Decision Support Systems 139 (2020) 113403

Contents lists available at ScienceDirect

Decision Support Systems

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

An empirical investigation of online review helpfulness: A big data perspective

Hoon S. Choi[a,*], Steven Leon[b,1]

[a] Department of Computer Information Systems, Walker College of Business, Appalachian State University, 416 Howard Street, 2113 Peacock Hall, Boone, NC 28608-2049, United States of America

[b] Department of Marketing & Supply Chain Management, Walker College of Business, Appalachian State University, 416 Howard Street, 4055 Peacock Hall, Boone, NC 28608-2037, United States of America

ARTICLE INFO

Keywords:
Review helpfulness
eWOM
amazon.com
Big data
Review/source/context factors

ABSTRACT

This study investigates the determinants of online review helpfulness, adopting various predictors from three dimensions of the online review management: source factors, review factors, and context factors. Based on a large, comprehensive dataset that includes 14,051,211 online reviews in 24 product categories from an ecommerce retailer, Amazon.com, this study provides empirical evidence on the effect of source and reviews factors on perceived review helpfulness, which the extant literature has reached inconsistent conclusions. In addition, this study considers the effects on review helpfulness created by context factors: product satisfaction, product popularity, product intangibility, and product variety. These factors are scarcely discussed in the existing literature. The major findings include (1) review extremity, review depth, and reviewer expertise have a positive effect on review helpfulness; (2) review inconsistency, product intangibility, product satisfaction, product popularity, product variety, and reviewer experience have negative effects; and (3) product intangibility moderates the effect of review extremity and depth on review helpfulness.

1. Introduction

Prior to purchasing products and services consumers often seek information about potential purchases from online reviews written by other consumers. Fullerton [27] states that 82% of consumers read reviews before making a purchase decision and 68% of consumers are willing to pay up to 15% more for the same product or service if they are assured that they will have a better experience. According to the theory of reasoned action, one of the determinants of intention to purchase online is trust in e-commerce retailers [64], which is substantially affected by consumer online reviews. Therefore, it is apparent that consumer reviews influence the sales of products and services, and thus that reviews are important to consumers and businesses.

Even though online reviews are deemed beneficial, not all of them actually are. Platforms such as Amazon, Yelp, and Trip Advisor list millions of online reviews, and many individual products and services have hundreds or thousands of reviews. With the significant volume of online reviews, finding helpful reviews therefore has become an arduous task for many consumers. Platforms that list product and service reviews can assist in this important endeavor, and indeed, many of them do because trust in e-commerce retailers is not simply built with any reviews but with reliable and useful reviews [5]. Additionally, as the theory of information diffusion argues, useful information tends to diffuse fast within the target audience [96], therefore, helpful online reviews can help quickly build trust, leading to a rapid revenue increase. However, countless unhelpful reviews are still a part of the digital landscape. The combination of the sheer number of available online reviews and the numerous unhelpful reviews results in information overload. Such overload strains consumers' cognitive capabilities and potentially reduces the impact of reviews that can be helpful to would-be purchasers. To help purchasers effectively find useful reviews, many online review platforms have adopted helpfulness voting systems, which eventually contribute to their business [46]. According to Hong, Xu, Wang and Fan [33]. Amazon.com could attain the extra revenue of $2.7 billion by adopting their online review system.

Although online review helpfulness has been prevalently investigated due to its importance, other researchers have omitted potential determinants of helpfulness: *context factors of online reviews*, although they explored review (content) factors and reviewer (source) factors. Review helpfulness research has not yet explored product

* Corresponding author.
E-mail addresses: choihs@appstate.edu (H.S. Choi), leonsm@appstate.edu (S. Leon).
[1] Co-Authors.

https://doi.org/10.1016/j.dss.2020.113403
Received 22 April 2020; Received in revised form 12 September 2020; Accepted 14 September 2020
Available online 20 September 2020

context factors, such as product satisfaction, product popularity, product intangibility, and product variety; therefore, this study intends to examine these factors. Prior review helpfulness studies regularly argue that helpfulness is a measure of the perceived value of online reviews that reduces product uncertainty. Due to the large number of online reviews, however, consumers often encounter high search costs [74] and significant information overload [62] while seeking information that is reliable and helpful to diminish product uncertainty. If that information can be obtained at a lower search cost, consumers are less likely to continue their information search. The least-effort principle supports this viewpoint, suggesting that consumers who are economically minded seek ways to make decisions quickly and with minimal effort [5,6]. According to Park and Lee [62], consumers encounter information overload more quickly with attribute-value reviews than simple-recommendation reviews, even though the attribute-value reviews are generally rated as helpful. This suggests convenient information that reduces product uncertainty may make consumers perceive individual reviews less helpful. For example, if a product has high review valence (e.g., 4.9/5.0) created by a substantial number of reviewers (e.g., more than 1000 reviewers), this may be enough to reduce product uncertainty for a consumer. Subsequently, consumers may then perceive individual reviews as less helpful or even overlook them entirely. However, extant research on helpfulness rarely considers the presence of these convenient information such as product satisfaction and product popularity. Previous studies have examined the effect of product types (search vs. experience) on review helpfulness [17,45]. However, product intangibility has been largely overlooked, although it has been identified as an important factor that influences the relationship between eWOM (Electronic Word-Of-Mouth) and sales [1,3,37]. Services are inherently intangible, which often increases both the quality uncertainty and the perceived risk. Accordingly, it is reasonable that a product's intangibility influences the perceived review helpfulness. A greater level of uncertainty carried by intangible products should make the purchase decision riskier to the consumer, thus making online reviews even more of a factor in purchase decisions. In a contested market with high product variety, the importance of eWOM increases because consumers tend to rely on it to make efficient purchase decisions [7,57,91]. Although product variety may increase the importance of review valence, the variety is known to decrease the effect of individual reviews on purchase decisions; because individual reviews contain textual information, there are higher search costs required to obtain the information [93]. Even though it can be expected that product variety would affect perceived helpfulness of individual reviews, little research has examined this relationship.

Another issue in the extant review helpfulness literature is conflicting findings from different studies. The first area is related to reviewer expertise, which has been operationalized and measured in various ways in previous studies, including defining expertise by experience. Therefore, not only are expertise results difficult to interpret due to inconsistent operationalization of the expertise factor, so are experience results. Subsequently, examining expertise and experience individually and concurrently in the same study could bring greater clarity to both factors. Second, it is unclear how certain factors – including review inconsistency, extremity, and depth – influence review helpfulness due to conflicting results from previous studies. One of the possible reasons for the inconsistent findings is truncated review data, where the review focused on certain product categories or covered short periods of time. In order to fill in the gap, this research adopts an unusually large dataset from Amazon.com (14,051,211 reviews), including all available product categories (24 categories) at the ecommerce platform. This approach is expected to provide more comprehensive and generalizable findings, which can contribute to both theory and practice in the ecommerce arena.

This study attempts to expand on the body of knowledge and conceptual understanding of online review helpfulness, 1) adding context factors, such as product satisfaction, product popularity, product intangibility, and product variety, 2) investigating reviewer experience and expertise further, and 3) examining previous conflicting findings with a larger, more comprehensive dataset. The remainder of this paper is organized as follows: 2. Literature review; 3. Hypothesis Development; 4. Research Methodology; 5. Analysis and Results; 6. Discussion; and 7. Conclusion.

2. Literature review

2.1. Review helpfulness

In a research setting, a sizable amount of attention has been given to review helpfulness [43,67,82], though largely from review (content) and reviewer (source) perspectives. Previous research has shown that review helpfulness is influenced by review factors such as extremity [55,73], depth, [5,9,67], and rating [16,44]. Additionally, review helpfulness is influenced by reviewer factors such as expertise [67,95], identity [45,74], and rank and reputation [5,24,44]. However, there is one area that has not seen the same attention: product (context) perspectives. While search and experience goods have been compared [5,67,73], product attributes such as satisfaction, popularity, intangibility, and variety have barely been examined in any great detail; although, satisfaction and popularity have limited examination in the hotel industry [43,95]. Upon further scrutiny of the literature, other areas that need reexamination relate to the inconclusive results surrounding review inconsistency, extremity and depth, and reviewer expertise and experience.

2.2. Review factors

Review inconsistency's relationship with review helpfulness has shown conflicting results. Siering and Muntermann [73] found a negative effect, which is contradictory to that of Lee and Choeh [45] examining the relationship between review inconsistency and review helpfulness. With regard to the relationship between review extremity and review helpfulness, previous studies also reported mixed results. Using review data from Amazon.com, Siering and Muntermann [73] reported a negative effect of review extremity on review helpfulness, whereas Cao, Duan and Gan [9] found a positive effect of review extremity on review helpfulness from semantic analysis using text review data from CNET Download.com. Whereas, Cao, Duan and Gan [9] found that reviews with extreme opinions receive more helpfulness votes because such reviews tend to garner greater attention from consumers. Forman, Ghose and Wiesenfeld [26] and Lee and Choeh [44] also found the positive effect of extremity on helpfulness. However, Mudambi and Schuff [55], Chua and Banerjee [16], and Lee and Choeh [45] reported no significant effect of extremity on helpfulness.

The extant literature also shows inconsistent conclusions related to the effect of review depth. Most research reported a positive effect of review depth on helpfulness; review depth is operationalized as the number of words in reviews [5,16,33,55,73]. Researchers commonly maintain that consumers prefer reviews with more words because they tend to obtain more information about specific products without searching for more reviews [33,55]. However, other studies found that there is no significant relationship between review depth and review helpfulness. In a study on top reviewers at Amazon.com, Huang, Chen, Yen and Tran [36] reported no significant relationship between review depth and helpfulness, arguing that the number of words may not determine the helpfulness, particularly in the case of reviews created by prolific reviewers. Chua and Banerjee [16] also did not find a significant effect of the depth on review helpfulness. They suggested that if a review has a number of words that crosses a certain threshold, then readability becomes a concern, requiring extra effort from the reader (or higher search costs). In the same vein, Dong, Schaal, O'Mahony and Smyth [20] found that prediction accuracy for review helpfulness rarely increases by factoring in the number of words or the number of sentences. Interestingly, Racherla and Friske [67] reported a negative effect of review depth, implying that consumers who are overwhelmed by numerous reviews may prefer concise reviews.

2.3. Reviewer factors

Prior research substantiates the relationship between helpfulness and reviewer attributes such as reviewer's identity exposure [35], reputation [58,67] and ranking in review communities [5]. However, the impact of reviewer expertise on helpfulness has shown inconsistent results. A number of studies suggest a positive relationship between reviewer expertise and review helpfulness [17,49,95]. Because reviewers with more expertise tend to have more knowledge and formulate two-sided reviews, discussing both the pros and cons of items, their reviews tend to be perceived as more helpful. However, Racherla and Friske [67] reported a negative effect. This may be attributable to the strong certainty embedded in the reviews, which may introduce resistance to the expert opinions, and therefore, their reviews may be considered less helpful. Moreover, it appears that previous studies have not examined reviewer expertise and experience concurrently. The studies tend to examine expertise over experience. However, prior studies adopted measures for reviewer expertise such as helpfulness scores [43], expertise claims within textual reviews [17], volume of reviews contributed by a reviewer [50], peers' evaluation of a reviewer [95], and previous review history [67]. Given that expertise and experience are not synonymous, previous studies that examine expertise by measuring review volume and review history as expertise may not be measuring expertise but experience instead. Rather than measuring expertise as one factor, where experience could be construed as expertise, it may be prudent to measure two factors, expertise and experience concurrently.

3. Hypothesis development

3.1. Theoretical framework

This study adopts an integrated theoretical framework for online reviews proposed by Hlee, Lee and Koo [32]. The framework defines various factors related to utilization and management of online review platforms, including three dimensions: 1) source factor, 2) review factor, and 3) context factor. The source factor dimension includes reputation, identity, expertise, bandwagon, and authority, which are primarily concerned with reviewer attributes. The review factor dimension includes signals in individual reviews, such as attribute, visual, textual, and lexical cues. In addition to these two dimensions, which are typically covered by heuristic-systematic models, this framework includes a context factor dimension. The context factor dimension includes environmental factors of online review dynamics, such as types of products or services and types of businesses (e.g., restaurant vs. hotel). Thus, product intangibility and product variety, which are examined in this study, are categorized as context factors since they are environmental factors. Hlee, Lee and Koo [32] considered hotel classes in a five-star rating system, which represents the potential quality of hotels, as a context factor in this typology. Therefore, it is reasonable to expect that product satisfaction and product popularity are context factors since they highly depend on perceived product quality and sales volume.

3.2. Review factors

3.2.1. Review inconsistency and review extremity

Although many studies about review helpfulness investigated the relationship between review extremity and review helpfulness, the studies offer mixed results. Using different measures for extremity is one reason for the inconclusive results. Some studies defined extremity as the absolute difference between the rating of the review and the average rating of the product [9,44,45,73], while others operationalized it as extreme ratings in a 5-star rating system [16,26,55]. As Baek, Ahn and Choi [5] indicated, however, the absolute difference should be considered as review inconsistency rather than extremity, because it represents the difference between a reviewer's opinion and the other reviewers' opinions concerning the same product. The definition of extremity in communications refers to the tendency to include overly positive or negative ideas [8,80], thus the later measure (i.e., 1 or 5 star-ratings) is likely to estimate the extremity in online reviews.

Therefore, we distinguish review inconsistency and extremity in the current study. A negative relationship between review inconsistency and review helpfulness is expected. When individuals receive information inconsistent with their prior knowledge, they are less likely to accept information that conflicts with their prior belief [89,90]. In the context of online recommendation systems, similarly, opinions inconsistent with other reviewers tend to be perceived as less credible [11] and as less helpful by review readers. However, extremity is expected to have a positive relationship with review helpfulness for the following reasons. First, extreme opinions garner more attention from audiences [9]. Second, they can weaken or strengthen the current position of the audience, helping consumers make purchase decisions [42]. Lastly, moderate opinions (e.g., 3 in a 5 star-rating system) are often perceived as indifference by reviewers [42] and neither confirms nor weakens the current beliefs of review readers about desired products. Therefore, consumers are more likely to perceive such vague opinions less helpful. In addition, when review inconsistency is controlled, reviews with extremity should be perceived more helpful. This discussion introduces the following hypotheses:

H1. . *Review inconsistency has a negative relationship with review helpfulness.*

H2. . *Review extremity has a positive relationship with review helpfulness.*

3.2.2. Review depth

In this research, it is expected that review depth will have a positive effect on review helpfulness. Consumers read online reviews to reduce quality uncertainty; thus, a specific description about the product should help consumers imagine its potential quality [55]. When a review has more words, it would have a higher likelihood of including useful information for potential buyers. Although some studies reported an insignificant relationship between the depth and the helpfulness, they focused on a specific reviewer group or items in their analysis. For instance, Huang, Chen, Yen and Tran [36] limited their empirical analysis to reviews created by the top 60 reviewers. For such reviews, the quantity of comments would not be a critical determinant of review helpfulness because they tend to be well organized with an adequate amount of words. Chua and Banerjee [16] focused their analysis on the reviews of the top 100 bestselling items in search and experience goods categories. Popular products should have a lower level of quality uncertainty for buyers, which would lessen the effect of review depth on review helpfulness, leading to the following hypothesis:

H3. . *Review depth has a positive relationship with review helpfulness.*

3.3. Context factors

3.3.1. Product intangibility

As many prior studies indicate, product characteristics affect review helpfulness. Likewise, product tangibility can be another important characteristic to determine review helpfulness. Extant eWOM literature shows that its effect depends on product tangibility. The studies that define intangible products as service-oriented products such hotels [48,86], education [47], and restaurants [37] commonly suggested a greater effect of eWOM. The studies argue that because service-oriented have more complicated product attributes, which leads to an inherently higher risk, potential consumers rely more on reviews to reduce uncertainty and risk [1,37,48]. Whereas, a recent study [3] found that the effect of eWOM on product sales is less for intangible products because reviews on intangible products is less likely to be informative than reviews on tangible products, describing objective features of the items [70]. Therefore, we predict that consumers perceive reviews on intangible products less helpful. Due to the intangibility, it should be more difficult for consumers to estimate quality of intangible products even after experiencing them [79], as well as difficult to describe their experience in text reviews. In addition, it would be more difficult to capture important features of intangible products in photos or videos, which are critical determinants of review helpfulness [51,92].

H4. . *Product intangibility has a negative relationship with review helpfulness.*

3.3.2. Moderation of product intangibility on review extremity

The relationship between review extremity and review helpfulness may differ by product tangibility. Because products with more intangible attributes have higher quality uncertainty [41], consumers need clearer opinions with certainty to make purchase decisions. As discussed, extreme reviews tend to contain more certainty about product quality, helping consumers to quickly move toward buying the products or moving toward alternatives [42]. For intangible products with high product uncertainty, therefore, they would perceive reviews with extremity as more helpful, leading to the following hypothesis:

H5. . *The positive relationship between review extremity and review helpfulness is greater for intangible products.*

3.3.3. Moderation of product intangibility on review depth

The positive effect of review depth may increase for intangible products. As review depth increases, the review is likely to contain more information about product features, increasing its perceived helpfulness [33,55]. The relationship between review depth and helpfulness ought to be greater because reviewers presumably would have more difficulty describing their experience with an intangible product [79]. Therefore, reviewers would need to construct their reviews with more textual information to be useful for intangible products than for tangible products. Additionally, consumers tend to be reluctant to purchase intangible products from the Internet because it is more difficult to assure product quality with information only available through web pages [39]. This implies that online reviews for intangible products would need more information to be perceived as useful than those for tangible products. This introduces the following hypothesis:

H6. . *The positive relationship between review depth and review helpfulness is greater for intangible products.*

3.3.4. Product satisfaction and popularity

Numerous studies have revealed the importance of average review valence on consumer purchase decisions [18,21,34,81]. Although the valence may include diverse aspects of satisfaction such as consumers' satisfaction with their online shopping experience, with seller's attitude, and with shipping processes, it generally represents overall product satisfaction, which can be translated as consumer satisfaction with product quality [4,12,15,88]. According to Weisfeld-Spolter, Sussan and Gould [83], product satisfaction, often presented as an aggregate rating in a five-star rating system, is one of the leading eWOM factors that consumers focus on to understand potential product quality, because it is a single convenient measure that describes the quality without further cognitive effort. This implies that it could affect helpfulness perceptions and thus how readers vote reviews as helpful. It is reasonable to expect that consumers would be less likely to pay attention to individual reviews when a product has high product satisfaction ratings since consumers can grasp the overall quality without increasing their search costs to read individual reviews.

Product popularity plays a significant role in reducing uncertainty or confirming potential concerns and doubts about a desired product [61]. Similar to product quality, it can influence perceived helpfulness of a review. According to Zhu and Zhang [94], the impact of online reviews on purchase decision decreases for popular products. This implies that when a product has a high level of popularity (e.g., highly ranked in sales ranking list or purchased by a large number of consumers), consumers would consider it as a quality product without further information. Consumers would be less likely to read individual reviews, or if they read the reviews, they would perceive the reviews as less helpful. Therefore, we predict the following hypotheses:

H7. . *Product satisfaction has a negative relationship with review helpfulness.*

H8. . *Product popularity has a negative relationship with review helpfulness.*

3.3.5. Product variety

Product variety has been recognized as one of the critical determinants of customer satisfaction in online shopping [13,59]. Zhou and Duan [93] found that the variety of products affects the relationship between online reviews and purchase decisions by consumers in the software market. As there are more available products (i.e. product variety), it is more difficult to access and diagnose all of the individual online reviews concerning the products, which leads to information overload. Therefore, the effect of each individual review on purchase decisions decreases. When there are too many diverse products, consumers may rely on more convenient measures, such as rank and average review valence to estimate and compare potential quality of the products rather than individual reviews. This suggests that online reviews in product categories with greater variety are less likely to be read by potential customers. Although consumers may access the reviews, they would be less likely to perceive the reviews as helpful in making their purchase decision. This discussion suggests the following hypothesis:

H9. . *Product variety has a negative relationship with review helpfulness.*

3.4. Source factor

3.4.1. Reviewer expertise and experience

Extant literature of review helpfulness reports conflicting findings on the effect of reviewer expertise. One of the reasons for the inconsistent findings can be traced to using dissimilar measures for reviewer expertise. Past measures included helpfulness scores of a reviewer [49], expertise claims in a textual review [17], peers' evaluation on a reviewer [95], and the total number of reviews created by a reviewer in the online community [63,67]. Particularly, the number of reviews created by a reviewer would not accurately represent expertise of a reviewer. Even if a reviewer has contributed a large number of reviews (i.e., more experienced), it does not necessarily mean the reviewer has expertise in the topic or has the skills to write quality reviews. According to NBC News, there are numerous questionable reviewers who have posted an extraordinary number of reviews on Amazon, Yelp, Facebook, and Google, known as "fake online reviewers" who are paid to increase or decrease average rating scores on certain products [23]. Review readers are not likely to perceive such reviews as helpful, because they generally contain short generic textual reviews such as "I love it" to meet the minimum number of words required in the platform. In addition, some frequent reviewers may write reviews to attain rewards, rather than to provide productive, helpful information to the community. For example, many online review communities employ "incentive hierarchies" that offer members points for their review activities (e.g., writing review, voting for other reviews' helpfulness) and assign the members a higher level of status when they have collected the required amount of points for the promotion [29]. Therefore, some of the reviewers may simply focus on generating a large number of reviews without useful knowledge concerning the products, which does not reflect reviewer expertise. This indicates that the level of experience presented in the online review platforms (e.g., the number of reviews contributed by a reviewer) should be distinguished from perceived expertise (e.g., peer's evaluation, helpfulness score).

In the current study, reviewer expertise and experience are separate factors and are both predicted to have positive relationships with review helpfulness for the following reasons. First, there are a large number of reviewers in the e-commerce ecosystem. When there are a large number of information sources, the information generated by those with a higher degree of expertise is perceived more credible and reliable [67]. Second, in order to limit search cost, consumers may filter out reviews created by reviewers that lack expertise or experience if they encounter information overload in the search process [72]. In such a situation, the reviews may not have an opportunity to attain a helpfulness vote from consumers. Lastly, consumers generally perceive experienced reviewers as more reliable [95] and often consider them as opinion leaders in the online community. When a similar opinion is presented by individuals with different levels of the reliability, the opinion submitted by an expert or experienced source tends to be perceived more credible [68]. The discussion above suggests the following hypotheses:

H10. . *Reviewer expertise has a positive relationship with review helpfulness.*

H11. . *Reviewer experience has a positive relationship with review helpfulness.*

Fig. 1 summarizes the hypotheses discussed above.

4. Research methodology

4.1. Data acquisition, transformation and selection

This research adopts online review data from Amazon.com used for previous studies by McAuley, Targett, Shi and Van Den Hengel [53] and He and McAuley [30]. The initial raw dataset from May 1996 to July 2014 contained approximately 83 million online reviews (82,456,877 reviews) across 24 product categories. In order to prepare the dataset for testing the proposed hypotheses, extensive data processing was performed, including data transformation, creating variables, and data selection. Because the raw data was in a json format, which is not suitable for statistical analysis, data transformation was performed to produce a structured dataset. The structured dataset contained variables such as, reviewer ID, product ID (i.e., ASIN code), helpfulness and unhelpfulness votes, textual review, five-star rating score, and review date. Using the existing variables, variables for the current study were created. Duplicate reviews and reviews that had neither helpfulness nor unhelpfulness votes were removed from the dataset. Surprisingly, 50,008,536 reviews (60.65% of the initial dataset) did not have a vote for helpfulness or unhelpfulness. Lastly, considering the robustness of the analysis, reviews that have fewer than five total votes, either helpfulness and unhelpfulness, were deleted from the dataset [5,74]. Accordingly, approximately 14 million review data (14,051,211 reviews) remained for the analysis (Table 1).

4.2. Variable operationalization

4.2.1. Dependent measure: Review helpfulness

As in previous studies, review helpfulness is operationalized using the percentage of helpfulness votes [36,55,74]. It is estimated by dividing the number of helpfulness votes by the total number of votes.

4.2.2. Review factors: Review inconsistency, review extremity, and review depth

Review inconsistency is operationalized as the absolute difference between the rating score of an individual review and the average rating score of a product. Review extremity is measured by the star rating score of an individual review [26,55]. We coded review extremity as a binary variable, noting that if a review has one of the two extreme rating scores in the five-star rating system (i.e., 1 or 5), it is coded as 1 for review extremity (0 for non-extreme reviews). Review depth is the number of words in a review [55,67,74], not including special characters and spaces between words.

4.2.3. Context factors: intangibility, satisfaction, popularity, and variety

Product intangibility refers to "what cannot be seen, tasted, felt, heard or smelled" [54]. All products or services available in an online shopping environment have more intangibility than products in an in-person shopping environment [56]; however, the core of digital products, including software, digital music, and digital movies, is purely intangible [71]. Therefore, in the dataset we define digital content products as intangible products, including instant videos, mobile apps, digital music, Kindle e-books, movies, and video games. We coded product intangibility as a binary variable, assigning 1 if a review described a digital product (0 for other product reviews). Product

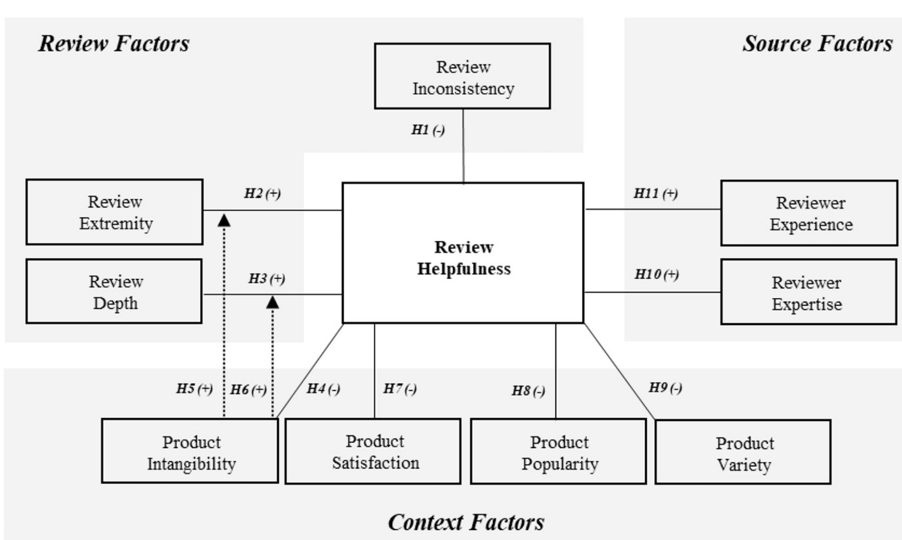

Fig. 1. Research Model.

Table 1
The Number of Items and Review Frequencies by Product Categories.

Category	# of products	# of reviews	Category	# of products	# of reviews
Instant video	23,965	583,934	Health	252,331	2,982,326
Apps	61,277	2638,172	Home	410,243	4,253,926
Automotive	320,115	1,373,767	Kindle	430,530	3,205,437
Baby	64,428	915,446	Movie	200,941	4,607,047
Beauty	249,274	2,023,070	Musical	83,047	500,176
Books	2,330,066	22,507,155	Office	130,006	1,243,186
CD & Vinyl	486,361	3,749,005	Patio	105,984	993,490
Cellphone	319,678	3,447,249	Pet	103,288	1,235,316
Clothing	1,136,004	5,748,920	Sports	478,898	3,268,695
Digital music	266,415	836,007	Tools	260,659	1,926,047
Electronics	476,002	7,824,482	Video Games	50,210	1,324,753
Grocery	166,049	1,297,156	Toys	327,698	2,252,771

satisfaction is considered consumer satisfaction with a product and is measured as the average star rating of a product [4,22,78]. Likewise, product popularity is operationalized as the number of ratings for a product [10,52,66,85]. Product variety is measured by the number of available products in a certain product category [14,25] that has at least one online review up to the date of review generation.

4.2.4. Reviewer factors: reviewer expertise and experience

Expertise is operationalized as the cumulative helpfulness of a reviewer, which is defined as the ratio of helpfulness votes to total votes received for all reviews generated by a reviewer. According to Herling [31], expertise refers to the possession of outstanding skills or knowledge in a certain domain, implying proficiency with creating a desired outcome. This suggests that expertise can exist when it is recognized and respected by knowledgeable people in the community. Although an individual has a great amount of knowledge and efficient skills, they will not be recognized as an expert without respect from other people [38]. In the context of online reviews, cumulative helpfulness can reflect recognition and respect of review readers as they acknowledge the reviewer's expertise, which may include domain knowledge, communication skills, and altruistic attitude. Meanwhile, reviewer experience is measured as the number of reviews generated by a particular reviewer [2,28,36].

4.2.5. Control variables: days lapsed and number of total votes

Two control variables related to perceived review helpfulness are included in the model. The first control variable is days lapsed. Day lapsed is the number of days a review has been available to consumers in an online community [67]. It negatively affects helpfulness votes [67,77] because online review systems generally present the latest reviews on the first page, giving them more visibility than older reviews on subsequent pages, thus require a greater level of effort from the reader to access older reviews. In addition, the latest reviews describe current quality of a product, which could be different from the initial quality described in older reviews. Therefore, older reviews tend to have fewer helpfulness votes. The second control variable is the total number of helpfulness and unhelpfulness votes for a review. As Mudambi and Schuff [55] indicated, because the dependent variable (i.e., review helpfulness) is a percentage, it hides the total number of votes on a review, which should affect perceived helpfulness.

5. Analysis and results

5.1. Analysis method

An empirical model was developed to test the proposed hypotheses. Due to the large size of the dataset, several variables have skewed and large distributions. A log-transformation was applied to enhance the normality of these variables: review depth, reviewer experience, product popularity, product diversity, days lapsed, and total votes. Because two hypotheses examine the interaction effect of product intangibility on the relationship between review extremity and review helpfulness and review depth and review helpfulness, interaction terms were created for these hypotheses. The research model including the variables is below:

$$\begin{aligned} \textit{Review Helpfulness} = \alpha_0 \; &+ \; \alpha_1 \textit{Review Inconsistencyy} + \alpha_2 \textit{Review Extremity} \\ &\quad + \alpha_3 \textit{Review Depth} \\ &+ \; \alpha_4 \textit{Product Intangibility} + \alpha_5 \textit{Review Extremity} \\ &\quad {}^{*}\textit{Product Intangibility} \\ &+ \; \alpha_6 \textit{Review Depth}^{*}\textit{Product Intangibility} \\ &\quad + \alpha_7 \textit{Product Satisfaction} \\ &+ \; \alpha_8 Ln(\textit{Product Popularity}) + \alpha_9 Ln(\textit{Product Variety}) \\ &\quad + \alpha_{10} \textit{Reviewer Expertise} \\ &+ \; \alpha_{11} Ln(\textit{Reviewer Experience}) + \alpha_{12} Ln(\textit{Days Lapsed}) \\ &\quad + \alpha_{13} Ln(\textit{Total Votes}) + \varepsilon \end{aligned}$$

To test the proposed model, Tobit regression was adopted as the primary analysis method for the following reasons [42,55,74]. First, because the dependent variable of this study is the ratio between the number of helpfulness votes and the total votes for an online review, it has an upper and a lower limit in its values from 0% to 100%. Second, the dataset may be subject to a selection bias because not all readers of reviews participate in the helpfulness vote. In addition to Tobit regression, the proposed hypotheses were tested using OLS regression for a robustness check [55]. Further, Tobit and OLS regression were conducted first with the full sample. Next the data were categorized by tangibility and a split sample analysis was performed, showing group comparisons. Finally, regressions were performed with various groups defined by the number of total helpfulness vote counts to check the robustness of our analysis.

5.2. Empirical results

5.2.1. Descriptive statistics

Table 2 illustrates the descriptive statistics for the variables. In the dataset, 9,335,872 (66.4%) out of 14,051,211 reviews were defined as having review extremity (a score of 1 or 5). Digital content products, which are defined as intangible products, represented 1,933,870 (13.8%) reviews. In the correlation matrix for the independent variables (Appendix 1), there are only low correlations between the independent variables. The highest correlation is between review inconsistency and product satisfaction (−0.321), which is far lower than 0.9 [19]. In addition, all Variance Inflation Factor (VIF) values of the independent variables are less than 1.24, indicating no multicollinearity in the variables [40].

5.2.2. Hypothesis test results

In the Tobit regression (Table 3), the coefficient of Review

Table 2
Descriptive statistics.

Variable	Mean	Standard deviation	Minimum	Maximum
Review helpfulness	0.41	0.11	0	0.91
Review inconsistency	1.06	0.91	0	3.97
Review depth	154.28	184.73	1	6808
Product satisfaction	4.06	0.66	1	5
Product popularity	285.25	1024.07	1	25,366
Product variety	1,087,138	971,205.60	23,965	2,330,066
Reviewer expertise	0.42	0.09	0	0.91
Reviewer experience	93.82	1076.79	1	44,557
Days lapsed	1864.19	1425.59	8	6619
Total votes	21.13	38.45	5	510
Frequency ($n = 14{,}051{,}211$)				
Review extremity	9,335,872 (66.4%)			
Product intangibility	1,933,869 (13.8%)			

Table 3
Results of Tobit and OLS Regression Tests (Total Votes ≥5).

Dependent: review helpfulness	Tobit		OLS ($R^2 = 0.5859$)	
	coefficient	*p*-value	coefficient	p-value
Constant	0.1618	< 0.001	0.1743	< 0.001
Review inconsistency	−0.0257	< 0.001	−0.0257	< 0.001
Review extremity	0.0160	< 0.001	0.0159	< 0.001
Review depth	0.0001	< 0.001	0.0001	< 0.001
Product intangibility	−0.0226	< 0.001	−0.0229	< 0.001
Review Extremity*Product Intangibility	0.0046	< 0.001	0.0046	< 0.001
Review Depth*Product Intangibility	0.0001	< 0.001	0.0001	< 0.001
Product satisfaction	−0.0021	< 0.001	−0.0019	< 0.001
Ln(Product Popularity)	−0.0023	< 0.001	−0.0024	< 0.001
Ln(Product Variety)	−0.0049	< 0.001	−0.0050	< 0.001
Reviewer Expertise	0.8705	< 0.001	0.8461	< 0.001
Ln(Reviewer Experience)	−0.0018	< 0.001	−0.0019	< 0.001
Ln(Days Lapsed): control	−0.0042	< 0.001	−0.0044	< 0.001
Ln(Total Votes): control	0.0095	< 0.001	0.0093	< 0.001

Inconsistency ($\alpha_1 = -0.0257$) is negative and statistically significant ($p < 0.001$), supporting Hypothesis 1. Hypothesis 2 is supported by the positive coefficient of Review Extremity ($\alpha_2 = 0.0160, p < 0.001$). The coefficient of Review Depth indicates a positive relationship between review depth and review helpfulness ($\alpha_3 = 0.0001, p < 0.001$), corresponding to the prediction of Hypothesis 3. Product Intangibility has a significant and negative coefficient ($\alpha_4 = -0.0226, p < 0.001$), supporting Hypothesis 4.

Concerning the interaction effect of product intangibility, the interaction term between review extremity and product intangibility is significant and positive ($\alpha_5 = 0.0046, p < 0.001$), confirming Hypothesis 5. The interaction term between review depth and product intangibility is positive and significant ($\alpha_6 = 0.0001, p < 0.001$), supporting Hypothesis 6. To check the robustness of the test results for Hypotheses 5 and 6, an additional split-sample analysis was performed. The simple slopes between the product categories were analyzed to further examine the interaction effect because hypothesis testing using interaction terms can be less powerful than a test of the simple slope differences [69] and the main effects may be affected by the interaction terms [55]. All of the variables related to the main effects are statistically significant and consistent with the regression results for interaction terms (Appendix 2). The coefficient of Review Extremity is higher in the analysis results for the intangible products ($\alpha_5 = 0.0245, p < 0.001$) than for the tangible products ($\alpha_5 = 0.015661, p < 0.001$). The simple slope analysis indicates the difference is significant ($\alpha_{difference} = 0.0089, p < 0.001$). Similarly, the coefficient of Review Depth is different for intangible products ($\alpha_6 = 0.0001, p < 0.001$) and tangible products ($\alpha_6 = 0.0001, p < 0.001$) and the difference is found to be significant ($\alpha_{difference} = 0.0001, p < 0.001$).[2] These findings are consistent with the results from the regressions for interaction terms for Hypothesis 5 and Hypothesis 6.

The hypothesis test results for Product Satisfaction is significant and negative ($\alpha_7 = -0.0021, p < 0.001$), consistent with the prediction of Hypothesis 7. Concerning Hypothesis 8, Ln(Product Popularity) has a significant and negative coefficient ($\alpha_8 = -0.0023, p < 0.001$). Thus, Hypothesis 8 is supported. Similarly, the significant and negative coefficient ($\alpha_9 = -0.0049, p < 0.001$) of Ln(Product Variety), lending support to Hypothesis 9. Reviewer Expertise has a significant and positive relationship with review helpfulness ($\alpha_{10} = 0.8705, p < 0.001$), supporting Hypothesis 10. However, Ln(Reviewer Experience) is negative and significant ($\alpha_{11} = -0.0018, p < 0.001$), which is contradictory to the prediction of Hypothesis 11. Two control variables in the models are statistically significant: Ln(Days Lapsed) is negative ($\alpha_{12} = -0.0042, p < 0.001$) while Ln(Total Votes) is positive ($\alpha_{13} = 0.0095, p < 0.001$).

As summarized in Tables 3, overall results from OLS regression are consistent with those from the Tobit regression. To check the robustness of the models, additional tests were performed with different sample sizes in terms of total helpfulness votes: with at least 10 total votes and with at least 15 total votes [74] (Appendix 3). There are 7,811,768 observations with at least 10 votes and 5,000,942 observations with at least 15 votes. The overall hypothesis test results remain constant across the analyses while the R-square increases as the minimum number of total votes increases.

6. Discussion

6.1. Review factors

Prior studies were not conclusive concerning the impact of review factors such as review inconsistency, review extremity, and review depth, in measuring online review helpfulness. We found review inconsistency has a negative relationship with perceived review helpfulness, corresponding to the finding of Siering and Muntermann [73] while contradictory to that of Lee and Choeh [45]. This finding indicates that when an online review is far from the overall average review rating, it may be perceived as less credible [11] and hence, less helpful. With regard to review extremity, our finding suggests it has a positive effect on review helpfulness, which is consistent with Forman, Ghose and Wiesenfeld [26]. This implies that extreme reviews are more useful because they help consumers confirm or reconsider their purchase decision more than moderate reviews [42]. Review depth is found to have a positive effect on the helpfulness, which supports the findings of most extant literature [5,16,33,55,73], though it differs from that of some current studies [16,36]. This suggests that online reviews with more textual information are perceived to be more helpful because they tend to deliver more explanations for star rating scores and product details.

6.2. Context factors

For intangible products, online reviews tend to be perceived as less helpful. This implies that it is more difficult to estimate the potential quality of intangible products such as movies, music, video games, and mobile apps [79]. Additionally, the description about quality tends to be more subjective and based on personal preference and feelings, which may not be useful for potential buyers. Further, a positive interaction effect between review extremity and review helpfulness was

[2] The coefficients of Review Depth for tangible/ intangible products look the same because they were rounded up at the fourth decimal. Their actual coefficients are 0.00000921 (tangible) and 0.0000137 (intangible) respectively. The split-sample analysis indicated the two coefficients were statistically different.

found. This implies that intangible products tend to have more quality uncertainty [41] and hence, extreme reviews can provide more clear and helpful suggestions for review readers. Similarly, the positive effect of review depth on helpfulness is found to be greater for intangible products. This suggests that the reviews for intangible products may need to contain more textual information to explain the reasons for their ratings and to describe the intangible features.

This study showed that both product satisfaction and popularity weaken perceived helpfulness of online reviews. In the presence of high product satisfaction and popularity, these findings imply that perceived helpfulness of individual reviews may decrease because satisfaction and popularity may be enough to assure product quality. However, when product satisfaction, which is presented in review valence, signals poor quality, individual reviews would be perceived as more helpful because they inform consumers of potential problems and use risks, helping consumers avoid such problems [87]. Similarly, our finding indicates that individual reviews for less popular products are perceived as more helpful. Because less popular products have more quality uncertainty [94], consumers would rely on individual reviews to make their purchase decision. Product variety is found to have a negative relationship with review helpfulness. This finding echoes the result by Zhou and Duan [93] reporting that the impact of online review on product sales decreases as more products are available in the market. In the presence of variety of products, the finding suggests that they may not perceive reviews as helpful unless it provides highly valuable information.

6.3. Source factors

The findings for source factors, including reviewer expertise and experience, provide an interesting conclusion. Not surprisingly, expertise is found to have a positive relationship with helpfulness, corresponding to the extant review helpfulness literature [17,49,95]. This verifies that consumers perceive online reviews created by expert reviewers as more credible and helpful [67]. Whereas, we found that reviewer experience has a negative relationship with helpfulness. This is contradictory to common findings in some studies [17,49,95], but consistent with Racherla and Friske [67], suggesting several possible explanations for this counterintuitive finding. One convincing explanation is that experienced reviewers may write reviews for diverse product categories but may fail to produce helpful reviews in categories where they do not have expertise. The dataset adopted in this study supports this explanation. Reviewers with more than 100 reviews had submitted their reviews to approximately 11.50 product categories (e.g., 12.31 categories for more than 200 reviews, 12.96 categories for more than 300 reviews, 13.33 categories for more than 400 reviews, 13.63 categories for more than 500 reviews), presenting a proportional relationship between the number of reviews contributed by a reviewer and the diversity of review domains. This implies that prolific reviewers at Amazon.com tend to review a broad range of product categories, rather than focusing on a few specific categories. This could produce less helpful reviews in some categories because it is difficult to achieve and maintain expertise in all categories. This explanation is further supported by the difference in the datasets of studies reporting contradictory findings. Studies that used online review data for a single product category [49,95] or a few similar categories [17] reported a positive effect of reviewer experience on review helpfulness, while a study adopting review data for diverse categories [67] found a negative effect. In addition, this finding may indicate the presence of “fake reviewers,” who simply participate in online reviews for unethical purposes [23], where consumers do not perceive these reviews as helpful. In the dataset for this study, for instance, there were a number of suspicious reviewers, producing a large number of reviews in a short period of time (e.g., 46,819 out of 22,728,542 reviewers generated more than 10 reviews per day on average). These findings suggest that the number of reviews contributed by a reviewer may not be an appropriate factor for determining online reviewer quality.

6.4. Theoretical and practical contributions

6.4.1. Theoretical contributions

This study provides theoretical contributions to the research of online review helpfulness. First, it introduces several novel predictors of online review helpfulness, such as product intangibility, product satisfaction, product popularity, and product variety, with empirical evidence of their relationships with review helpfulness. In particular, although many prior studies used a taxonomy of search/experience goods [55,67,73] or hedonic/utilitarian goods [60] to examine how the effect of review extremity and review depth differ by product types, product intangibility has been absent in previous studies. This study suggests that product intangibility is another significant factor that determines not only review helpfulness, but also the effect of review extremity and review depth on helpfulness. Second, it is one of the first studies that distinguishes between reviewer expertise and reviewer experience, illustrating that the two predictors have contradictory effects on review helpfulness. We highlight that these two factors should not be considered interchangeable, and that they should be separated in future research. This study also suggests that the total number of reviews contributed by a reviewer, which was used as a measure of review expertise in prior studies [63,67], may not be adequate to measure reviewer expertise. Third, this study provides reliable findings for the mixed results in the extant literature concerning the effect of review depth, review extremity, reviewer expertise, and review experience on review helpfulness. This study uses empirical evidence from analyzing a larger, more comprehensive dataset than those adopted in prior studies. This study also provides rigor to the findings, illustrating that the hypothesized relationships remain constant when tested with different levels of total helpfulness votes (i.e., total votes ≥ 5, 10, and 15) and with different analytical approaches (i.e., Tobit and OLS regression).

6.4.2. Practical contributions

This study contains several important insights for practitioners. As noted through the data preparation process, 60.65% of online reviews have neither helpfulness nor unhelpfulness votes. In addition, 38.47% of the reviews with no vote at all were in the review platform for more than 365 days. This demonstrates that a tremendous number of online reviews are ignored in the Amazon.com e-commerce platform, simply wasting data capacity. Practitioners of online review platforms need to consider how to manage the reviews that provide little value for their data resource efficiency. They may consider discarding useless reviews after a certain time period or consider not displaying them on product pages.

Concerning review factors, the positive relationship between review extremity and helpfulness (Hypothesis 2) suggests that user interface designers of online review platforms may consider presenting extremely negative or positive reviews (i.e., 1 or 5 star-ratings) along with sufficient review depth (Hypothesis 3) at the top of product pages to help consumers make effective purchase decisions. This approach also can reduce consumers' search cost and consequently, make the online platform more user friendly and increase website stickiness, which leads to sales increases [55]. In particular, this may be more effective for digital or intangible products, where the effect of review extremity is more substantial (Hypothesis 5).

The findings concerning context factors provide important practical implications. The negative effect of product intangibility on review helpfulness (Hypothesis 4) suggests that managers of online review platforms may consider discarding the individual text review system for digital products, such as movies, music, and software. Since reviews with textual information for digital products are perceived to be less helpful by readers, it would be more effective to focus on the five-star rating system. This corresponds to the decision by Netflix, which is one of the largest digital content providers, removing their text review system, instead, adopting a thumbs up and thumbs down vote because consumers pay little attention to individual text reviews [76]. If online review platforms want to keep the text reviews for digital products, they need to consider strategies to collect more in-depth text reviews, which are perceived to be more helpful (Hypothesis 6). For example, they may more actively invite verified expert reviewers for each digital product category and place their reviews at the top of each digital product page, to help potential consumers make purchase decisions. In addition, they might require a minimum number of words in the text reviews for digital products to discourage reviewers from producing vague, useless reviews, such as “It is cool”, or “I don't like it.” The negative effect of product satisfaction (Hypothesis 7), product popularity (Hypothesis 8), and product variety (Hypothesis 9) on review helpfulness indicates that the importance of the reviews depends on product attributes, which are context factors. Therefore, practitioners need to employ different strategies in collecting online reviews from individual consumers. For example, they may less actively encourage buyers of products with high satisfaction and high popularity, and more actively encourage buyers of products with low satisfaction and low popularity. For prolific online reviewers, the finding of a negative relationship between product variety and helpfulness (Hypothesis 9) suggests that reviewers need to focus their reviews on less competitive product categories if they want to obtain more helpfulness votes.

Lastly, the findings from Hypothesis 10 and Hypothesis 11 suggest that online review platforms should not solely use the number of reviews contributed by a reviewer (Hypothesis 11) to estimate reviewer quality and status but instead, focus on the cumulative helpfulness of a reviewer (Hypothesis 10). In particular, online platforms with reviews of diverse products and services (e.g., Amazon.com, Google.com, Yelp.com) should seriously consider this suggestion, because it is likely that there are numerous reviewers who are writing reviews for a broad range of product categories, including categories in which the reviewers are not providing helpful information. However, this suggestion does not imply that the number of reviews written by a reviewer should be ignored, only utilized in a different way. For instance, review platforms might consider using the number of reviews to weigh reviewer expertise to estimate the consistency of the expertise. It is reasonable to suggest that reviewers who attained 80% of helpfulness votes in 100 reviews are consistently more reliable and their expertise valued more than those who attained the same votes in 10 reviews.

7. Conclusion

As e-commerce gains in popularity, so does the amount of online reviews. Online reviews have increased tremendously, introducing information overload to consumers. Because of this, online review systems need to provide reviews in a more effective way. The findings in this study suggest that online review system management strategies should differ by context factors, such as product intangibility, product satisfaction, product popularity, and product variety. Additionally, when evaluating the quality of reviewers, review systems ought to estimate quality of reviewers by product category, emphasizing helpfulness votes attained in each category, and provide reviews of such expert reviewers in the category prior to the other reviews.

Although this study provides significant contributions to the theoretical foundation of review helpfulness research and it offer numerous practical implications, it has limitations that present opportunities for future research. First, although our dataset includes a large number of online reviews and contains many diverse product categories, it includes only reviews that were rated by consumers. This resulted in a reduction of 60.65% of reviews (50,008,536 reviews) from the initial dataset. As Mudambi and Schuff [55] pointed out, reviews that do not have helpfulness votes might still be helpful or unhelpful. For example, reviews in the categories with high product variety might have not received helpfulness votes because there are too many reviews. Second, although our empirical model explains a fairly large variance of helpfulness, between 58% and 70% depending on the size of the dataset, there may be other relevant predictors of review helpfulness. As prior studies reported, the number of photos, review readability [24], identity disclosure [67], reviewer innovativeness [60], and average sentence length [65] can be important predictors that can provide a more comprehensive understanding of the dynamics of review helpfulness. Moreover, even though this study was not able to incorporate consumer-provided visual content such as photos and videos due to technical limitations of data collection, it is an important review factor that can predict perceived review helpfulness [51,84]. Third, the dataset adopted in this study is somewhat dated. Due to the complexity and the size of raw data, a long period of time was spent preparing data for the analysis. Data preparation is one of the most significant challenges when research includes big data [75]. Future studies need to consider this issue in their research design to deliver more timely research outcomes. For example, researchers may want to adopt cloud computing platforms (e.g., Amazon Web Service, Google Cloud) for initial data processing, transforming data from an unstructured data format to a structured data format, merging data, and creating new data fields. This would provide a more stable and accurate data processing result without unexpected system crashes since cloud computing platforms allow for scaling processing powers without upgrading computing technologies. In particular, when researchers do not have advanced computing systems with adequate processing power, it would be advantageous to select a cloud computing platform for data preparation. When performing statistical analysis with big data, a useful analysis tool is STATA, though R and Python are known as more reliable for big data analytics. In the current study analyzing approximately 14 million observations, STATA showed faster and more stable performance than R and Python. Lastly, our findings are based on a dataset collected from a single online review platform, Amazon.com. Consequently, the findings of this study may not remain consistent if data are gathered from other review platforms that provide different review environments. Future studies may attempt to use our methodology, while adopting data from multiple sources and platforms, and comparing the differences in the results.

Acknowledgement

We deeply appreciate Dr. Julian McAuley sharing his data for our research and Mr. Fredrick Woolard helping process the data for analysis. This research would have not been possible without their exceptional help and contribution. We would also like to show our gratitude to the editor and anonymous reviewers for their valuable suggestions that significantly improved our manuscript.

Appendix 1 Correlation matrix

Variable	RI	RD	RE1	PS	PP	RE2	DL	TV	PV	RE3
RI	1.000									
RD	0.124	1.000								
RE1	−0.071	−0.092	1.000							
PS	−0.321	0.010	0.145	1.000						
PP	−0.253	0.114	0.029	0.069	1.000					
RE2	0.109	0.019	−0.011	−0.068	−0.010	1.000				

DL	0.056	−0.015	0.009	−0.106	−0.086	0.110	1.000			
TV	−0.033	0.011	0.039	0.026	0.022	−0.015	−0.007	1.000		
PV	−0.152	−0.014	0.084	0.026	0.061	−0.102	0.009	0.014	1.000	
RE3	0.064	0.022	0.199	0.064	−0.025	0.047	0.018	−0.005	0.090	1.000

Note: RI (Review Inconsistency), RD (Review Depth), RE1 (Reviewer Expertise), PS (Product Satisfaction), PP (Product Popularity), RE2 (Reviewer Experience), DL (Days Lapsed), TV (Total Votes), PV (Product Variety), RE3 (Review Extremity)

Appendix 2 Results of Group Comparison (total votes ≥5, * > 0.01)

Dependent: review helpfulness	Intangible products (n = 1,933,869)		Tangible products (*n* = 12,117,342)	
	Tobit	OLS (R^2 = 0.5787)	Tobit	OLS (R^2 = 0.5819)
Constant	0.1674*	0.1817*	0.1548*	0.1668*
Review inconsistency	−0.0373*	−0.0370*	−0.0238*	−0.0239*
Review extremity	0.0245*	0.0243*	0.0156*	0.0156*
Review depth	0.0001*	0.0001*	0.0001*	0.0001*
Product satisfaction	−0.0020*	−0.0014*	−0.0019*	−0.0017*
Ln(Product Popularity)	−0.0037*	−0.0039*	−0.0021*	−0.0022*
Ln(Product Variety)	−0.0042*	−0.0043*	−0.0049*	−0.0050*
Reviewer expertise	0.8329*	0.8017*	0.8763*	0.8532*
Ln(Reviewer Experience)	−0.0033*	−0.0034*	−0.0016*	−0.0016*
Ln(Days Lapsed): control	−0.0054*	−0.0056*	−0.0041*	−0.0042*
Ln(Total Votes): control	0.0120*	0.0116*	0.0091*	0.0090*

Appendix 3 Tobit and OLS Regression Results with Different Samples (* > 0.01)

Dependent: review helpfulness	Votes ≥ 10 (*n* = 7,811,768)		Votes ≥ 15 (*n* = 5,000,942)	
	Tobit	OLS (R^2 = 0.6580)	Tobit	OLS (R^2 = 0.6997)
Constant	0.1582*	0.1655*	0.1384*	0.1437*
Review Inconsistency	−0.0250*	−0.0251*	−0.0239*	−0.0239*
Review Extremity	0.0143*	0.0143*	0.0128*	0.0128*
Review Depth	0.0000*	0.0000*	0.0000*	0.0000*
Product Intangibility	−0.0213*	−0.0215*	−0.0193*	−0.0195*
Review Extremity*Product Intangibility	0.0040*	0.0040*	0.0027*	0.0028*
Review Depth*Product Intangibility	0.0000*	0.0000*	0.0000*	0.0000*
Product Satisfaction	−0.0029*	−0.0027*	−0.0031*	−0.0030*
Ln(Product Popularity)	−0.0018*	−0.0018*	−0.0015*	−0.0016*
Ln(Product Variety)	−0.0044*	−0.0045*	−0.0039*	−0.0040*
Reviewer Expertise	0.8894*	0.8764*	0.9053*	0.8963*
Ln(Reviewer Experience)	−0.0013*	−0.0013*	−0.0014*	−0.0014*
Ln(Days Lapsed): control	−0.0035*	−0.0036*	−0.0031*	−0.0031*
Ln(Total Votes): control	0.0047*	0.0046*	0.0054*	0.0054*

References

[1] M.E. Abd-Elaziz, W.M. Aziz, G.S. Khalifa, M. Abdel-Aleem, Determinants of Electronic word of mouth (EWOM) influence on hotel customers' purchasing decision, Int. J. Heritage, Tour. Hospital. 9 (2/2) (2015).

[2] A. Agnihotri, S. Bhattacharya, Online review helpfulness: role of qualitative factors, Psychol. Mark. 33 (11) (2016) 1006–1017.

[3] A. Babić Rosario, F. Sotgiu, K. De Valck, T.H. Bijmolt, The effect of electronic word of mouth on sales: a meta-analytic review of platform, product, and metric factors, J. Mark. Res. 53 (3) (2016) 297–318.

[4] S. Baccianella, A. Esuli, F. Sebastiani, Multi-facet rating of product reviews, European Conference on Information Retrieval, Springer, 2009, pp. 461–472.

[5] H. Baek, J. Ahn, Y. Choi, Helpfulness of online consumer reviews: Readers' objectives and review cues, Int. J. Electron. Commer. 17 (2) (2012) 99–126.

[6] S.J. Basinger, H. Lavine, Ambivalence, information, and electoral choice, Am. Polit. Sci. Rev. 99 (2) (2005) 169–184.

[7] S. Boo, J. Kim, Comparison of negative eWOM intention: an exploratory study, J. Qual. Assur. Hosp. Tour. 14 (1) (2013) 24–48.

[8] E. Brummelman, J. Crocker, B.J. Bushman, The praise paradox: when and why praise backfires in children with low self-esteem, Child Dev. Perspect. 10 (2) (2016) 111–115.

[9] Q. Cao, W. Duan, Q. Gan, Exploring determinants of voting for the "helpfulness" of online user reviews: a text mining approach, Decis. Support. Syst. 50 (2) (2011) 511–521.

[10] P.-Y. Chen, S.-y. Wu, J. Yoon, The impact of online recommendations and consumer feedback on sales, ICIS 2004 Proceedings, 58 2004.

[11] M.Y. Cheung, C. Luo, C.L. Sia, H. Chen, Credibility of electronic word-of-mouth: informational and normative determinants of on-line consumer recommendations, Int. J. Electron. Commer. 13 (4) (2009) 9–38.

[12] P.K. Chintagunta, S. Gopinath, S. Venkataraman, The effects of online user reviews on movie box office performance: accounting for sequential rollout and aggregation across local markets, Mark. Sci. 29 (5) (2010) 944–957.

[13] N. Cho, S. Park, Development of electronic commerce user-consumer satisfaction index (ECUSI) for internet shopping, Ind. Manag. Data Syst. 101 (8) (2001) 400–406.

[14] J.-K. Chong, T.-H. Ho, C.S. Tang, Product structure, brand width and brand share, Product Variety Management, Springer, 1998, pp. 39–64.

[15] A. Chou, L. Shu, Towards extracting affordances from online consumer product reviews, ASME Paper No. DETC2014–35288, 2014.

[16] A.Y. Chua, S. Banerjee, Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth, J. Assoc. Inf. Sci. Technol. 66 (2) (2015) 354–362.

[17] A.Y. Chua, S. Banerjee, Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality, Comput. Hum. Behav. 54 (2016) 547–554.

[18] G. Cui, H.-K. Lui, X. Guo, The effect of online consumer reviews on new product sales, Int. J. Electron. Commer. 17 (1) (2012) 39–58.

[19] I.R. Dohoo, C. Ducrot, C. Fourichon, A. Donald, D. Hurnik, An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies, Prevent. Vet. Med. 29 (3) (1997) 221–239.

[20] R. Dong, M. Schaal, M.P. O'Mahony, B. Smyth, Topic extraction from online reviews for classification and recommendation, Twenty-Third International Joint Conference on Artificial Intelligence, 2013.

[21] W. Duan, B. Gu, A.B. Whinston, Informational cascades and software adoption on the internet: an empirical investigation, MIS Q. (2009) 23–48.

[22] T.H. Engler, P. Winter, M. Schulz, Understanding online product ratings: a customer satisfaction model, J. Retail. Consum. Serv. 27 (2015) 113–120.

[23] J. Enoch, Can you trust online reviews? Here's how to find the fakes, NBC News, 2019.
[24] B. Fang, Q. Ye, D. Kucukusta, R. Law, Analysis of the perceived value of online tourism reviews: influence of readability and reviewer characteristics, Tour. Manag. 52 (2016) 498–506.
[25] L. Felipe Scavarda, A. Reichhart, S. Hamacher, M. Holweg, Managing product variety in emerging markets, Int. J. Oper. Prod. Manag. 30 (2) (2010) 205–224.
[26] C. Forman, A. Ghose, B. Wiesenfeld, Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets, Inf. Syst. Res. 19 (3) (2008) 291–313.
[27] L. Fullerton, Online reviews impact purchasing decisions for over 93% of consumers, report suggests, The Drum, 2017.
[28] B. Gao, N. Hu, I. Bose, Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews, Decis. Support. Syst. 95 (2017) 1–11.
[29] P.B. Goes, C. Guo, M. Lin, Do incentive hierarchies induce user effort? Evidence from an online knowledge exchange, Inf. Syst. Res. 27 (3) (2016) 497–516.
[30] R. He, J. McAuley, Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering, Proceedings of the 25th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2016, pp. 507–517.
[31] R.W. Herling, Operational definitions of expertise and competence, Adv. Dev. Hum. Resour. 2 (1) (2000) 8–21.
[32] S. Hlee, H. Lee, C. Koo, Hospitality and tourism online review research: a systematic analysis and heuristic-systematic model, Sustainability 10 (4) (2018) 1141.
[33] H. Hong, D. Xu, G.A. Wang, W. Fan, Understanding the determinants of online review helpfulness: a meta-analytic investigation, Decis. Support. Syst. 102 (2017) 1–11.
[34] N. Hu, N.S. Koh, S.K. Reddy, Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales, Decis. Support. Syst. 57 (2014) 42–53.
[35] N. Hu, L. Liu, J.J. Zhang, Do online reviews affect product sales? The role of Reviewer Characteristics and Temporal Effects, Information Technology and Management 9 (3) (2008) 201–214.
[36] A.H. Huang, K. Chen, D.C. Yen, T.P. Tran, A study of factors that contribute to online review helpfulness, Comput. Hum. Behav. 48 (2015) 17–27.
[37] E. Jeong, S.S. Jang, Restaurant experiences triggering positive electronic word-of-mouth (eWOM) motivations, Int. J. Hosp. Manag. 30 (2) (2011) 356–366.
[38] B.I. Johanna, M. van der Heijden, The development and psychometric evaluation of a multidimensional measurement instrument of professional expertise, High Ability Stud. 11 (1) (2000) 9–39.
[39] C. Katawetawaraks, C. Wang, Online shopper behavior: influences of online shopping decision, Asian J. Business Res. 1 (2) (2011).
[40] P. Kennedy, A Guide to Econometrics Cambridge, The MIT Press, MA, 2003.
[41] Y. Kim, R. Krishnan, On product-level uncertainty and online purchase behavior: An empirical analysis, Manag. Sci. 61 (10) (2015) 2449–2467.
[42] N. Korfiatis, E. GarcíA-Bariocanal, S. SáNchez-Alonso, Evaluating content quality and helpfulness of online product reviews: the interplay of review helpfulness vs. review content, Electron. Commer. Res. Appl. 11 (3) (2012) 205–217.
[43] M. Lee, M. Jeong, J. Lee, Roles of negative emotions in customers' perceived helpfulness of hotel reviews on a user-generated review website, Int. J. Contemp. Hosp. Manag. 29 (2) (2017) 762–783.
[44] S. Lee, J.Y. Choeh, Predicting the helpfulness of online reviews using multilayer perceptron neural networks, Expert Syst. Appl. 41 (6) (2014) 3041–3046.
[45] S. Lee, J.Y. Choeh, The determinants of helpfulness of online reviews, Behav. Inform. Technol. 35 (10) (2016) 853–863.
[46] S. Lee, J.Y. Choeh, The interactive impact of online word-of-mouth and review helpfulness on box office revenue, Manag. Decis. 56 (4) (2018) 849–866 (ISSN: 0025-1747).
[47] C. Li, X. Wang, The power of eWOM: a re-examination of online student evaluations of their professors, Comput. Hum. Behav. 29 (4) (2013) 1350–1357.
[48] S.W. Litvin, R.E. Goldsmith, B. Pan, Electronic word-of-mouth in hospitality and tourism management, Tour. Manag. 29 (3) (2008) 458–468.
[49] Y. Liu, X. Huang, A. An, X. Yu, HelpMeter: a nonlinear model for predicting the helpfulness of online reviews, Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01, IEEE Computer Society, 2008, pp. 793–796.
[50] Z. Liu, S. Park, What makes a useful online review? Implication for travel product websites, Tour. Manag. 47 (2015) 140–151.
[51] Y. Ma, Z. Xiang, Q. Du, W. Fan, Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep leaning, Int. J. Hosp. Manag. 71 (2018) 120–131.
[52] E. Maslowska, E.C. Malthouse, V. Viswanathan, Do customer reviews drive purchase decisions? The moderating roles of review exposure and price, Decis. Support. Syst. 98 (2017) 1–9.
[53] J. McAuley, C. Targett, Q. Shi, A. Van Den Hengel, Image-based recommendations on styles and substitutes, Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2015, pp. 43–52.
[54] B. Mittal, The advertising of services: meeting the challenge of intangibility, J. Serv. Res. 2 (1) (1999) 98–116.
[55] S.M. Mudambi, D. Schuff, What makes a helpful review? A study of customer reviews on Amazon. com, MIS Q. 34 (1) (2010) 185–200.
[56] M.V. Nepomuceno, M. Laroche, M.-O. Richard, How to reduce perceived risk when buying online: the interactions between intangibility, product knowledge, brand familiarity, privacy and security concerns, J. Retail. Consum. Serv. 21 (4) (2014) 619–629.
[57] K. Nusair, eWOM Behavior in Social Networking Websites: An Empirical Investigation of Antecedents, GAI International Academic Conferences Proceedings, 2016, p. 296.
[58] J. Otterbacher, 'Helpfulness' in online communities: a measure of message quality, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2009, pp. 955–964.
[59] C. Page, E. Lepkowska-White, Web equity: a framework for building consumer value in online companies, J. Consum. Mark. 19 (3) (2002) 231–248.
[60] Y. Pan, J.Q. Zhang, Born unequal: a study of the helpfulness of user-generated product reviews, J. Retail. 87 (4) (2011) 598–612.
[61] D.-H. Park, S. Kim, The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews, Electron. Commer. Res. Appl. 7 (4) (2008) 399–410.
[62] D.-H. Park, J. Lee, eWOM overload and its effect on consumer behavioral intention depending on consumer involvement, Electron. Commer. Res. Appl. 7 (4) (2008) 386–398.
[63] S. Park, J.L. Nicolau, Asymmetric effects of online consumer reviews, Ann. Tour. Res. 50 (2015) 67–83.
[64] P.A. Pavlou, Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model, Int. J. Electron. Commer. 7 (3) (2003) 101–134.
[65] A. Qazi, K.B.S. Syed, R.G. Raj, E. Cambria, M. Tahir, D. Alghazzawi, A concept-level approach to the analysis of online review helpfulness, Comput. Hum. Behav. 58 (2016) 75–81.
[66] L. Qiu, D. Li, Effects of aggregate rating on eWOM acceptance: an attribution theory perspective, PACIS, 2010, p. 147.
[67] P. Racherla, W. Friske, Perceived 'usefulness' of online consumer reviews: An exploratory investigation across three services categories, Electron. Commer. Res. Appl. 11 (6) (2012) 548–559.
[68] M.-A. Reinhard, S.L. Sporer, Content versus source cue information as a basis for credibility judgments, Soc. Psychol. 41 (2) (2010) 93–104.
[69] C.D. Robinson, S. Tomek, R. Schumacker, Tests of moderation effects: difference in simple slopes versus the interaction term, Multiple Linear Regr. Viewpoints 39 (1) (2013) 16–24.
[70] A. Rossmann, K.R. Ranjan, P. Sugathan, Drivers of user engagement in eWoM communication, J. Serv. Mark. 30 (5) (2016) 541–553.
[71] J. Rowley, Understanding digital content marketing, J. Mark. Manag. 24 (5–6) (2008) 517–540.
[72] R. Savolainen, Filtering and withdrawing: strategies for coping with information overload in everyday contexts, J. Inf. Sci. 33 (5) (2007) 611–621.
[73] M. Siering, J. Muntermann, What Drives the Helpfulness of Online Product Reviews? From Stars to Facts and Emotions, 7 Wirtschaftsinformatik, 2013.
[74] M. Siering, J. Muntermann, B. Rajagopalan, Explaining and predicting online review helpfulness: the role of content and reviewer-related signals, Decis. Support. Syst. 108 (2018) 1–12.
[75] U. Sivarajah, M.M. Kamal, Z. Irani, V. Weerakkody, Critical analysis of big data challenges and analytical methods, J. Bus. Res. 70 (2017) 263–286.
[76] D. Smith, Why Netflix's 'thumbs up, thumbs down' ratings system is fundamentally flawed, Business Insider businessinsider.com (2017).
[77] M.D. Smith, E. Brynjolfsson, Consumer decision-making at an internet shopbot: brand still matters, J. Ind. Econ. 49 (4) (2001) 541–558.
[78] E.I. Sparling, S. Sen, Rating: how difficult is it? Proceedings of the fifth ACM Conference on Recommender Systems, ACM, 2011, pp. 149–156.
[79] W.-C. Tsao, M.-T. Hsieh, eWOM persuasiveness: do eWOM platforms and product type matter? Electron. Commer. Res. 15 (4) (2015) 509–541.
[80] T.K. Van Dijk, F. Datema, S. Welten, F.J. Van de Vijver, Acquiescence and extremity in cross-national surveys: Domain dependence and country-level correlates, (2009).
[81] I.E. Vermeulen, D. Seegers, Tried and tested: the impact of online hotel reviews on consumer consideration, Tour. Manag. 30 (1) (2009) 123–127.
[82] X. Wang, L. Tang, E. Kim, More than words: do emotional content and linguistic style matching matter on restaurant review helpfulness? Int. J. Hosp. Manag. 77 (2019) 438–447.
[83] S. Weisfeld-Spolter, F. Sussan, S. Gould, An integrative approach to eWOM and marketing communications, Corp. Commun. Int. J. 19 (3) (2014) 260–274.
[84] P. Xu, L. Chen, R. Santhanam, Will video be the next generation of e-commerce product reviews? Presentation format and the role of product type, Decis. Support. Syst. 73 (2015) 85–96.
[85] J. Yang, R. Sarathy, J. Lee, The effect of product review balance and volume on online Shoppers' risk perception and purchase intention, Decis. Support. Syst. 89 (2016) 66–76.
[86] C.-L.A. Yen, C.-H.H. Tang, Hotel attribute performance, eWOM motivations, and media choice, Int. J. Hosp. Manag. 46 (2015) 79–88.
[87] D. Yin, S. Mitra, H. Zhang, 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) (2016) 131–144.
[88] M. Zeelenberg, R. Pieters, Beyond valence in customer dissatisfaction: a review and new findings on behavioral responses to regret and disappointment in failed services, J. Bus. Res. 57 (4) (2004) 445–455.
[89] V.A. Zeithaml, Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence, J. Mark. 52 (3) (1988) 2–22.
[90] W. Zhang, S. Watts, Knowledge adoption in online communities of practice, ICIS 2003 Proceedings, 2003, p. 9.
[91] Z. Zhang, Q. Ye, R. Law, Y. Li, The impact of e-word-of-mouth on the online popularity of restaurants: a comparison of consumer reviews and editor reviews, Int. J. Hosp. Manag. 29 (4) (2010) 694–700.

[92] S. Zhou, B. Guo, The order effect on online review helpfulness: a social influence perspective, Decis. Support. Syst. 93 (2017) 77–87.
[93] W. Zhou, W. Duan, Online user reviews, product variety, and the long tail: An empirical investigation on online software downloads, Electron. Commer. Res. Appl. 11 (3) (2012) 275–289.
[94] F. Zhu, X. Zhang, Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics, J. Mark. 74 (2) (2010) 133–148.
[95] L. Zhu, G. Yin, W. He, Is this opinion leader's review useful? Peripheral cues for online review helpfulness, J. Electron. Commer. Res. 15 (4) (2014) 267.
[96] X. Zhu, Y. Kim, H. Park, Do messages spread widely also diffuse fast? Examining the effects of message characteristics on information diffusion, Comput. Hum. Behav. 103 (2020) 37–47.

Hoon S. Choi is an assistant professor of the Department of Computer Information Systems at Appalachian State University. He received his Ph.D. in Information Systems and Cyber Security at the University of Texas at San Antonio. His research interests include e-commerce, electronic Word-of-Mouth (eWOM), digital piracy, data analytics, information security, and human-computer interactions. Dr. Choi published his papers in journals such as Journal of Management Information Systems (JMIS), Decision Support Systems (DSS), and Journal of Electronic Commerce Research (JECR), and Journal of Information Systems Applied Research (JISAR).

Steven Leon is an Associate Professor of Supply Chain Management in the Department of Marketing and SCM in the Walker College of Business, Appalachian State University. His research interests are in the areas of air transportation, service systems, and technology. His research has appeared in a variety of journals including Industrial Management and Data Systems, SCM: An International Journal, International Journal of Logistics Management, and Transportation Journal. He is the author of Financial Intelligence for Supply Chain Managers: Understand the Link between Operations and Corporate Financial Performance. His PhD is in Transportation and Logistics from North Dakota State University and his MBA is in International Business from Loyola University Maryland.