumer buzz and traffic in social media marketing predict the value of the firm? Journal of Management Information Systems 30 (2) (2013) 213–238.
[30] X. Luo, S. Raithel, M.A. Wiles, The impact of brand rating dispersion on firm value, Journal of Marketing Research 50 (3) (2013) 399–415.
[31] X. Ma, L. Khansa, Y. Deng, S.S. Kim, Impact of prior reviews on the subsequent review process in reputation systems, Journal of Management Information Systems 30 (3) (2013) 279–310.
[32] A.S. Mattila, The role of culture in the service evaluation process, Journal of Service Research 1 (3) (1999) 250–261.
[33] A.S. Mattila, The impact of culture and gender on customer evaluations of service encounters, Journal of Hospitality & Tourism Research 24 (2) (2000) 263–273.
[34] J. Meyers-Levy, Gender differences in the use of message cues and judgments, Journal of Marketing Research 28 (1) (1991) 84–96.
[35] W.W. Moe, M. Trusov, The value of social dynamics in online product ratings forums, Journal of Marketing Research 48 (3) (2011) 444–456.
[36] S.M. Mudambi, D. Schuff, What makes a helpful online review? A study of customer reviews on amazon.com, MIS Quarterly 34 (1) (2010) 185–200.
[37] P. Racherla, W. Friske, Perceived 'usefulness' of online consumer reviews: an exploratory investigation across three services categories, Electronic Commerce Research and Applications 11 (6) (2012) 548–559.
[38] W. Shen, Y.J. Hu, J. Rees, Competing for attention: an empirical study of online reviewers' strategic behavior, MIS Quarterly 39 (3) (2015) 683–696.
[39] B.A. Sparks, V. Browning, The impact of online reviews on hotel booking intentions and perception of trust, Tourism Management 32 (6) (2011) 1310–1323.
[40] S. Sridhar, R. Srinivasan, Social influence effects in online product ratings, Journal of Marketing 76 (5) (2012) 70–88.
[41] J. Klayman, Y.W. Ha, Confirmation, disconfirmation, and information in hypothesis testing, Psychological Review 94 (2) (1987) 211–228.
[42] M. Sun, How does the variance of product ratings matter? Management Science 58 (4) (2012) 696–707.
[43] S. Tirunillai, G.J. Tellis, Does chatter really matter? Dynamics of user-generated content and stock performance, Marketing Science 31 (2) (2012) 198–215.
[44] N.K.F. Tsang, J. Ap, Tourists' perceptions of relational quality service attributes: a cross-cultural study, Journal of Travel Research 45 (3) (2007) 355–363.
[45] M. Van Birgelen, K. de Ruyter, A. de Jong, M. Wetzels, Customer evaluations of after-sales service contact modes: an empirical analysis of national culture's consequences, International Journal of Research in Marketing 19 (1) (2002) 43–64.
[46] R. Vázquez-Carrasco, G.R. Foxall, Influence of personality traits on satisfaction, perception of relational benefits, and loyalty in a personal service context, Journal of Retailing and Consumer Services 13 (3) (2006) 205–219.
[47] D. Yin, S.D. Bond, H. Zhang, Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews, MIS Quarterly 38 (2) (2014) 539–560.
[48] D. Yin, S. Mitra, H. Zhang, When do consumers value positive versus negative reviews? An empirical investigation of confirmation bias in online word of mouth, Information Systems Research 27 (1) (2016) 131–144.
[49] G. Yin, L. Wei, W. Xu, M. Chen, Exploring Heuristic Cues For Consumer Perceptions of Online Reviews Helpfulness: The Case of yelp.com, Proceedings of the 18th Pacific Asia Conference on Information Systems, Chengdu, China, 2014 Available at: http://www.pacis-net.org/file/2014/2206.pdf (last accessed April 22, 2016).
[50] Y. Yu, W. Duan, Q. Cao, The impact of social and conventional media on firm equity value: a sentiment analysis approach, Decision Support Systems 55 (4) (2013) 919–926.
[51] W. Zhou, W. Duan, An empirical study of how third-party websites influence the feedback mechanism between online word-of-mouth and retail sales, Decision Support Systems 76 (2015) 14–23.

Baojun Gao is Associate Professor of Management Science at Wuhan University. He received his BE, MSc and PhD from Xian Jiaotong University. His research interests are in the areas of business analytics, simulation and information systems. His research has appeared in IEEE Transactions on SMC:Systems, Electronic Commerce Research and Applications, China Economic Review, Service Science.

Nan Hu is an Associate Professor at Stevens Institute of Technology. He received his Ph.D. from the University of Texas at Dallas. Nan's research focuses on investigating the value implications and market efficiency of both traditional information (e.g. company financial report, analyst forecast, corporate governance) and non-traditional information (e.g. blog opinion, online consumer reviews), using a combination of theories from accounting, finance, marketing, information economics, sociology, psychology, and computer science. Nan's research has appeared at MIS Quarterly, Journal of Management Information Systems, Communications of the ACM, Journal of Computer Security, and Information Technology and Management, among others.



Indranil Bose is Professor of Management Information Systems at the Indian Institute of Management, Calcutta. He acts as Coordinator of IIMC Case Research Center. He holds B. Tech. from the Indian Institute of Technology, MS from the University of Iowa, MS and Ph.D. from Purdue University. His research interests are in business analytics, telecommunications, information security, and supply chain management. His publications have appeared in Communications of the ACM, Communications of AIS, Computers and Operations Research, Decision Support Systems, Ergonomics, European Journal of Operational Research, Information & Management, International Journal of Production Economics, Journal of Organizational Computing and Electronic Commerce, Journal of the American Society for Information Science and Technology, Operations Research Letters etc. He is listed in the International Who's Who of Professionals, Marquis Who's Who in the World, Marquis Who's Who in Asia, Marquis Who's Who in Science and Engineering, and Marquis Who's Who of Emerging Leaders 2007. He serves as Senior Editor of Decision Support Systems and as Associate Editor of Communications of AIS, Information Technology & Management, and several other IS journals.