



Examining the impact of review tag function on product evaluation and information perception of popular products

Zhuolan Bao¹ · Wenwen Li² · Pengzhen Yin³ · Michael Chau⁴

Received: 1 May 2018 / Revised: 22 December 2019 / Accepted: 18 January 2020 /
Published online: 27 May 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Since online reviews have become an increasingly important information source for consumers to evaluate products during online shopping, many platforms started to adopt review mechanisms to maximize the value of such massive reviews. In recent years, the review tag function has been adopted in practices and leading the research of sentiment and opinion extraction techniques. However, the examination of its impact has been largely overlooked. In this paper, by proposing a framework through the lens of attribution theory, we look into the effect of the review tag function on two focal outcomes. One is the evaluation of highly-rated popular products, the other is the helpfulness perception of product reviews. Experimental methods and qualitative analysis were utilized to test our hypotheses. Our findings demonstrate the importance of tag function application as it further increases consumers' product evaluation for popular products. We also found that different tag function appearances influence consumers' cognitive biases in review helpfulness perception.

Keywords Online reviews · Review tag · Product evaluation · Perceived bias

1 Introduction

User-generated product reviews are very popular and widely adopted in online markets. Because of their great value in reducing information asymmetry of the Internet, online reviews are considered as a facilitating tool for consumers to

✉ Zhuolan Bao
baozhuolan@cuhk.edu.cn

¹ School of Management and Economics, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China

² School of Management, Fudan University, Shanghai, China

³ School of Management, Hefei University of Technology, Hefei, China

⁴ Faculty of Business and Economics, The University of Hong Kong, Pok Fu Lam, Hong Kong

make purchase decision (Mudambi and Schuff 2010; Dimoka et al. 2012; Dellarcas 2003). Hence, the impact of online reviews is increasingly important and has been intensively investigated by researchers (Lu et al. 2013; Forman et al. 2008; Yin et al. 2016).

Many research focuses on the role of product's average review ratings, as it serves as a salient signal for potential consumers to learn about the product (Sun 2012). Prior research have found the impact of product ratings on consumers' perceptual and behavioral outcomes in the shopping process, such as product evaluation, sales, consumer revisit intention, and perception of product review information (Lu et al. 2013; Ba and Pavlou 2002; Yin et al. 2016; Yan et al. 2015).

However, for consumers, the mere average rating information might not suffice. The plenty of information embedded in review content is also important in providing consumers with different opinions on the product (Archak et al. 2011). To assist consumers in reading the massive review content, market platforms introduce new mechanisms to help them identify the most important or valuable information (Chen et al. 2017; Mudambi and Schuff 2010).

The automatic review tagging system is one of such attempts. The automatic tagging system utilizes and extracts the content of consumer reviews to automatically generate products' feature-related tags using text-mining technologies. In current practices, both TripAdvisor.com and Tmall.com have been presenting the most frequently mentioned review content on top of all the reviews. While TripAdvisor only shows the tag label (shown in Fig. 1), i.e., the most frequent features that people comment on, Tmall displays the tag labels, the number of mentions, as well as the corresponding sentiment using different colors (shown in Fig. 2).

Thus, besides an average rating, the tags derived from review content also facilitate consumers to instantly grasp the keywords, or collective opinions from prior consumers. However, as most prior research has focused on examining the impact of review rating on consumers' product evaluation or information perception (Sun 2012; Pan and Zhang 2011; Yin et al. 2016), little is known about how the review tag function would matter.

2,041 Reviews from our TripAdvisor Community

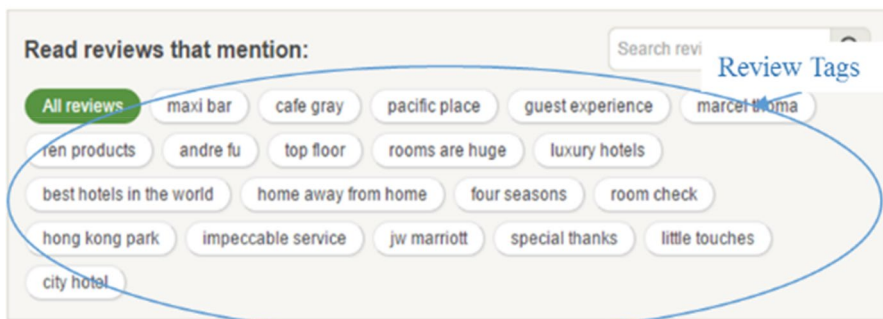


Fig. 1 Review tag function on TripAdvisor.com



Fig. 2 Review tag function on Tmall.com

To understand the role of the review tag function, an intuitive starting point is to learn how the function is applied. The review tag function is mostly used for popular products (Hu and Liu 2004) because of two reasons. First, only sufficient texts could afford a basis for extracting tags. For an unpopular product, its number of reviews could be too small to produce tags. Second, reviews for hot-pick products are often too numerous to be processed by humans in an effective way. For such products, the tag function could also bring the most benefits for their potential consumers.

Our research questions rise naturally from the above discussions. From a practical aspect, for a highly-rated popular product, its high rating might have already brought a positive impact on consumers' perception towards the product and its reviews. It would be beneficial for managers or sellers to understand how the review tag function would influence their potential consumers in evaluating the product. In particular, in this study, we aim at answering the following two questions. First, will the tag function influence consumers' evaluation towards highly-rated popular products? Second, will the presented tags influence consumers' perception towards product reviews?

To answer the questions, we tested our hypotheses by utilizing online experiments using mock product webpages. By assigning respondents to conditions of different tag function settings, we collected data on their responses of product evaluation and review perception. We also conducted focus group to further analyze the process of consumers' product evaluation and review perception. Our findings revealed that, tag function adoption increase the consumers' evaluation towards a product. Meanwhile, we found that people would prefer negative reviews for a popular product when all its tags appear positive, while people's preference for positive reviews does not change by the adoption of tag function.

The remaining part of the paper is arranged as follow. We first introduce the current usage of the review tag function based on data collected from Tmall.com, and present our theory background. After we propose our hypotheses on the impact of the review tag function, we present our experiment process and analysis, which is followed by the process and analysis of our focus group. In the last section, we discuss the contributions and limitations of the research.

2 Research background

2.1 Review tag function

Archak et al. (2011) described a simple version of a tag-generating tool. The tag-generation process includes and is not limited to feature extraction, sentiment classification and text summarization based on reviews (Hu and Liu 2004). As a result, a set of noun phrases and the respective sentiment are produced, which corresponds to product features and their evaluation.

Tag function could be presented with different appearances. We take the current tag function on Tmall.com as our target prototype. Emerged from Taobao.com, Tmall.com focuses on B2C transactions and provides higher quality and compliances. Current practice in Tmall.com limits the total number of review tags to ten, so that the function at most displays the top ten features and their sentiment according to the feature frequency ranking. On Tmall.com, the positive tags are presented in red color and negative tags are in green color. Both types of tags display the frequency of the features being mentioned in reviews. And normally on Tmall.com, if there are negative tags, they are often placed after all other positive tags, being less noticeable.

To obtain a direct understanding of the review tag function for popular products on Tmall.com, we selected products from different categories and gave an overview of the tag function usage on Tmall.com. We used the name of each product category as keyword and randomly selected ten products for each keyword from the top 100 most-sold items in the search results. We collected their price, sales, promotion, tag function usage information and review information from their pages.

In total, we collected information for 340 popular products, and around 67% (228 out of 340) of them were using the tag function. Table 1 shows the descriptive statistics of the collected data. For those using the tag function, the average number of tags is 9.57. The average numbers of positive and negative tags are 8.12 and 1.45 respectively, which partially supported that negative reviews are much fewer than positive ones (Dellarocas 2003; Hu et al. 2017).

As shown, some of the features showed differences between the two subsamples, with tag function and without tag function. We performed t-tests and found that the average review rating in Subsample 1 is significantly higher than the rating in Subsample 2 ($M=4.812$ vs 4.598 , $t(338)=3.19$, $p=0.0015$). Also, the average ratings for Subsample 1 products are above 4.4, indicating that products using the tag function are highly rated.

In Subsample 1, price is marginally lower than the price in Subsample 2 ($M=223.522$ vs 376.267 , $t(338)=1.95$, $p=0.052$). The sales and the number of reviews did not show significant differences across the two subsamples.

With the above being shown, we identify a typical appearance of the review tag function for popular products on Tmall.com. The typical appearance consists of nine or ten tags, including one or two negative tags and eight or nine positive tags. Figure 3 depicts a sample of typical tag function appearance on Tmall.com, which includes nine positive tags and one negative tag.

Table 1 Descriptive statistics of review tag function usage on Tmall.com

Product features	Subsample 1: with Tag function					Subsample 2: without Tag function				
	N	Mean	Std. Dev	Min	Max	N	Mean	Std. Dev	Min	Max
Price	228	223.52	618.17	0.43	5,299	112	376.27	789.18	1	3,999
Sales	228	16,953.61	23,516.47	17	160,436	112	13,301.85	25,433.46	1	115,150
No of reviews	228	51,860.64	84,039.34	61	480,625	112	39,075.18	78,821.53	0	410,389
Average rating	228	4.81	0.08	4.40	5	112	4.60	1.01	0	5
No of Tags	228	9.57	1.85	2	10	112	n/a	n/a	n/a	n/a
No of Positive Tags	228	8.12	1.65	2	10	112	n/a	n/a	n/a	n/a
# of Negative Tags	228	1.45	0.93	0	4	112	n/a	n/a	n/a	n/a



Fig. 3 Typical tag appearance on Tmall.com

For popular products, an all-positive appearance of the review tag function provides a strong support to the product from prior buyers, which intuitively could be an ideal situation for sellers to promote their products. Therefore, besides a typical review tag appearance, we are also interested in the all-positive appearance, so that to understand their difference in affecting consumers' perceptions towards the product as well as its reviews.

2.2 Attribution theory

Attribution theory was developed in the field of social psychology for understanding how people perceive and evaluate the behaviors of others (Heider 1958). Attribution refers to the perception or the inference of cause. Attribution research is concerned with all aspects of causal inferences.

There are three assumptions in the theory (Försterling 1986). The first one is that people interpret behavior in terms of its causes and that these interpretations play an important role in determining reactions to the behavior (Kelley and Michela 1980; Heider 1958). The second assumption is that people are generally motivated to gain a realistic understanding of the causes that have led to different events in their personal domain (Heider 1958). Third, it is assumed that a causal understanding serves the function of attaining personal goals and survival (Kelley 1987).

At first, when Heider proposed the attribution theory in his book (Heider 1958), he distinguished causes of actions into two basic types, personal or internal causes, and situational or external causes. For example, if Tom recommends a movie to others, his action might be due to his internal taste for this movie, or to other external factors, e.g. every audience of the movie on that day is given a voucher.

Later, Kelley extended and elaborated on how individuals infer causes (Kelley 1967, 1973). According to his topology of person–stimulus–circumstances, general attributions could be made to the person (Tom's taste for the movie), the stimulus (the movie quality), and circumstances (special gifts for the audiences).

Information is used to facilitate an observer's attribution of a behavior. One important piece of information is consensus information. Kelley (1973) proposed that when people make attribution to an actor's behavior, they would take into consideration how others behave in the same situation. The term, consensus information, is used to refer to the way in which other people respond to the stimulus. Take Tom's recommending the movie as an example. If everyone who has watched the

movie recommends it, we would observe high consensus. So, when most others behave in a similar way to Tom, i.e., there is high consensus, we, as observers, tend to attribute to product-related causes, which are external to Tom. But as consensus decreases, our attribution would be more internal to him (McGill 1989).

The impact of consensus varies with a number of mediating factors, such as the salience of the consensus, representativeness of the sample, relevance of the consensus and so on (Kassin 1979). Among them, sample representativeness seems to be of greater importance. As noted in Kassin (1979), one of the most severe limitations of consensus is that, the consensus is based on the observation of a limited sample. The consensus utilization requires observers' beliefs in the sample representativeness (Wells and Harvey 1977), otherwise, the validity of consensus would be violated.

3 Hypotheses development

The theoretical development of attribution theory had enabled consumer research to explore a variety of studies, specifically the research line which examines the process by which individuals assign causal agency to outcomes experienced by others in online reviews (He and Bond 2015; Sun 2012).

When making an online purchase decision, individuals would observe others' product experience expressed in reviews. Online review context readily fits into the principle proposed by Kelley under the assumption that consumers interpret a distribution of review assessment to understand the possible outcomes that they may experience in the product purchase situation (He and Bond 2015). When processing the review information, they would attribute the review content to either product-related features or reviewers' characteristics. To avoid potential loss in their purchase decision, consumers tend to seek for options with favorable and critical consensus from others.

There have been studies using attribution theory within the context. For example, Chen and Lurie (2013) studied the effect of temporary contiguity on reviews' causal attribution. They found that when reviews' writing closely follows consumption, positive reviews would be more attributed to products and hence be more valued. In examining the review dispersion, He and Bond (2015) found that consumers taste similarity moderates the relationship between review dispersion and the attribution to reviewers.

In online markets, potential consumers obtain consensus on product evaluation from two sources, the overall rating and the summarized review tags. The overall rating represents the average numeric evaluation towards the product, reflecting a simple holistic assessment of the product. As for review tags, each of the tags represents a set of product-related features and the respective sentiment, either positive or negative. Since review tags demonstrate positive or negative evaluation of the most frequent features, they can also be regarded as consensus information. Hence the dispersion in review tags' positivity/negativity indicates the strength of the consensus opinions on various attributes.

According to the attribution theory, the consensus information affords a basis for confidence in one's judgment (Kelley 1973). If a product has a high rating, potential consumers tend to attribute the consensus to product-related causes, prevailing confidence in a positive product evaluation (Qiu and Li 2010).

Similarly, if a popular product is shown with all-positive tags or typical tag appearance, the dominant positive consensus would be further strengthened, yielding a higher evaluation towards the product comparing to the situation when only overall rating is shown. Therefore, we hypothesize the following.

Hypothesis 1 When the highly-rated popular product is shown with all-positive tags or typical tag appearance, consumers' product evaluation is more likely to be higher comparing to when there is no tag function.

Other than affecting product evaluation, the presence of the review tag function could also influence consumers' utilizing review information.

For review systems without tag function, consensus comes from the overall rating information. Suppose a popular product is highly-rated, its high overall rating could result in consumers' tendency of favoring positive reviews, as positive reviews are more likely to be attributed to product-related causes. On the other hand, a consumer might tend to attribute negative reviews to reviewer-attributed causes, and perceive them as less helpful in reflecting the product's true evaluation. Therefore, a positive bias could be yielded.

For review system with tag function, besides a high rating, all tags for a product are positive, hence consumers' preference for positive reviews would be further strengthened due to the more salient positive consensus. In terms of the negative reviews, since negative evaluation is not included in consensus information, consumers tend to regard negative reviews as attributable to reviewers and perceive them as less helpful. We hypothesize the following.

Hypothesis 2a When a highly-rated popular product is shown with all positive tags, consumers are more likely to perceive positive reviews as helpful comparing to when there is no tag function.

Hypothesis 2b When a highly-rated popular product is shown with all positive tags, consumers are less likely to perceive negative reviews as helpful comparing to when there is no tag function.

However, the above discussion for all-positive situation neglects the role of the consensus' perceived validity (Wells and Harvey 1977). When a product is shown with only positive tags, potential consumers would assume the consensus sample to be positively biased. In such cases, the effect of high positive consensus might perish. Consumers might be willing to obtain negative opinions of the product, so as to acquire more non-biased views derived from product-related reflection. Therefore, a negativity bias in reviews could emerge for popular products with

all-positive tag function appearance. So, contradicting Hypothesis 2b, we also hypothesize the following.

Hypothesis 2c When a highly-rated popular product is shown with all positive tags, consumers are more likely to perceive negative reviews as helpful comparing to when there is no tag function.

Different from all-positive tag appearance, a typical tag appearance might draw consumers' attention to the negative tag and lessen their confidence in the high product evaluation. Therefore, the preference for positive reviews would be reduced, but the preference for negative reviews prevail due to its value in reflecting product-related information. We hypothesize the following.

Hypothesis 3a When a highly-rated popular product is shown with typical tag appearance, consumers are less likely to perceive positive reviews as helpful comparing to when there is no tag function.

Hypothesis 3b When a highly-rated popular product is shown with typical tag appearance, consumers are more likely to perceive negative reviews as helpful comparing to when there is no tag function.

4 Study 1. Experiment

We first utilized an online experiment to test our hypotheses on the review tag function. We manipulated the tag function appearances with a between-subjects design.

4.1 Manipulation

We selected a computer product as our manipulation target for the experiment. Computer is a common product among students. As we would use student sample to conduct our survey, we expected the selected product to be a possible choice considered by the sample group.

A computer product with the tag function from Tmall.com was selected as our prototype to create our mock pages for manipulation. The experimental manipulation used in the study was developed according to the above overview of the randomly collected products in Table 1.

To test our hypotheses, we used two appearances of the tag function. The first one, the appearance with ten positive tags (we call it 10PT and thereafter), and the second one, the typical appearance with nine positive tags and one negative tag (we call it 9PT1NT). Also, as a control, we create an extra group with no tag function shown in the webpages. Therefore, in total, three appearances of tag function were designed, and participants were randomly assigned to each of the three groups.

For each group, we kept every element the same except for the tag appearances. The product descriptions were copied from the prototype's webpage, while

eliminating irrelevant information such as recommended products offered by the seller.

In terms of the product reviews, we first decided the number of reviews we would use for the mock page. Since normally, people would not read all the reviews for a popular product. As each page of reviews contains only 20 pieces of reviews, consumers would not try to click through all the pages to obtain information. We arbitrarily decided to collect around 200 reviews from the prototype webpage, which would induce at least nine or ten clicks to finish reading all the reviews.

Tag function had been used in our prototype webpage. It contained ten positive tag labels. For each tag in the prototype webpage, we collected the proportional number of reviews. So, in our mock webpage, the frequency of each review tag showed the number of reviews collected from the corresponding review tag.

Besides the reviews from all the positive tags, we created a negative tag. To avoid bias for any specific product feature, the negative tag was made with a general label on purpose. After considering the normality of review tag frequency distribution and comparing the current frequencies of positive tags, we set the frequency of the only negative tag as the same as the second lowest frequency of positive tags collected above. And the reviews for the negative tags were selected manually. Two coders independently judged whether the selected reviews matched with the negative tag and whether their evaluations were consistent.

Therefore, 210 reviews for a total of eleven tags were prepared. They were all displayed in each of the three groups. But in 10PT group, only ten positive tags were shown in the tag function area, while in 9PT1NT, nine positive tags and a negative tag were shown.

4.2 Stimulus preparation

In preparation of the stimuli for our experiment, we needed to identify the sentiment of the text reviews. First, we randomly selected 60 reviews from the prototype's Tmall webpage, with the intention of selecting five from them as our stimulus reviews.

To assess the sentiment of the reviews, we also recruit two coders to evaluate the sentiment of the reviews. The coders were unaware of the study's purpose. Each coder was presented review texts and were asked to rate whether the review is positive, negative or neutral. To prevent potential biases, we did not present any description of the computer product. Second, to prevent ordering bias, each coder received a different random order of reviews. We also made sure the two coders complete the tasks independently.

For the reliability of the coded review sentiment, two reliability scores were calculated for each of the product reviews (Lombard et al. 2002). We obtained 0.8208 on Krippendorff's alpha, which exceeded the recommended value 0.70 (Krippendorff 2004). We also had 96.67% on Cohen's kappa which also exceeded the recommendation value of 0.80 (Cohen 1968).

As the sentiment coding is deemed reliable, we dropped the two reviews with coders' disagreed evaluation, and grouped the remaining ones by their sentiment.

Then we randomly selected two from positive reviews, one from neutral reviews and two from negative reviews as our stimulus.

4.3 Experiment procedure

Participants were recruited from an IS course at a Chinese university. Participants received extra credit for their participation, and they were randomly assigned to three experimental conditions.

First, after giving their consent, participants were instructed to read the product information from a webpage of Tmall.com for at least 3 minutes. All participants were able to read Chinese to participate in the survey. They were free to browse all the descriptions or reviews, or to read selected reviews by clicking each review tag of the product. Next, we used an attention test to ensure whether participants did read the product information.

After that, participants received the survey questions, followed by demographic items. They were first asked to state their overall evaluation of the product. Next, posing randomly in the sequence, five stimulus reviews are presented to participants, one at a time. By reading each review, participants were asked to report their perceptions of the review helpfulness. The helpfulness perception was measured by using three items adapted from Sen and Lerman (2007) and Yin et al. (2014). All items are of 9-point semantic differential-scale, as presented in Table 2. Also, to understand the participants' causal attribution, we asked an additional question measured by a 9-point semantic differential scale: "To what extent are the contents of the consumer review based on the product?"

4.4 Results

In total, 101 participants completed the experiment. Before we conducted further analyses, we dropped four responses that failed to answer correctly in the attention test. We also dropped five responses for that their response duration was less than 3 minutes.

Table 2 Measurement items

Variable	Items
Product Evaluation	
PE1	How do you think of the product reviewed? [Very good/Very bad]
PE2	How do you think of the product reviewed? [Very desirable/Not at all desirable]
Helpfulness perception	
HP1	How do you think of the review? [Very useful/Not at all useful]
HP2	How do you think of the review? [Very accurate/Not at all accurate]
HP3	Assuming that you were thinking of buying this product, how likely would you be to use the above consumer review in your decision-making? [Very likely/Very unlikely]

Table 3 Demographics

Participants	Items	Percentage	Participants	Items	Percentage
Gender	Male	63.04%	Age	19–20	21.74%
	Female	36.96%		21–22	65.22%
				23–24	13.04%
Monthly Expense	Less than 500 rmb	1.09%	Years of Using Taobao.com	Less than 1 year	1.09%
	500–999 rmb	25.00%		Around 3 years	45.65%
	1000–1499 rmb	53.26%		Around 4 years	27.17%
	1500–1999 rmb	11.96%		Around 5 years	13.04%
	2000–2499 rmb	4.35%		Around 6 years	6.52%
	2500–2999 rmb	1.09%		Around 7 years	4.35%
	3000 rmb or more	3.26%		Around 8 years	1.09%
			10 years or more	1.09%	

Table 4 Descriptive statistics and correlation matrices

	Mean (Std.Dev)	PE	PE1	PE2	HP	HP1	HP2	HP3
PE	6.57 (1.32)	1						
PE1	6.88 (1.41)	0.88	1					
PE2	6.25 (1.56)	0.90	0.58	1				
HP	4.97 (2.30)	0.14	0.13	0.12	1			
HP1	5.33 (2.50)	0.15	0.14	0.13	0.94	1		
HP2	4.69 (2.28)	0.15	0.13	0.13	0.94	0.83	1	
HP3	4.88 (2.56)	0.10	0.11	0.06	0.94	0.83	0.83	1

After removing all the invalid responses, we had a total of 92 complete responses. Demographics of the participants are shown in the Table 3.

Table 4 shows the descriptive statistics and correlation matrix of constructs.

4.4.1 Measurement reliability and validity

For exploratory factor analysis, we used principal components analysis with both varimax and oblimin rotations (DeVellis 2016). The result consistently provided two factors. All of our items loaded as expected on their focal factors and loaded less than 0.4 on the other factor. Therefore, we retained all indicators.

Next, we conducted confirmatory factor analysis to examine the reliability and validity of the two constructs in the study. Cronbach's alpha for product evaluation was 0.734 and that for perceived helpfulness was 0.935. Also, the composite rho values for the two constructs were 0.882 and 0.958 respectively, indicating sufficient internal consistency and reliability (Nunnally 1967; Hair et al. 2010). To establish convergent validity, we tested the average variances extracted (AVEs) for the two constructs, yielding 0.789 and 0.885, which were well above 0.5 and demonstrated

convergent validity (Hair et al. 2010). Finally, the data also passed the Fornell and Larcker's test, indicating discriminant validity (Fornell and Larcker 1981).

4.4.2 Hypotheses testing

Our first hypothesis is whether product evaluation varies between with- and without-tag-function groups. A t-test was performed to examine the difference in product evaluations across different conditions.

In line with our hypotheses, the result showed that the difference in product evaluation between with-tag-function group and control group was significant ($M=6.775$ vs 6.172 , $t(90)=2.12$, $p=0.037$). Evaluation to product shown with tag function is significantly higher than the evaluation to product shown without tag function, despite that all the other information is the same in both conditions.

Next, we inspected the three groups with more details. We did separate t-tests to examine the two with-tag-function groups comparing to the control group. For the 10PT group, we found that the product evaluation is higher than the evaluation in control group with marginal significance ($M=6.722$ vs 6.172 , $t(57)=1.545$, $p=0.064$). In the 9PT1NT group, the product evaluation is significantly higher than that in control group ($M=6.818$ vs 6.172 , $t(63)=1.997$, $p=0.025$). Figure 4 shows the difference in product evaluation at with 95% confidence intervals.

Therefore, we conclude that our Hypothesis 1 is supported. When product is shown with the two types of tag function, it is evaluated higher comparing with when it is shown without tag function.

To test our hypotheses on tag function's impact on reviews, we converted our data into review level. First of all, we tested whether the positivity bias or negativity bias exist. We performed an ANOVA test to investigate the difference in perception level across positive, neutral and negative sentiment. The results confirmed that both two biases exist ($M_{\text{positive}}=4.780$ vs $M_{\text{neutral}}=2.554$ vs $M_{\text{negative}}=6.350$, $F(2,$

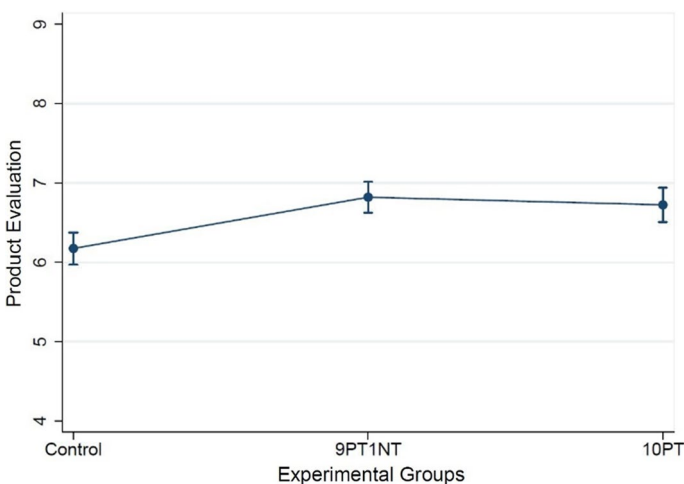


Fig. 4 Product evaluation

457)=132.43, $p < 0.001$). Reviews' helpfulness perception increases when reviews are either positive or negative comparing to when reviews are neutral. Figure 5 shows the differences at with 95% confidence intervals.

The responses to our supplementary question on participants' causal attribution also supported our assumption that the more the reviews are attributed to product features, the more they are perceived as helpful. Review helpfulness and the causal attribution were highly correlated and an additional ANOVA test reported significant and consistent increases in helpfulness perception across different attribution level.

In order to test Hypothesis 2a and 3a together, we selected the responses for positive stimulus reviews. As each participant responded to more than one positive review, a repeated-measure ANOVA was performed to examine the helpfulness perception across the three groups. The results showed that the helpfulness perception for positive reviews is not significantly different between 10PT and control groups ($M_{\text{control}}=4.771$, $M_{10\text{PT}}=5.08$, $F(1,115)=0.47$, $p=0.6255$), or between 9PTINT and control groups ($M_{\text{control}}=4.771$, $M_{9\text{PTINT}}=4.571$, $F(1,127)=0.49$, $p=0.6135$). Thus, Hypothesis 2a and Hypothesis 3a are rejected, yet we still see the difference in their marginal effect as shown in Fig. 6—helpfulness perception for positive reviews is slightly higher (but insignificant) for products shown with all positive tags.

We use the responses for negative stimulus reviews to test Hypothesis 2b, 2c and 3b simultaneously. Again, we performed a repeated-measure ANOVA. The results showed the helpfulness perception for negative reviews is significantly different across the three versions ($M_{\text{control}}=6.307$, $M_{9\text{PTINT}}=6.071$, $M_{10\text{PT}}=6.741$, $F(2,180)=3.16$, $p=0.045$). Two separate t-tests were performed to examine the two with-tag-function groups comparing to the control group. Results showed that in 10PT group, the helpfulness perception is significantly higher than that in control group ($t(116)=1.8026$, $p=0.037$). And it showed no difference for helpfulness perception in between the 9PTINT group and the control group (t

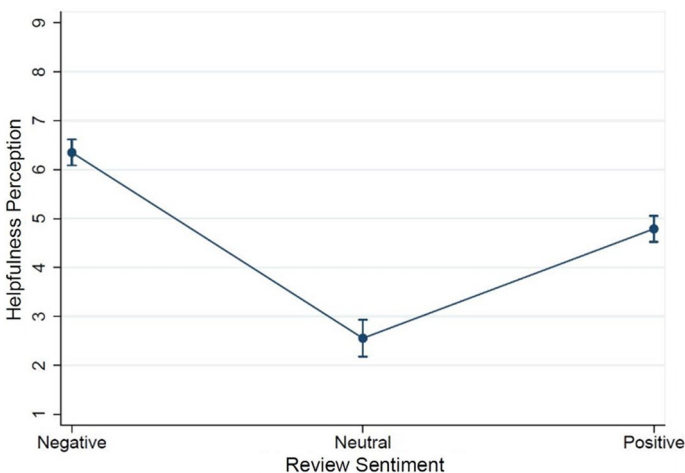


Fig. 5 Positivity bias and negativity bias

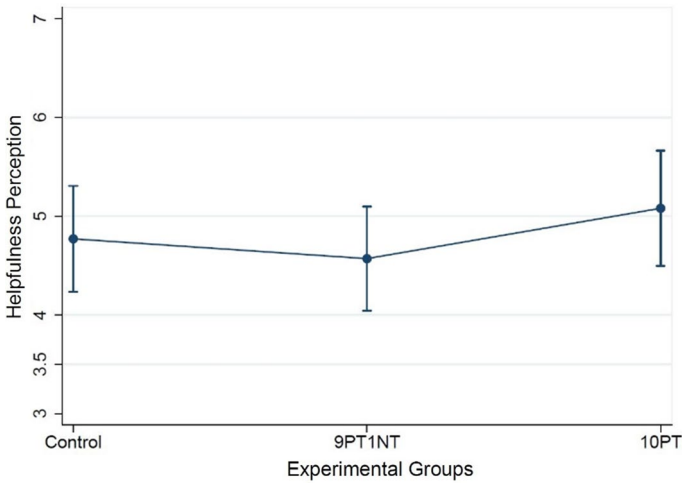


Fig. 6 Helpfulness perception for positive reviews

(128) = 0.8995, $p = 0.815$). Figure 7 shows the differences in helpfulness perception at with 95% confidence intervals.

Therefore, Hypothesis 2b and 3b are rejected, but Hypothesis 2c is supported. The helpfulness perception for negative reviews did not show difference between a typical review tag setting and no-tag-function setting. In contrast, when review tags are shown and all positive, negativity bias is larger comparing to it is in no-tag-function setting.

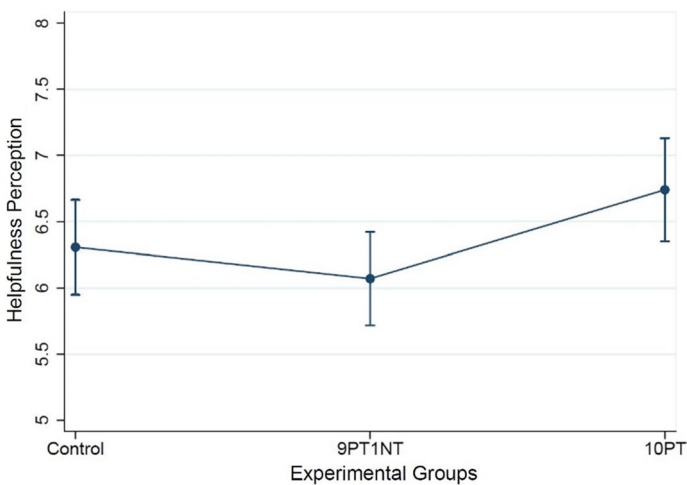


Fig. 7 Helpfulness perception for negative reviews

We also tested on the neutral reviews. As presented in Fig. 8, results of an ANOVA test showed no significant difference in helpfulness perception among three groups ($M_{\text{control}}=2.885$, $M_{9\text{PT}1\text{NT}}=2.172$, $M_{10\text{PT}}=2.630$, $F(2,89)=1.47$, $p=0.236$).

5 Study 2. Focus group

Next, to help further understand the results shown in our study 1, a qualitative study was chosen. We used focus groups to understand consumers thoughts of product evaluation and information perception in online markets, as consumers might not be used to talking about the topic though they are used to shopping online (Morgan 1997).

5.1 Design and procedure

Eligible participants are those that speak Chinese, and had experience in shopping on Tmall.com. We recruited participants in a university and selected four to join our focus group (Ryan et al. 2014).

Focus questions developed based on our research questions, as shown in Table 5. Questions were related to the three versions of tag-function appearances (as in Study 1) on mattress product. The focus group was conducted in a classroom and lasts for 70 min. It was conducted by two researchers: a moderator and an observer.

The moderator first asked a few broad question relevant to information search in online shopping. Then the questions were asked following the focus question list. And all participants expressed their opinions to the key questions and they were also encouraged to talk and interact with each other. In the last fifteen minutes,

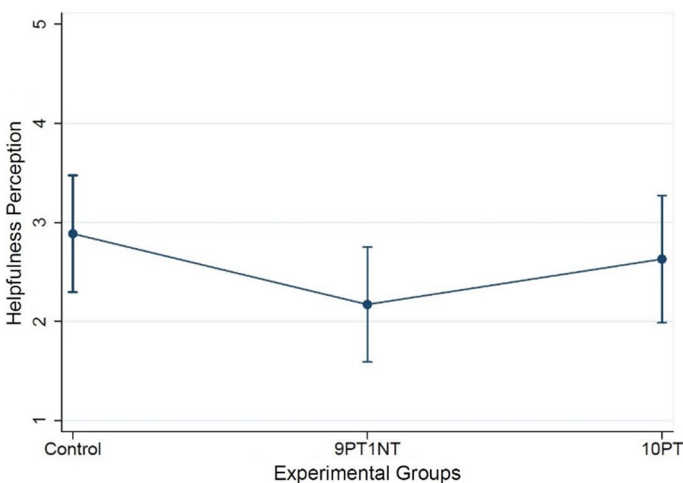


Fig. 8 Helpfulness perception for neutral reviews

Table 5 Focus questions*Introduction*

How long have you been shopping online?

Which online shopping platform do you usually use? Do you use Tmall.com?

What types of products do you often shop online?

Key questions

Can you describe your product evaluation process when you shop online?

What types of information do you particularly look for when you shop online?

Suppose you are considering buying a mattress online, please read the product page and browse the information as you want

How do you like the product?

How do you use the review tags, and how do you think of them?

How do you think of the product if the review tags are shown with all positive tags, and if they are shown with nine positive but one negative tags?

How do you think of the positive/negative reviews? Why is the reviewer holding such opinion?

Review questions

Do you want to add anything related to the process of product evaluation during online shopping?

Do you want to add anything related to the process of information search during online shopping?

Can you help review all the aspects that you would consider in online product evaluation and information search?

participants were asked to review the key aspects that they considered when evaluating a product online.

5.2 Data analysis

The focus group was audiotaped and transcribed by verbatim. The transcripts were used as texts for analysis. One member of the research team who has experience in focus group studies analyzed the data. Another member checked the findings and reviewed the coding of the data.

5.3 Findings

On average, the four participants had eight years of online shopping experience and were aged between 21 and 25 years old, including one male and three female, which is very similar to our sample used in Study 1. The participants all had slight knowledge about mattress product. When they were asked to evaluate the product page, they intuitively expressed their opinions. As there was no additional topics found in the last fifteen minutes, the data saturation was likely to be reached.

5.3.1 Benefits of tag function

The analysis showed that tag function is beneficial to the information search in online markets. The usefulness of the function is consistently agreed by all the

participants. Positive tags indeed provide confidence in their reasoning process for the overall positive evaluation, confirming our H1. In addition, they use the tag function not only to learn from prior buyers' opinions on the product, but also to obtain an idea of the key product attributes.

It is very convenient for me to evaluate a product, especially when I have many options. It helps me to identify the product with the most number of red (positive) tags... [Participant 2]
(with the function)... I can learn about the product and to see whether the tags are of my interests. If yes, then I'll take a look at the relevant reviews. Otherwise, I'll just ignore it. [Participant 3]

5.3.2 Negativity bias

It was seen that participants considered the positive information as less important as the negative information. All participants expressed their reliance on negative comments. But the positive reviews are largely neglected by them.

When there is no negative tag, I would start read the reviews and search for negative review comments on purpose... I normally do not read positive comments [Participant 1]

When there are negative tags, I'll only click on the negative tags and read the reviews... most products are positive nowadays... If a product has one or more negative tags, I would probably put the product in my 'waiting' list. Once I found one with better tag appearances, I might never go back. [Participant 3]

With the analysis, we found a possible resolution of the rejection of our H2a, H2b and H3. As people would ignore the positive reviews, the positive comments might not generate difference in consumers' perception among the three versions in Study 1.

Also, for a typical tag appearance, the product's negative comments could be easily found, therefore, the negative comments are easy to be located. But for all positive tag appearance, the negative comments are unnoticeable at the first glance. So, consumers are more likely to search for it purposely to support the overall positive evaluation, which enhances the importance of negative comments. Thus, our H2c is also supported in Study 2.

5.3.3 Attribute features

It was found that participants do not merely obtain the information of the number of positive tags and negative tags, they would also take the attributes' features into their consideration when using the tag function. If the attribute is unimportant, or its comments rely much on subjective judgment, participants do not attach a larger weight to the attribute being mentioned.

My taking of negative reviews depends on whether the attribute is important to me. If the attribute is not important, then I don't care... if many people com-

ment that a specific product looks ugly, I won't take the negative tags that seriously as I do know how it looks and I don't think it is ugly. [Participant 4]

Therefore, if there is a negative tag on an unimportant attribute of a product, the seller need not to worry about it. Consumers may not rely on its judgment, and there is also no need for them to particularly look for negative comments.

5.3.4 Attribution process

The analysis showed that the participants inferred the causes of a product opinion. But again, the inference is dependent on the attribute features, i.e. the importance and whether or not can be judged objectively.

I will put more weights on the product-side (causes). [Participants 1]
if I found that a reviewer had very similar needs or requirements for it (mattress product), I may particularly rely on her judgments. [Participants 3]
It depends. If the reviewer says the mattress is too hard, I may not believe it. It is too subjective. But if the reviewer says the mattress has a strange odor, I will definitely take it and won't consider the product any more. [Participant 4]

6 Discussion

6.1 Key Findings and contributions

Using the two studies above, we concluded our findings of hypotheses testing. In study 1, by directly manipulating the tag function usage, we provide evidence supporting part of our hypotheses. And in Study 2, we used qualitative methods to supplement the results in Study 1 and explored the consumers' product evaluation and information search process.

The research has three major findings. First, in general, participants give higher product evaluation when tag function is shown, providing stronger confidence in attributing the product quality. Also, this finding echoes with other theories, e.g. the signalling theory (Connelly et al. 2011) and information processing theory (McGuire 1968), that an augmented perception/attitude could be formed when a strong signal/information cue is sent via information systems. Second, tag function does not lead to significant difference in reviews' helpfulness perceptions for positive reviews. Closer inspection using interview revealed that positive comments in review platform are too common to draw consumers' attention, yielding their indifference perception towards positive reviews. Third, overwhelming positivity in tags leads to a higher negativity bias in review perceptions comparing to less positivity in tags (nine positive tags and one negative tag) and no-tag-function situation. Theory development and later interview suggested that validity of extremely positive information could be questioned and require consumers' additional examinations.

Our study makes the following contributions. First, we fill in the research gap of investigating the impact of the review tag function in online markets. Prior studies on tag function largely focused on algorithms and techniques for review summarization

(Hu and Liu 2004; Tsaparas et al. 2011), our research contributes to the understanding of the its role in online markets and shows its impact on product evaluation as well as product review's helpfulness perception.

Second, our findings add value to the current understanding of consumers' product and review perception during review consumption. As IS scholars have increasingly recognized the important impact of review texts (Chevalier and Mayzlin 2006; Archak et al. 2011; Bao and Chau 2016), tag function, as a tool integrating the textual power in reviews, shows its effectiveness in influencing consumers information consumption. Also, our conclusions of helpfulness perception supplements the existing research on the role of review variance. As prior studies reaches inconclusive results on the impact of review variance on review helpfulness perception, we found that sellers could benefit from a slight variance of product opinions.

Third, by adapting the attribution theory, our research provided a new theoretical lens to understand the role of tags function in online marketing. While previous studies in the domain are largely focusing on examining online reviews, our paper took both product evaluation and reviews into consideration under the context of tag functions.

6.2 Practical implications

Though the review tag function is not popularly used in current online markets, our findings provide insights for sellers or platforms in deciding its adoption on popular products. A tag function can not only eliminate information overload and inform sellers of consumers opinions in an aggregated level, but also has the potential of offering consumers the most important aspects for considering a product, which might in turn influence the product evaluation and purchase intention.

More specifically, a typical tag function appearance will produce the best performance of product reviews, enhancing product evaluation and preventing overreaction to the positive or negative reviews. As consumers have the tendency of looking for negative comments even with all positive situations, sellers shall not concern too much about a few negative comments on unimportant features or attributes. To the contrary, they should avoid presenting only positive tags on their product pages, in case their mere negative comments would backfire on the good product reputation.

In addition, various empirical results have found the helpfulness perception of reviews are influenced by many factors (Yin et al. 2016; Hu et al. 2017; Chen and Lurie 2013). Supplementing their research, our findings focus on tag function, which is a promisingly useful application in online markets. Since we found that the adoption of tag function would bring higher negativity bias when all tags are positive, practitioners are encouraged to reconsider their adoption decision, so that they would not be backfired by their well-managed positive reviews.

6.3 Limitations

Our study is not without limitations. Although we studied two appearances of the review tag function in online markets which are more commonly seen in online

markets, other appearances could also be common and worth examining. Based on our result, it is interesting to learn how an increased portion of negative review tags influence the consumers' product and review perceptions comparing to no-tag-function situation. In addition, currently we use the count of positive/negative tags for describing the tag function appearances, whereas the ratio of the positive and negative tags might be more appropriate for further analyses based on larger data sets.

In studying the impact of consensus information in consumers' causal attribution on product reviews, we emphasized only on the sample representativeness for consensus, leaving other possible elements being unexamined, such as the magnitude and relevance of consensus (Kassin 1979).

Besides, from the findings of Study 2, it is revealed that the attribution process of review opinions is dependent on the importance of attributes being commented in the tags. Subjective perception, preference or consumer knowledge could potentially influence the process, of which the findings may contribute to the product review research.

Also, our target prototype is the review tag function for a computer product and a mattress product on Tmall.com. Generalizability issue of the research could be raised due to the differences existing in product type (Huang et al. 2009), culture and languages as well as in shopping conventions. Though we found consistent results for the two product categories, future research could draw on the function usage on other products, or other platforms to explore its role in perceptual and behavioral outcomes of consumers.

Acknowledgements This research is supported in part by the General Research Fund from the Hong Kong Research Grants Council (#17514516B), the Seed Funding for Basic Research from the University of Hong Kong (#104003314), and the grants from National Natural Science Foundation of China (Project No. 71701061).

References

- Archak N, Ghose A, Ipeirotis PG (2011) Deriving the pricing power of product features by mining consumer reviews. *Manage Sci* 57(8):1485–1509. <https://doi.org/10.1287/mnsc.1110.1370>
- Ba S, Pavlou PA (2002) Evidence of the effect of trust building technology in electronic markets: price premiums and buyer behavior. *MIS Q* 26(3):243–268. <https://doi.org/10.2307/4132332>
- Bao Z, Chau M (2016) A schema-oriented product clustering method using online product reviews. In: International Conference on Information Systems (ICIS 2016)
- Chen Z, Lurie NH (2013) Temporal contiguity and negativity bias in the impact of online word of mouth. *J Mark Res* 50(4):463–476. <https://doi.org/10.1509/jmr.12.0063>
- Chen P-Y, Hong Y, Liu Y (2017) The value of multi-dimensional rating systems: evidence from a natural experiment and randomized experiments. *Manage Sci*. <https://doi.org/10.1287/mnsc.2017.2852>
- Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *J Mark Res* 43(3):345–354. <https://doi.org/10.1509/jmkr.43.3.345>
- Cohen J (1968) Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull* 70(4):213. <https://doi.org/10.1037/h0026256>
- Connelly BL, Certo ST, Ireland RD, Reutzel CR (2011) Signaling theory: a review and assessment. *J Manag* 37(1):39–67
- Dellarocas C (2003) The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Manage Sci* 49(10):1407–1424. <https://doi.org/10.1287/mnsc.49.10.1407.17308>

- DeVellis RF (2016) Scale development: Theory and applications, vol 26. Sage publications, Thousand Oaks
- Dimoka A, Hong Y, Pavlou PA (2012) On product uncertainty in online markets: theory and evidence. *MIS Q* 36(2):395–426
- Forman C, Ghose A, Wiesenfeld B (2008) Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets. *Inf Syst Res* 19(3):291–313. <https://doi.org/10.1287/isre.1080.0193>
- Fornell C, Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. *J Mark Res* 18(1):39–50. <https://doi.org/10.2307/3151312>
- Försterling F (1986) Attributional conceptions in clinical psychology. *Am Psychol* 41(3):275. <https://doi.org/10.1037/0003-066X.41.3.275>
- Hair JF, Anderson RE, Babin BJ, Black WC (2010) Multivariate data analysis: a global perspective, vol 7. Pearson, Upper Saddle River, NJ
- He SX, Bond SD (2015) Why Is the crowd divided? Attribution for dispersion in online word of mouth. *J Consum Res* 41(6):1509–1527. <https://doi.org/10.1086/680667>
- Heider F (1958) The psychology of interpersonal relations. Psychology Press
- Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp 168–177. <https://doi.org/10.1145/1014052.1014073>
- Hu N, Pavlou PA, Zhang J (2017) On self-selection biases in online product reviews. *MIS Quart* 41(2):449. <https://doi.org/10.25300/MISQ/2017/41.2.06>
- Huang P, Lurie NH, Mitra S (2009) Searching for experience on the web: an empirical examination of consumer behavior for search and experience goods. *J Mark* 73(2):55–69. <https://doi.org/10.1509/jmkg.73.2.55>
- Kassin SM (1979) Consensus information, prediction, and causal attribution: a review of the literature and issues. *J Pers Soc Psychol* 37(11):1966. <https://doi.org/10.1037/0022-3514.37.11.1966>
- Kelley HH (1967) Attribution theory in social psychology. In: Nebraska symposium on motivation. University of Nebraska Press
- Kelley HH (1973) The processes of causal attribution. *Am Psychol* 28(2):107. <https://doi.org/10.1037/h0034225>
- Kelley HH (1987) Attribution in social interaction. In: Ones EE, Kanouse DE, Kelley HH, Nisbett RE, Valins S, Weiner B (eds) Attribution: Perceiving the causes of behavior. Lawrence Erlbaum Associates Inc, Hillsdale
- Kelley HH, Michela JL (1980) Attribution theory and research. *Annu Rev Psychol* 31(1):457–501. <https://doi.org/10.1146/annurev.ps.31.020180.002325>
- Krippendorff K (2004) Content analysis: an introduction to its methodology. Sage, Thousand Oaks
- Lombard M, Snyder-Duch J, Bracken CC (2002) Content analysis in mass communication: assessment and reporting of intercoder reliability. *Hum Commun Res* 28(4):587–604. <https://doi.org/10.1111/j.1468-2958.2002.tb00826.x>
- Lu X, Ba S, Huang L, Feng Y (2013) Promotional marketing or word-of-mouth? Evidence from online restaurant reviews. *Inf Syst Res* 24(3):596–612. <https://doi.org/10.1287/isre.1120.0454>
- McGill AL (1989) Context effects in judgments of causation. *J Pers Soc Psychol* 57(2):189. <https://doi.org/10.1037/0022-3514.57.2.189>
- McGuire WJ (1968) Personality and attitude change: An information-processing theory. In: Greenwald AG, Brock TC, Ostrom TM (eds) Psychological foundations of attitudes. Academic Press, New York, p 196
- Morgan D (1997) The focus group guidebook, vol 1. Sage publications, Thousand Oaks
- Mudambi SM, Schuff D (2010) What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quart* 34(1):185–200. <https://doi.org/10.2307/20721420>
- Nunnally JC (1967) Psychometric theory, 3rd edn. McGraw-Hill, New York
- Pan Y, Zhang JQ (2011) Born unequal: a study of the helpfulness of user-generated product reviews. *J Retail* 87(4):598–612. <https://doi.org/10.1016/j.jretai.2011.05.002>
- Qiu L, Li D (2010) Effects of aggregate rating on eWOM acceptance: an attribution theory perspective. The Pacific Asia Conference on Information Systems (PACIS 2010)
- Ryan KE, Gandha T, Culbertson MJ, Carlson C (2014) Focus group evidence: Implications for design and analysis. *Am J Eval* 35(3):328–345
- Sen S, Lerman D (2007) Why are you telling me this? An examination into negative consumer reviews on the web. *J Interact Mark* 21(4):76–94. <https://doi.org/10.1002/dir.20090>

- Sun M (2012) How does the variance of product ratings matter? *Manage Sci* 58(4):696–707. <https://doi.org/10.1287/mnsc.1110.1458>
- Tsaparas P, Ntoulas A, Terzi E (2011) Selecting a comprehensive set of reviews. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data Mining. ACM, pp 168–176. <https://doi.org/10.1145/2020408.2020440>
- Wells GL, Harvey JH (1977) Do people use consensus information in making causal attributions? *J Pers Soc Psychol* 35(5):279. <https://doi.org/10.1037/0022-3514.35.5.279>
- Yan X, Wang J, Chau M (2015) Customer revisit intention to restaurants: Evidence from online reviews. *Inf Syst Front* 17(3):645–657. <https://doi.org/10.1007/s10796-013-9446-5>
- Yin D, Bond S, Zhang H (2014) Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Q* 38(2):539–560. <https://doi.org/10.25300/MISQ/2014/38.2.10>
- Yin D, Mitra S, Zhang H (2016) Research Note—when do consumers value positive vs. negative reviews? An empirical investigation of confirmation bias in online word of mouth. *Inform Syst Res* 27(1):131–144. <https://doi.org/10.1287/isre.2015.0617>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.