ICIS 2022 Proceedings | Online Reviews and Recommendations

Dec 12th, 12:00 AM

How is the review helpfulness evaluated?

Tien Thuy Nguyen
University of Auckland, t.nguyen@auckland.ac.nz

Arvind Tripathi
University of Auckland, aktri@hotmail.com

How is the Review Helpfulness Evaluated?

Short Paper

Tien Thuy Thi Nguyen
The University of Auckland
Auckland, 1010, New Zealand
t.nguyen@auckland.ac.nz

Arvind K. Tripathi
The University of Kansas
Lawrence, KS, USA
arvind.tripathi@ku.edu

Abstract

A user-generated review that is perceived as helpful is valuable for both customer and the retailer, and that is why online markets such as Amazon.com collect public opinion on reviews that are perceived more helpful. Review platforms allow customers to vote for reviews they deem helpful. While prior literature has examined what drives the helpfulness of reviews, many of these studies have looked at drivers of perceived helpfulness of reviews in isolation. Using the lens of dual process theory, this research examines how consumers evaluate the helpfulness of a review. We propose a framework and provide empirical evidence for the evaluation of the review helpfulness process. We find that extreme reviews have a higher effect on review helpfulness compared to moderate reviews, and this effect is mediated by the depth and sentiment of the review content.

Keywords: dual-process theory, review helpfulness, peripheral cues, central cues

Introduction

User-generated reviews (UGRs) play a key role in consumers' purchase decision-making (Hu, Chen, and Lee 2017; Hu, Koh, and Reddy 2014; Li and Huang 2020; Mudambi and Schuff 2010). Since customers cannot look at all the available reviews, reviews voted as helpful, play an outstanding role and significantly stimulate effective review adoption (Qahri-Saremi and Montazemi 2019). Understanding the role and the value of UGR, retailers put significant effort into managing the review system to help prospective customers utilize the peer reviews, especially the reviews that are perceived as helpful by users. Indeed, prominent e-commerce platforms allow prospective customers to filter and access the most helpful reviews to assist in their purchase decision. For instance, Amazon.com allows customers to filter reviews by recent reviews and top reviews, with top reviews as the default option. The top reviews option exposes customers to reviews perceived as most helpful out of all the reviews.

A large body of literature on review helpfulness has emerged in various disciplines, such as information systems, marketing, and management (Huang et al. 2018; Kwok and Xie 2016). While this literature broadens our understanding of review helpfulness, this literature is fragmented and riddled with contradictory findings, as we explain next. Most of the studies have focused on the antecedents of perceived review helpfulness. Extant literature highlights the role of rating or review extremity (extreme versus moderate star ratings given by a reviewer) (Mudambi and Schuff 2010) of UGRs in predicting the review helpfulness (Baek, Ahn, and Choi 2012; Huang et al. 2018). Given that previous research has advocated a nonlinear relationship between review rating (review extremity) and review helpfulness (Chua and Banerjee 2015; Forman, Ghose, and Wiesenfeld 2008; Mudambi and Schuff 2010), the results are, however, contradictory. Some scholars find that moderate rating drive the helpfulness of reviews (Cao, Duan, and Gan 2011; Mudambi and Schuff 2010). Meanwhile, others have claimed that extreme rating (either extremely negative or extremely positive) are more influential than moderate ones (Forman et al. 2008; Pavlou and Dimoka 2006) to explain the helpfulness of reviews. Likewise, though many scholars have confirmed the impact of the word count in review (review depth) on the product review helpfulness, the findings remain in conflict. Several papers advocate the linear relationship between review depth and review helpfulness (Baek et al. 2012; Mudambi and Schuff 2010). On the other hand, some other studies confirm the inverted U-shaped relationship between them (Li and Huang 2020). Alternatively, even a few findings posit that a shorter review is more helpful than a lengthy review (Kwok and Xie 2016). Finally, content review cues (Huang et al. 2018) such as sentiment score, are another major predictor that has a significant role in predicting helpful votes. However, the embedded sentiment in the review, which leads to the persuasiveness of the review, has been addressed with inconsistent impacts (Rietsche et al. 2020).

Research by Mudambi and Schuff (2010) has identified the product type (experience vs. search goods) that can explain why the abovementioned inconsistency exists. However, if these results are applicable to all kinds of experienced goods and search goods, more evidence is needed. For instance, the literature almost provides evidence for only a few typical experience goods such as books, movies, and electronic devices (Li and Huang 2020; Mudambi and Schuff 2010). As suggested by prior research (Li and Huang 2020), other product categories, such as skin care products (cosmetics category), are worth considering to generalize the findings on perceived review helpfulness. Because despite belonging to experienced goods, cosmetics also fall under the category of sensitive items. Further, the borderline between search goods, experience goods, and credence goods is sometimes blurred. Therefore, our study opted for skin care products to additionally check to fill the gap in this regard.

Moreover, despite the rich exploration of crucial indicators of review helpfulness, the existing studies mainly focus on a monotonous relationship to seek answers for what makes a review helpful (Kwok and Xie 2016; Mudambi and Schuff 2010). The previous researchers examined different factors in isolation but ignored or paid very little attention to the combination or connection between them (Rietsche et al. 2020; Zhu, Yin, and He 2014), which could deepen our understanding. It is believed that the inter-relationship between the determinants may offer alternative explanations of how potential customers evaluate a review by providing a substantial understanding of the process of quality-based judgment and evaluation in more diversified scenarios. For instance, the explicit interaction between review content and star rating enhances the prediction of review helpfulness (Du et al. 2019). Such confirmation highlights the logical assessment behavior, which is likely to level up the perceived helpfulness as readers jointly evaluate the star rating and review content (Baek et al. 2012).

Underpinned by dual-process theory, particularly the elaboration likelihood model (ELM) (Petty and Cacioppo 1986), we argue that review readers rely on different sources of information (e.g., message content factors, non-content factors, see more at Petty & Cacioppo (1981)) that match different elaboration levels to process the information, subsequently benefiting the evaluation of review helpfulness. Accordingly, the prominent factors highly correlated with review helpfulness are categorized into two primary sources of information: peripheral cues and central cues (Baek et al. 2012). Peripheral cues contain the primary signals for heuristic information processing, whereas central cues involve the secondary signals used for systematic information processing (Baek et al. 2012; Chaiken 1980).

Following that, we posit that UGR extremity (Hu et al. 2017; Mudambi and Schuff 2010), which is defined as the degree of deviation from the midpoint rating of a review, is the peripheral cue of review. UGR extremity is represented by extreme negative ratings, extreme positive ratings, and moderate ratings. It is categorized as a non-content factor that readers quickly seize at first sight without much cognitive effort (Pelsmacker, Dens, and Kolomiiets 2018). While review content features (length of review and review sentiment) are considered central cues (Baek et al. 2012; Zhu et al. 2014), which need more cognitive effort to assess (Pelsmacker et al. 2018). Given the characteristics and relationship between peripheral and central cues, we hypothesize that the central cues inherently play the mediating role, facilitating the relationship between the peripheral cues and review helpfulness. In other words, the prospective review readers anchor the star rating, and content cues chronologically to evaluate the helpfulness of a review (Huang et al. 2018).

While we are beginning to understand how key antecedents interact in their impact on review helpfulness, we are yet to understand how these peripheral and central cues are interplayed and processed by the readers and eventually affect review helpfulness (Rietsche et al., 2020; Zhu, Yin, & He, 2014). This research aims to examine that. We constructed the model of review helpfulness emphasizing the mediating role of the textual content (review sentiment) and review depth to explain the behavior of the UGR helpful voting. By focusing on the direct and indirect relationship between review helpfulness and review extremity, our paper examines whether there exists a diminishing or increasing return relationship between review helpfulness and star rating; or whether moderate or extreme reviews are more helpful. More importantly, it examines whether this typical relation remains in both direct and indirect impact cases. Likewise, the mediation model emphasizing the intermediary role of review content advocates the possibility of looking into the content cues (sentiment and review depth) and their significance in the transmitter's position.

To validate the proposed model of review helpfulness, we employ path-analysis in SEM (structural equation modeling) to fit the data of 5190 online reviews collected from Amazon.com. We find that UGR extremity influences review helpfulness partially mediated by negative sentiment and content length. In more detail, a review with an extreme star rating (either negative extreme or positive extreme) is more likely to be perceived as helpful compared to a review with a moderate rating. The results assert the role of numerical rating and review content as major factors in predicting review helpfulness.

The study contributes to the existing literature on review helpfulness by developing a novel mediation model that considers the link between rating signals and information content signals in explaining review helpfulness. In alignment with the findings, the study reinforced the significant role of review rating as the most salient determinant direct and indirectly influences the perceived review helpfulness (Baek et al. 2012; Mudambi and Schuff 2010; Rietsche et al. 2020). Consistent with several previous papers, our study demonstrates the curvilinear relationship between review rating and review helpfulness (Mudambi and Schuff 2010; Park and Nicolau 2015). Nevertheless, instead of showing an inverted U-shaped relationship between them, our study demonstrates a U-shaped relation, which indicates the more influential impact of reviews with extreme ratings rather than moderate ratings. This finding addresses the difference between the findings in a few earlier research for experienced goods in the marketplace. In addition, findings also highlight the mediating role of content characteristics of a review, unveiling and indicating the possibility that customers on the e-commerce sites, especially Amazon, not just heuristically process the review information but also systematically and logically evaluate it.

Literature Review and Conceptual Development

Dual-process Theory

Dual-process theory (William James 1890) is a social and cognitive psychology theory that explains the different degrees of information processing in each individual. Information is processed via two execution phases called type 1 and type 2, including intuitive and deliberate reasoning, respectively (Thompson 2014). Type 1 processes initiate the preliminary response's foundation, often shortly executed that may or may not be adjusted by consequential deliberation. In comparison, type 2 counterparts are deliberative reason-based processes requiring longer time and effort to accomplish (Brewer 1988; Thompson 2014).

Our study is closely aligned with the implication perspective of the theory (Petty and Cacioppo 1981). In accordance, people in the effortless processing mode, with low capacity or low motivation, tend to be less involved in elaboration. In contrast, people with full capacity and motivation tend to experience a high engagement in elaboration. Linking with the information sources in the processing model (Chaiken 1980; Petty and Cacioppo 1986), the degree of elaboration is collated with the two information processing modes, including peripheral and central processing. Peripheral processing on elaboration likelihood, also known as heuristic information processing (Chaiken 1980), is particularized in peripheral cues, whereas central processing reflects the systematic information processing (Chaiken 1980) and requires the subjects to exert comparatively more effort in judgment (Chaiken 1980).

Hypotheses Development

We use ELM grounded on dual-process theory as a lens in conjunction with the literature review on information diagnosticity (Filieri 2015; Mudambi and Schuff 2010; Zhanfei Lei et al. 2021) to develop our review helpfulness model and evoke corresponding hypotheses.

User-generated Review Extremity

A user-generated review star rating echoes the total valence of the review (Chua and Banerjee 2015; Li 2018). It represents a favorable valence of online reviews (Li 2018) and a salient signal demonstrating the review quality (Li 2018). Since review ratings reflect the total experience of the reviewer toward a purchase rather than just the product experience, eventually representing the total review's quality (Li 2018), it should be highly correlated with the helpful perceptions of customers when they diagnose the information (Filieri 2015).

Literature on extreme and two-sided arguments in review has suggested the significant variation effect of star rating. An extreme rating indicates a very low or a very high star rating demonstrating an extremely negative product view (one star – low extremity) or extremely positive product view (five star – high extremity), respectively (Filieri 2015; Liu and Karahanna 2017). In comparison, a three-star rating demonstrates a moderate product view. Star rating represents the attitude extremity (Mudambi and Schuff 2010; Van der Pllgt, Ester, and van der Linden 1983). And various star ratings correlate with various review extremities; consequently, user-generated review extremity reflects the variation from the central point of the attitude scale (Mudambi and Schuff 2010). Indeed, the two types of star ratings (extreme rating vs. moderate rating) perform distinct effects. The possible explanation can be fundamentally based on the equivocal verse unequivocal emotional situations, so-called review equivocality (Forman et al. 2008). For example, since extreme reviews (either negative or positive) often provide clear implications (unequivocal information) which actively drive customers to action for a purchase decision, it is proposed to be more helpful than a two-sided user-generated review (moderate review), where information is provided ambiguous and equivocal, and therefore is impossible to offer a clear guide to action (Forman et al. 2008).

The tremendous impact of the review score has been documented at an early stage of review establishment. The prior online review helpfulness literature suggested that online review rating significantly impacts the review helpfulness in terms of either moderate or extreme rating reviews, positing a curvilinear relationship (Huang et al. 2015; Liu and Park 2015; Mudambi and Schuff 2010). There have been mixed arguments on the helpfulness of extreme verse moderate user-generated reviews. Many scholars have advocated the significance of unequivocal information rather than equivocal information and then favor the role of extreme reviews (Forman et al. 2008; Pavlou and Dimoka 2006). In contrast, others have posited the significant impacts of reviews with moderate ratings (Cao et al. 2011; Mudambi and Schuff 2010).

We argue that potential customers tend to follow the extreme rating as the first salient signal because it enables them to promptly classify the product as bad or good. Such a clear opinion helps save time and effort in evaluation, attributing customers to a cost-effective purchase decision (Pavlou, Liang, and Xue 2007). Compared to the moderate rating review, which is usually accompanied by two-sided arguments with pros and cons, it is likely to offer mixed opinions, resulting in hesitation in judgment and evaluation. In brief, scholars who advocate the importance of unequivocal UGR suggest that extreme star rating reviews are more persuasive, therefore being regarded as more helpful than moderate star rating reviews.

In light of dual-process theory, customers' judgment and the decision can be pinched on the processing of peripheral cues (review score) and central cues (content cues) (Baek et al. 2012). We can indeed claim their use of both peripheral cues and central cues since potential customers generally count on user-generated reviews for both information search and evaluation of alternatives that have been found in the phase of information search (Baek et al. 2012). In the absence of motivation or capability, the prospective customers will not use the central cues to vote for the reviews' helpfulness. Instead, they will likely go with the peripheral cues to perceive the helpfulness of the review. In this study, we argue that there is a high association between user-generated review extremity and review helpfulness. In addition, in case of lacking motivation or capability, customers will skip the content evaluation but only rely on the star rating (peripheral cue) and rest at the information search stage to judge the helpfulness of the review. Consequently, we posit the following hypothesis

H1: User-generated reviews with extreme ratings are more helpful than reviews with moderate ratings.

Review Sentiment and Review Depth

Literature on review helpfulness supports the strong effects of positive and negative sentiment score features on review helpfulness. In general, a review with a stronger sentiment polarity will lead to a higher possibility of being evaluated as helpful (Cao et al. 2011). Although a high negative sentiment review is often associated with a low star rating (Hu et al. 2014), it tends to be more attractive to information seekers and subsequently more powerful in driving their perception of helpfulness (Cao et al. 2011).

Review depth is another aspect of review content that positively impacts the helpfulness of the review (Chua and Banerjee 2015; Li and Huang 2020; Mudambi and Schuff 2010). Since review depth enhances information diagnosticity (Mudambi and Schuff 2010), its examination can affect the evaluation and judgment of customers (Li and Huang 2020). Also, review depth is the most straightforward quantitative dimension of content quality (Zhu et al. 2014). A longer review is supposed to provide more and deeper information than a shorter one, especially details of the product description and product performance (Mudambi and Schuff 2010). As review content diagnosticity requires systematic information processing, prospective customers need to make more judgments and be fully motivated to achieve a high likelihood of elaboration (Petty and Cacioppo 1986) in the role of mediator transmitting the perceptiveness. Thus, sentiment score and review depth reflecting different aspects of content quality of reviews are categorized into central cues when considered in the model of review helpfulness (Baek et al. 2012; Zhu et al. 2014).

Our study inherits the results from previous research by Poston & Speler (2005) and Alavi, Maryam, E. Leidner (2001) regarding the evaluation process and decision performance. Generally, their papers investigated how content ratings and credibility indicators influence the evaluation of the content, thereby affecting the decision-making performance. The content search and evaluation process turn out to be a mediating factor, transmitting the effect from the salient factor – rating to the final decision-making. Although their results were analyzed and discussed based on the different contexts of knowledge management systems applied in cooperations, significant similarities can be identified. First, user-generated review systems and knowledge management systems are sources of wisdom (Liu and Karahanna 2017), containing countless WOM facilitating the user's evaluation and decision-making. Second, both systems use star rating scores as well as review cues (content cues and contextual review cues) as the key facilitators leveraging the usage of information. Finally, due to the vast volume of information, users of both systems may find overwhelmed when assessing the content to sort the quality and helpful content for their purchase decision or collaborating decision; therefore, understanding which review is helpful is essential and of interest. Thus, their findings are consistent and feasible for explaining our context's evaluation process and decision-making. In this regard, the diagnosticity of the review content is seen as the content search and evaluation process. This diagnosticity requires systematic information processing, which means prospective customers need to make more effort for judgment and be fully motivated to achieve a high likelihood of elaboration (Petty and Cacioppo 1986) after achieving the initial perception anchoring on review ratings. Past research has also proved that rating is more important in the stage of searching while content factors (e.g., review sentiment and review depth) are more influential in the stage of evaluation (Hu et al. 2014). Therefore, review sentiment and review depth can be considered to be the role of mediators transmitting the perceptiveness of review helpfulness. Thus, review sentiment and review depth are categorized into central cues when considered in the model of review helpfulness.

Additionally, dual-process theory (e.g., ELM model) suggests that because ratings and review sentiment are different in customers' engagement in elaboration, they accompany two distinct phases of the information process when coping with decision-making. These two processes propose certain heuristics strategies to attain the best results of the judgment. Customers may stick to the peripheral cues (review ratings) in the first stage of assessment when they are initially dealing with a user-generated review since it is intuitively obtained without any effort. Indeed, review rating works efficiently as a salient first signal, giving the initial perception of the helpfulness of the review (Li and Huang 2020; Mudambi and Schuff 2010). Afterward, customers will rely more on the qualitative content of the reviews, using central cues (review sentiment, review depth). Due to its requirement for more effort and enough motivation, these cues are considered to take at the latter since customers can decide to pay more to gain more or make a decision right at the first phase. Another key point is that associative processing (processing the peripheral cues) is faster than ruled-based processing (processing the central cues) (Brewer 1988). Therefore, the first mode of processing operates and finishes initially due to effortless catching and fast processing. Indeed, with just a quick glance at the star ratings, readers can capture the valence of reviews. This process may take just a few seconds to complete. In comparison, the latter model of the process requires effort and time to analyze, resulting in a longer process and completeness. Then, it is likely to expect the sequential theoretical processing of these two modes in the context of user-generated reviews. These arguments suggest that the proximity to the final decision of central cues is closer than peripheral cues once customers take into account central cues to assess their judgment. Further, there is a high correlation between these two information process phases because the latter evaluation or judgment results are highly dependent on the initial belief establishment. We advocate this argument, and therefore, we expect mediation effects of the central cues in the relationship between peripheral cues and review helpfulness. Our research model is shown in Figure 1.

H2: A change in perceived review helpfulness caused by user-generated review extremity (peripheral cues) is associated with a change in sentiment of the review content (central cues).

H3: A change in perceived review helpfulness caused by user-generated review extremity (peripheral cues) is associated with a change in the review depth of the review content (central cues).

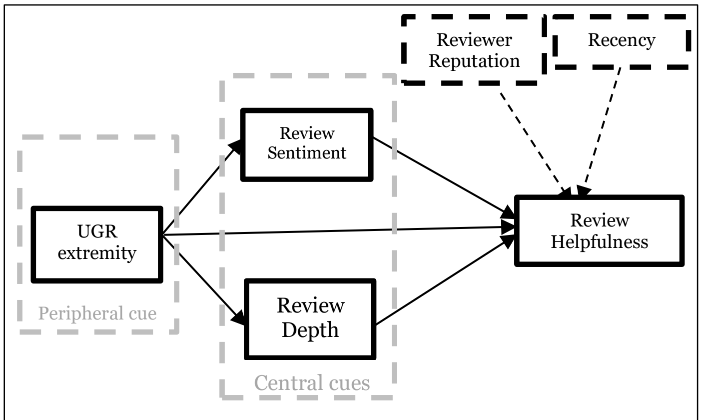

Figure 1. Research model

Method

Our study used an online dataset collected from Amazon.com. We employed Structure Equation Modelling (SEM) to estimate the model and the bootstrapping method as a conventional method to test the multiple mediation processes. SEM is a proper method to apply to test the effects of multiple mediators at the same time on the same outcome (Li 2011).

Data Collection

We utilized a dataset including 5190 actual online user-generated reviews of the top bestseller skincare products on Amazon.com, which were scrapped in July 2021. We chose Amazon as the source of our data because it is well-known, one of the largest online retailers in the world, where their user-generated content systems are fully functional and entirely made good use of. Thus, it enables us the best proper sample as we investigate the review helpfulness vote behavior of those using it. Besides, avoiding using the lesser ranking site to avert the risk of skewed data was another benefit.

Variables

Different from the previous papers, such as Mudambi & Schuff (2010), which used review helpfulness as the value derived by dividing the number of helpful votes by the total number of votes (unhelpful votes plus helpful votes), the dependent variable in our study is review helpfulness, measured by the total number of helpful votes (Zhanfei Lei, Yin, and Zhang 2021; Zhu et al. 2014). The explanatory variables in our study are user-generated review extremity, review depth, and review sentiment. User-generated review extremity (UGR.Ex) is measured as the review's star rating, with the value ranging from one to five (Hu et al. 2017; Mudambi and Schuff 2010). Adopted from the previous papers, we categorized star rating into two corresponding terms, linear (RR) and quadratic, to account for the nonlinear relationship (RR^2) between user-generated review extremity and review helpfulness (Mudambi and Schuff 2010; Yin, Bond, and Zhang 2014). Review depth (RD) is the number of words counted in a review (Cao et al. 2011; Mudambi and Schuff 2010). Both of these measures are captured directly for each review from Amazon data. Review sentiment, however, has been operationalized based on the content of the review and needs a more meticulous step to extract. We employed a lexicon-based method (a sentiment analysis) (Hartmann et al. 2019) to yield the value of customers' sentiment polarity and determined a negative sentiment score (Senti-Nega) to represent the review sentiment variable (Hu et al. 2017). In more detail, we utilized the NRC emotion lexicon (Mohammad and Turney 2013), which comprises a list of words and their connotations with eight emotions (e.g., anger, fear) and two sentiment valences (negative and positive). Following NRC algorithm, after the texts had been broken down into separate sentences and words, they were then assessed and collated with the best match of the embedded polarity dictionaries. The work of extracting the sentiment values was carried out using the R program, in which we applied the NRC method implemented in syuzhet package.

We included the reviewer reputation and the review recency as the two control variables in the model. Since there exists a positive association between the reviewer's reputation and online review (Chua and Banerjee 2015), the differences in the reviewer's reputation may cause differences in the number of helpful votes each review receives. Specifically, it is measured by the total number of helpful votes a reviewer achieved. Likewise, recency is the elapsed time and was quantified by the number of days since the date the review was posted (Lee and Choeh 2014). Assuming the perceived review helpfulness may be influenced by the recency effect, which describes the more recent information as more weighted than earlier-presented information, we, therefore, account for review recency as the second control variable in our theoretical model. The descriptive statistics for all the main variables in the complete data set are presented in table 1.

	Mean	SD	Min	Max	RD	Recency	RerHV	RR	RH	RS
RD	48.62	63.84	0	1,069	--					
Recency	1,265.28	876.10	5	4,896	0.014	--				
RerHV	228.24	1,734.38	0	66,738	.228**	0.008	--			
RR	4.34	1.28	1	5	-.049**	.148**	0.009	--		
RH	27.95	188.63	0	6,939	.306**	-0.015	.153**	-.048**	--	
RS	0.09	0.12	0	1	-.168**	0.011	-.029*	0.018	-.028*	--

***. Correlation is significant at the 0.01 level (2-tailed).*
**. Correlation is significant at the 0.05 level (2-tailed). N= 5190*

Table 1. Descriptive Statistics

Empirical Analysis and Preliminary Results

The model specification is presented in Figure 1. In our model, user-generated review extremity and reviewer-helpful vote are exogenous variables, whereas review sentiment, review depth, and review helpfulness are endogenous variables. We hypothesized a nonlinear relationship between review extremity and perceived review helpfulness. We expect that reviews with extreme ratings (either strongly positive extreme or strongly negative extreme) are more helpful than reviews with moderating ratings. In order to test this first hypothesis and confirm the argument on this, we, therefore, construct a model including a linear term of star rating and a quadratic term of star rating. We expect the linear term to be negative whereas the quadratic term to be positive, representing a U-shaped relationship, indicating that moderate reviews are less helpful than extreme reviews (Mudambi and Schuff 2010).

Goodness of Fit of the Model

The results are based on the maximum likelihood method with lavaan using R programming. The fit indices in our structural model reveal that our model fits the data well. In detail, DF is equal to 4, indicating an over-identified model in which the number of observations is greater than the number of parameters and that the path analysis model is executable (DF>=0). The CFI is = 0.988, TLI = 0.938, thus, all above the suggested threshold of 0.9 (Hu & Bentler 1999). Besides, RMSEA = 0.121, which is generally considered an unacceptable fit (RMSEA should be <0.08); while SRMR = 0.057, indicating a close fit (Hu and Bentler 1999) or our model captured the data well. According to Kenny, Kaniskan, & McCoach (2015), a model with a small degree of freedom, RMSEA is not meaningful because it can falsely imply a poor-fitting model while in fact, the model fits the data well. And consequently, we can make the decision on the acceptance of our model based on CFI, TLI, and SRMR (Kenny et al. 2015). In general, based on the listing fit indices, SEM presents a good fit.

We tested the first hypothesis using path analysis in SEM. The output shows that there is a statistically significant association between the user-generated review extremity and review helpfulness (p_value < 0.001). Specifically, as expected, the linear term (star rating) demonstrates a negative value (-59.704), and the quadratic term (star rating^2) shows a positive value (8.806), therefore indicating a U-shaped relationship, elaborating that extreme user-generated reviews are more helpful than moderate user-generated reviews. Our first hypothesis (H1) is, therefore, supported. Yet, the finding confirms our expectation about the nonlinear relationship and asserts the previous conclusion of existing literature on this antecedent (Mudambi and Schuff 2010). However, in contrast to the study of Mudambi & Schuff (2010), who found the inverted U-shaped relationship, we support the U-shape relation, implying a more valuable role of extreme reviews other than moderate reviews.

Bootstrapping

We conducted 5000 bootstrapped samples. Generally, results revealed that the effect of review rating on review helpfulness through review depth was negative and marginally significant (a12*b21: $\beta = -6.514$; SE = 4.165; 95%CI = [-18.725, -1.112]). Similar results were found for the effect of star rating through sentiment negative value variable (a11*b11: $\beta = -4.820$; SE = 3.000; 95%CI = [-12.541, -0.051]). With regards to the quadratic term of RR, the results also show a statistical significant role of negative sentiment in mediating the relationship between review extremity and review helpfulness (a21*b11: $\beta = 0.485$; SE = 0.326; 95%CI = [0.018, 1.436]). Nevertheless, no significant conclusion is confirmed for the review depth mediator because the results show that zero is inside the confidence intervals (95%CI = [-0.031, 2.369]). The results accept the alternative hypotheses that review sentiment and review depth mediate the relationship between star rating and helpful votes. We finally confirm that H2 and H3 are supported.

Conclusion

This research in progress has provided a framework and evidence through preliminary data analysis. We have successfully confirmed a systematic and logical evaluation of the review helpfulness process and elucidated a possibility of how prospective customers perceive the review helpfulness. However, the study needs further examination and elaboration on the relationship between content features (central cues) and review helpfulness. Our next step would be to validate the model and make it more robust.

References

Baek, Hyunmi, Joongho Ahn, and Youngseok Choi. 2012. "Helpfulness of Online Consumer Reviews: Readers' Objectives and Review Cues." *International Journal of Electronic Commerce* 17(2):99–126.

Brewer, M. B. 1988. "A Dual Process Model of Impression Formation." *In T. K. Srull & R. S. Wyer, Jr. (Eds.), Advances in Social Cognition* 1:1–36.

Cao, Qing, Wenjing Duan, and Qiwei Gan. 2011. "Exploring Determinants of Voting for the 'Helpfulness' of Online User Reviews: A Text Mining Approach." *Decision Support Systems* 50(2):511–21.

Chaiken, Shelly. 1980. "Heuristic versus Systematic Information Processing and the Use of Source versus Message Cues in Persuasion." *Journal of Personality and Social Psychology* 39(5):752–66.

Chua, Alton Y. K. and Snehasish Banerjee. 2015. "Understanding Review Helpfulness as a Function of Reviewer Reputation, Review Rating, and Review Depth." *Journal of the Association for Information Science and Technology* 66(2):354–62.

Du, Jiahua, Jia Rong, Hua Wang, and Yanchun Zhang. 2019. "Helpfulness Prediction for Online Reviews with Explicit Content-Rating Interaction." *Web Information Systems Engineering (WISE)*.

Filieri, Raffaele. 2015. "What Makes Online Reviews Helpful? A Diagnosticity-Adoption Framework to Explain Informational and Normative Influences in e-WOM." *Journal of Business Research*.

Forman, Chris, Anindya Ghose, and Batia Wiesenfeld. 2008. "Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets." *Information Systems Research* 19(3):291–313.

Hartmann, Jochen, Juliana Huppertz, Christina Schamp, and Mark Heitmann. 2019. "Comparing Automated Text Classification Methods." *International Journal of Research in Marketing* 36(1).

Hu, Li Tze and Peter M. Bentler. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives." *Structural Equation Modeling: A Multidisciplinary Journal* 6(1):1–55.

Hu, Nan, Noi Sian Koh, and Srinivas K. Reddy. 2014. "Ratings Lead You to the Product, Reviews Help You Clinch It? The Mediating Role of Online Review Sentiments on Product Sales." *Decision Support Systems* 57(1):42–53.

Hu, Ya Han, Kuanchin Chen, and Pei Ju Lee. 2017. "The Effect of User-Controllable Filters on the Prediction of Online Hotel Reviews." *Information and Management* 54(6):728–44.

Huang, Albert H., Kuanchin Chen, David C. Yen, and Trang P. Tran. 2015. "A Study of Factors That Contribute to Online Review Helpfulness." *Computers in Human Behavior* 48:17–27.

Huang, Liqiang, Chuan Hoo Tan, Weiling Ke, and Kwok Kee Wei. 2018. "Helpfulness of Online Review Content: The Moderating Effects of Temporal and Social Cues." *Journal of the Association for Information Systems* 19(6):503–22.

Kenny, David A., Burcu Kaniskan, and D. Betsy McCoach. 2015. "The Performance of RMSEA in Models With Small Degrees of Freedom." *Sociological Methods and Research* 44(3):486–507.
Kwok, Linchi and Karen L. Xie. 2016. "Factors Contributing to the Helpfulness of Online Hotel Reviews: Does Manager Response Play a Role?" *International Journal of Contemporary Hospitality Management* 28(10):2156–77.
Lee, Sangjae and Joon Yeon Choeh. 2014. "Predicting the Helpfulness of Online Reviews Using Multilayer Perceptron Neural Networks." *Expert Systems with Applications* 41(6):3041–46.
Li, Mengxiang and Pu Huang. 2020. "Assessing the Product Review Helpfulness: Affective-Cognitive Evaluation and the Moderating Effect of Feedback Mechanism." *Information and Management*.
Li, Spencer D. 2011. "Testing Mediation Using Multiple Regression and Structural Equation Modeling Analyses in Secondary Data." *Evaluation Review* 35(3):240–68.
Li, Xitong. 2018. "Impact of Average Rating on Social Media Endorsement: The Moderating Role of Rating Dispersion and Discount Threshold." *Information Systems Research* 29(3):739–54.
Liu, Qianqian Ben and Elena Karahanna. 2017. "The Dark Side of Reviews: The Swaying Effects of Online Product Reviews on Attribute Preference Construction." *MIS Quarterly: Management Information Systems* 41(2):427–48.
Liu, Zhiwei and Sangwon Park. 2015. "What Makes a Useful Online Review? Implication for Travel Product Websites." *Tourism Management* 47:140–51.
Mohammad, Saif M. and Peter D. Turney. 2013. "Crowdsourcing a Word-Emotion Association Lexicon." *Computational Intelligence* 29(3):436–65.
Mudambi, Susan M. and David Schuff. 2010. "What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.Com." *MIS Quarterly* 34(1):185–200.
Park, Sangwon and Juan L. Nicolau. 2015. "Asymmetric Effects of Online Consumer Reviews." *Annals of Tourism Research* 50:67–83.
Pavlou, Paul A. and Angelika Dimoka. 2006. "The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums and Seller Differentiation." *Information Systems Research* 17(4):392–414.
Pavlou, Paul A., Huigang Liang, and Yahjiong Xue. 2007. "Understanding and Mitigating Uncertainty in Online Exchange Relationships: A Principal Agent Perspective." *MIS Quartely* 31(1):105–36.
Pelsmacker, Patrick, Nathalie Dens, and Alona Kolomiiets. 2018. "The Impact of Text Valence, Star Rating and Rated Usefulness in Online Reviews." *International Journal of Advertising* 37(3):340–59.
Petty, Richard E. and John Cacioppo. 1986. "The Elaboration Likelihood Model of Persuasion." *Advances in Experimental Social Psychology* 19:123–205.
Petty, Richard E. and John T. Cacioppo. 1981. "Issue Involvement as Moderator of the Effects on Attitude of Advertising Content and Context." *Advances in Consumer Research* (8):20–24.
Van der Pllgt, Joop, Peter Ester, and Joop van der Linden. 1983. "Attitude Extremity, Consensus and Diagnosticity." *European Journal of Social Psychology* 13(4):437–39.
Qahri-Saremi, Hamed and Ali Reza Montazemi. 2019. "Factors Affecting the Adoption of an Electronic Word of Mouth Message: A Meta-Analysis." *Journal of Management Information Systems*.
Rietsche, Roman, Daniel Frei, Emanuel Stöckli, and Matthias Söllner. 2020. "Not All Reviews Are Equal - A Literature Review on Online Review Helpfulness." *27th European Conference on Information Systems - Information Systems for a Sharing Society, ECIS 2019* (June).
Thompson, Valerie A. 2014. "Chapter Two - What Intuitions Are… and Are Not" edited by Brian H. Ross. *Psychology of Learning and Motivation* 60:35–75.
William James. 1890. "The Principles of Psychology." *New York: Henry Holt and Company* 1.
Yin, Dezhi, Samuel D. Bond, and Han Zhang. 2014. "Anxious or Angry? Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews." *MIS Quarterly* 38(2):539–60.
Yin, Dezhi, Samuel D. Bond, and Han Zhang. 2017. "Keep Your Cool or Let It out: Nonlinear Effects of Expressed Arousal on Perceptions of Consumer Reviews." *Journal of Marketing Research* 54(3).
Zhanfei Lei, Dezhi Yin, and Han Zhang. 2021. "Focus Within or On Others: The Impact of Reviewers' Attentional Focus on Review Helpfulness." *Information System Research* 32(3):801–19.
Zhu, Ling, Guopeng Yin, and Wei He. 2014. "Is This Opinion Leader's Review Useful? Peripheral Cues for Online Review Helpfulness." *Journal of Electronic Commerce Research* 15(4):267–80.