How Does the Review Tag Function Benefit Highly-Rated Popular Products in Online Markets?

Zhuolan Bao[1], Wenwen Li[1], Pengzhen Yin[2], and Michael Chau[1](✉)

[1] The University of Hong Kong, Pokfulam, Hong Kong
{mbao, liwwen}@connect.hku.hk, mchau@business.hku.hk
[2] Hefei University of Technology, Hefei, Anhui, China
yinpz@hfut.edu.cn

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, we specifically look into the effect of the tag function on the evaluation of highly-rated popular products and helpfulness perception of their reviews by proposing a framework through the lens of attribution theory. Experimental methods 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 make purchase decision [1–3]. Hence, the impact of online reviews is increasingly important and has been intensively investigated by researchers [4–6].

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 [7]. 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 [4, 6, 8, 9].

W. Cho et al. (Eds.): WEB 2017, LNBIP 328, pp. 47–63, 2018.
https://doi.org/10.1007/978-3-319-99936-4_5

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 [10]. To assist consumers in reading the massive review content, market platforms introduce new mechanisms to help them identify the most important or valuable information [1, 11].

The automatic review tagging system is one of such attempts. The automatic tagging system utilizes and extracts the content of consumer-generated reviews to generate automatically 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 corresponding sentiment as well as the number of mentions using different colors (shown in Fig. 2).

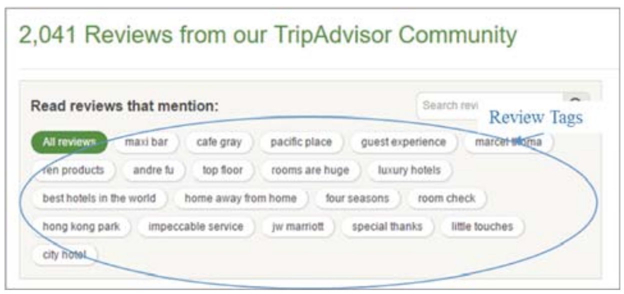

Fig. 1. Review tag function on TripAdvisor.com

Fig. 2. Review Tag function on Tmall.com

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 [6, 7, 12], little is known about how the review tag function would matter.

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 [13] 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. 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 as well as the results. In the last section, we discuss the contributions and limitations of the research.

2 Research Background

2.1 Review Tag Function

Archak, Ghose [10] 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 [13]. 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 give 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 [3, 14].

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

N = 340	Subsample 1: with tag function					Subsample 2: without tag function				
Product features	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
# 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
# of tags	228	9.57	1.85	2	10	112	n/a	n/a	n/a	n/a
# 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

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.

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 [15]. 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 [16]. 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 [15, 17]. 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 [15]. Third, it is assumed that a causal understanding serves the function of attaining personal goals and survival [18].

At first, when Heider proposed the attribution theory in his book [15], 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 [19, 20]. 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 [20] 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. In the meantime, 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 [21].

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 [22]. Among them, sample representativeness seems to be of greater importance. As noted in Kassin [22], 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 [23], otherwise, the validity of consensus would be violated.

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 [24].

3 Hypotheses Development

Our context of online reviews readily fits into the principle proposed by Kelley. When making a purchase decision, individuals would observe others' product experience expressed in reviews. When processing the review information, they would attribute the review content to either product-related features or reviewers' characteristics.

There have been studies using attribution theory within the context. For example, Chen and Lurie [25] 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 [24] 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 the evaluation of the most frequent features, they can also be regarded as high consensus information.

According to the attribution theory, the consensus information affords a basis for confidence in one's judgment [20]. 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.

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, if all tags for a product are positive, consumers' preference for positive reviews would be further strengthened due to the more salient positive consensus. But 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. So, 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, 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.

However, the above discussion neglects the role of the consensus' perceived validity [23]. As shown in Sect. 2.1, a typical appearance of tag function contains not only positive, but also negative tags. 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 4. 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.

4 Methodology

We first utilized an online experiment in order 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.

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 times of 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 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 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 [26]. We obtained 0.8208 on Krippendorff's alpha, which exceeded the recommended value 0.70 [27]. We also had 96.67% on Cohen's kappa which also exceeded the recommendation value of 0.80 [28].

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 three 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, positing 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 [29] and Yin, Bond [30] 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?"

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]

Note: The measurements are developed based on the prior work of Sen and Lerman [29] and Yin, Bond [30]. All items are of 9-point semantic differential-scale.

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?"

5 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 three minutes. 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 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 shows the descriptive statistics and correlation matrix of constructs.

5.1 Measurement Reliability and Validity

For exploratory factor analysis, we used principal components analysis with both varimax and oblimin rotations [31]. The result consistently provided two factors. All 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

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

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 [32, 33]. 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 [33]. Finally, the data also passed the Fornell and Larcker's test, indicating discriminant validity [34].

5.2 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 3 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 ($M_{positive} = 4.780$ vs $M_{neutral} = 2.554$ vs $M_{negative} = 6.350$, $F\ (2,\ 457) = 132.43$, $p < 0.001$). Reviews' helpfulness perception increases when reviews are either

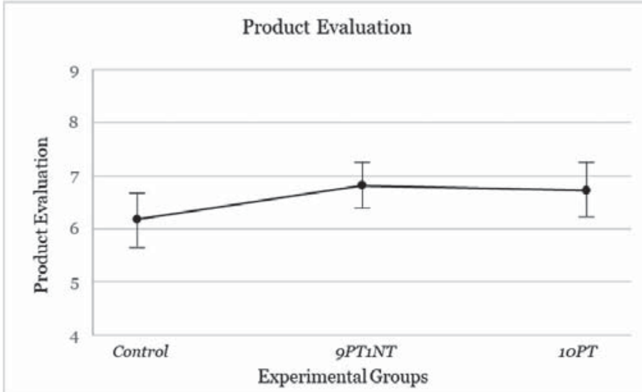

Fig. 3. Product evaluation

positive or negative comparing to when reviews are neutral. Figure 4 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.

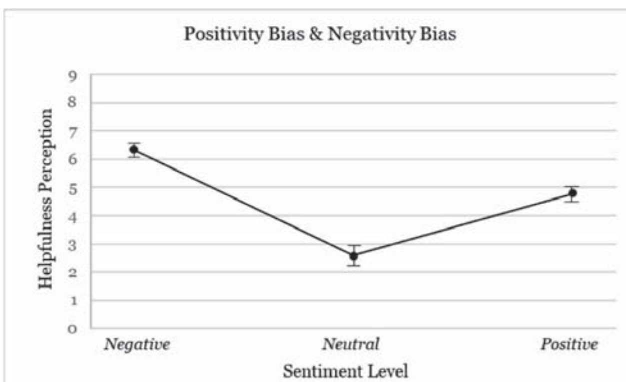

Fig. 4. Positivity bias & negativity bias

In order to test Hypotheses 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 ($M_{control} = 4.771$, $M_{10PT} = 5.08$, $F(1,115) = 0.47$, $p = 0.6255$), or between 9P1NT and control groups ($M_{control} = 4.771$, $M_{9PT1NT} = 4.571$, $F(1,127) = 0.49$, $p = 0.6135$). Thus, Hypotheses 2a and 3a are rejected, yet we still see the difference in their marginal effect as shown

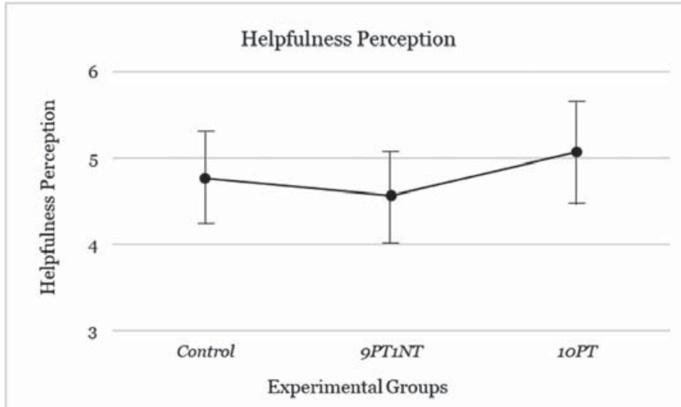

Fig. 5. Helpfulness perception for positive reviews

in Fig. 5 – 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 Hypotheses 2b, 3b and 4 simultaneously. Again, we performed a repeated-measure ANOVA. The results showed the helpfulness perception for negative reviews is significantly different across the three versions ($M_{control} = 6.307$, $M_{9PT1NT} = 6.071$, $M_{10PT} = 6.741$, $F(2,180) = 3.16$, $p = 0.045$). 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 9PT1NT group and the control group ($t(128) = 0.8995$, $p = 0.815$). Figure 6 shows the differences in helpfulness perception at with 95% confidence intervals.

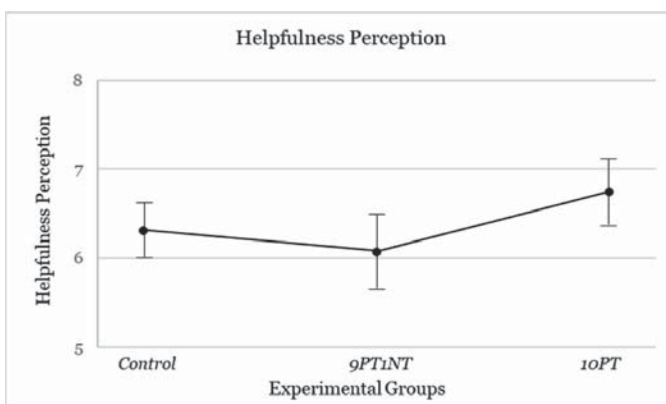

Fig. 6. Helpfulness perception for negative reviews

Therefore, Hypotheses 2b and 3b are rejected, but Hypothesis 4 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.

We also tested on the neutral reviews. Results of an ANOVA test showed no significant difference in helpfulness perception among three groups ($M_{control} = 2.885$, $M_{9PT1NT} = 2.172$, $M_{10PT} = 2.630$, $F\ (2{,}89) = 1.47$, $p = 0.236$).

6 Discussion

6.1 Key Findings and Contributions

By directly manipulating the tag function usage, we provide evidence supporting part of our hypotheses. The research has three major findings. First, in general, participants give higher product evaluation when tag function is shown. Second, for positive reviews, tag function does not lead to significant difference in reviews' helpfulness perception. Third, for negative reviews, overwhelming positivity in tags leads to a higher negativity bias comparing to typical tag function appearance (nine positive tags and one negative tag) and no-tag-function situation.

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 [13, 35], 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 [10, 36, 37], 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 supplement the existing research on the role of review variance. As prior studies reach 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 [6, 14, 25]. 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 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 [22].

Also, our target prototype is the review tag function for a computer product on Tmall.com. Generalizability issue of the research could be raised due to the differences existing in product type [38], culture and languages as well as in shopping conventions. Future research could draw on the function usage on other products or platforms to explore its role in perceptual and behavioral outcomes of consumers.

Acknowledgement. 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

1. Mudambi, S.M., Schuff, D.: What makes a helpful online review? A study of customer reviews on Amazon.com. MIS Q. **34**(1), 185–200 (2010). https://doi.org/10.2307/20721420
2. Dimoka, A., Hong, Y., Pavlou, P.A.: On product uncertainty in online markets: theory and evidence. MIS Q. **36**(2), 395–426 (2012)
3. Dellarocas, C.: The digitization of word of mouth: promise and challenges of online feedback mechanisms. Manag. Sci. **49**(10), 1407–1424 (2003). https://doi.org/10.1287/mnsc.49.10.1407.17308
4. Lu, X., et al.: Promotional marketing or word-of-mouth? Evidence from online restaurant reviews. Inf. Syst. Res. **24**(3), 596–612 (2013). https://doi.org/10.1287/isre.1120.0454

5. Forman, C., Ghose, A., Wiesenfeld, B.: Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets. Inf. Syst. Res. **19**(3), 291–313 (2008). https://doi.org/10.1287/isre.1080.0193
6. Yin, D., Mitra, S., Zhang, H.: Research note—when do consumers value positive vs. negative reviews? An empirical investigation of confirmation bias in online word of mouth. Inf. Syst. Res. **27**(1), 131–144 (2016). https://doi.org/10.1287/isre.2015.0617
7. Sun, M.: How does the variance of product ratings matter? Manag. Sci. **58**(4), 696–707 (2012). https://doi.org/10.1287/mnsc.1110.1458
8. Ba, S., Pavlou, P.A.: Evidence of the effect of trust building technology in electronic markets: price premiums and buyer behavior. MIS Q. **26**(3), 243–268 (2002). https://doi.org/10.2307/4132332
9. Yan, X., Wang, J., Chau, M.: Customer revisit intention to restaurants: evidence from online reviews. Inf. Syst. Front. **17**(3), 645–657 (2015). https://doi.org/10.1007/s10796-013-9446-5
10. Archak, N., Ghose, A., Ipeirotis, P.G.: Deriving the pricing power of product features by mining consumer reviews. Manag. Sci. **57**(8), 1485–1509 (2011). https://doi.org/10.1287/mnsc.1110.1370
11. Chen, P.-Y., Hong, Y., Liu, Y.: The value of multi-dimensional rating systems: evidence from a natural experiment and randomized experiments. Manag. Sci. (2017). https://doi.org/10.1287/mnsc.2017.2852
12. Pan, Y., Zhang, J.Q.: Born unequal: a study of the helpfulness of user-generated product reviews. J. Retail. **87**(4), 598–612 (2011). https://doi.org/10.1016/j.jretai.2011.05.002
13. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM (2004)
14. Hu, N., Pavlou, P.A., Zhang, J.: On self-selection biases in online product reviews. MIS Q. **41**(2) (2017). https://doi.org/10.25300/misq/2017/41.2.06
15. Heider, F.: The Psychology of Interpersonal Relations. Psychology Press, London (1958)
16. Försterling, F.: Attributional conceptions in clinical psychology. Am. Psychol. **41**(3), 275 (1986). https://doi.org/10.1037/0003-066X.41.3.275
17. Kelley, H.H., Michela, J.L.: Attribution theory and research. Annu. Rev. Psychol. **31**(1), 457–501 (1980). https://doi.org/10.1146/annurev.ps.31.020180.002325
18. Kelley, H.H.: Attribution in social interaction. In: Jones, E.E., et al. (eds.) Attribution: Perceiving the Causes of Behavior. Lawrence Erlbaum Associates, Inc., Hillsdale (1987)
19. Kelley, H.H.: Attribution theory in social psychology. In: Nebraska Symposium on Motivation. University of Nebraska Press (1967)
20. Kelley, H.H.: The processes of causal attribution. Am. Psychol. **28**(2), 107 (1973). https://doi.org/10.1037/h0034225
21. McGill, A.L.: Context effects in judgments of causation. J. Pers. Soc. Psychol. **57**(2), 189 (1989). https://doi.org/10.1037/0022-3514.57.2.189
22. Kassin, S.M.: Consensus information, prediction, and causal attribution: a review of the literature and issues. J. Pers. Soc. Psychol. **37**(11), 1966 (1979). https://doi.org/10.1037/0022-3514.37.11.1966
23. Wells, G.L., Harvey, J.H.: Do people use consensus information in making causal attributions? J. Pers. Soc. Psychol. **35**(5), 279 (1977). https://doi.org/10.1037/0022-3514.35.5.279
24. He, S.X., Bond, S.D.: Why is the crowd divided? Attribution for dispersion in online word of mouth. J. Consum. Res. **41**(6), 1509–1527 (2015). https://doi.org/10.1086/680667
25. Chen, Z., Lurie, N.H.: Temporal contiguity and negativity bias in the impact of online word of mouth. J. Mark. Res. **50**(4), 463–476 (2013). https://doi.org/10.1509/jmr.12.0063

26. Lombard, M., Snyder-Duch, J., Bracken, C.C.: Content analysis in mass communication: assessment and reporting of intercoder reliability. Hum. Commun. Res. **28**(4), 587–604 (2002). https://doi.org/10.1111/j.1468-2958.2002.tb00826.x
27. Krippendorff, K.: Content Analysis: An Introduction to Its Methodology. Sage, Thousand Oaks (2004)
28. Cohen, J.: Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. Psychol. Bull. **70**(4), 213 (1968). https://doi.org/10.1037/h0026256
29. Sen, S., Lerman, D.: Why are you telling me this? An examination into negative consumer reviews on the web. J. Interact. Mark. **21**(4), 76–94 (2007). https://doi.org/10.1002/dir.20090
30. Yin, D., Bond, S., Zhang, H.: Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. MIS Q. **38**(2), 539–560 (2014). https://doi.org/10.25300/MISQ/2014/38.2.10
31. DeVellis, R.F.: Scale Development: Theory and Applications, vol. 26. Sage Publications, Thousand Oaks (2016)
32. Nunnally, J.C.: Psychometric Theory, 3rd edn. McGraw-Hill, New York (1967)
33. Hair, J.F., et al.: Multivariate Data Analysis: A Global Perspective, vol. 7. Pearson, Upper Saddle River (2010)
34. Fornell, C., Larcker, D.F.: Evaluating structural equation models with unobservable variables and measurement error. J. Mark. Res. **18**(1), 39–50 (1981). https://doi.org/10.2307/3151312
35. Tsaparas, P., Ntoulas, A., Terzi, E.: Selecting a comprehensive set of reviews. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM (2011)
36. Chevalier, J.A., Mayzlin, D.: The effect of word of mouth on sales: online book reviews. J. Mark. Res. **43**(3), 345–354 (2006). https://doi.org/10.1509/jmkr.43.3.345
37. Bao, Z., Chau, M.: A schema-oriented product clustering method using online product reviews. In: International Conference on Information Systems (ICIS 2016) (2016)
38. Huang, P., Lurie, N.H., Mitra, S.: Searching for experience on the web: an empirical examination of consumer behavior for search and experience goods. J. Mark. **73**(2), 55–69 (2009). https://doi.org/10.1509/jmkg.73.2.55