Elaboration likelihood model, endogenous quality indicators, and online review helpfulness

Yen-Chun Chou [a], Howard Hao-Chun Chuang [a,*], Ting-Peng Liang [b,1]

[a] College of Commerce, National Chengchi University Taipei, Taiwan
[b] Electronic Commerce Research Center, National Sun Yat-Sen University Kaohsiung, Taiwan

ARTICLE INFO

ABSTRACT

Given strong influences of online customer reviews on consumer purchase decisions, identifying helpful reviews has received broad attention from practitioners and researchers. The elaboration likelihood model (ELM) has been adopted to explain the review feature–helpfulness link. However, when analyzing reviews from websites, existing studies tend to ignore that quality indicators such as length and readability are merely cues and have not circumvented endogeneity induced by unseen argument quality. Hence, we propose an extended ELM application to observational data on review helpfulness. We develop a research model that integrates relevant quality indicators and sentiment features based on a circumplex model of affect. To test our hypotheses, we use publicly available review datasets from three platforms (Amazon.com, Drugs.com, and Yelp.com) and adopt an instrument-free method that allows for arbitrary correlations between unseen argument quality and multiple endogenous indicators. Our analysis shows that ignoring endogeneity would result in invalid effect size and hypothesis-testing. In addition to identifying effects of endogenous quality indicators on review helpfulness, we find asymmetric effects of positive and negative valence contingent on low or high arousal. By articulating conceptual pitfalls and illustrating empirical remedies, our study aims to be a prototypical example of performing ELM-grounded analyses of online customer reviews.

1. Introduction

User-generated customer reviews are increasingly influential on consumers' purchase decisions [46]. However, in the presence of a large number of reviews, quickly extracting useful information is a non-trivial task. Hence, identifying features of online reviews that affect the perceived usefulness by consumers has received wide attention from both the industry and academia. On the industrial front, platforms such as Amazon.com and Yelp.com engage in distributing useful reviews to customers for a wide variety of products/services. They allow users to rate whether a review is useful or not. Top-listed positive and negative reviews for each item are based on the helpful votes received by those reviews. Online reviews have been shown to affect sales performance and consumer choice of service providers [13]. Besides affecting consumer decisions, online reviews help firms to glean customer needs for new product development [64]. The value of helpful reviews is recognized in different industry sectors such as retailing, tourism, and dining.

To echo the prevalence of online reviews, numerous studies collect website data and perform exploratory analyses of the link between various review features and votes of helpfulness. We find that most of the review features studied such as length, readability, and reviewer credibility, are quality indicators of reviews. Besides, sentiment-related variables represent a popular category of review features, such as widely-studied positive and negative valence of review content. Such empirical investigations into the effects of quality indicators and sentiment, however, usually lack theoretical perspectives for categorizing review features and rationalizing their effects. Thus, some studies adopt elaboration likelihood model (ELM) to construct a dichotomous classification for review features (e.g., [1,34,54]).

According to ELM [57], when people receive messages that are intended to be persuasive, they use two routes to process messages. The

* Corresponding author.
E-mail addresses: yenchun@nccu.edu.tw (Y.-C. Chou), chuang@nccu.edu.tw (H.H.-C. Chuang), liang@mis.nsysu.edu.tw (T.-P. Liang).
[1] Deceased May 2021. In memory of TP, a great mentor and wonderful human being.

central route is analogous to rational thinking[2] by which people pay attention to content of the message and evaluate *quality of the arguments*, i.e., plausibility of persuasive argumentation [4]. In contrast, the peripheral route is analogous to intuitive judgment by which people use heuristic rules to process information cues from the message. That is, the strength of persuasive arguments and information cues of a review determine one's attitude toward a message, i.e., perceived helpfulness [7]. Review features such as *length* and *readability* related to argument quality are processed by *central route* for rational thinking. In contrast, review features not related to argument quality such as *review sentiment* are heuristic rules processed by the *peripheral route* for intuitive judgements. Fig. 1(a) illustrates the prevalent ELM application in existing studies using observational data, in which quality indicators serve as proxies to argument quality and take effects through the central route. Also, it is commonly accepted that features related to sentiment affect review helpfulness via the peripheral route.

Despite that ELM is a conceptually appropriate theory for review helpfulness, we observe two common pitfalls of applying ELM. First, according to the ELM theory, both quality indicators and sentiment factors are heuristic cues processed by the peripheral route for intuitive judgements. Other than heuristic cues, Petty and Cacioppo [57] further state that quality indicators, such as the *number of arguments* in a message, can enhance issue-relevant thinking when a person is highly involved in an issue. That is, quality indicators could take effects through the central route in high-involvement cases, but also serve as intuition-provoking cues for the peripheral route. Hence, classifying quality indicators exclusively as central cues in Fig. 1(a) is not theoretically appropriate. Second, the central route is analogous to rational thinking and hence is mainly driven by argument quality (i.e., the plausibility of persuasive messages that affects the audience's subjective perception), rather than review features related to quality. While high argument quality stands for messages being perceived as strong and cogent [55] and thus contributes to review helpfulness, argument quality is typically unobserved for archival data studies. Despite their correlations with argument quality, quality indicators such as length are heuristic cues and not equivalent to argument quality. An analogy is that an intelligence quotient is an indicator related to but non-identical to unseen individual capability. As a result, unobserved argument quality would cause the omitted-variable bias (endogeneity) that invalidates regression estimates on the effects of review feature.

In response to the two outstanding issues, we propose a corrected ELM application for review helpfulness in Fig. 1(b), where we correct the misperception that quality indicators merely drive the central route, and posit that researchers should explicitly address unobserved argument quality. We try to answer the first research question: *how do review features take effects on review helpfulness with the consideration of unobserved argument quality?* Based on the conceptualization in Fig. 1(b), we carefully review the literature and develop an integrated model on the review feature–helpfulness link. Moreover, for emotion-oriented sentiment, we employ the circumplex model [58] to go beyond positive and negative emotions in the peripheral route. The overarching theoretical perspective – a circumplex of valence and arousal – leads us to the second research question: *how do different combinations of emotional polarities and activation levels take effects on review helpfulness?*

After articulating the theoretical model and deriving research hypotheses, we empirically test the model using datasets from three different e-commerce platforms – Amazon.com [65], Drugs.com [76], and Yelp.com [77]. The use of publically available data from multiple sites enhances reproducibility of our work and helps triangulate the effects of review features. Moreover, our empirical analysis explicitly tackles the omitted-variable bias/endogeneity [68] incurred by argument

quality. This is fairly challenging because multiple quality indicators are subject to endogeneity. Instrumental variable (IV) is a common remedy, but the adequacy/validity of an IV is often debated [11]. We illustrate that an instrument-free approach [52] is a promising technique for researchers to tackle the bias in online review studies. Indeed, addressing endogeneity alters the estimated effects of quality indicators. Also, we find that well-known effects of negative valence are contingent on the under-studied arousal. Taken together, we contribute to the literature by elucidating ELM for review helpfulness on the conceptual front and endogeneizing unobserved argument quality on the empirical front.

2. Literature review

Research in online customer reviews started with the online word of mouth (eWOM). As online reviews are the most influential representations of eWOM [8,35], studies on eWOM examined the relationship between product reviews and sales. For example, the effect of negative review ratings on sales is found to be more influential than positive ratings (e.g., [12]). Alternatively, the volume of reviews (e.g., [33]) and the disclosure of reviewer identify (e.g., [20]) are found to have a positive relationship with product sales. Liang et al. [37] categorized review comments on mobile apps into products and services, both of which were found to have significant effects on app sales ranking.

An implicit assumption underlying these studies is that product reviews can help consumers to infer product quality and reduce uncertainties in purchase decision-making processes [24]. A consensus in the literature is that the number of helpful reviews tends to have a significant positive association with customers' buying decisions [18] and firms' sales performance [33]. Following the premise, follow-up research explored the issue of what makes online reviews helpful? That is, understanding *antecedents of review helpfulness*. The first two antecedents – *length* and *rating* – are descriptive, where the former is perceived to be a quality indicator and the latter is perceived to be an emotion indicator. For example, Hong et al. [24] suggest that length (number of words) is reflective of review depth. Rating, on the other hand, is associated with review extremity [66]. Due to their simplistic and nature, the two variables are widely studied (e.g., [17,51]). Some studies (e.g., [9,45,50]) further examine the moderation effects of product types on the relationships between length/rating and review helpfulness.

The next two descriptive antecedents are *duration* since a review posted (review age) and *reviewer information* (e.g., ranking, cumulative helpful votes, identity). The former is obvious as an older review has longer duration to accumulate helpful votes due to its timeliness [66]. The latter is related to source credibility and could affect readers' perceived helpfulness [23,74]. Huang et al. [25] find that a reviewer's place of origin (as a social connection with readers) and the duration since a review posted are significant antecedents of review helpfulness. Karimi et al. [28] show that even reviewer profile image has effects on helpfulness analyzing a sample from Google Play.

Unlike *length* that is merely a count of words, the nature of review content, such as *readability* and *valence,* are also included with the aid of text mining techniques. *Readability* affects people's comprehension and hence their perceptions toward the text. Some studies (e.g., [23,30]) apply various readability indices that include the number of characters and words to quantify the difficulty of text comprehension, and assess the effects of readability on helpfulness. As for *valence*, some studies (e.g., [38,50,51]) control for valence using *rating*.

Furthermore, text mining techniques enable researchers to extract *positive* and *negative sentiment* based on word identification and go beyond rating numbers. For instance, Baek et al. [2] apply the sentiment analysis to extract the percentage of negative words from a review and examine the influence on review helpfulness. Chen et al. [6] investigate how the mixture of positive and negative attitudes of a review affects the proportion of helpful votes to total votes.

The aforementioned factors – length, rating, duration, reviewer information, readability, and valence – are essentially content- and re

---

[2] We refer rational thinking specifically to a person's scrutiny of the messages presented. Judgements by a simple cue like credible reviewers, while reasonable, are dependent on heuristic rules and denoted as intuitive responses.
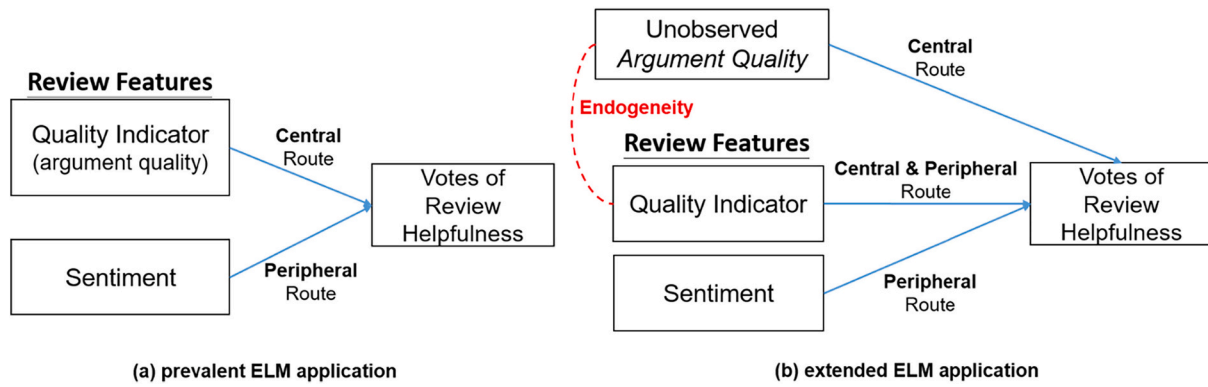
Fig. 1. Prevalent versus extended ELM application.

viewer-related signals that have effects on review helpfulness [61] and the effects may vary with product/platform types [8,9]. Note that such a list of six antecedents is not exhaustive and mainly based on econometric analyses of archival data from websites. Fewer psychometric studies (e.g., [7,18]) use survey data to assess nuanced antecedents/moderators of review helpfulness. That said, the rapidly increasing number of econometric and psychometric studies on review helpfulness in the past 5–10 years results in a large body of literature and leads to the appearance of meta-analytical studies on effects of the foregoing antecedences [24,66].

The above discussion illustrates the popularity and the importance of the research stream. Among those studies, a substantial proportion of research models are rather data-driven without a solid theory to interpret their findings, particularly those conducting archival data analyses of online review helpfulness. Among the comparatively small proportion of theory-based studies, the ELM has been the predominant perspective for review helpfulness (e.g., [1,2,7,10,44,54,72–74]). ELM is a theory of informational influence and provides a framework explaining how messages are processed for effective communication [57].

While ELM seems to be an appropriate theory for studies on antecedents of review helpfulness, as discussed in the introduction, we identify the two common issues of the application of ELM: *unobserved argument quality* and *quality indicators' effects through both central and peripheral routes*. We posit that addressing argument quality in accordance with the assumptions imposed by ELM should be noticed by many e-commerce researchers who still actively conduct empirical analyses of review helpfulness (e.g., [1,36]). Our study aims to be a prototypical example of addressing the two non-trivial issues for empirical analyses of review helpfulness. In the next section, we develop a research model grounded on ELM and the circumplex model [58].

3. Research model and hypotheses

3.1. Theoretical foundation and research model

According to ELM, when people receive messages that are intended to be persuasive, they look into messages in two ways. On one hand, quality of arguments is processed by the central route for rational thinking. On the other hand, informational cues from the message are processed by the peripheral route for intuitive judgment. On top of the initial dichotomous classification, ELM further suggests that informational cues related to quality such as number of arguments can trigger issue-relevant thinking when a person is highly involved in an issue. For example, in one experiment, undergraduate students were given messages regarding their senior comprehensive exams (high involvement), Petty and Cacioppo [56] found that increasing the number of strong arguments enhances persuasion, but increasing the number of weak arguments reduces persuasion. That is, those quality-related cues can take effects through the central route as well as the peripheral route.

In the context of online review, we note that all review features are *informational cues* observed from messages. Different from experiments where argument quality can be manipulated, the strength of persuasive argumentation of an online review and how it affects perceived helpfulness is unobserved by researchers. Nevertheless, as explained earlier, argument quality is correlated with quality indicators (heuristic cues) such as length, as they both involve in issue-relevant thinking. As a result, when measuring effects of review features using observational data, we have to consider unobserved argument quality and the associated endogeneity. Fig. 2 is our research model that is grounded on the extended ELM specification in Fig. 1(b). Below we articulate features included in the model, and will discuss how to handle the endogeneity issue in the method section.

Based on ELM, we specify unobserved argument quality (plausibility of review) processed by the central route, and review features as informational cues processed by both central and peripheral routes. The proposed model integrates review features summarized the literature review. First, we specify *review length*, *readability*, and *reviewers' ranking* as quality indicators. Relatively few studies include the number of semantic topics as quality indicator because semantic identification traditionally involves a lot of human interventions (e.g., manual categorization in [44]). While time-consuming, Cao et al. [5] report that the inclusion of semantic topics can add new insights in analyzing quality-related cues of review content. Combining computer scripts and human judgements, Chen et al. [6] predefine a list of product topics and examine the breadth of a review measured by the number of covered topics. Instead of human encoding, we leverage topic modeling techniques to computationally assess review breadth [41,42].

For sentiment on the peripheral route, many prior studies use simple review rating or positive and negative valence to capture sentiment. We
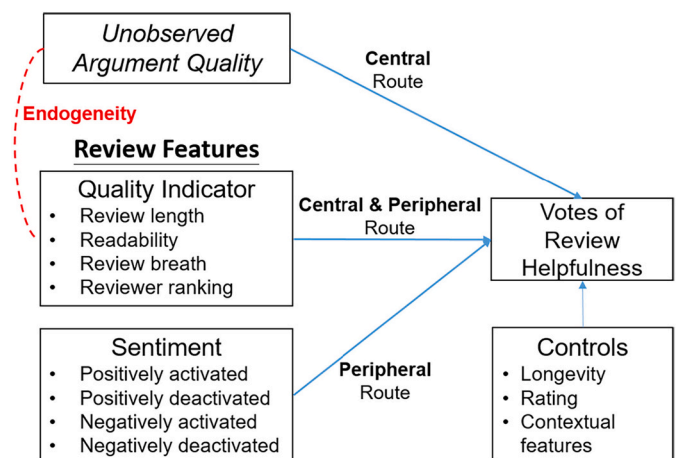

Fig. 2. Research model.

use the circumplex model [58] to characterize review sentiment not only by *valence* but also through *arousal*, i.e., the extent to which an actor is activated/deactivated. We note that each emotion should be represented via a point on the circumplex composed of valence polarity and activation level. As shown in Fig. 3, sentiment of a review can be categorized into four quadrants on the circumplex – positively activated (upper right), negatively activated (upper left), negatively deactivated (lower left), and positively deactivated (lower right). The circumplex perspective implies that differences in levels should be taken into account when assessing emotion intensity reflected in review text. Take the first quadrant (positively activated) for instance, while "DELIGHTED" and "SATISFIED" have valence almost at the same level, the former has a significantly higher level of arousal (activated) and deviates a lot more from neutral emotions. Similar cases can be observed in other quadrants that represent different emotions. In response to the call for a better understanding of the sentiment-review helpfulness link [16], our model goes beyond one-dimensional polarity and explicitly differentiates emotions constituted by valence and arousal.

Aside from afore-mentioned information cues related to quality and sentiment from review content, we also include longevity of a review (i. e., the time duration since a review was posted) to control for unfair comparisons among reviews posted early and those posted later, since reviews that exist longer have longer periods to elicit helpful votes. Our model also includes product rating and other contextual features to control for product/service heterogeneity that may affect popularity of a review and its likelihood to receive helpful votes [72,74].

3.2. Hypotheses development

According to the two blocks of review features in our research model (Fig. 2), we develop two sets of hypotheses: H1a-H1d related to quality indicator and H2a-H2c related to review sentiment.

Hypothesis 1. Quality Indicator.

A message with better quality has been shown to be more credible and receptive by viewers [7,63]. In our conceptual model, review features related to unobserved argument quality are called quality indicators, including length, breadth, readability, and reviewer ranking. When people are less involved with the topic, they tend to use intuitive rules, i.e., the peripheral route, to process those quality indicators. For example, longer reviews or a review with more topics covered are easily perceived to be richer in depth and information content compared to shorter reviews [6,45]. Similarly, people perceive easy-to-read reviews to be efficient for information-seeking, and reviews from higher ranking

reviewers to be more credible. [8,30]. In the low involvement case, people use the foregoing heuristics to link quality indicators with argument quality. The intuitive rules of content richness, information-processing easiness, and source credibility in turn are likely to contribute to perceived helpfulness. On the contrary, when people are highly involved or motivated with the context, they will scrutinize all possible cues that can help them to assess review quality. In general, quality indicators are positively correlated with argument quality. Therefore, quality indicators can serve as persuasive arguments processed by the central route, leading to review helpfulness. Overall, quality indicators can serve as persuasive arguments (central route) or intuitive rules (peripheral route) for review processing. We thus posit that quality indicators are proxies for review quality, and in general positively associated with perceived helpfulness of reviews.

H1a. Review length is positively associated with review helpfulness.

H1b. Review readability is positively associated with review helpfulness.

H1c. Reviewer ranking is negatively associated with review helpfulness.

H1d. Review breath is positively associated with review helpfulness.

Hypothesis 2. Review Sentiment.

Prior studies on psychology have suggested various dimensions to classify emotions. Among these dimensions, researchers have consistently sorted emotions into two basic dimensions: valence and arousal [48,58]. Valence describes the extent to which an experience is pleasant (positive valence) or unpleasant (negative valence), while arousal describes the extent to which an actor is activated or deactivated [48]. For example, anger, anxiety, and sadness are all negative emotions. However, while anger and anxiety are characterized by states of heightened arousal or activation, sadness is characterized by low arousal/activation [3].

When discussing emotional expressions in an online product review, prior studies mainly focus on valence, and arousal is under-studied [71]. As discussed in section 2, we follow the circumplex model to assess emotions located at positively activated, positively deactivated, negatively activated, and negatively deactivated quadrants. Emotions scattered at the circumplex represent different levels of valence and arousal (see Fig. 3). Emotional words are generally processed faster and more efficiently than non-emotional words [27,31], and processing can even occur automatically [21,22]. Through the lens of ELM, sentiment-related features are processed via the peripheral route for intuitive judgements.

Negativity bias suggests that negative information is generally considered more diagnostic than positive information. Thus, people are
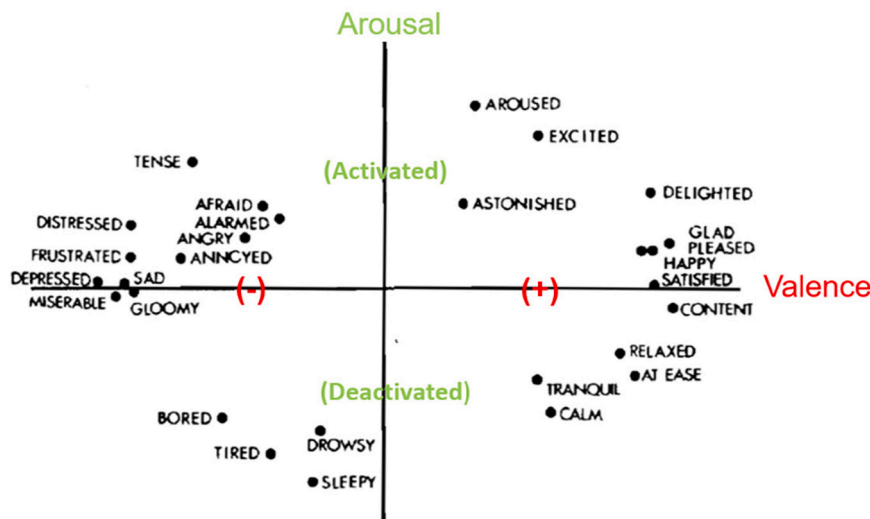

Fig. 3. Circumplex model of valence and arousal (adapted from [58]).

inclined to believe those who express negative feelings [26,62]. Yin et al. [70] confirm that reviews with stronger negative emotions are considered more helpful. We take one step further to distinguish negatively deactivated from negatively activated emotions. We argue that negativity bias holds only for negatively deactivated expressions in a review. In contrast, overmuch negative as well as high levels of emotion activation appear to be strong and subjective expressions. Hence, we posit that negatively activated expressions in a review are perceived as less helpful than negatively deactivated expressions that are critical but not agitated. Similarly, while prior researchers have found that positive sentiment has positive impacts on review helpfulness [16], we further differentiate positively deactivated from positively activated emotions. Given that high arousal shows irrationality [71], we argue negative influences from high arousal cancel out positive effects from positive emotions. Consequently, the positive relationship between positive emotions with review helpfulness holds only for positively deactivated expressions. We thus form the following hypotheses on three emotion quadrants with significant effects on review helpfulness:

H2a. Positively deactivated emotions in a review are positively associated with review helpfulness.

H2b. Negatively activated emotions in a review are negatively associated with review helpfulness.

H2c. Negatively deactivated emotions in a review are positively associated with review helpfulness.

4. Data and variables

4.1. Data processing and summary

We use review datasets from three different e-commerce platforms – Amazon.com (Want et al. 2013), Drugs.com (Gräßer et al. 2018), and Yelp.com (Yelp 2020) – to evaluate our research model. The first dataset is published by Wang et al. [65] who scrape tablets' review from Amazon.com. The dataset includes product features and customer reviews on tablets for 24 weeks since February 2012. After removing missing values and abnormal entries such as helpful votes greater than total votes, we end up with 40,485 reviews customer reviews. The second dataset is published by Gräßer et al. (2018) who scrap data from Drugs.com, a professional website on drugs with reviews shared by patients. After removing abnormal entries, we obtain 23,459 reviews posted in 2017. The third dataset is a sample of production data published by Yelp.com in 2020. We focus on 152,751 customer reviews posted in 2019 for food and restaurants, the most prominent category in in Yelp.

For the three datasets, the key dependent variable is the number of helpful votes received by a review, which is a direct indicator of perceived helpfulness of the review. We also obtain descriptive features of a review, including review length, rating associated with a review, and duration of a review since posting. For the Amazon.com tablet reviews, we have further access to reviewers' ranking and brands of tablets. For the Drug.com drug reviews, we have extra information on what specific kind of medical condition a posted review is about. For the Yelp.com restaurant reviews, we have access to the number of reviewers' previous posts, overall rating of the focal business, and the number of reviews written for the restaurant.

We further perform text mining and sentiment analysis of unstructured text content through the following protocol. We first perform text preprocessing: (1) transforming all text into lowercase (2) removing stop words (e.g., the, and, of), punctuation marks, numbers, and spaces (3) stemming words (e.g., values, valued and valuing are all replaced with value) [41]. The cleaned data is then used to calculate the readability index, and to examine textual content via sentiment analysis based on the circumflex of emotions and NMF topic modeling (to be elaborated in the next section). After the text mining process, we match the text data with descriptive features of reviews mentioned above.

Table 1 illustrates the list of variables used in the study. *Helpfulness* is the dependent variable measured by the number of helpful votes received by a review. Quality indicators and review sentiment – our main focuses of the study – is discussed in Sections 4.2 and 4.3 separately. We further include longevity and rating as the control variables. *Longevity* measures the number of days from the posting date of a review to the date reviews are scrapped. *ReviewRating* is the numerical rating assigned by each reviewer to the product in a review. Note that a 5-point scale is used by Amazon.com and Yelp.com, and a 10-point scale instead is applied by Drugs.com.

For the Amazon.com dataset, we are able to collect tablet brands (*Brand*), e.g. Apple and ASUS, as additional controls. Similarly, for the Drug.com dataset, we derive 19 dummy variables (*Condition*) for 20 different medical conditions (e.g. birth control, insomnia) revealed by patients in their reviews. Finally, for the Yelpc.om dataset, the rating of a restaurant (*BizRating*) and the number of reviews a restaurant received at the time of data collection (*BizReviewCounts*) are included as extra control variables. As the average rating of a restaurant is increased by 0.5 from 1 to 5, we model 9 levels of *BizRating* using 8 dummy variables. The summary statistics of key variables are reported in Tables 2.

4.2. Operationalization of quality indicators

Quality indicators are composed of length, readability, breadth, and reviewer ranking. *Length* is measured by the number of words of a

Table 1
Variable definition.

Types of Variable	Variable	Description
Dependent Variable	Helpfulness	The number of helpful votes of a review
Quality Indicators	Length	The number of words of a review
	Breadth	The number of topics of a review obtained from NMF topic modeling
	Readability	Readability of a review measured by Flesch Reading Ease Index
	ReviewerRanking	Reviewer's ranking from Amazon.com
	ReviewerPosts	The number of reviews a user had posted at the time of data collection by Yelp.com
Review Sentiment	PositivelyActivated	The intensity of emotions measured by scores of positive and high arousal words in a review
	PositivelyDeactivated	The intensity of emotions measured by scores of positive and low arousal words
	NegativelyActivated	The intensity of emotions measured by scores of negative and high arousal words
	NegativelyDeactivated	The intensity of emotions measured by scores of negative and low arousal words
Control Variables	Longevity	The number of days since a review posted
	ReviewRating	The rating assigned to the product in a review, from 1 to 5 (modeled as 4 dummy variables) for Amazon.com and Yelp.com. As for Drugs.com, it follows a 10-point scale (modeled as 9 dummy variables).
	Brand	For Amazon.com dataset, twenty tablet brands (e.g., Apple, Acer) are modeled as 19 dummy variables.
	Condition	For Drugs.com dataset, one of twenty possible patients' conditions (e.g., birth control, depression) for a review is modeled as 19 dummy variables.
	BizRating	For Yelp.com dataset, the average rating of a restaurant increases by 0.5 from 1 to 5, and is modeled as 8 dummy variables.
	BizReviewCounts	For Yelp.com dataset, the number of reviews a restaurant acquired at the time of data collection.

review. *Readability* is calculated via the Flesch Reading Ease Index [19]. The higher the index is, the better the readability of a review. The equation of the index is as follows:

$$Reading\ Ease\ (Readability) = 206.835 - 1.015 \left(\frac{total\ words}{total\ sentences} \right)$$
$$- 84.6 \left(\frac{total\ syllables}{total\ words} \right)$$

Breadth is defined as the number of topics a review covers. Instead of counting on human coders to manually process reviews in order identify common topics covered by reviews, we applied the non-negative matrix factorization (NMF) methodology [32] that is similar to the oft-used principal componential analysis and computationally easy to implement. The core idea is to decompose a document-term matrix A (m words by n reviews) into the product of two non-negative matrices K and W. The former is a m by k matrix that reveals words associated with topic, whereas the latter is a k by n weight matrix that informs us the weights of each review for the k topics. Both K and W matrices are subject to non-negativity constraints that have been shown to result in more interpretable topics for various types of data [42]. Through NMF topic modeling, we were able to identify common semantic topics covered by our reviews.

We performed NMF analysis using scikit-learn in Python [53] and identified representative topics from the three datasets respectively. Due to the unsupervised nature of NMF, we tested different numbers of topics and examined the resulting coherence scores. The intuition behind coherence is to assess the interpretability of each topic based on co-occurrence of its keywords. We adopted the UCI [47] and UMass [43] measures for coherence scores and larger scores have been found to be more interpretable. We stayed with four topics for tablet reviews of Amazon.com, five topics for drug reviews of Drugs.com, and seven topics for restaurant reviews of Yelp.com based on coherence scores and the fact that these topics emerged from NMF seemed more interpretable. For instance, Fig. 4 illustrates the four topics and top 10 words in each topic derived from running the NMF procedure on tablet reviews of Amazon.com.

Given the extracted k by n weight matrix W, we attempted to assess the total number of topics covered by each review. Each entry w_{ij} in W stands for the weight (loading) of a topic i for review j. We then computed a normalize weight $\widetilde{w_{ij}} = w_{ij} / \sum_i w_{ij}$ for each review j such that weights would be more comparable across reviews. Similar to Mankad et al. [41], a topic i is considered to be covered in a review j if $\widetilde{w_{ij}}$ is greater than the median of all $\widetilde{w_{ij}}$ for the given i. Accordingly, *breadth* of a review is defined as the total number of topics whose $\widetilde{w_{ij}}$ satisfying the foregoing requirements.

$$Breadth_j = \sum_{i=1}^{k} Ind\left(\widetilde{w_{ij}} > median\left(\widetilde{w_{i1}}, \widetilde{w_{i2}}, ..., \widetilde{w_{in}} \right) \right)$$

where *Ind* is an indicator function. The last quality indicator – *ReviewerRanking* – is a measure to reflect a reviewer's credibility. Amazon.com provides such ranking to reflect the number of posts and helpful votes received of a reviewer. Instead, Yelp.com shows the number of posts a reviewer has written (*ReviewerPosts*). *ReviewerRanking* in Amazon.com is relatively holistic, as *ReviewerPosts* only measures quantity without considering helpfulness perceived by readers. That said, the metric reflects how prolific/active a reviewer is and thus should be correlated to unseen argument quality.

4.3. Operationalization of sentiment features

For emotional expressions, we used the sentiment analysis to extract positive and negative valences of words in a review, as well as the level of emotion activation (arousal) from review content. We applied the dictionary by Warriner et al. [67] who defined 13,915 words with scores from 1 to 9 to indicate levels of valence and arousal. Traditionally, researchers mainly focus on valence: words with valence scores greater than 5 into the category of positive valence, and those with scores smaller than 5 into the category of negative valence. A review's valence was measured as the number of positive words and the number of negative words.

The widely-adopted practice in prior studies, however, consider only counts of words without recognizing that each word has its

Table 2
Summary statistics of online reviews.

Variables	Amazon.com (N = 404,485)				Drugs.com (N = 23,459)			
	Mean	Stdev	Min	Max	Mean	Stdev	Min	Max
Helpfulness	4.91	24.29	0.00	950.00	6	14.18	0.00	328.00
Length	88.15	114.40	4.00	2925.00	52.82	25.14	1.00	279.00
Breadth	2.00	0.66	1.00	4.00	2.50	0.80	0.00	5.00
Readability	78.97	13.90	−375.19	116.91	79.98	13.88	−132.58	121.47
PositivelyActivated	13.61	16.33	0.00	447.47	4.71	4.60	0.00	36.35
PositivelyDeactivated	118.06	153.38	0.00	4003.49	46.15	24.47	0.00	294.54
NegativelyActivated	4.41	6.93	0.00	251.76	6.66	6.78	0.00	67.45
NegatvielyDeactivated	15.03	21.80	0.00	479.90	16.27	10.32	0.00	81.03
ReviewerRanking	5,191,811.92	5,285,042.81	3.00	15,668,643	–	–	–	–
Longevity	226.80	163.15	1.00	3173.00	191.46	98.22	0.00	345.00
ReviewRating	3.80	1.45	1.00	5.00	5.99	3.64	1.00	10.00

| Variables	Yelp.com (N = 152,751)			
	Mean	Stdev	Min	Max
Helpfulness	1.18	3.95	0.00	321.00
Length	57.03	49.59	1.00	558.00
Breadth	3.53	1.08	0.00	7.00
Readability	78.34	14.90	−2883.09	121.22
PositivelyActivated	7.75	8.36	0.00	130.96
PositivelyDeactivated	75.89	65.29	0.00	804.11
NegativelyActivated	1.93	3.30	0.00	89.68
NegatvielyDeactivated	8.17	9.34	0.00	141.74
ReviewerPosts	131.64	537.09	1.00	14,691.00
Longevity	186.63	103.07	0.00	364.00
ReviewRating	3.76	1.47	1.00	5.00
BizRating	3.84	0.65	1.00	5.00
BizReviewCounts	403.17	639.57	5.00	6221.00
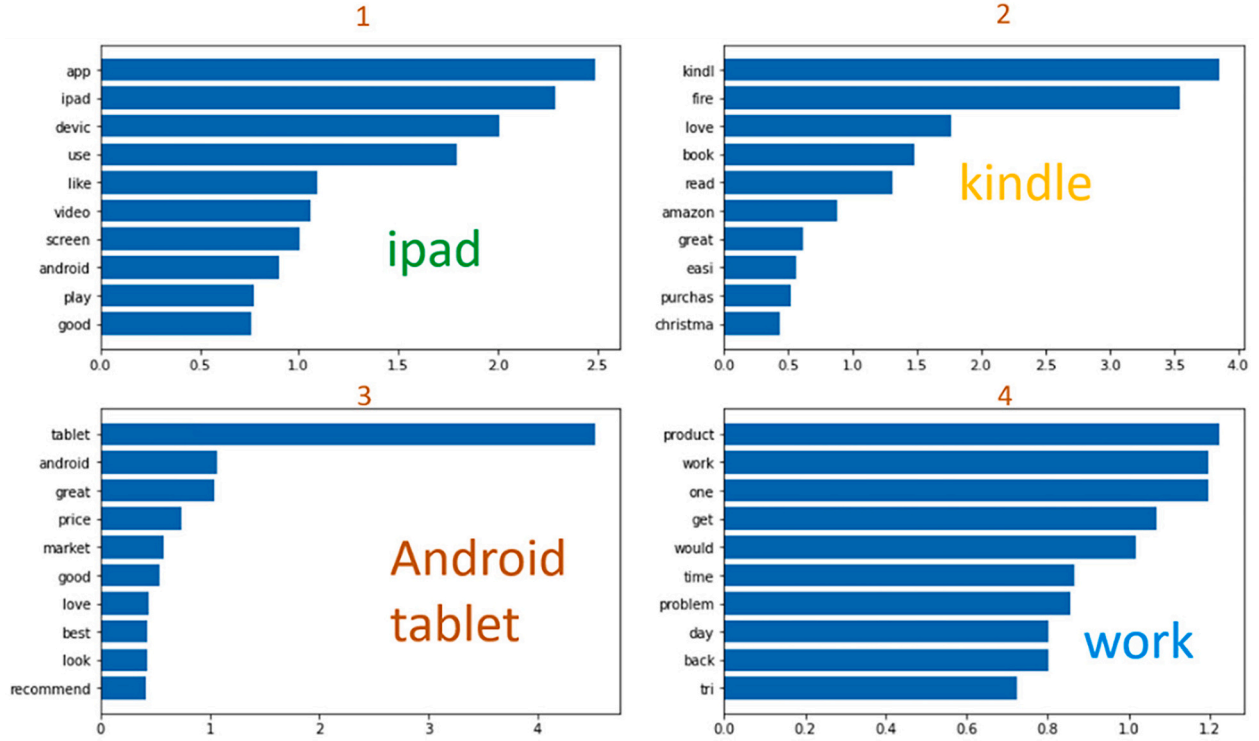
Fig. 4. Top 10 words in each topic identified by NMF (tablets on Amazon.com).

emotion level. To better capture the intensity of emotion, we followed the circumplex model to define emotions via a circumplex with different levels of valence and arousal (see Fig. 3). For each emotional word in a review, we used scores by Warriner et al. [67] to reflect different levels of sentiment. Since the score scale is from 1 to 9 on the horizontal axis (valence) and on the vertical axis (arousal), the circumplex has a center point (5, 5) being the neutral origin of four emotion quadrants. Therefore, we define the score of each emotional word as the Manhattan distance from the neutral origin. The Manhattan distance (L1 norm) is preferred over the Euclidean distance (L2 norm) in our setting because L2 norm over-weighs outlying words deviating relatively far away from the origin. We then aggregate scores of words within each quadrant of the circumplex to form four variables *PostivelyActivated*, *PostivelyDeactivated*, *NegativelyAactivated*, *Negatively Deactivated* that capture emotional intensity in each review. Specifically, the four variables of a review *i* is computed in the following way, where *j* is the index of emotion words in a review.

$$PositivelyActivated_i = \sum_{j \in 1stQudrant} \left(Valence\,of\,Word_{ij} - 5\right) + \left(Arousal\,of\,Word_{ij} - 5\right)$$

$$NegativelyActivated_i = \sum_{j \in 2ndQudrant} \left(5 - Valence\,of\,Word_{ij}\right) + \left(Arousal\,of\,Word_{ij} - 5\right)$$

$$Negatively\,Deactivated_i = \sum_{j \in 3rd\,Qudrant} \left(5 - Valence\,of\,Word_{ij}\right) + \left(5 - Arousal\,of\,Word_{ij}\right)$$

$$Positively\,Deactivated_i = \sum_{j \in 4th\,Qudrant} \left(Valence\,of\,Word_{ij} - 5\right) + \left(5 - Arousal\,of\,Word_{ij}\right)$$

5. Empirical analysis and findings

5.1. Model specification

To address endogeneity due to unobserved argument quality and its correlation with quality indicators, we adopt an instrument-free method proposed by Park and Gupta [52]. Considering the following regression model with one endogenous regressor:

$$Y = X\beta + P\alpha + \varepsilon$$

where *X* is an exogenous variable, *P* is an endogenous regressor, and *ε* is the structural error term. The non-zero correlation between *P* and *ε* leads to biased estimates from ordinary least squares (OLS) regression that demands exogeneity of predictor variables (cov(X, ε) = 0). Nonetheless, one can obtain consistent estimates of model parameters if the correlation of regressors and error term can be specified. Park and Gupta [52] use the copula model to construct a flexible joint distribution from the marginal distributions of *P* and *ε*, while allowing for a wide range of correlations between the two variables. In most regression settings, the structural error term is assumed to be a normal distribution. The distribution of *P* can be empirically estimated from the observational data. Given the two marginal distributions, the copula model produces the joint distribution $f(P, \varepsilon)$ that properly capture the correlation between the endogenous variables and the error term. Because the method does not require any valid instrument variables (IV) that need to be exogenous (cov(IV, ε) = 0) and non-weak (cov(IV, P) $\neq$ 0), it has gained increasing tractions recently in marketing (e.g., [14]) and information systems (e.g., [29]).

This method enables us to obtain consistent estimates for hypotheses testing by adding additional regressors, which are the inverse normal of the marginal distribution of the endogenous variables. In our study, the endogenous variables are the four quality indicators that are arguably correlated with unobserved argument quality. Specifically, we construct four additional regressors as follows:

$$log(Length_i)^E = \varnothing^{-1}(H(log(Length_i)))$$

$$Breadth_i{}^E = \varnothing^{-1}(H(Breadth_i))$$

$$Readability_i{}^E = \varnothing^{-1}(H(Readability_i))$$

$$log(Ranking_i)^E = \varnothing^{-1}(H(log(Ranking_i)))$$

where $H(.)$ is the empirical cumulative distribution function, and $\varnothing^{-1}(.)$ is an inverse normal cumulative distribution function. The superscript E denotes the error-correction nature of the four variables estimated from copula joint distributions. The method is advantageous in its flexibility to simultaneously accommodate multiple endogenous variables, which is a methodologically daunting task [39]. Below is the regression model for our first sample of tablet reviews.

$$log(Helpfulness_i) =$$

$\alpha + \beta_1 log(Length_i) + \beta_2 Breadth_i + \beta_3 Readability_i + \beta_4 log(ReviewerRanking_i) + \gamma_1 PositivelyActivated_i + \gamma_2 PostivelyDeactivated_i + \gamma_3 NegativelyActivated_i + \gamma_4 NegativelyDeactivated_i + \tau_1 log(Length_i)^E + \tau_2 Breadth_i{}^E + \tau_3 Readability_i{}^E + \tau_4 log(ReviewerRanking_i)^E + \delta_1 Longevity_i + \boldsymbol{\delta_{2\sim5}ReviewRating_i} + \boldsymbol{\delta_{6\sim24}Brand_i} + \varepsilon_i.$

where the index i stands for each review, and the dependent variable *Helpfulness* represents the number of helpful votes of a review. *Length* reflects the number of words in a review. *Breadth* shows the number of topics covered in a review using NMF topic modeling. *Readability* is the Flesh Reading Ease Index to indicate how easily a review can be understood by an average person. *ReviewerRanking* is the variable showing the reviewer's rank of a review from Amazon.com. We note that *Helpfulness*, *Length* and *Ranking* are highly right-skewed with large scale differences. Hence, we take the natural log transformation of the three variables to ease the estimation of the model. Parameters β_1 to β_4 are coefficients for the four quality indicators that are endogenous and correlated with the unobserved argument quality. We include the four correction terms with coefficients τ_1 to τ_4 to tackle endogeneity induced by unseen argument quality, such that we can obtain consistent β_1-β_4 estimates for testing H1. *PositivelyActivated, PositivelyDeactivated, NegativelyActivated,* and *NegativelyDeactivated* are the four variables from the circumplex model of affect. Parameters γ_1 to γ_4 are the coefficients for these four peripheral cues and their estimates are used for testing H2.

We also include control variables in our model. *Longevity* indicates the duration of a review from its posting date of a review to the date it is scrapped. We use *Longevity* to control the accumulation of helpful votes due to long periods. Other than review content, the star rating of a review is another important cue. We then use four dummy variables to reflect the rating from 1 to 5, and control for the influence of the star rating. As customers have preferences over product brands, we also include tablet brands as control variables. We model them as 19 dummy variables in the regression model. Parameters δ_1 to δ_{24} are coefficients for the control variables.

For our second sample of drug reviews, we specify the following regression model that differs from the first model in two minor ways. First, we include three rather than four endogeneity correction terms for *Length, Breadth,* and *Readability* because reviewer *Ranking* is unavailable in this website. Second, the star rating ranges from 1 to 10 as opposed to 1 to 5. So we introduce 9 dummy variables (δ_2 to δ_{10}) as control variables. Also, we include 19 dummy variables (δ_{11} to δ_{29}) to accommodate heterogeneities among 20 different patients' conditions when writing reviews.

$$log(Helpfulness_i) = \alpha + \beta_1 log(Length_i) + \beta_2 Breadth_i + \beta_3 Readability_i$$

$$+ \gamma_1 PositivelyActivated_i + \gamma_2 PostivelyDeactivated_i$$

$$+ \gamma_3 NegativelyActivated_i + \gamma_4 NegativelyDeactivated_i$$

$$+ \tau_1 log(Length_i)^E + \tau_2 Breadth_i{}^E + \tau_3 Readability_i{}^E +$$

$$+ \delta_1 Longevity_i + \boldsymbol{\delta_{2\sim10}ReviewRating_i} + \boldsymbol{\delta_{11\sim29}Condition_i} + \varepsilon_i$$

For our last sample of restaurant reviews, we specify a regression model with few minor tweaks. First, while reviewer *ReviewerRanking* is unavailable, we include *ReviewerPosts* with coefficient β_4 as an alternative quality indicator to capture the activeness of a reviewer. Second, we include 8 dummy variables (δ_6 to δ_{13}) for *BizRating* as well as *BizReviewCounts* (i.e., number of accumulated reviews for the restaurant in review i) to control for the effects of restaurant popularity on review helpfulness.

$$log(Helpfulness_i) = \alpha + \beta_1 log(Length_i) + \beta_2 Breadth_i + \beta_3 Readability_i +$$
$\beta_4 log(ReviewerPosts_i) + \gamma_1 PositivelyActivated_i + \gamma_2 PostivelyDeactivated_i + \gamma_3 NegativelyActivated_i + \gamma_4 NegativelyDeactivated_i + \tau_1 log(Length_i)^E +$

$\tau_2 Breadth_i{}^E + \tau_3 Readability_i{}^E + \tau_4 log(ReviewerPosts_i)^E + \delta_1 Longevity_i +$

$\boldsymbol{\delta_{2\sim5}ReviewRating_i} + \boldsymbol{\delta_{6\sim13}BizRating_i} + \delta_{14} log(BizReviewCounts_i) + \varepsilon_i$

5.2. Empirical findings

Table 3 shows the estimation results of tablet reviews (Amazon.com) using two modeling approaches. Model 1 (OLS) is the common estimation approach that does not formally address endogeneity, whereas Model 2 (OLS Copula) is the main model that corrects for endogeneity using the Park and Gupta [52] method, Accordingly, we shall discuss our empirical findings based on Model 2. One key assumption of the copula method is that endogenous variables should not be normally distributed [52], such that endogeneity-correction terms can be identified. We test the normality of the four quality indicators using Pearson chi-square normality test, and reject all null hypotheses (all with $p < 0.001$) showing support for the assumed non-normality. Note that ordinary parameter inference of the four generated regressors are incorrect, and thus a bootstrap method is applied to compute standard errors [52].

From Model 2, we can see that the additional correction terms are mostly significant (τ_1 to τ_4), indicating the need to tackle endogeneity. The estimated effects of review length ($\beta_1 = 0.3775$, $p < 0.001$), readability ($\beta_3 = 0.0024$, $p < 0.001$) and reviewer ranking ($\beta_4 = -0.0424$, $p < 0.001$) are as expected and consistent with the literature. The effect size is substantially different from their effects in Model 1 that ignores endogeneity. Also, one quality indicator, *Breadth*, is insignificant in both Models 1 and 2. Accordingly, the empirical findings support hypotheses H1a-H1c, and confirm the role of quality indicators on review helpfulness, after controlling for endogeneity.

As for the four sentiment-related features, Model 2 suggests that only negatively deactivated emotions ($\gamma_4 = 0.0023$, $p < 0.001$) are positively associated with review helpfulness, whereas negatively activated emotions exhibit negative effects ($\gamma_3 = -0.0059$, $p < 0.001$). In the domain of positive emotions, as hypothesized in H2, positively deactivated ($\gamma_2 = 0.0007$, $p < 0.001$) emotions are found to be positively associated with review helpfulness. To ensure that the results are not artifacts of how we operationalize the four sentiment features, we compute the four emotion intensity variables using the Euclidean distance instead of the Manhattan distance. The results remain qualitatively the same. The findings offer firm support for extending review sentiment into four quadrants composed of valence and arousal.

Table 4 shows the estimation results of drug reviews (Drugs.com) using two modeling approaches. The significant correction term for *Length* in Model 4 ($p < 0.01$) consistently suggests that endogeneity ought to be addressed. For the three quality indicators, only *Length* shows strong and significant effects ($\beta_1 = 0.1902$, $p < 0.001$), whilst estimates of *Breath* and *Readability* are not statistically significant. The estimation renders partial support for H1. As for sentiment features, positively deactivated emotions ($\gamma_2 = 0.0025$, $p < 0.001$) consistently show positive associations with helpfulness, in line with H2. That said, negative emotions in low and high arousal exhibit different effects from

Table 3
Estimation results of tablet reviews from Amazon.com ($n = 40{,}485$).

Variables	Model 1 OLS	Model 2 OLS Copula	Variables	Model 1 OLS	Model 2 OLS Copula
β_1:**log(Length)**	0.2056***	0.3775***	γ_1:**PositivelyActivated**	2.885e-04	3.132e-04
	(0.0071)	(0.0416)		(4.710e-04)	(5.362e-04)
β_2:**Breadth**	0.0115	0.0199	γ_2:**PositivelyDeactivated**	7.445e-04***	7.354e-04***
	(0.0062)	(0.0134)		(8.131e-05)	(1.117e-04)
β_3:**Readability**	0.0014***	0.0024***	γ_3:**NegativelyActivated**	−0.0059***	−0.0059***
	(2.815e-04)	(4.906e-04)		(8.310e-04)	(1.026e-03)
β_4:**log(ReviewerRanking)**	−0.0624***	−0.0424***	γ_4:**NegativelyDeactivated**	0.0023***	0.0022***
	(0.0021)	(0.0067)		(4.390e-04)	(6.187e-04)
τ_1:**log(Length)^E**	–	−0.1688***	δ_1:**Longevity**	0.0013***	0.0013***
		(0.0368)		(2.481e-05)	(3.484e-05)
τ_2:**Breadth^E**	–	−0.0061	δ_2:**ReviewRating = 2**	−0.1783***	−0.1830***
		(0.0087)		(0.0170)	(0.0185)
τ_3:**Readability^E**	–	−0.0177*	δ_3:**ReviewRating = 3**	−0.2790***	−0.2849***
		(0.0074)		(0.0160)	(0.0176)
τ_4:**log(ReviewerRanking)^E**	–	−0.0442***	δ_4:**ReviewRating = 4**	−0.3526***	−0.3632***
		(0.0128)		(0.0137)	(0.0155)
$\delta_6\sim\delta_{24}$: **Brand**	Included	Included	δ_5:**ReviewRating = 5**	−0.1570***	−0.1685***
				(0.0126)	(0.0144)
Adj R^2	OLS: 0.2647		OLS Copula: 0.2562		

Standard errors are reported in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 4
Estimation results of drug reviews from Drugs.com ($n = 23{,}459$).

Variables	Model 3 OLS	Model 4 OLS Copula	Variables	Model 3 OLS	Model 4 OLS Copula
β_1:**log(Length)**	0.1630***	0.1902***	γ_1:**PositivelyActivated**	0.0132***	0.0141***
	(0.0165)	(0.0207)		(0.0014)	(0.0014)
β_2:**Breadth**	−0.0102	−0.0108	γ_2:**PositivelyDeActivated**	0.0017***	0.0025***
	(0.0072)	(0.0188)		(4.090e-04)	(5.394e-04)
β_3:**Readability**	7.152e-04	0.0017	γ_3:**NegativelyActivated**	0.0070***	0.0077***
	(4.078e-04)	(0.0011)		(9.153e-04)	(9.875e-04)
τ_1:**log(Length)^E**	–	−0.0508**	γ_4:**NegativelyDeactivated**	−0.0038***	−0.0030***
		(0.0180)		(7.327e-04)	(7.728e-04)
τ_2:**Breadth^E**	–	0.0003	δ_5:**ReviewRating = 5**	−0.1014***	−0.1012***
		(0.0149)		(0.0258)	(0.0244)
τ_3:**Readability^E**	–	−0.0126	δ_6:**ReviewRating = 6**	−0.1664***	−0.1653***
		(0.0149)		(0.0320)	(0.0316)
δ_1:**Longevity**	0.0047***	0.0047***	δ_7:**ReviewRating = 7**	−0.0414	−0.0400
	(5.567e-05)	(5.71e-05)		(0.0276)	(0.0257)
δ_2:**ReviewRating = 2**	−0.0418	−0.0430	δ_8:**ReviewRating = 8**	0.1016***	0.1020***
	(0.0251)	(0.0242)		(0.0213)	(0.0218)
δ_3:**ReviewRating = 3**	−0.0940***	−0.0946***	δ_9:**ReviewRating = 9**	0.2238***	0.2224***
	(0.0277)	(0.0271)		(0.0198)	(0.0208)
δ_4:**ReviewRating = 4**	−0.1936***	−0.1933***	δ_{10}:**ReviewRating = 10**	0.3375***	0.3370***
	(0.0315)	(0.0280)		(0.0162)	(0.0160)
$\delta_{11}\sim\delta_{29}$: **Condition**	Included	Included			
Adj R^2	OLS: 0.4128		OLS Copula: 0.4118		

Standard errors are reported in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

the first analysis of tablet reviews. Negatively activated are positively and deactivated emotions are negatively associated with review helpfulness. Moreover, positively activated emotions with no significance in tablet samples show significantly positive effects. In these reviews written for drug users, higher arousal (regardless of valence) contributes to review helpfulness. The negative effects of negatively deactivated emotions on perceived helpfulness seem to suggest that readers seeking for healthy/medical-related experiences are less fond of reviews with low valance-low arousal mixes, which are possibly related to lower morale and slight depression.

Finally, Table 5 shows estimation results of restaurant reviews (Yelp.com). Several observations can be made. First, three of the four correction terms (τ_1 to τ_4) are significant and endogeneity concerns are arguably higher in this case, because *Readability* reveals its expected effects only after correcting for potential endogeneity biases in Model 6. The effects of *Length* on review helpfulness also seem to be significantly under-estimated in the naïve OLS regression model. Second, like

ReviewerRanking in Amazon.com, *ReviewerPosts* in Yelp.com exhibits positive effects with review helpfulness, implying that reviews written by relatively productive reviewers are more likely to be perceived useful. The foregoing estimates render firm support for H1a-H1c. Third, in line with H2, positively deactivated emotions persistently show positive effects on review helpfulness. Similar to analysis of drug reviews, high arousal in either low or high valence manifests significantly positive effects. This implies that passionate/energetic expressions in restaurant reviews also tend to be receptive to readers.

Table 6 summarizes the results of hypotheses-testing. Our analysis leads to two key observations. First, among the oft-tested quality indicators, review length exhibits significant and substantial effects across the three samples generated between 2012 and 2020, suggesting length persistently contributes to review helpfulness. The positive effects of readability and reviewer ranking/reviewer posts are consistent with literature. Readability, despite showing no significant effects in drug reviews, recovers its hypothesized effects in restaurant reviews until we

Table 5
Estimation results of restaurant reviews from Yelp.com ($n = 152{,}751$).

Variables	Model 5 OLS	Model 6 OLS Copula	Variables	Model 5 OLS	Model 6 OLS Copula
β_1:**log(Length)**	0.1002*** (0.0039)	0.3830*** (0.0261)	γ_1:**PositivelyAcativated**	0.0033*** (2.286e-04)	0.0032*** (2.740e-04)
β_2:**Breadth**	−4.938e-04 (0.0013)	0.0038 (0.0046)	γ_2:**PositivelyDeacativated**	9.190e-04*** (4.934e-05)	8.992e-04*** (6.511e-05)
β_3:**Readability**	−1.198e-04 (9.272e-05)	4.409e-04** (1.450e-04)	γ_3:**NegativelyAcativated**	0.0045*** (5.038e-04)	0.0045*** (6.044e-04)
β_4:**log(ReviewerPosts)**	0.1359*** (9.496e-04)	0.0810*** (0.023)	γ_4:**NegativelyDeactivated**	2.692e-04 (2.504e-04)	2.486e-04 (3.119e-04)
τ_1:**log(Length)**[E]	–	−0.2167*** (0.0189)	δ_6:**BizRating = 1.5**	−0.0580 (0.0402)	−0.0616 (0.0389)
τ_2:**Breadth**[E]	–	−0.0049 (0.0050)	δ_7:**BizRating = 2**	−0.0287 (0.0388)	−0.0315 (0.0378)
τ_3:**Readability**[E]	–	−0.0101* (0.0050)	δ_8:**BizRating = 2.5**	−2.029e-04 (0.0380)	−0.0041 (0.0376)
τ_4:**log(ReviewerPosts)**[E]	–	0.0960* (0.0411)	δ_9:**BizRating = 3**	0.0247 (0.0380)	0.0209 (0.0369)
δ_1:**Longevity**	−2.430e-05 (1.340e-05)	−2.447e-05 (1.349e-05)	δ_{10}:**BizRating = 3.5**	0.0472 (0.0379)	0.0433 (0.0370)
δ_2:**ReviewRating = 2**	−0.1278*** (0.0062)	−0.1288*** (0.0063)	δ_{11}:**BizRating = 4**	0.0476 (0.0379)	0.0436 (0.0368)
δ_3:**ReviewRating = 3**	−0.2226*** (0.0061)	−0.2229*** (0.0065)	δ_{12}:**BizRating = 4.5**	0.0910* (0.0380)	0.0870* (0.0370)
δ_4:**ReviewRating = 4**	−0.2213*** (0.0056)	−0.2212*** (0.0059)	δ_{13}:**BizRating = 5**	0.1567*** (0.0390)	0.1522*** (0.0379)
δ_5:**ReviewRating = 5**	−0.1844*** (0.0051)	−0.1842*** (0.0054)	δ_{14}:**log(BizReviewCounts)**	−0.0256*** (0.0011)	−0.0261*** (0.0011)
Adj R^2	OLS: 0.2459		OLS Copula: 0.2010		

Standard errors are reported in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 6
Summary of hypotheses-testing results.

Hypothesis Testing	Amazon.com (Tablet)	Drugs.com (Drug)	Yelp.com (Restaurant)
Length (H1a)	+ (supported)	+ (supported)	+ (supported)
Readability (H1b)	+ (supported)	not supported	+ (supported)
Reviewer Ranking (H1c)	+ (supported)	unavailable	+ (supported)
Breadth (H1d)	not supported	not supported	not supported
Positive Deactivation (H2a)	+ (supported)	+ (supported)	+ (supported)
Negative Activation (H2b)	- (supported)	not supported	not supported
Negative Deactivation (H2c)	+ (supported)	not supported	not supported

employ the robust Copula method, indicating the necessity for tackling endogeneity issues. In contrast, breadth – the coverage of semantic topics from NMF – turns to be insignificant in all corrected regression models. The finding is different from the literature (e.g., [5,41]) and reveals the need for extra empirical tests.

Second, different from quality indicators that tend to show consistent effects across platforms, emotions in the valence-arousal circumflex exhibit divergent associations with review helpfulness. On top of positive valence from prior studies (e.g., [2,16]), our research model finds that positively deactivated expressions are positively associated with review helpfulness, which is consistent in the three datasets across platforms. As high arousal is linked to agitation and irrationality [71], positively activated expressions are found to have insignificant effects for tablet reviews. Interestingly, for drug and restaurant reviews, we instead find positive effects of positively activated expressions. For positive sentiment, it seems that the effect of activation varies by contexts, and such variations are also present for negative expressions. For drug and restaurant reviews, high arousal combined with negative valence lead to positive effects on review helpfulness. Nevertheless, for tablet reviews, agitation of high arousal and negative expressions exhibit negative effects. The asymmetric effects of positive and negative valence on review helpfulness – contingent on low/high arousal as well as product/service categories, are worth exploring in subsequent studies.

6. Discussion and implications

Using 40,485 reviews on tablets from Amazon.com, 23,459 reviews on drugs from Drug.com, and 152,751 reviews on restaurants from Yelp.com, we perform an empirical analysis of the link between review features and the number of helpfulness votes. Rooted in the ELM, we argue that argument quality is the key driver of rational evaluation (central route). However, for observational data from websites, only review features or information cues are extractable and argument quality is typically unseen. We extend the ELM to elaborate on how quality indicators of a review are correlated with unobserved argument quality, resulting in endogeneity that interfere empirical results. We adopt an instrument-free method to formally address this issue and find most of the endogeneity-correction terms are statistically significant, supporting our conjecture.

Our estimation using samples from three different types of products and platforms consistently suggests that endogenous quality indicators should be econometrically tackled instead of being treated as ordinary predictors in OLS regression. We show that ignoring endogeneity would lead to under- or over-estimation of effects and invalid conclusions (e.g., *Readability* of Yelp data). Despite the sharp differences in the sampling settings, the three models with endogeneity corrections show full or partial support for our research hypotheses respectively. We observe that review length, reviewer ranking, and positively deactivated expressions are the influential features with consistent effects. While prior studies on eWOM generally treat review length as the basic information, our finding indicates that review length is the major variable with substantial impacts on review helpfulness. In addition, we supplement the literature on influences of emotion intensity by distinguishing positive/negative deactivation from activation. Our analysis reveals non-unidirectional effects of the negativity bias, i.e., some people perceive negative reviews as more authentic/helpful [10,70], in combination with arousal. The inconclusiveness calls for investigations into how activation levels in combination with valence would vary by types of products and platforms.

From the theoretical perspective, the ELM provides a conceptual foundation for researchers to categorize review features into issue-

relevant thinking ones (central route) and the ones coupled with intuitive responses (peripheral route). Those studies in general derive review- or reviewer-related antecedents from collected online reviews and explore either quality indicators and/or sentiments of reviews (e.g., [59,69,74,75]). Even though those studies fit with the framework of ELM and contribute to the body of literature on online review, we point out two outstanding issues: (1) the influence of unobserved argument quality (2) the dual effects of quality indicators on the central as well as peripheral routes. We acknowledge the issues and propose an extended ELM application to illuminate the mechanism between review features and review helpfulness. The mechanism further explains the correlation between unobserved argument quality and quality indicators (cues) due to their common link to the central route.

To statistically address the theoretically-rationalized endogeneity, we apply the copula method by Park and Gupta [52] to specify joint distributions of endogenous quality indicators and error term, which contains unobserved argument quality and random shocks. IV regression such as two stage least squares (2SLS) is the traditional approach to address endogeneity. Yet, the method has been debated in the literature as a decent IV is required to be uncorrelated with the error term (exclusion restriction) and sufficiently correlated with endogenous regressors (relevant instruments). The exclusion restriction is not directly testable and associated tests mostly rely on selected IVs. Strong IVs for the second condition (relevance), however, carry the risk of violating the first condition (exclusion). Identifying a set of valid IVs for multiple endogenous variables, in fact, is conceptually and practically difficult for empirical researchers [39]. We concur that the instrument-free method adopted in our study is particularly suitable for empirical studies on review helpfulness using archival data, where one could easily find more than one endogenous variable like we do.

Among a large body of studies on antecedents of review helpfulness using website data, our empirical study is the first, to the best of our knowledge, to examine the relationship between review features and review helpfulness under an explicit consideration of quality indicators being endogenous. We integrate relevant review features from the literature and make mild improvements on emotion measurement. In the literature, the level of emotion is commonly measured by percentage of emotion words, or one step further, by word counts of positive and negative valence. In this study, we apply Russel's [58] circumplex model to capture emotions not only by emotion valence, but also by the level of affect activation. For instance, while anxiety and sadness both represent negative valence, we further consider their differences in terms of the level of emotion activation. In addition, we go beyond binary indicators and word counts to measure emotions in a review into levels of four quadrants on the affect circumplex – positively activated, negatively activated, negatively deactivated, and positively deactivated. Such measurements reveal that the negativity bias effect from the literature is not universal but varies by the level of activation. Specifically, drugs and restaurants are service-oriented and require more information to reduce risk and uncertainty [9], and hence the associated review with more negatively activated emotions are perceived as more helpful. In contrast, reviews on tablets with tangible product features are perceived as helpful when negatively deactivated emotions are expressed. As the feelings expressed by reviewers can influence thousands of consumers who read the reviews [54], the interpersonal impact of emotions in online reviews is likely to be pervasive and long-lasting. We encourage researchers to build upon our work and conduct subsequent investigations into the emotion-review helpfulness link.

Our study also carries practical implications as articulating components of helpful reviews is crucial for store owners and online platform operators, and is also beneficial to reviewers and bloggers to construct helpful reviews. Many consumers may perceive online reviews as more important than generic information provided by a product seller or service provider. Hence, identifying and providing helpful reviews for customers is economically valuable. Knowing what constitutes a helpful review can help people/business incorporate certain helpful features in their product description and/or marketing campaign to increase customers' intention to buy. Based on our analyses, in addition to writing adequately long and readable reviews, people/business who aim to provide helpful reviews for reputation, reflection, or reward are supposed to carefully express emotions in terms of valence and arousal, such that the perceived helpfulness of reviews can be enhanced.

While our study enhances the understanding about how quality indicators, sentiment features, and other signals contribute to perceived helpfulness, our empirical findings are subject to limitations. First, in spite of our data collection efforts, our empirical analysis, like many others, is based on a set of finite review samples in certain time periods. Hence, we have no intent to overstate the generalizability of our results. Instead, we just aim to show theoretical and econometrical enhancements that could be leveraged by numerous researchers, who draw on the ELM theory and perform observational data analyses of online reviews (from various types of products, services, and platforms). Second, although our ELM-based framework includes arguably comprehensive review features, it cannot be exhaustive due to data availability. That said, the value of our empirical analysis is to bring attention to peer researchers (especially those who use online review data) the necessity of considering endogeneity, when attempting to estimating how review quality indicators associate with review helpfulness. Note that we focus on theoretical implications of review helpfulness, and are distinct from predictive modeling aiming for accurate prediction of helpful votes using machine learning algorithms (e.g., [15]).

Third, identification of semantic topics traditionally involves human coders to have predefined topics and manually process customer reviews (e.g., [44]). We instead use the NMF topic modeling to identify topics from each review and assess the coverage of topics as a proxy for breadth. While text mining enables automation of topic identification, we should recognize the variable is only the efficient approximation of human-encoded topics. Note that the coverage of topics proposed by Mankad et al. [41] is not a perfect measure of breadth and the concept of review breadth can be a challenging research topic alone.

Finally, the way we operationalize emotional intensity of each word in a review is by no means exhaustive. For instance, negation, amplification, or irony is also present in online product reviews (e.g., [60]). Different classifications of positive emotions – optimism, resilience, hope, and confidence – have been proposed too [40]. Our sentiment analysis is confined to four quadrants of the circumplex model of affect. Subsequent studies are encouraged to develop a comprehensive and even unified framework for emotional intensity in online product/service reviews.

References

[1] N. Aghakhani, O. Oh, D.G. Gregg, J. Karimi, Online review consistency matters: an elaboration likelihood model perspective, Inf. Syst. Front. (2020) (in press).

[2] 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.

[3] F.L. Barrett, J.A. Russell, Independence and bipolarity in the structure of current affect, J. Pers. Soc. Psychol. 74 (4) (1998) 967–984.

[4] A. Bhattacherjee, C. Sanford, Influence processes for information technology acceptance: an elaboration likelihood model, MIS Q. 30 (4) (2006) 805–825.

[5] 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.

[6] X. Chen, J. Sheng, X. Wang, J. Deng, Exploring determinants of attraction and helpfulness of online product review: a consumer behaviour perspective, Discret. Dyn. Nat. Soc. 2016 (1) (2016) 1–19.

[7] C.M.Y. Cheung, C.L. Sia, K.Y.Y. Kuan, Is this review believable? A study of factors affecting the credibility of online consumer reviews from an ELM perspective, J. Assoc. Info. Syst. 13 (8) (2012) 618–635.

[8] C.M.K. Cheung, D.R. Thadani, The impact of electronic word-of-mouth communication: a literature analysis and integrative model, Decis. Support. Syst. 54 (1) (2012) 461–470.

[9] H.S. Choi, S. Leon, An empirical investigation of online review helpfulness: a big data perspective, Decis. Support. Syst. 139 (2020) 113403.

[10] H.C. Chung, H. Lee, H. Koo, N. Chung, Which is more important in online review usefulness, heuristic or systematic Cue?, in: Proceedings of ENTER2017 eTourism Conference Springer, Rome, Italy, 2017.

[11] T. Conley, C. Hansen, P.E. Rossi, Plausibly exogenous, Rev. Econ. Stat. 94 (1) (2012) 260–272.

[12] G. Cui, H.K. Lui, S. Guo, The effect of online consumer reviews on new product sales, Int. J. Electron. Commer. 17 (1) (2012) 39–57.

[13] R. Cui, J. Li, D. Zhang, Reducing discrimination with reviews in the sharing economy: evidence from field experiments on Airbnb, Manag. Sci. 66 (3) (2020) 1071–1094.

[14] H. Datta, K.L. Ailawadi, H.J. van Heerde, How well does consumer-based brand equity align with sales-based brand equity and marketing mix response? J. Mark. 81 (3) (2017) 1–20.

[15] J. Du, J. Rong, H. Wang, Y. Zhang, Helpfulness prediction for online reviews with explicit content–rating interaction, in: Web Information Systems Engineering (WISE), Hong Kong, China, 2019.

[16] W. Fan, Contributions to Online Review Helpfulness: A Literature Review and Agenda for Future Research, Working Paper, Alto University, 2018.

[17] B. Fang, Y. Qiang, K. Deniz, R. Law, Analysis of the perceived value of online tourism reviews: influence of readability and reviewer characteristics, Tour. Manag. 52 (2016) 498–506.

[18] R. Filieri, F. McLeay, B. Tsui, Z. Lin, Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services, Inf. Manag. 55 (2018) 956–970.

[19] R. Flesch, A new readability yardstick, J. Appl. Psychol. 32 (3) (1948) 221.

[20] 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.

[21] M. Gendron, K.A. Lindquist, L. Barsalou, L.F. Barrett, Emotion words shape emotion percepts, Emotion 12 (2) (2012) 314–325.

[22] M.A. Gernsbacher, B.M. Hallada, R.R.W. Robertson, How automatically do readers infer fictional characters' emotional states? Sci. Stud. Read. 2 (3) (1998) 271–300.

[23] A. Ghose, P.G. Ipeirotis, Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics, IEEE Trans. Knowl. Data Eng. 23 (10) (2011) 1498–1512.

[24] H. Hong, D. Xu, A. Wang, W. Fan, Understanding the determinants of online review helpfulness: a meta-analytical investigation, Decis. Support. Syst. 102 (2017) 1–11.

[25] L. Huang, C.H. Tan, W. Ke, K.K. Wei, Helpfulness of online review content: the moderating effects of temporal and social cues, J. Assoc. Info. Syst. 19 (6) (2018) 503–522.

[26] E.E. Jones, K.E. Davis, From acts to dispositions the attribution process in person perception, in: L. Berkowitz (Ed.), Advances in Experimental Social Psychology vol. 2, Elsevier, Amsterdam, 1966, pp. 219–266.

[27] P. Kanske, S.A. Kotz, Concreteness in emotional words: ERP evidence from a Hemifield study, Brain Res. 1148 (2007) 138–148.

[28] S. Karimi, F. Wang, Online review helpfulness: impact of reviewer profile image, Decis. Support. Syst. 96 (2017) 39–48.

[29] N. Kim, W. Kim, Do your social media lead you to make social deal purchases? Consumer-generated social referrals for sales via social commerce, Int. J. Inf. Manag. 39 (2018) 38–48.

[30] 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.

[31] S.T. Kousta, D.P. Vinson, G. Vigliocco, Emotion words, regardless of polarity, have a processing advantage over neutral words, Cognition 112 (3) (2009) 473–481.

[32] D.D. Lee, H.S. Sueng, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.

[33] B. Li, E. Ch'ng, A.Y. Chong, H. Bao, Predicting online e-marketplace sales performances: a big data approach, Comput. Ind. Eng. 101 (2016) 565–571.

[34] B. Li, F. Hou, Z. Guan, A.Y. Chong, X. Pu, Evaluating online review helpfulness based on elaboration likelihood model: The moderating role of readability, in: PACIS 2017 Proceedings, 2017, p. 257.

[35] X. Li, C. Wu, F. Mai, The effect of online reviews on product sales: a joint sentiment-topic analysis, Inf. Manag. 56 (2) (2019) 172–184.

[36] M. Li, P. Huang, Assessing the product review helpfulness: affective-cognitive evaluation and the moderating effect of feedback mechanism, Inf. Manag. 57 (7) (2020) 103359.

[37] T.P. Liang, X. Li, C.T. Yang, M. Wang, What in consumer reviews affects the sales of mobile apps: a multifacet sentiment analysis approach, Int. J. Electron. Commer. 20 (2) (2015) 236–260.

[38] Z.W. Liu, S. Park, What makes a useful online review? Implication for travel product websites, Tour. Manag. 47 (2015) 140–151.

[39] G. Lu, X.D. Ding, D.X. Peng, H.H. Chuang, Addressing endogeneity in operations management research: recent developments, common problems, and directions for future research, J. Oper. Manag. 64 (2018) 53–64.

[40] F. Luthans, K.W. Luthans, B.C. Luthans, Positive psychological capital: beyond human and social capital, Business Horizons 47 (1) (2004) 45–50.

[41] S. Mankad, H.S. Han, J. Goh, S. Gavirneni, Understanding online hotel reviews through automated text analysis, Serv. Sci. 8 (2) (2016) 124–138.

[42] J. Mejia, S. Mankad, A. Gopal, Service quality using text mining: measurement and consequences, Manuf. Serv. Oper. Manag. (2021) forthcoming.

[43] D. Mimno, H. Wallach, E. Talley, M. Leenders, A. McCallum, Optimizing semantic coherence in topic models, in: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, 2011, pp. 262–272. Edinburgh, Scotland, UK.

[44] M. Mousavizadeh, M. Koohikamali, M. Salehan, The Effect of Central and Peripheral Cues on Online Review Helpfulness: A Comparison between Functional and Expressive Products, Int. Conference on Information Systems, Fort Worth, U.S., 2015.

[45] S.M. Mudambi, D. Schuff, What makes a helpful online review? A study of customer reviews on Amazon.com, MIS Q. 34 (1) (2010) 185–200.

[46] R. Murphy, Local Consumer Review Survey 2020, BrightLocal Ltd., 2020. Retrieved from, https://www.brightlocal.com/learn/local-consumer-review-survey/.

[47] D. Newman, Y. Noh, E. Talley, S. Karimi, T. Baldwin, Evaluating topic models for digital libraries, in: Proceedings of the 10th Annual Joint Conference on Digital Libraries, 2010, pp. 215–224. New York, NY, USA.

[48] P.M. Niedenthal, Emotion concepts, in: Handbook of Emotions 3rd Edition, The Guilford Press, New York, N.Y, 2008, pp. 587–600.

[50] Y. Pan, J.Q. Zhang, Born unequal: a study of the helpfulness of user-generated product reviews, J. Retail. 87 (4) (2011) 598–612.

[51] S. Park, J.L. Nicolau, Asymmetric effects of online consumer reviews, Ann. Tour. Res. 50 (January) (2015) 67–83.

[52] S. Park, S. Gupta, Handling endogenous regressors by joint estimation using copulas, Mark. Sci. 31 (4) (2012) 567–586.

[53] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[54] C.H. Peng, Z. Yin, C.P. Wei, H. Zhang, How and when review length and emotional intensity influence review helpfulness: Empirical evidence from Epinions.com, in: Int. Conference on Information Systems, Auckland, New Zealand, 2014.

[55] R.E. Petty, J.T. Cacioppo, Attitudes and Persuasion: Classic and Contemporary Approach 1st Edition, Routledge, New York, N.Y, 1981.

[56] R. Petty, J.T. Cacioppo, The effects of involvement on responses to argument quantity and quality: central and peripheral routes to persuasion, J. Pers. Soc. Psychol. 46 (1) (1984) 69–81.

[57] R. Petty, J.T. Cacioppo, Elaboration likelihood model, in: L. Berkowitz (Ed.), Advances in Experimental Social Psychology, Academic Press, San Diego, CA, 1986, pp. 123–205.

[58] J.A. Russell, A circumplex model of affect, J. Pers. Soc. Psychol. 39 (6) (1980) 1171–1178.

[59] M. Salehan, D.J. Kim, Predicting the performance of online consumer reviews: a sentiment mining approach to big data analytics, Decis. Support. Syst. 81 (C) (2016) 30–40.

[60] P. Savov, R. Nielek, Ridiculously expensive watches and surprisingly many reviewers: A study of irony, in: Proceedings of 2016 IEEE/WIC/ACM Int. Conference on Web Intelligence (WI), 2016.

[61] 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.

[62] J.J. Skowronski, D.E. Carlston, Social judgment and social memory—the role of cue diagnosticity in negativity, positivity, and extremity biases, J. Pers. Soc. Psychol. 52 (4) (1987) 689–699.

[63] S.W. Sussman, W.S. Siegal, Informational influence in organizations: an integrated approach to knowledge adoption, Inf. Syst. Res. 14 (1) (2003) 49–65.

[64] A. Timoshenko, How Companies Can Mine Online Reviews for Product-Development Gold, Kellogg Insight. Retrieved from, https://insight.kellogg.northwestern.edu/article/how-companies-can-mine-online-reviews-for-product-development-gold, 2019.

[65] X.S. Wang, F. Mai, R.H.L. Chiang, Database submission – market dynamics and user-generated content about tablet computers, Mark. Sci. 33 (3) (2013) 449–458.

[66] Y. Wang, J. Wang, T. Yao, What makes a helpful online review? A meta-analysis of review characteristics, Electron. Commer. Res. 19 (2) (2019) 257–284.

[67] A.B. Warriner, V. Kuperman, M. Brysbaert, Norms of valence, arousal, and dominance for 13,915 English lemmas, Behav. Res. Methods 45 (4) (2013) 1191–1207.

[68] J.M. Wooldridge, Econometric Analysis of Cross Section and Panel Data 2nd Edition, MIT Press, Cambridge, M.A, 2010.

[69] J. Wu, Review popularity and review helpfulness: a model for user review effectiveness, Decis. Support. Syst. 97 (2017) 92–103.

[70] D. Yin, S.D. Bond, H. Zhang, Anxious or angry? Effects of discrete emotions on the perceived usefulness of online reviews, MIS Q. 38 (2) (2014) 539–560.

[71] G. Yin, Q. Zhang, Y. Li, Effects of Emotional Valence and Arousal on Consumer Perceptions of Online Review Helpfulness, 2014. Americas Conference on Information Systems, Savannah, US.

[72] G. Yin, L. Wei, W. Xu, M. Chen, Exploring heuristic cues for consumer perceptions of online reviews helpfulness: The case of yelp.com, in: Pacific Asia Conference on Information Systems, Chengdu, China, 2014.

[73] T. Zhou, Y. Lu, B. Wang, Examining online consumers' initial trust building from an elaboration likelihood model perspective, Inf. Syst. Front. 18 (2) (2016) 265–275.

[74] 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) 267–280.

[75] S. Zhou, B. Gou, The order effect on online review helpfulness: a social influence perspective, Decis. Support. Syst. 93 (2017) 77–87.

[76] Felix Gräßer, Surya Kallumadi, Hagen Malberg, Sebastian Zaunseder, Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning, in: Proceedings of the 2018 International Conference on Digital Health (DH '18), ACM, New York, NY, USA, 2018, pp. 121–125.

[77] Yelp. Yelp Open Dataset, 2020 available at https://www.yelp.com/dataset.

Yen-Chun Chou is an Associate Professor of Management Information Systems in the College of Commerce at National Chengchi University in Taipei, Taiwan. She received her B.A. in management information systems from National Chengchi University, Taiwan, her M.S. in information systems management from Carnegie Mellon University, and Ph.D.

from Arizona State University. Her current research interests are the value of information technology, service science, and online commerce, and data analytics. She serves as an associate editor of *Journal of Electronic Commerce Research*. Her research papers have been published in *Journal of Operations Management, Decision Support Systems, Information & Management, European Journal of Operational Research*, etc.

Howard Hao-Chun Chuang is an Associate Professor in the College of Commerce at National Chengchi University. He received his Ph.D. (major: supply chain operations; minor: statistics) from the Mays Business School at Texas A&M University. Prior to his academic career, he served as a supply corporal for a year in Taiwan army. His research interests are retailing, supply chain operations, data analytics, applied econometrics, and quantitative modeling. He is an associate editor of *Decision Sciences*, a senior editor of *Journal of Electronic Commerce Research*, and serves on the editorial board of *Journal of Operations Management* as well as *Production and Operations Management*. His papers have appeared in *Journal of Operations Management, Production and Operations Management, Decision Sciences, Decision Support Systems, European Journal of Operational Research*, etc.

Ting-Peng Liang is a Life-time National Chair Professor and director of the Electronic Commerce Research Center of National Sun Yat-sen University. He is an AIS Fellow (2003) and has taught at University of Illinois, Purdue University, Chinese University of Hong Kong, and City University of Hong Kong. In 2014, he was awarded an LEO by the Association for Information Systems for lifetime achievement in the field of information systems. He is the founding Editor-in-Chief of *Pacific Asia Journal of the Association for Information Systems*, co-editor-in-Chief of *Journal of Electronic Commerce Research*, and serves on the editorial board of *Journal of the AIS, Decision Support Systems*, and several other journals. His papers have appeared in *MIS Quarterly, Management Science, Journal of Management Information Systems, Operations Research, Decision Support Systems, International Journal of Electronic Commerce*, and other journals.