A Critical Assessment of Consumer Reviews: A Hybrid NLP-based Methodology

Dr. Baidyanath Biswas*
Enterprise and Innovation Group,
DCU Business School,
Dublin City University, Ireland
Email: baidyanath.biswas@dcu.ie
(*Corresponding Author)

Dr. Pooja Sengupta
MIS and Analytics Area
Indian Institute of Management, Ranchi, India
Email: pooja.sengupta@iimranchi.ac.in

Dr. Ajay Kumar
AIM Research Center on Artificial Intelligence in Value Creation
EMLYON Business School, France
Email: akumar@em-lyon.com

Prof. Dursun Delen
Department of Management Science & Information Systems
Spears School of Business, Oklahoma State University, USA
Email: dursun.delen@okstate.edu

Prof. Shivam Gupta
Department of Information Systems, Supply Chain Management & Decision Support
NEOMA Business School, France
Email: shivam.gupta@neoma-bs.fr

A Critical Assessment of Consumer Reviews: A Hybrid NLP-based Methodology

Abstract

Online reviews are integral to consumer decision-making while purchasing products on an e-commerce platform. Extant literature has conclusively established the effects of various review and reviewer related predictors towards perceived helpfulness. However, background research is limited in addressing the following problem: *how can readers interpret the topical summary of many helpful reviews that explain multiple themes and consecutively focus in-depth?* To fill this gap, we drew upon Shannon's Entropy Theory and Dual Process Theory to propose a set of predictors using NLP and text mining to examine helpfulness. We created four predictors - *review depth*, *review divergence*, *semantic entropy* and *keyword relevance* to build our primary empirical models. We also reported interesting findings from the interaction effects of the *reviewer's credibility*, *age of review*, and *review divergence*. We also validated the robustness of our results across different product categories and higher thresholds of helpfulness votes. Our study contributes to the electronic commerce literature with relevant managerial and theoretical implications through these findings.

Keywords: *Online Reviews; Natural Language Processing (NLP); Shannon's Entropy; Text Analytics; Zero-truncated Regression*

1. Introduction and Motivation

Online reviews, also known as electronic word-of-mouth (e-WOM), are user-generated contents that enable customers to present their first-hand assessments about products and services after consumption. Most users treat online reviews as personal recommendations based on first-hand usage experiences from previous buyers and acknowledge them to be more effective than traditional marketing communication mix while describing the *pros* and *cons* of a product or service [1, 10, 18, 45, 68, 69]. According to the 2021 Statista Global Consumer Survey conducted in the

United States, 40 per cent of buyers consulted online reviews to search product-specific information, ranking second only after search engines (measured at 62 per cent)[1]. Therefore, online reviews act as free "sales assistants" for prospective buyers while also enabling sellers to improve their sales volume [11].

Recently, many independent websites have been developed, such as *Trustpilot*, *Reviews.io* and *PowerReviews*, which are useful for buyers and sellers to read and publish online product reviews. Before the final purchase, potential customers also refer to platforms such as *Moz Local* and *BrightLocal*, which publish reviews for local businesses, restaurants, and hotels. Additionally, the helpfulness votes received by online reviews help potential customers by focussing on the most voted helpful reviews only. These findings signify that "online reviews", especially the "helpful" ones, strongly influence the online sale of products. Scholarly studies also confirm that online reviews help to boost product sales on electronic marketplaces [12, 13, 22, 27, 40, 41].

Extant literature has reported that *review-based* attributes [7, 22, 45, 46] and *reviewer-based* attributes [32, 56, 66] are suitable for examining the helpfulness of online reviews. However, many challenges remain with these traditional predictors. For instance, using title length and review length (measured by words) [3,15,31] to enumerate the effect of review richness towards its helpfulness is an incomplete measurement. Further, when reviewers write online reviews, they often discuss a particular product feature in-depth or talk about many elements within the scope of a single review. To elaborate on the existing gap, we present an example in Figure 1: online reviews for *Kaspersky Anti-Virus* extracted from Amazon. It shows three reviews, where Review#1 (marked in RED) discusses four topics - "security", "license", "packaging", and "features." Review#2 (marked in BLUE) discusses three topics - "price", "seller", and "activation." In comparison, Review#3 (marked in GREEN) discusses three topics - "installation", "support", and "price" – indicating multiple reviews for the same product with varying topic depth and breadth.

---

[1] Statista.com: https://www.statista.com/forecasts/997051/sources-of-information-about-products-in-the-us

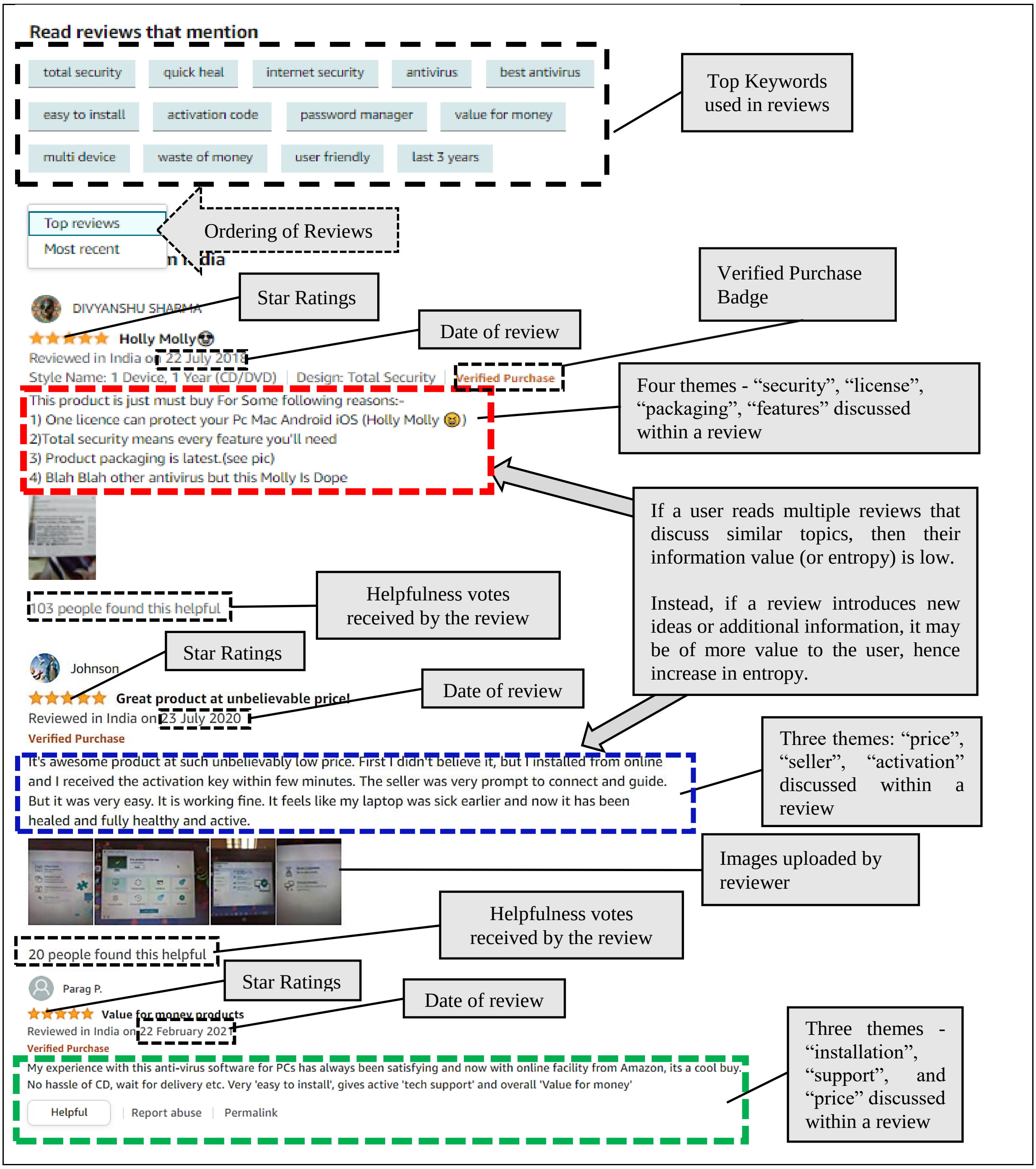

Fig. 1 – Online reviews related to Kaspersky Anti-Virus explaining various predictors, embedded topics and information gain (Amazon India)

When users on an e-commerce platform seek new information about a product, they read multiple online reviews for the product [65,66]. If users read multiple reviews that discuss similar topics or merely repeat previously used words, their informational value (or entropy) is low. Instead, if a review introduces new ideas, it may be of more informational value to the user, leading to an increase in entropy. Therefore, it needs to be examined what the major topics (or themes) presented within online reviews for each product are, and then build upon this knowledge to identify which type of reviews (i.e. those with “in-depth” discussion on a particular topic or “distributed” across multiple topics) are more useful to readers. We propose two predictors to measure “in-depth” and “distributed” reviews, namely “review depth” and “review divergence”, respectively. These are based on the concept of information entropy borrowed from Shannon’s Entropy Theory [54]. We add to the literature [23,24,66] by using correlated topic modelling to build these entropy-based measures. For instance, based on Fig. 1, each review text can be vectorially represented by a mixture of topics. If a topic is absent from a review, its coefficient is zero. Thus, topic-modelling helps reduce the multiple features of a product discussed within online reviews and represented by measurable values. Therefore, we present our first research question: **RQ1: How do “review depth” and “review divergence” affect the helpfulness of online reviews?**

However, the uniqueness of information as perceived by a user when reading multiple reviews can also vary with their semantic themes [68]. For instance, when comprehending the information coverage of two similar reviews, “the quality of printing is very nice and gives more quality printing than other refill inks” and “highly recommended printer cartridge product,” the user analyses their information richness in combination with their semantic content and linguistic style of presentation [25]. Therefore, it is also important to understand how readers perceive the newness of information across multiple reviews. This semantic (dis)similarity can be computed by natural language processing techniques such as latent semantic analysis (LSA). While related studies have applied the concept of entropy and incremental information gain by counting each word occurring in an online review [58, 66], our study improves upon Fresneda and Gefen [23,24] by combining the entropy-based measure with semantic similarity. Thus, we present our second research question: **RQ2: How does “semantic entropy” affect the helpfulness of online reviews?**

Next, we note that e-commerce platforms may present the relevant keywords occurring across the online reviews related to a particular product. Users find it more convenient to select online reviews based on these important keywords and vote the associated online reviews as more helpful [1,47]. For instance, Fig. 1 shows the relevant keywords for Kaspersky Anti-virus on Amazon are “total security”, “activation code”, “user friendly”, and “internet security”. Yang et al. [69] reported that users often deemed reviews with matching titles and content more helpful. Similarly, it needs to be studied what are the major keywords within the online reviews for each product, and then build a predictor based on these relevant keywords that enable reviews to be perceived more helpful to readers. Consequently, we apply a weighted overlap-score [43] to measure this relevance and present our third research question: **RQ3: How does “keyword relevance” affects the helpfulness of online reviews?**

Next, drawing from the Dual Process Theory [17,18,20,66], we posit that *review depth* measures the informational signals of multidimensional textual content in the review and, therefore, represents *informational influence*. In contrast, the *review divergence* measures the consistency of the content of a review compared to other reviews available for the same product and therefore represents *normative influence*. Hence, reviews that have more diversity and discuss lots of features might appear less helpful to readers than those discussing fewer topics but are more focused. So, we expect negative moderation effects of *review divergence* for *review depth*. Besides, the reputation of a reviewer on Amazon can magnify a user’s perceptions toward information in the review. Users expect highly-ranked reviewers to write more in-depth reviews using more topic-relevant keywords. So, we expect positive moderation effects of *credibility* for *review depth* and *keyword relevance*. Next, the *age of review* measures the time difference between data collection and review submission. According to the Dual Process Theory, the *age of review* reduces the user’s uncertainty by allowing renewed information and enhancing the informational influence [66]. Thus, among in-depth reviews, older ones will become less relevant than more recent ones. So, we expect negative moderation effects of *age of review* for *review depth* and *keyword relevance*. Thus, we present our fourth research question: **RQ4: How does “age of review”, “credibility of the reviewer”, and “review divergence” moderate the main effects on review helpfulness?**

Therefore, the helpfulness needs to be re-examined with a fresh set of predictors that (i) accommodates the need for measuring "newness of information" among multiple reviews, (ii) adjust for the tradeoff between the in-depth discussion and diversity of topics discussed in a review. We aim to address these challenges by proposing four predictors of helpfulness using NLP, text mining and Shannon's Entropy Theory, as follows: (i) **"keyword relevance"** (ii) **"review depth" -** (iii) **"review divergence"**, and (iv) "**semantic entropy"**.

We use a data-driven approach to build the empirical framework to address the research questions. The primary contributions of this study are threefold: (i) application of correlated topic modelling techniques to extract the latent predictors from online reviews. Next, we applied Shannon's Entropy Theory to create "review depth" and "review divergence". They represent the informational influence and normative influence, respectively, derived from the Dual Process Theory [17]; (ii) build a "semantic entropy" measure. We compute the incremental entropy to measure uniqueness across consecutive reviews and then normalize with semantic similarity, thereby, it extends Wu et al. [66]; (iii) test the interaction effects of the *age of review*, reviewer's *credibility*, and *review divergence* on the relationships of these predictors towards helpfulness.

Section 2 presents the literature review and theoretical foundations. Section 3 describes the data and methodology. Section 4 presents the empirical modelling, and Section 5 discusses the results. Section 6 presents our study's theoretical, methodological, and managerial implications. Finally, Section 7 concludes this study with limitations and directions for future research.

2. Background Work and Theoretical Foundation

2.1 Online reviews and determinants of helpfulness

Online consumer reviews (OCR)-s are perceived as the *wisdom of the crowd*. They have proved to be a reliable and valid source of "purchase information" that influence the users of electronic platforms [23,27, 41]. While there is no comprehensive structure of OCRs that users abide by, most of them consist of: (i) **review content**: qualitative text consisting of a *review title*, the detailed *review-text*, or multimedia files such as images or videos describing the product (*unstructured* data);

(ii) **review-valence**: overall number of stars assigned to the product by the past consumer (*structured* data); (iii) **reviewer-details**: historical information about the consumer who submitted a review, such as *credibility* (e.g. reviewer belongs to the Top 500 ranks), *total reviews posted*, *reviewer expertise* (*structured* data). These elements of an OCR help to generate many relevant determinants of helpfulness, namely *title length* [7, 53]; *title sentiments* [7, 45, 53, 69]; *review length* [7, 26, 45, 46]; *review sentiments* [26, 45, 53, 69]; *review readability* [7, 22, 42, 45]; *review valence* or *star-ratings* [2, 45,46]. Additionally, various attributes of a reviewer play an important role in predicting the helpfulness of online reviews, such as: *identity* [32], *profile* [32], *actual name/non-anonymity* [2, 56], *credibility* [15, 56, 71] and *reputation* [15, 47, 66].

2.2 Information Theory and Entropy as the Theoretical Lens

Recently, scholars have applied entropy from Information Theory [54] and combined it with text mining to explain the helpfulness of an OCR [23, 24, 58, 66]. According to Fresneda and Gefen [23, 24], entropy represents a measure of information uniqueness of an online review. When users on an e-commerce platform seek new information about a product, they read online reviews to reduce the uncertainty during the purchase decision-making [64,65]. A recent survey by Statista reports that nearly 70 per cent of online shoppers typically read between one and six customer reviews before making a purchasing decision. In contrast, less than one in ten shoppers did not read customer reviews before buying[2]. Also, we note that users on Amazon can sort the order of reviews based on "top reviews" and "most recent" before reading them (see Fig.1). While reading, if users encounter reviews that discuss similar topics or merely repeat previously used words or themes, their informational value (or entropy) is low. Instead, if a review introduces new ideas or features, it may be of more informational value to the user, leading to entropy changes. Our methodology approximates this conceptualization. Therefore, given the huge number of online reviews available for every product on Amazon, proposing an entropy-based predictor to measure the review

---

[2] Statista Report on "Share of Shoppers reading reviews before purchase"

helpfulness will be more useful for potential customers, allowing them to read information-rich reviews quicker. For example, based on Fig. 1, when a user reads a set of reviews about Kaspersky Anti-Virus, with every review that mentions a unique attribute of the product, entropy increases, and so does the helpfulness of the review for the user. Similarly, a review with fewer unique ideas might not be helpful to the user. In this manner, the entire process leads to a change in information entropy.

Within the scope of information systems, entropy can be defined as a "measure of the amount of information the system contains" (Belzer, 1973, p.301) [4] [3]. According to Hausser and Strimmer [29], Shannon's entropy is given as Eq. (1)

$$H = -\sum_{i=1}^{n} \Phi_i log(\Phi_i) \tag{1}$$

where the associated probability for i^{th} event is Φ_i; $\Phi_i > 0$ and $\sum_i \Phi_i = 1$.

2.3 Elaboration Likelihood Model as the Theoretical Lens

Our study draws motivation from the *Elaboration Likelihood Model* and the *Dual Process Theory* to explain how information received from online media influences users during subsequent decision-making. Petty and Cacioppo [50] proposed the Elaboration Likelihood Model as a form of the Dual Process Theory. ELM Theory suggests two primary routes to persuasion with the help of communication cues while explaining the helpfulness of OCRs [1,10,45,68,69]. The first route involves thoughtful consideration of *central cues* directly related to an OCR's helpfulness. The second route involves *peripheral cues* that implicitly affect an OCR's helpfulness when high cognitive processing is required. Consistent with the ELM theory, we identify central cues as those predictors primarily derived from the textual content of online reviews because processing the

[3] Tossing a biased coin can be a good example, where the probability of showing a head is "p" and that of showing a tail is "q", s.t. $(p + q) = 1$ and $p \neq q$. Intuitively, we understand that a person, who tosses the coin, is less surprised when the biased side of the coin (say "head") appears successively in consecutive tosses. Therefore, the associated information entropy in each trial is low. If, the coin is strongly biased towards a particular side $(i.e.\ p = 1;\ or\ q = 1)$, the person is least surprised after each trial, and therefore, the associated information entropy value is close to 0. On the contrary, when an unbiased coin is tossed, the person is always surprised because each side is equally likely, thereby leading to an information entropy of almost 1.

review text requires high cognitive effort by online users [10, 45]. Examples of central cues include *readability*, *total title sentiments*, *keyword relevance*, and *total review sentiments*). In addition, entropy-related predictors proposed in this study, such as *review depth*, *review divergence* and *semantic incremental entropy*, require enormous cognitive effort to comprehend and therefore comprise central cues. Next, we identify peripheral cues as those simpler attributes (such as *high star ratings*, *is real customer*, *reviewer expertise*, the *credibility of the reviewer*, *age of review*, *superlatives used in the review*), all those which may simplify the search process for online users while reading reviews. Table 1 summarises the extant literature on review helpfulness. Fig. 2 presents the proposed conceptual framework for this study.

Table 1. - Summary of relevant literature on OCRs employing ELM and DPT

Academic Source	Theory	Context	Response	Major Findings
Cheung et al. [10]	ELM	Online survey Epinions.com	RC	Argument quality, source credibility for RC
Yang et al. [68]	ELM	TripAdvisor	RH	Topic regularity, semantic style affect RH
Mousavizadeh et al. [45]	ELM	Amazon	RP, RH	Longer reviews, extreme star ratings are more RP.
Aghakhani et al. [1]	ELM	Yelp	RH	Propose “review consistency” for RH
Wang and Karimi [62]	ELM	Amazon	RH	Linguistic choice → perception of RH
Yang et al. [69]	ELM	Amazon	RH	Review title resembles content, are more RH
Baek et al. [3]	ELM	Amazon	RH	Peripheral/central cues→RH
Meek et al. [44]	DPT	Zomato	RH	Contextual and descriptive attributes towards RH
Wu et al. [66]	DPT	Amazon	RV, RH	Review depth, review divergence on RH
Filieri et al. [20]	DPT	Restaurant customers	RV, PI	Popularity, two-sided reviews, experts RH
Filieri [18]	DPT	TripAdvisor users	ID, IA/RA	Quality of information and ratings → online users
Chua and Banerjee [15]	DPT	Amazon	RH, RP	Antecedents of “review efficacy” vs RH
Lee and Hong [37]	ELM, DPT	Online survey of hotel users	IA/RA	Trust transfer mechanism within review platforms
This Study	ELM, DPT	Amazon	RH	Review depth/divergence, semantic entropy → RH

RH=Review Helpfulness; RP=Review Popularity; RV = Review Voting; PI=Purchase Intentions; ID = Information Diagnosticity; IA/RA = Information/Review Adoption; RC=Review Credibility ELM = Elaboration Likelihood Model; DPT = Dual Process Theory

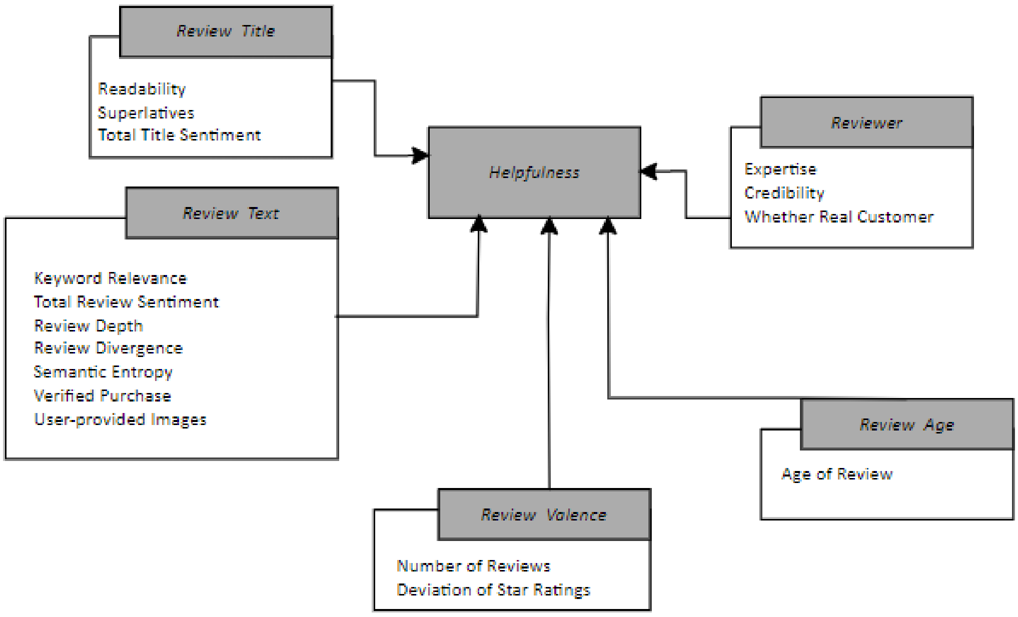

Fig. 2 - The proposed framework to examine the effect of semantic and NLP-based features on the "count of helpfulness votes" received by an online review

3. Data and Methodology used for this study

3.1 Data Description and Feature-Engineering

We extracted online reviews from Amazon India for the following products to build the proposed empirical framework. Books (2976), Groceries (2146), Electronics (3265), Apparels (3770), Travel Accessories (2881), E-gift cards (2890), Insurance plans for mobiles (2385), Kindle e-books (2777), Anti-virus software (2250), Operating Systems software (2454), and Video games (2629)[4]. We accomplished the web-scraping task with the help of an R program and retrieved a total of 30,423 reviews published up to December 2020. Several studies have successfully used Amazon data to examine the helpfulness of online reviews [1,7,14,46,66]. However, the categories have highly different product attributes in terms of product quality evaluability (i.e., search vs experience) and product tangibility (i.e., tangible vs intangible), which significantly affect review helpfulness votes [3,14,21,38,46,64]. Therefore, we also decided to run robustness checks separately across these product categories with split-samples (see Table 8). Next, many reviews on the Amazon platform

[4] The count of records for each category is indicated in the parentheses.

did not receive any helpful votes. Therefore, in line with extant studies [2], we removed the review records with zero helpful votes, leading to 930 final records for our final analysis.

Next, according to our proposed framework, we characterized those features into the following groups: (i) **review title**: *readability*, *superlatives* and *total title sentiments*, (ii) **review text**: *keyword relevance*, *total review sentiments*, *review depth*, *review divergence*, *semantic entropy*, *verified purchase badge* and *user-provided images*; (iii) ***product details***: *total number of ratings* and *deviation of star ratings*; and (iv) **reviewer attributes**: *reviewer expertise*, *credibility*, *real customer* and *age of review*.

To facilitate feature-engineering of the variables in our empirical framework, we employed: (i) Shannon's Information Theory using *entropy*[5] package in R; (ii) correlated topic modelling using *stm*[6] package in R; (iii) text analysis using LIWC [49]; (iv) semantic analysis using *lsa*[7] package in R. Next, we computed the readability of each review title as: $Readability = 1/\ GunningFogIndex$ where $GunningFogIndex = 0.4 * (ASL + PCW)$; $ASL = average\ sentence\ length$; and $PHW = percentage\ of\ compound\ words$ [5]. To measure the superlatives, we computed the number of adjectives in the review title with LIWC [2;5]. Finally, we applied *Information Retrieval Theory* to consider top keywords from the document-term matrix (DTM) and build the *keyword relevance* measure. We used the bag-of-words model for the texts retrieved from OCRs and computed an aggregate score created from *TF-IDF* [43,47,57]. We measured the reviewer identity as a dummy variable, with a value of one if the reviewer's name was mentioned alongside a review. Instead, if simply "Amazon customer" was mentioned, we coded it as zero [2]. Similarly, we measured the presence of user-provided images of a product in the review as a dummy variable, with a value of one if images were attached alongside a review and zero without them [21,38].

---

[5] Entropy Package in R: https://cran.r-project.org/web/packages/entropy/entropy.pdf
[6] Structural Topic Models Package in R: https://cran.r-project.org/web/packages/stm/stm.pdf
[7] Latent Semantic Analysis Package in R: https://cran.r-project.org/web/packages/lsa/lsa.pdf

3.2 Correlated Topic Models (CTM)

Probabilistic topic models are a type of unsupervised machine-learning technique where the embedded topics in a corpus of text can reveal with the help of hidden random variables, using Bayesian techniques [8]. A correlated topic model is a hierarchical representation of a collection of documents. For example, CTM models permits the reviews for a particular product to build upon a joint cluster of *topics*, while the contribution of each topic may vary within each review text (see Fig. 1). Therefore, a *topic* is represented as a probability distribution over a *corpus* of *words*, while the *words* within a topic represent its *topical content*. As a result, every document can be associated with multiple topics but with different contributions.

3.2.1 Calculating topic-level depth for each review

First, we built the CTM models and identified the topic probabilities (or contributions) for each review using the *stm* package in R. The *stm* package extracts $n = 7$ topics (Fig. 3) for the entire corpus of reviews for a typical product "Kaspersky Antivirus" and the corresponding topic contributions (given as theta-values) θ_i from each review r_i are given in Table 2. Using the functions in the *stm* package, we reported the top associated words for each topic using four types of measure: highest probability, FREX metric[8], lift[9], and score[10].

Next, we considered a typical review for "Kaspersky Antivirus", which can be vectorially represented as $\overrightarrow{r_1} = [0.086,\ 0.154,\ 0.117, 0.246,\ 0.120,\ 0.114, 0.164]$. We applied Shannon's Entropy using Eq. (1) with the CTM results to combine the topic-level contributions for the review $\overrightarrow{r_1}$ and computed the per-review topic-level depth as:

$$H(\overrightarrow{r_1})_{Kaspersky} = -\sum_{i=1}^{7}\{0.086 * log(0.086) + \cdots\cdots + 0.164 * log(0.164)\} = 0.822$$

[8] For more information on the FREX metric, see Bischof and Airoldi [6]
[9] For more information on lift, see Taddy [59]
[10] For more information on score, see the documentation for *lda* package in R.

Topic 1 Top Words:
Highest Prob: kasperski, best, year, can, comput, quick, get
FREX: best, can, satisfi, devic, get, comput, job
Lift: hai, hii, issuesi, nowit, possess, reload, working, can
Score: best, kasperski, year, quick, can, comput, get

Topic 2 Top Words:
Highest Prob: work, total, easi, got, also, deliveri, user
FREX: deliveri, easi, much, alreadi, fast, got, cant
Lift: global, establish, enterpris, grow, level, most, recognit
Score: work, easi, total, got, deliveri, also, user

Topic 3 Top Words:
Highest Prob: use, protect, virus, time, like, updat, fine
FREX: time, updat, virus, like, use, mani
Lift: producf, remov, monitor, period, averag, adequ
Score: use, protect, time, virus, like, updat, fine

Topic 4 Top Words:
Highest Prob: slow, softwar, system, dont, hour, version, heal
FREX: system, doesnt, slow, softwar, check, clean, help
Lift: config, jam, lose, therebi, undoubt, initi, fals
Score: system, slow, softwar, dont, heal, doesnt, hour

Topic 5 Top Words:
Highest Prob: nice, price, just, well, will, issu, recommend
FREX: nice, code, download, price, websit, just, seal
Lift: pro, latest, slowdown, great, system, custom, cool
Score: nice, price, just, well, code, genuin, will

Topic 6 Top Words:
Highest Prob: product, antivirus, laptop, excel, great, activ
FREX: product, great, excel, one, antivirus, expect, realli
Lift: blah, molli, dope, total, holli, latest, pic
Score: product, antivirus, one, excel, great, laptop, activ

Topic 7 Top Words:
Highest Prob: good, secur, instal, buy, money, key, featur
FREX: instal, secur, buy, featur, key, thank, perfect
Lift: self, promis, reliabl, fulli, instal, secur, buy
Score: good, secur, instal, buy, featur, key, money

Fig. 3 - Correlated topics and top keywords using *stm* package (Kaspersky reviews)

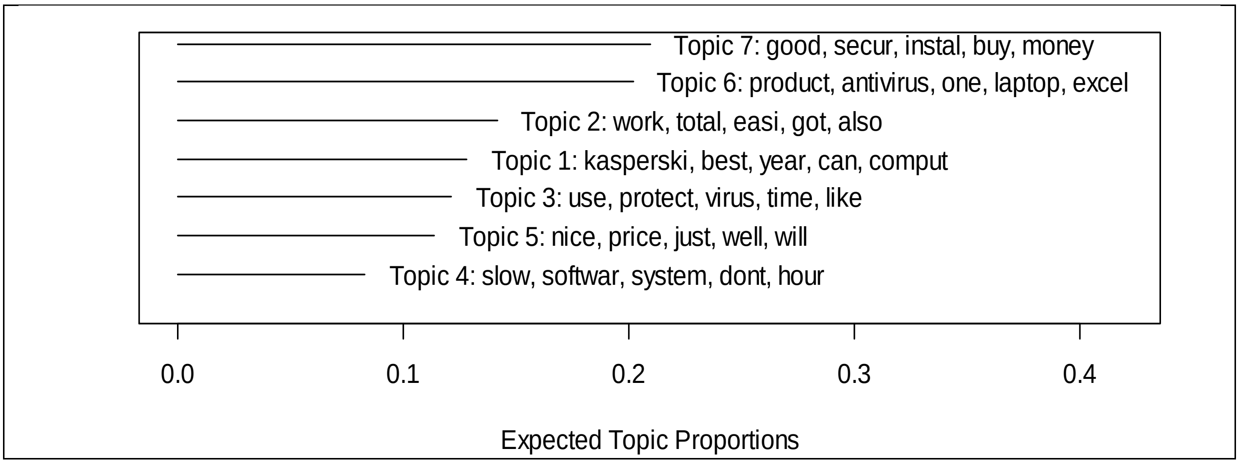

Fig. 4 - Top 5 keywords based on expected probabilities for each of 7 topics (Kaspersky reviews)

Table 2: Aggregate topic proportions for Kaspersky reviews

Topic #	Mean Theta Θ_i
Topic 1	0.12847
Topic 2	0.14164
Topic 3	0.12128
Topic 4	0.08312
Topic 5	0.11396
Topic 6	0.20196
Topic 7	0.20954

3.2.2 Calculating "review divergence" across various topics

The *Kullback-Leibler divergence* (Kullback and Leibler, 1951), also known as the *KL divergence,* has been popularly used to measure the difference between two probability distributions over a random variable x. While Shannon's Entropy measures how much information is in the data, KL divergence or *information divergence* measures the *relative entropy*, i.e. the difference between two probability distributions that map the same information signal (here the review vector). Let $p(x)$ and $q(x)$ be two probability distributions of a discrete random variable x. That is, both $p(x)$ and $q(x)$ add up to 1, and $p(x) > 0$ and $q(x) > 0$ for any x in X. Therefore, ***KL divergence*** of $q(x)$ from $p(x)$ is a measure of the information lost when $q(x)$ is used to approximate $p(x)$, denoted by $D_{KL}\,(p||q)$ and given as $D_{KL}\,(p||q) = \sum_{x\in X} p(x) log \frac{p(x)}{q(x)}$.

We apply *relative entropy*, which KL Divergence measured among a typical review and a representative review for each product sold on Amazon. Now, a baseline review $\overrightarrow{R_p}$ for any product p sold on Amazon can be represented as the aggregate (over N reviews) of individual topic contribution θ_i from each review r_i and given by $\overrightarrow{R_p} = \frac{1}{N}\sum_{i=1}^{n} \theta_i|\overrightarrow{r_i}$. From Fig. 4 and Table 2, the baseline review vector for the product "Kaspersky Antivirus" is given as $\overrightarrow{R_p} =$ [0.128, 0.142, 0.121, 0.083, 0.114, 0.202, 0.210] such that $\sum_i \Theta_i = 1$, where Θ_i is the topic-proportion for the aggregate vector $\overrightarrow{R_p}$. Let us again get back to the review ($\overrightarrow{r_1}$) for this calculation. Therefore, the KL-divergence between the baseline review $\overrightarrow{R_p}$ and $\overrightarrow{r_1}$ is denoted as $D_{KL}\left(\overrightarrow{r_1}||\overrightarrow{R_p}\right)$, which ideally represents the *review divergence* or *relative entropy* of $\overrightarrow{r_1}$ measured against the baseline review $\overrightarrow{R_p}$. To elucidate, we show the sample calculation of *review divergence*.

$$D_{KL}\left(\overrightarrow{r_1}||\overrightarrow{R_p}\right)_{Kaspersky} = -\sum_{i=1}^{7} 0.086 * \left(\frac{0.086}{0.128}\right) + \cdots\cdots + \ 0.164 * \left(\frac{0.164}{0.210}\right)$$

3.2.3 Semantic Similarity and Incremental Entropy

Next, we computed the incremental information entropy offered by each review for a particular product sold on the e-commerce platform. It was measured with the *count of unique words* from each OCR besides the *count of words* given in {"product description" presented by the manufacturer, reviewer expertise} [22]. For instance, a review for *Kaspersky Anti-virus* consists of 40 words [11]. *"Best product to protect your PC/ laptops. Total security helps to clean your PC from all viruses. Very genuine product. I am using it since last five years. Highly recommended product. Thanks Amazon to available this product at reasonable price"* Therefore, the simple Shannon Entropy of the review is given by Eq. (2):

$$E(156, 40) = -\left[\frac{40}{156+40} * log\left(\frac{40}{156+40}\right) + \frac{156}{156+40} * log\left(\frac{156}{156+40}\right)\right] \quad (2)$$

which we computed for each review available for a particular product. Next, we time-sorted the reviews for a given product beginning from the earliest post-date to the most recent date and computed the difference in entropy for each review similar to Eq. (2). Then, we normalized this incremental entropy to incorporate the semantic similarity of the n^{th} review and the previous $(n-1)$ reviews. In this manner, we could accurately compare the information gain (measured by *incremental entropy*) achieved after reading a new review with unique words while also considering its semantic similarity (measured by cosine distance) within the semantic space generated by a Latent Semantic Analysis (LSA) of the entire corpus of reviews for that particular product. The normalization step ensured that any randomly written review-text did not appear in

---

[11] Kaspersky Total Security - 1 User, 1 Year (CD): https://www.amazon.in/Kaspersky-Total-Security-Latest-Version/dp/B01AD36M8C/

our calculations as unique content. We present the descriptive statistics and predictors used to build our framework in Table 3 and Table 4.

3.3 Pairwise correlation and multicollinearity checks

We computed the pairwise correlations and the variance inflation factors (VIF) for the numerical predictors. Subsequently, we also verified whether these values stayed within permissible limits - pairwise correlation within ±0.5 and VIF-s within 10. The pairwise correlation fluctuated between 0.602 and -0.455, while the VIF fluctuated from 1.008 to 2.451, presented in Table 5.

Table 3. Descriptive statistics for the variables in our framework

		N	**Mean**	**Std. Dev.**	**Max**	**Min**
	Independent Variable					
	Readability	930	0.075	0.206	0.556	0.0151
Review Title	Superlatives	930	8.287	12.783	100.00	0.000
	Total Title Sentiment	930	0.068	0.413	1.818	-1.272
	Keyword Relevance	930	0.252	0.113	1.000	0.036
	Total Review Sentiment	930	0.107	0.310	50.000	-1.250
Review Text	Review Depth	930	0.379	0.302	9.680	0.010
	Review Divergence	930	-0.173	0.180	0.133	-1.259
	Semantic Entropy	930	0.089	0.124	0.270	-0.257
	Verified Purchase Badge	930	-	-	1 (805)#	0 (125)#
	User-provided Images	930	-	-	1 (492)#	0 (438)#
Product Details	Number of Reviews	930	4.143	1.258	7.824	0.693
	Deviation of Star Ratings	930	3.011	1.603	2.103	0.000
Review Age	Age of Review	930	3.827	0.318	7.750	0.693
	Reviewer Expertise	930	37.292	84.797	1025.000	1.000
Reviewer Attributes	Credibility	930	-	-	1 (39)#	0 (891)#
	Is Real Customer	930	-	-	1 (115)#	0 (815)#
	Dependent Variable					
	Helpfulness	930	5.000	16.388	277.000	1.000

Verified Purchase Badge, User-provided Images, Is Real Customer, Credibility are dummy variables, hence no Mean or Std. Dev.; #Numbers in parentheses indicate the counts

Table 4. Variables used to build our framework – brief descriptions and literature sources

S. No.	Variable	Brief Description	Literature Source
Independent Variable			
Review Title			
1	Readability	Text readability of the Review Title (FOG) (N)	[2, 22,42]
2	Superlatives	Adjectives used in the Review Title (N)	Developed from [5]
3	Total Title Sentiment	Positive and Negative sentiments of Review Title (N)	[45,53,69]
Review Text			
4	Keyword Relevance	Keyword Overlap Score based on TF-IDF (N)	Developed from [43]
5	Total Review Sentiment	Positive and Negative sentiments of Review Text (N)	[2, 45,53,57,69]
6	Review Depth	Entropy-based score for detailed topic discussion (N)	Developed from [66]
7	Review Divergence	Entropy-based score for coverage across related topics (N)	Developed from [66]
8	Semantic Entropy	Semantic similarity combined with Shannon's entropy (N)	Developed from [22,24]
9	Verified Purchase Badge	Whether the purchase had been verified by Amazon (D)	[24,30]
10	User-provided Images	Whether review contains a product image (D)	[21,38]
Product Details			
11	Number of Reviews	Total number of reviews the product has (N)	[66]
12	Deviation of Star Ratings	Difference between stars and average rating (N)	[26,38,66]
Review Age			
13	Age of Review	Difference between the dates *when data was collected* and *when a review was submitted* (N)	[48, 53]
Reviewer Attributes			
14	Reviewer Expertise	Total Helpful Votes received by the reviewer in the past (N)	[2,3,28]
15	Credibility	Whether Top reviewer (e.g. within Top 100) (D)	[67]
16	Is Real Customer	Whether the real name of a user or "Amazon Customer" (D)	[2,3,56]
Dependent Variable			
17	Helpfulness	Count of Helpful Votes received by a review (N)	[2, 5,22,24,46,66]

Note: **N**: Numeric variable; **D**: Dummy variable

Table 5. Pairwise correlation among the variables in our empirical framework

	VIF	[01]	[02]	[03]	[04]	[05]	[06]	[07]	[08]	[09]	[10]	[11]	[12]	[13]
Independent Variable														
Review Title														
Readability [01]	2.330	1.000												
Superlatives [02]	1.269	0.437	1.000											
Total Title Sentiment [03]	1.363	-0.008	0.009	1.000										
Review Text														
Keyword Relevance [04]	2.166	0.573	0.341	0.008	1.000									
Total Review Sentiment [05]	1.536	0.042	0.072	0.437	0.043	1.000								
Review Depth [06]	2.194	0.354	0.191	0.014	0.561	0.016	1.000							
Review Divergence [07]	1.848	0.070	-0.069	-0.016	0.123	-0.022	0.172	1.000						
Semantic Entropy [08]	2.451	0.104	0.022	-0.075	0.129	-0.051	0.423	0.602	1.000					
Product Details														
Number of reviews [09]	1.879	0.233	0.203	-0.341	0.123	0.111	-0.211	0.312	0.091	1.000				
Deviation of Star Ratings [10]	1.598	0.009	0.067	0.452	-0.031	0.545	-0.014	0.036	-0.040	-0.004	1.000			
Reviewer Attributes														
Age of Review [11]	1.709	-0.123	-0.455	0.381	0.165	-0.218	0.405	0.125	-0.394	-0.411	0.045	1.000		
Reviewer Expertise [12]	1.065	-0.047	-0.010	0.051	-0.123	0.080	-0.116	-0.148	-0.048	0.127	0.432	0.313	1.000	
Dependent Variable														
Helpfulness [13]	1.008	-0.070	-0.022	0.004	-0.155	0.034	-0.109	-0.106	0.008	0.007	0.238	0.019	0.287	1.000

N=930 observations; *Verified Purchase Badge*, *User-provided Images*, *Is Real Customer*, *Credibility* are dummy variables. Hence, no pairwise correlations were calculated for these variables.

4. Empirical Modelling

After extracting the predictors, we used count-data models to investigate the helpfulness of OCRs. The primary statistical models available for examining non-negative integer outcomes are Poisson and Negative Binomial distributions [45]. Now, with count data, the outcome of zero may be due to: (i) i**nflation -** when excess zeroes are present compared to the expected number based on count data distribution, (ii) t**runcation -** when systematically zeros are non-existent.

In recent studies, such aberration types in count response have been dealt with by using zero-inflated count regression models [34, 37, 67] or zero-truncated count regression models [49, 60]. In this study, we chose the zero-truncated form of the Poisson and Negative Binomial regression model to examine the effect of predictors extracted from the semantic content of OCRs. When OCRs receive zero helpful votes, it is difficult to gauge whether they have been (i) *unhelpful* or (ii) whether they have been read, understood, but simply not voted. Therefore, we removed the review records with zero helpful votes, leading to 930 final records for our modelling. The minimum outcome in a zero-truncated model is $Y = 1$. Thus, the zero-truncated count data probability distribution has the following general form:

$$P(Y = y) = \frac{f(Y=y)}{f(Y>0)} = \frac{f(Y=y)}{[1-f(Y=0)]} \quad \text{for} \quad y = 1,2,3,\ldots$$

The above functional form ensures that the zero-truncated distribution's probability mass function is normalized by dividing all probabilities for y greater than zero by $(1 - f(Y = 0))$. Therefore, a truncated Poisson distribution has the following probability mass function:

$$P(Y = y|Y > 0) = \frac{f(y)}{1-f(0)} \quad \text{where } f(y) = \frac{e^{-\lambda}\lambda^{y}}{y!} \text{ and } f(0) = e^{-\lambda}.$$

The parameter λ is parametrized using the predictor variables as $\mu = log(\lambda) = X'\beta$.

The above equation can also be written as; $\mu_i = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki})$

where X is the design matrix, and β is the vector of regression coefficients. Consistent with extant studies [14,15], we also decided to run robustness checks with higher thresholds (up to 2, 5, 10 votes) for the number of review helpfulness votes (see Table 9).

4.1 Moderation Analysis

Next, we conducted a post-hoc moderation analysis to assess whether the *age of review*, the *credibility of the reviewer* and *review divergence* moderated the main effects. First, according to the Dual Process Theory, *credibility of the reviewer* represents normative influence on the user to conform to the expectations of others (i.e. the subjective opinion of a reputed reviewer) [17,18,66]. Therefore, the reputation of a reviewer on Amazon can magnify the user's perceptions toward information in the review. Also, the readers on an e-commerce platform expect highly-ranked reviewers to write more in-depth reviews using more topic-relevant keywords than discuss more diverse product information across multiple reviews [66]. So, we expect stronger positive moderation effects of *credibility* with *review depth* and *keyword relevance* while weaker negative moderation effects for *review divergence* and *semantic entropy*.

Second, the *age of review* expresses the time difference between *when data was collected* and *when a review was submitted*. Therefore, by the Dual Process Theory [17,18,66], the *age of review* will enhance the informational influence toward processing the relevant review content, allow renewed information about a product, and reduce the user's uncertainty [66]. Thus, in case of more in-depth reviews, older ones will become less relevant than more recent ones. So, we expect stronger negative moderation effects of *age of review* with *review depth* and *keyword relevance*.

Third, the *review divergence* measures the variety of topical contents discussed in a review compared to other reviews available for the same product. Therefore, based on the Dual Process Theory [20,44,66], it represents normative influence. Also, readers prefer more in-depth reviews due to their stronger informational influence than those with more breadth. Therefore, we expect negative moderation effects of *review divergence* for *review depth* and positive moderation effects for *semantic entropy*.

5. Discussion of Results

5.1 Main Results

From the main effects reported in Table 6, we find that *review depth*, *semantic entropy* and *keyword relevance* have a significant positive influence on helpfulness votes. In contrast, *review divergence* has a significant negative effect on helpfulness votes. The robustness checks performed with multiple product types (Table 8) and threshold reviews (Table 9) also confirm that *review depth*, *semantic entropy* and *keyword relevance* help to make a review more helpful. In contrast, the *review divergence* hurts its helpfulness.

First, we found that the coefficients of *keyword relevance* are positive and significant. It measures the degree of similarity of keywords used in an online review to describe the different attributes of a product. We calculated the overlap score (to measure keyword relevance) using a weighted sum of TF-IDF across each review document based on the Information Retrieval Theory [43]. Therefore, when users intend to search for quick information on a product and its features, they look for keywords associated with the reviews. The Amazon platform also provides top keywords as a convenient way to filter online reviews, read them and therefore, allow users to mark them helpful (see Figure 1). Thus, the higher usage of relevant product-specific keywords within the review-text serve as peripheral cues to the reader to make it more helpful, drawing from the ELM Theory. Our findings coincide with past studies using corroboration keywords [24]; informativeness of reviews measured by the count of keywords [33]—at the same time, extending past studies that applied text-regression [47;48] and TF-IDF based measures [1] to examine the predictors to review helpfulness.

Second, we found that the coefficients of *review depth* are positive and significant in our results. The *review depth* measures informational cues of the multidimensional textual content of a review while describing the product, and it can increase information diagnosticity for the reader. Therefore, according to the Dual Process Theory, it represents informational influence. When users seek new information on a product and its features on online platforms, they look for detailed and more in-depth reviews, making them more helpful in purchasing decisions. Findings from our study are in-line with other studies that examined similar predictors, such as review length [51] and open-ended textual content of reviews [46] and Hong et al. [31]. Our findings also extend past studies that measured review depth by the number of words [15,27,58].

Third, we found that the coefficients of *review divergence* are negative and significant in our results. According to the Dual Process Theory, *review divergence* is a normative predictor for helpfulness. It represents the consistency of a review's textual content in comparison to the textual content of other reviews for the same product. We computed the divergence for a given review with the help of the KL-divergence metric, which compared the review with a representative review (generated from the best set of topics for the product). Therefore, our results convey that readers do not find such a review helpful. Findings from our study are in line with valence consistency [51]; content deviation [66], while it extends past related studies that examined consistency, such as rating deviation [33], rating inconsistency [1] and the number of attributes in a product [70].

Fourth, we found that the coefficients of *semantic entropy* are positive and significant in our results. Our measurement of semantic entropy was based on a combination of incremental entropy and normalization with semantic similarity. It also ensured that any non-overlapping review-text did not appear as unique content. According to Shannon's Entropy Theory, an online review text which consists of more unique words and is written with an inimitable linguistic style than the previous reviews may convey additional information to the reader. Thus, such a review possesses a higher information entropy, and users find it more helpful. Findings from our study are consistent with past studies that explored incremental entropy [23] and semantic characteristics of reviews [47]. At the same time, it extends Fresneda and Gefen [24], who had examined unique corroboration entropy and recommendation entropy for helpful reviews.

5.2 Moderation Results

First, we examined the interaction effects of the *reviewer's credibility*. It strongly amplifies the overall positive impact of *review depth* and weakens the negative effect of *review divergence* on the review helpfulness. Thus, readers on an e-commerce platform expect highly-ranked reviewers to write more in-depth reviews than discuss more diversified product information. Also, in-depth reviews do not appear much helpful to the readers when the reviewer is less respectable. Next, *keyword relevance* shows a strong positive interaction effect, indicating that readers expect more meaningful and product-related keywords from an expert reviewer. However, semantic entropy shows a weak negative effect, indicating that reviews with high entropy and semantic uniqueness are less likely to be helpful when the reviewer is less reputed and might consider them one-off.

Second, we examined the interaction effects of *age of review*. There are strong negative moderation effects for *review depth* and *keyword relevance* on the review helpfulness. As the time elapsed between the *date of the review* and the *date of data collection* increases, older but more in-depth reviews became less relevant than more recent reviews with a detailed, in-depth discussion about the product. So, those in-depth reviews are not as helpful to readers. Similarly, the relevance of keywords in the review content diminishes as time passes. Therefore, older reviews with more enriched discussions and relevant keywords within the textual content appeared less helpful when newer reviews for the same product were available on the e-commerce platform.

Third, we examined the interaction effects of *review divergence*. It strongly amplifies the overall positive impact of *review depth* and weakens the positive effect of *semantic entropy* on the review helpfulness. Therefore, “review depth” remaining more-or-less constant, reviews that have more breadth and discuss lots of topics might appear less helpful than reviews discussing fewer topics. This result reinstates our finding that readers prefer more in-depth reviews due to their stronger informational influence than those with more breadth, which exert a normative influence. Next, semantic entropy shows a negative effect, indicating that reviews with high information entropy and semantic style are less likely to be helpful to the readers when the review has less breadth and involves few discussion topics while describing the product.

Table 6. Explanatory models for "count of helpful votes"

	Dependent variable: *Count of helpful votes*				
	M1	**M2**	**M3**	**M4**	**M5**
Review Title					
Readability	0.013**	0.016 ***	0.066***	0.011***	0.027*
	(0.004)	(0.003)	(0.014)	(0.003)	(0.013)
Superlatives	0.001	0.005*	0.012	0.003*	0.006
	(0.003)	(0.002)	(0.006)	(0.002)	(0.007)
Total Title Sentiment	0.059	0.021	0.246*	0.067	0.612*
	(0.019)	(0.060)	(0.193)	(0.057)	(0.015)
Review Text					
Keyword Relevance	0.620**			0.514***	2.907***
	(0.208)			(0.171)	(0.450)
Total Review Sentiment	0.151*	0.171 ***	0.510**	0.261***	0.612***
	(0.122)	(0.083)	(0.066)	(0.080)	(0.088)
Review Depth	0.206*			0.426***	0.705***
	(0.145)			(0.105)	(0.110)
Review Divergence	-1.201***			-1.340***	-2.109***
	(0.220)			(0.149)	(0.560)
Semantic Entropy	0.464*			0.699***	1.922***
	(0.213)			(0.151)	(0.236)
Verified Purchase Badge	0.081*	0.125 **	0.624**	0.106*	0.709**
	(0.034)	(0.069)	(0.238)	(0.070)	(0.234)
User-provided Images	0.124***	0.561	1.175***	0.344***	1.467***
	(0.085)	(0.058)	(0.224)	(0.034)	(0.034)
Product Details					
Number of Reviews	-0.004**	-0.002**	-0.002**	-0.003**	-0.003**
	(0.005)	(0.022)	(0.034)	(0.034)	(0.034)
Deviation of Star Ratings	-0.075*	-0.137 ***	-0.256***	-0.113***	-0.232***
	(0.023)	(0.017)	(0.051)	(0.016)	(0.053)
Review Age					
Age of Review	0.450***	0.466 ***	1.033***	0.565***	2.011***
	(0.054)	(0.041)	(0.134)	(0.043)	(0.109)
Reviewer Attributes					
Credibility	0.0004***	0.0001***	0.014***	0.001***	0.005***
	(0.00003)	(0.00004)	(0.002)	(0.00006)	(0.001)
Is Real Customer	0.078	0.230**	0.233	0.193**	0.044
	(0.090)	(0.069)	(0.215)	(0.069)	(0.214)
Reviewer Expertise	0.309*	0.300**	0.601*	0.488**	0.817*
	(0.152)	(0.102)	(0.395)	(0.114)	(0.425)
Intercept	2.842***	2.570***	6.621***	3.540***	9.010***
	(0.450)	(0.281)	(0.333)	(0.366)	(0.522)
Observations	930	930	930	930	930
Adj. R2 / Pseudo R2	0.230	0.251	0.273	0.304	0.378
Log-Likelihood	-	-2840.51	-1720.58	-2681.39	-1681.27
AIC	-	5705.02	3467.16	5394.78	3396.50

*Note: *p<0.05; **p<0.01;***p<0.001; Robust Standard errors in parenthesis;*
M1=OLS with all variables; M2= Truncated Poisson; M3= Truncated Negative Binomial; M4= Truncated Poisson with all variables; M5= Truncated Negative Binomial with all variables

Table 7. Moderation Effects of "Credibility", "Age of Review" and "Review Divergence"

	Dependent variable: *Count of helpful votes*		
	M1	**M2**	**M3**
Keyword Relevance	3.530***	6.081*	
	(0.437)	(2.183)	
Review Depth	0.639**	0.794**	1.885***
	(0.308)	(0.253)	(0.413)
Review Divergence	-2.186***		-1.663
	(0.484)		(1.042)
Semantic Entropy	1.431***		1.036*
	(0.433)		(0.524)
Credibility	5.486*		
	(1.720)		
Age of Review		0.106	
		(0.839)	
Interaction Effects			
Credibility * Keyword Relevance	24.179**		
	(1.833)		
Credibility * Review Depth	12.756**		
	(2.179)		
Credibility * Review Divergence	-6.820**		
	(1.024)		
Credibility * Semantic Entropy	-3.083		
	(2.628)		
Age of Review * Keyword Relevance		-1.939*	
		(0.135)	
Age of Review * Review Depth		-0.299*	
		(0.078)	
Review Divergence * Review Depth			-8.859**
			(1.573)
Review Divergence * Semantic Entropy			0.294*
			(0.087)
Intercept	3.040***	3.379***	2.446***
	(0.383)	(0.388)	(0.473)
Observations	930	930	930
Adj. R2 / Pseudo R2	0.065	0.074	0.063
Log-Likelihood	-1800.32	-1783.26	-1803.96
AIC	3622.64	3588.52	3625.92

*Note: *p<0.05; **p<0.01;***p<0.001; Robust Standard errors in parenthesis; M1=Interaction with* ***Credibility****; M2=Interaction with* ***Age of Review****; M3=Interaction with* ***Review Divergence;*** *All models were run with Truncated Negative Binomial Regression*

Table 8. Robustness Checks for "count of helpful votes" across product-types

	Dependent variable: *Count of helpful votes*			
	TNB-M1	**TNB-M2**	**TNB-M3**	**TNB-M4**
Review Title				
Readability	0.033*	0.017	0.036*	0.022**
	(0.016)	(0.019)	(0.020)	(0.016)
Superlatives	0.003	0.032	0.003	0.033*
	(0.008)	(0.016)	(0.008)	(0.016)
Total Title Sentiment	0.309	0.063	0.114	0.246
	(0.138)	(0.025)	(0.080)	(0.043)
Review Text				
Keyword Relevance	2.007**	4.635**	2.171**	4.196***
	(0.261)	(0.974)	(0.183)	(0.296)
Total Review Sentiment	0.344	0.454	0.484	0.190
	(0.211)	(0.171)	(0.129)	(0.052)
Review Depth	0.191	0.123	0.178*	0.472
	(0.056)	(0.108)	(0.045)	(0.134)
Review Divergence	-1.122**	-1.958	-1.025**	-2.205**
	(0.569)	(0.887)	(0.255)	(0.502)
Semantic Entropy	1.243*	1.810	2.193*	1.662*
	(0.566)	(0.566)	(1.042)	(0.374)
Verified Purchase Badge	0.748***	5.801**	1.102*	0.505*
	(0.226)	(1.376)	(0.148)	(0.148)
User-provided Images	0.993***	1.555***	1.549***	0.843**
	(0.257)	(0.349)	(0.328)	(0.264)
Product Details				
Number of Reviews	-0.015*	-0.045**	-0.028	-0.019
	(0.011)	(0.011)	(0.012)	(0.015)
Deviation of Star Ratings	-0.124*	-0.281***	-0.231***	-0.154***
	(0.071)	(0.074)	(0.068)	(0.071)
Review Age				
Age of Review	0.902***	1.281***	1.009***	0.875***
	(0.164)	(0.307)	(0.260)	(0.175)
Reviewer Attributes				
Credibility	0.002*	0.021***	0.018*	0.002**
	(0.001)	(0.006)	(0.001)	(0.006)
Is Real Customer	0.275	0.112	0.125	0.206
	(0.138)	(0.004)	(0.078)	(0.085)
Reviewer Expertise	1.532*	0.041	1.255*	0.246
	(0.518)	(0.047)	(0.735)	(0.139)
Intercept	6.292***	8.768***	9.556***	5.979***
	(1.123)	(2.320)	(2.123)	(1.529)
Observations	467	463	421	509
Pseudo R2	0.221	0.277	0.216	0.262
Log-Likelihood	-825.965	-799.393	-745.623	-885.453
AIC	1685.93	1634.786	1525.246	1802.906

*Note: *p<0.05; **p<0.01;***p<0.001; Robust Standard errors in parenthesis; TNB=Truncated Negative Binomial; M1=Search; M2=Experience; M3=Tangible; M4=Intangible*

Table 9. Robustness Checks for "count of helpful votes" across various threshold levels

	Dependent variable: *Count of helpful votes*			
	TNB-M5	**TNB-M6**	**TNB-M7**	**TNB-M8**
Review Title				
Readability	0.014*	0.016*	0.007	0.004**
	(0.011)	(0.007)	(0.009)	(0.010)
Superlatives	0.001	0.004	0.005	0.007
	(0.006)	(0.005)	(0.006)	(0.016)
Total Title Sentiment	0.324	0.179	0.103	0.245
	(0.212)	(0.124)	(0.075)	(0.264)
Review Text				
Keyword Relevance	0.550**	0.555**	1.191**	1.705*
	(0.208)	(0.353)	(0.436)	(0.554)
Total Review Sentiment	0.336	0.076	0.067	0.352
	(0.295)	(0.185)	(0.141)	(0.132)
Review Depth	0.403*	0.021	0.444**	0.436**
	(0.263)	(0.069)	(0.284)	(0.127)
Review Divergence	-0.912**	-1.330***	-1.408**	-0.715*
	(0.220)	(0.339)	(0.454)	(0.307)
Semantic Entropy	0.213	0.342	0.421	1.127***
	(0.130)	(0.258)	(0.133)	(0.373)
Verified Purchase Badge	0.224	0.190	0.163	0.749*
	(0.034)	(0.046)	(0.087)	(0.144)
User-provided Images	0.587**	0.685***	1.100***	0.471**
	(0.221)	(0.128)	(0.170)	(0.099)
Product Details				
Number of Reviews	-0.016**	-0.020**	-0.024**	-0.045**
	(0.015)	(0.021)	(0.028)	(0.035)
Deviation of Star Ratings	-0.075*	-0.071*	-0.101*	-0.049*
	(0.023)	(0.036)	(0.046)	(0.062)
Review Age				
Age of Review	0.836***	0.852***	0.950***	0.294*
	(0.169)	(0.104)	(0.120)	(0.190)
Reviewer Attributes				
Credibility	0.001*	0.0002***	0.001***	0.0003*
	(0.0005)	(0.00006)	(0.0004)	(0.0001)
Is Real Customer	0.345	0.143	0.322	0.157
	(0.257)	(0.137)	(0.184)	(0.062)
Reviewer Expertise	0.291	0.426*	0.556	0.256*
	(0.152)	(0.231)	(0.324)	(0.132)
Intercept	6.463 ***	5.930***	6.615***	2.847***
	(0.421)	(0.872)	(1.067)	(1.528)
Observations	638	792	861	119
Pseudo R2	0.285	0.298	0.274	0.226
Log-Likelihood	-371.065	-853.213	-1189.590	-477.662
AIC	776.130	1740.426	2413.180	989.324

*Note: *p<0.05; **p<0.01;***p<0.001; Robust Standard errors in parenthesis; TNB=Truncated Negative Binomial; M5: up to 2 votes; M6: up to 5 votes; M7: up to 10 votes; M8:greater than 10 votes*

6. Implications of our study

6.1 Theoretical Implications

Our study offers two theoretical insights as follows. First, it contributes to the existing studies on review helpfulness that have applied Entropy Theory to examine online product reviews and their perceived helpfulness [18,23,24,58,66]. Using the theoretical lenses of Shannon's Entropy to build predictors suggests that not only the number of words [58] but also, how unique those words are, and whether they can provide new information to the reader, will improve the helpfulness of the review during decision-making. Additionally, this study proposed the "semantic entropy" predictor, which examined semantic uniqueness and the "keyword similarity", which reviewed product and feature-relevant keywords in online reviews. These two predictors allowed a balance of "unique" yet "thematically related" text consisting of "relevant keywords" within the reviews. Therefore, given the vast number of online reviews available on e-commerce platforms, the application of entropy-based predictors will improve the accuracy of the helpfulness voting scheme and be more useful for potential customers.

Second, our study contributes to the existing studies on review helpfulness that have applied Dual Process Theory [15,18,20,37,44,66]. Dual Process Theory tells that the *review depth* represents *informational influence* while the *review divergence* represents *normative influence*. Thus, online reviews that discuss a diverse array of features are less helpful than those consisting of in-depth discussion of fewer features, signifying the stronger effect of *informational influence* over *normative influence*. Also, the Dual Process Theory guides us in explaining the findings from the interaction effects in this study. Therefore, among reviews with similar *review depth*, those talking about multiple topics appeared less helpful than those talking about fewer topics. The interaction-effects of the *credibility* of the reviewer follow next. Product reviews written by highly-ranked reviewers but focussed on fewer features emerged more helpful to readers than those with varied product information. Also, the reviews written by low-ranked reviewers were less helpful among those with similar *review depth*.

6.2 Managerial Implications

Our study offers two important managerial insights as follows. First, the findings from our study suggest that the predictors "review depth" and "review divergence" have a differential effect on the perceived helpfulness of the reviews. While readers prefer online reviews that engage in a more in-depth discussion of a particular product feature, they may not always like reviews that discuss multiple features. Thus, e-commerce platforms may think of an improved way of engaging with the customers when they submit online reviews after a successful purchase. Currently, Amazon allows previous customers to write mostly open-ended textual statements about their purchased products. Such reviews are often written without focusing on a particular topic(s) and may not help future users in their purchase decision-making. Instead, e-commerce platforms can fine-tune this existing process by guiding customers – first, create a finite list of topics unique to each product category, build focused questions around those topics, and then ask customers to answer them. Recently, a few review management platforms such as Gominga[12] and Bazaarvoice[13] are using topic-based review selection and feedback management for Amazon and Costco.

Second, our study proposes the "semantic entropy" predictor, which can be an automated measure of helpfulness for online reviews in e-commerce platforms. This measure was built using the incremental information gained from multiple online reviews and also accounts for the unique semantic style of the textual content relevant to each product category. Currently, readers rate online reviews as helpful and require human intervention. Further, some reviews might not even contain meaningful product information or talk about its features. However, incorporating "semantic entropy" as a component of helpfulness scores can remove some of the subjective judgement involved in the existing manual process, avoid repetitive reviews or irrelevant product information by highlighting only relevant textual content. And e-commerce platforms can also

---

[12] Gominga, the review company: https://gominga.com/review-manager/
[13] Bazarvoice: https://www.bazaarvoice.com/products/

allow some screening time for each review before publishing it for all readers. In this manner, online reviews can become more useful and attractive to potential buyers.

7. Conclusion and Future Scope of Research

Our study identified several interesting insights, including NLP and text mining-based predictors that contribute to the popularity of successful online reviews on e-commerce platforms. In particular, we identified *keyword relevance*, *review depth*, *review divergence* and *semantic entropy* as important predictors for online reviews. Therefore, given the huge number of online reviews available for every product on Amazon, proposing an entropy-based predictor to measure the review helpfulness will be more useful for potential customers, - (i) allowing them to read information-rich reviews quicker, and (ii) reduce the potential challenges occurring from human intervention in the helpfulness voting mechanism.

Our study has a few limitations. First, few studies have examined the effect of user comments published in response to online reviews and its subsequent influence on helpfulness voting. Therefore, future research may build predictors based on these comments and examine their effects on helpfulness. Second, a few studies have employed mixed-method analysis of review helpfulness. Therefore, future research could examine review helpfulness and its predictors applying NLP-based text-mining in combination with interviews of successful customers.

Acknowledgements

The authors thank the anonymous referees for their constructive suggestions that significantly improved this manuscript. The authors also thank the former Editor-in-Chief of the journal, Professor James R. Marsden, for his continued guidance and support throughout the review process.

References

1. Aghakhani, N., Oh, O., Gregg, D. G., & Karimi, J. (2020). Online Review Consistency Matters: An Elaboration Likelihood Model Perspective. *Information Systems Frontiers*, 1-15.
2. Agnihotri, A., & Bhattacharya, S. (2016). Online review helpfulness: Role of qualitative factors. *Psychology & Marketing*, *33*(11), 1006-1017.
3. Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of online consumer reviews: Readers' objectives and review cues. *International Journal of Electronic Commerce*, *17*(2), 99-126.
4. Belzer, J. (1973). Information theory as a measure of information content. *Journal of the American Society for Information Science*, *24*(4), 300-304.
5. Biswas, B., Sengupta, P., & Chatterjee, D. (2020). Examining the determinants of the count of customer reviews in peer-to-peer home-sharing platforms using clustering and count regression techniques. *Decision Support Systems*, *135*, 113324.
6. Bischof, J., & Airoldi, E. M. (2012). Summarizing topical content with word frequency and exclusivity. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)* (pp. 201-208).
7. Biswas, B., Sengupta, P., & Ganguly, B. (2021). Your reviews or mine? Exploring the determinants of "perceived helpfulness" of online reviews: a cross-cultural study. *Electronic Markets*, 1-20.
8. Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of Science. *The Annals of Applied Statistics*, *1*(1), 17-35.
9. Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77-84.
10. Cheung, C. M. Y., Sia, C. L., & Kuan, K. K. (2012). Is this review believable? A study of factors affecting the credibility of online consumer reviews from an ELM perspective. *Journal of the Association for Information Systems*, *13*(8), 2.
11. Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, *54*(3), 477-491.
12. Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(3), 345-354.
13. Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science*, *29*(5), 944-957.
14. Choi, H. S., & Leon, S. (2020). An empirical investigation of online review helpfulness: A big data perspective. *Decision Support Systems*, *139*, 113403.
15. Chua, A. Y., & Banerjee, S. (2015). Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. *Journal of the Association for Information Science and Technology*, *66*(2), 354-362.
16. Chua, A. Y., & Banerjee, S. (2017). Analyzing review efficacy on Amazon. com: Does the rich grow richer? *Computers in Human Behavior*, *75*, 501-509.
17. Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal And Social Psychology*, *51*(3), 629.

18. Du, J., Rong, J., Wang, H., & Zhang, Y. (2021). Neighbor-aware review helpfulness prediction. *Decision Support Systems*, *148*, 113581.

19. Filieri, R. (2015). What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. *Journal of Business Research, 68*(6), 1261-1270.

20. Filieri, R., McLeay, F., Tsui, B., & Lin, Z. (2018). Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services. *Information & Management*, *55*(8), 956-970.

21. Filieri, R., Raguseo, E., & Vitari, C. (2018). When are extreme ratings more helpful? Empirical evidence on the moderating effects of review characteristics and product type. *Computers in Human Behavior*, *88*, 134-142.

22. Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., & Freling, T. (2014). How online product reviews affect retail sales: A meta-analysis. *Journal of Retailing*, *90*(2), 217-232.

23. Fresneda, J. E., and Gefen, D. (2019). A semantic measure of online review helpfulness and the importance of message entropy. *Decision Support Systems*, *125*, 113117.

24. Fresneda, J. E., & Gefen, D. (2020). Gazing at the stars is not enough, look at the specific word entropy, too!.*Information & Management*, *57*(8), 103388.

25. Hernández-Ortega, B. (2018). Don't believe strangers: Online consumer reviews and the role of social psychological distance. *Information & Management*, *55*(1), pp. 31-50.

26. Ghasemaghaei, M., Eslami, S. P., Deal, K., & Hassanein, K. (2018). Reviews' length and sentiment as correlates of online reviews' ratings. *Internet Research*.

27. Ghose, A., & Ipeirotis, P. G. (2010). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, *23*(10), 1498-1512.

28. Guo, B., & Zhou, S. (2017). What makes population perception of review helpfulness: an information processing perspective. *Electronic Commerce Research*, *17*(4), 585-608.

29. Hausser, J., & Strimmer, K. (2009). Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. *Journal of Machine Learning Research*, *10*(7).

30. He, J., Wang, X., Vandenbosch, M. B., & Nault, B. R. (2020). Revealed preference in online reviews: Purchase verification in the tablet market. *Decision Support Systems*, *132*, 113281.

31. Hong, H., Xu, D., Wang, G. A., & Fan, W. (2017). Understanding the determinants of online review helpfulness: A meta-analytic investigation. *Decision Support Systems, 102*, 1-11.

32. Karimi, S., & Wang, F. (2017). Online review helpfulness: Impact of reviewer profile image. *Decision Support Systems*, *96*, 39-48.

33. Kaushik, K., Mishra, R., Rana, N. P., & Dwivedi, Y. K. (2018). Exploring reviews and review sequences on e-commerce platform: A study of helpful reviews on Amazon. in. *Journal of retailing and Consumer Services*, *45*, 21-32.

34. Kuan, K. K., Hui, K. L., Prasarnphanich, P., & Lai, H. Y. (2015). What makes a review voted? An empirical investigation of review voting in online review systems. *Journal of the Association for Information Systems*, *16*(1), 1.

35. Kullback, S., and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), pp. 79-86.

36. Lee, I. (2018). Usefulness, funniness, and coolness votes of viewers: An analysis of social shoppers' online reviews. *Industrial Management & Data Systems*.

37. Lee, J., & Hong, I. B. (2019). Consumer's electronic word-of-mouth adoption: the trust transfer perspective. *International Journal of Electronic Commerce*, *23*(4), 595-627.

38. Lee, S., Lee, S., & Baek, H. (2021). Does the dispersion of online review ratings affect review helpfulness?. *Computers in Human Behavior*, *117*, 106670.

39. Lee, J., & Park, C. (2022). The effects of corporate, review and reviewer characteristics on the helpfulness of online reviews: the moderating role of culture. *Internet Research*.

40. Li, X., Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information & Management*, *56*(2), 172-184.

41. Li, K., Chen, Y., & Zhang, L. (2020). Exploring the influence of online reviews and motivating factors on sales: A meta-analytic study and the moderating role of product category. *Journal of Retailing and Consumer Services*, *55*, 102107.

42. Malik, M. S. I., & Hussain, A. (2017). Helpfulness of product reviews as a function of discrete positive and negative emotions. *Computers in Human Behavior*, *73*, 290-302.

43. Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval, 1(1), 496, Cambridge University Press. Cambridge.

44. Meek, S., Wilk, V., & Lambert, C. (2021). A big data exploration of the informational and normative influences on the helpfulness of online restaurant reviews. *Journal of Business Research*, *125*, 354-367.

45. Mousavizadeh, M., Koohikamali, M., Salehan, M., & Kim, D. J. (2020). An Investigation of Peripheral and Central Cues of Online Customer Review Voting and Helpfulness through the Lens of Elaboration Likelihood Model. *Information Systems Frontiers*, 1-21.

46. Mudambi, S. M., & Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly*, 34(1). pp. 185-200.

47. Ngo-Ye, T. L., & Sinha, A. P. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, *61*, 47-58.

48. Ngo-Ye, T. L., Sinha, A. P., & Sen, A. (2017). Predicting the helpfulness of online reviews using a scripts-enriched text regression model. *Expert Systems with Applications*, *71*, 98-110.

49. Pennebaker, J. W., Boyd, R. L., Jordan, K., Blackburn, K. (2015). The development and psychometric properties of LIWC2015, Tech. Rep., University of Texas at Austin.

50. Petty, R. E., & Cacioppo, J. T. (1986). "The elaboration likelihood model of persuasion". Advances in Experimental Social Psychology. London, England: Elsevier. 19: 124–129.

51. Quaschning, S., Pandelaere, M., & Vermeir, I. (2015). When consistency matters: The effect of valence consistency on review helpfulness. *Journal of Computer-Mediated Communication*, *20*(2), 136-152.

52. Rousseau, R. (1994). The number of authors per article in library and information science can often be described by a simple probability distribution. *Journal of Documentation*.

53. Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, *81*, 30-40.

54. Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379-423.

55. Short, J., Williams, E., and Christie, B. (1976). *The Social Psychology of Telecommunications*.

56. Siering, M., Muntermann, J., & Rajagopalan, B. (2018). Explaining and predicting online review helpfulness: The role of content and reviewer-related signals. *Decision Support Systems*, *108*, 1-12.

57. Siering, M. (2021). Leveraging online review platforms to support public policy: Predicting restaurant health violations based on online reviews. *Decision Support Systems*, *143*, 113474.

58. Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Roy, P. K. (2017). Predicting the "helpfulness" of online consumer reviews. *Journal of Business Research*, *70*, 346-355.

59. Taddy, M. (2013). Multinomial inverse regression for text analysis. *Journal of the American Statistical Association*, *108*(503), 755-770.

60. Topaloglu, O., & Dass, M. (2019). The impact of online review content and linguistic style matching on new product sales: The moderating role of review helpfulness. *Decision Sciences*.

61. Wan, Y., & Nakayama, M. (2014). The reliability of online review helpfulness. *Journal of Electronic Commerce Research*, *15*(3), 179.

62. Wang, F., & Karimi, S. (2019). This product works well (for me): The impact of first-person singular pronouns on online review helpfulness. *Journal of Business Research*, *104*, 283-294.

63. Wang, J., Yang, Z., & Brocato, E. D. (2018). An investigation into the antecedents of prepurchase online search. *Information & Management*, *55*(3), 285-293.

64. Weathers, D., Swain, S. D., & Grover, V. (2015). Can online product reviews be more helpful? Examining characteristics of information content by product type. *Decision Support Systems*.

65. Weiss, A. M., Lurie, N. H., & MacInnis, D. J. (2008). Listening to strangers: whose responses are valuable, how valuable are they, and why?. *Journal of Marketing Research*, *45*(4), 425-436.

66. Wu, C., Mai, F., and Li, X. (2021). The effect of content depth and deviation on online review helpfulness: Evidence from double-hurdle model. *Information & Management*, *58*(2), 103408.

67. Xu, Q. (2014). Should I trust him? The effects of reviewer profile characteristics on eWOM credibility. *Computers in Human Behavior*, *33*, 136-144.

68. Yang, S., Zhou, C., & Chen, Y. (2021). Do topic consistency and linguistic style similarity affect online review helpfulness? An elaboration likelihood model perspective. *Information Processing & Management*, *58*(3), 102521.

69. Yang, S., Yao, J., & Qazi, A. (2020). Does the review deserve more helpfulness when its title resembles the content? Locating helpful reviews by text mining. *Information Processing & Management*, *57*(2), 102179.

70. Yi, J., & Oh, Y. K. (2022). The informational value of multi-attribute online consumer reviews: A text mining approach. *Journal of Retailing and Consumer Services*, *65*, 102519.

71. Zhu, L., Yin, G., & He, W. (2014). Is this opinion leader's review useful? Peripheral cues for online review helpfulness. *Journal of Electronic Commerce Research*, *15*(4), 267.