Erasmus
School of
Economics

ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

MASTER THESIS DATA SCIENCE AND MARKETING ANALYTICS

A study of online review helpfulness

MIHAELA CRACIUN | 479483

SUPERVISOR: Michiel Van Crombrugge

SECOND ASSESSSOR: Anastasija Tetereva

Word count: 14488 | Datum: 28-10-2021

Table of contents

Introduction

The ever-increasing number of Internet users created the opportunity for online customers to share their product consumption experiences online. This way, the basic word-of-mouth (WOM) product discussions have now moved to online markets, where these electronic communities offer the customer both the chance to easily share impressions and access different types of product reviews (Ward & Ostrom, 2003). Moreover, besides providing a vast amount of helpful information to customers about products and services, these social structures are also perceived as the most trusted sources of valuable recommendations (Park et al., 2007), after the advice given by friends and family (Brown, 1987).

The emergence of electronic word-of-mouth (e-WOM) and the increasing number of customers both posting and seeking product evaluations on the Internet has considerably influenced customers' purchasing behavior worldwide (Grewal & Levy 2009). Product reviews have become a necessary tool for both customers and retailers looking to attract and retain buyers. Since in the physical shop, customers can test and analyze the product just before making the purchase decision, in the online market, the product value assessment is more restricted. Therefore, online buyers use other sources of information such as e-WOM, and more specifically, peer-generated product reviews, which are preferred over the vendor offered information (Bickart & Schindler 2001). One can argue that reviewer evaluations are subjective. However, most of the time, these are more credible than any other traditional source of information (Bickart & Schindler 2001). This and other prior findings indicate that peer-generated product reviews can have a significant effect on product popularity and sales (Duan et al.,2008; Dellarocas et al., 2007; Chevalier & Mayzlin,2006; Forman et al.,2008; Ghose & Ipeirotis, 2010; Liu, 2006; Archak et al., 2011; Godes & Mayzlin, 2004). Important to highlight, however, is that online product reviews could serve as either an opportunity or a threat for the business. Since these represent a valuable means for tracking consumer attitudes towards products in real time, an immediate adjustment of company strategies or product is necessary. Otherwise, the accumulation of negative product reviews could negatively affect company sales (Lee et al., 2008). This explains why both the retailer and the seller pay significant attention to different aspects of product reviews (Chevalier & Mayzlin, 2006; Anderson & Magruder, 2012; Hu et al., 2009; Li & Hitt, 2010; Chen & Xie, 2008).

On the other side of the coin, the exponentially increasing number of product reviews makes it hard for the customer to extract the most valuable insights and evaluate the underlying quality of the product. Therefore, the crucial focus of online retailers has switched from the mere presentation of reviews to the strategic use of online peer-generated content that offers greater value to the customer (Mudambi & Schuff, 2010). This indicates that online retailers are on the mission to show and arrange the most persuasive reviews, which, according to Jiang and Benbasat (2007), considerably improves the client's online buying experience and boosts company profits. For instance, Amazon.com tested different features in the general presentation of reviews between 2007 and 2010 (Pan & Zhang, 2011). Initially, the most recent reviews were presented first. Then product reviews were presented according to their valence, followed by Amazon's ranking of review importance. Finally, those were shown based on their "peer reviewing" ranking established on the given "helpful" votes of other buyers. Since then, Amazon.com still asks their customers: "Was this review helpful to you?" and positions reviews with the highest number of helpful votes regardless of their valence on the product's information page.

According to Dabholkar (2006), customers value helpful online reviews, which help them go through numerous available peer-generated product reviews more efficiently and make a more informed and thus better product choice. This most used method of presenting customer reviews diminishes customer purchase uncertainty, perceived risk, and information searching costs (Zhu & Zhang, 2010). This way, online retail websites with more helpful reviews provide greater value to the customers, leading to increased sales (Chen et al. 2008; Chevalier & Mayzlin 2006; Clemons et al. 2006; Ghose & Ipeirotis, 2006). However, according to Ghose and Ipeirotis (2010), the commonly used peer-reviewing feature is not that helpful for ranking recent reviews because helpful votes are accumulated in a relatively long period. Therefore, given the strategic potential of helpful reviews and the main drawback of the most used feature for review helpfulness assessment, it is of high importance to understand what aspects of peer-generated product reviews make them helpful. For instance, a more extended product review might make customers perceive it as more valuable than a shorter one because it might contain a more detailed product and experience description (Mundambi & Schuff, 2010). But, because of the law of diminishing marginal effect, an ever-increasing number of words might not have a constant positive impact on perceived review helpfulness. Huang et al. (2015) identified a significant effect of word count of a review until it reaches the average word count threshold. Thus, these findings might indicate that there is a non-linear relationship but to claim this there is a need for further investigation, which may solve the retailers' dilemma regarding the shape of the relationship.

Mudambi and Schuff (2010) focused more on analyzing review characteristics and found that review extremity, depth, and product type significantly affect perceived online review helpfulness. Similarly, in their investigation, Phan and Zhang (2011) identified that product type moderates the impact of review valance and length on review helpfulness. However, according to Korfiatis et al. (2012), style and, more specifically, stylistic elements of a review seem to significantly influence the degree of review helpfulness compared to other review characteristics. Besides this, Sussman and Siegal (2003), Liu and Park (2015) identified that reviewer identity disclosure and expertise significantly impact review perceived helpfulness.

Even if these research streams serve as building blocks in understanding multiple aspects of product reviews, they overlook some interesting avenues worth investigating. The studies mentioned above researched some of the most prominent direct relationships, but those were conducted in isolation, and somewhat fragmented conclusions have been drawn. Moreover, for some review aspects, previously conducted research does not reach a consensus (i.e., review valence). Thus, there is a call for further investigation to uncover the genuine relationships between review and reviewer characteristics and their perceived helpfulness. As a response, this research paper aims to document the impact of some essential reviews, reviewer characteristics, and product types on perceived review helpfulness and their interaction effect with other components that could explain the previously documented ambiguous or even spurious relationships.

Mudambi and Schuff (2010), Pan and Zhang (2011) found that word count positively impacts perceived online review helpfulness. But what they did not consider is the negative impact of information overload on customer information processing, which indicates that the positive relationship between review word count and helpfulness is linear to a certain extent. Thus, our academic contribution is to document the shape of the relationship between word count and perceived helpfulness. This might seem like what Huang et al. (2015) did, but they just investigated whether word count variable stays significant after the imposed threshold. However, we aim to refute the previous findings regarding the linear shape of the relationship. Besides this, Korfiatis et al. (2012) found readability as the qualitative measure of a review to have a significant impact on review helpfulness. However, there is no investigation conducted on the moderating effect that readability might have on the relationship between word count and review helpfulness. We expect that readability significantly diminishes the impact of review word count on review helpfulness. Another significant academic contribution that might prove previously identified spurious direct relationship between reviewer expertise and review helpfulness is the investigation of the moderating effect word count and readability might have on the relationship.

Moreover, since a consensus has not been reached regarding the effect of valence on perceived helpfulness (Zou et al., 2011), it would be a valuable academic contribution to investigate the moderating effect word count and readability might have on the relationship between valance and review helpfulness. Similarly, the investigation of multiple other moderating effects might explain somewhat mixed conclusions that have been previously drawn. Thus, to uncover the impact of specific qualitative and quantitative aspects of product reviews on perceived helpfulness the following research question is formulated:

What review characteristics moderate the effect of other review aspects on its perceived review helpfulness?

For a complete picture, our analysis focuses on investigating the moderating effect of product types such as experience and search goods as well as high and low involvement products on the relationships between qualitative and quantitative characteristics of reviews on perceived helpfulness. Multiple studies focused their attention on investigating the moderating effect of search and experience goods (Mudambi &

Schuff, 2010). But still, little research was conducted to uncover the impact of differences between these product types. Therefore, this research has the aim to fill the gap by answering the question:

To what extent does the effect of review and reviewer characteristics on perceived review helpfulness vary by product type?

This study aims to explain why some reviews are perceived as being more helpful than others and, in this manner, suggest some effective e-WOM management strategies for different product types. Retail websites might use these in developing product review writing guidelines that may foster more effective communication systems. This way, these online social interactions could be stimulated, which benefits the customer by reducing their purchasing uncertainty and the online retailer by increasing customer "stickiness" to provided services. Besides this, we aim to develop a machine learning algorithm that retailers could use in initially assessing the helpfulness of product reviews. This will considerably facilitate a better organization of reviews that will weaken the established Matthew effect, which implies that top reviews get more helpful votes due to their position within the review page compared to lower ones (Wan, 2015). Hence, this way, the customer burden of analyzing the ever-increasing enormous amount of review data will be diminished, leading towards a helpfulness assessment system that is not influenced anymore by the three previously identified types of bias: imbalanced vote bias, winner circle bias, and early-bird bias (Liu et al., 2007).

In the rest of the research paper, we first discuss the previous studies conducted within the WOM and e-WOM research fields. Then the research models are developed, around which hypotheses that are going to be tested are formulated. After this, we discuss in detail the data this study makes use of. Subsequently, we also discuss the methods that are implemented in analyzing our data. Then, we present the result of the analysis and the performance of the built-in models. And finally, we discuss the effects and initiate a discussion around those by summarizing our conclusions and limitations.

Literature review

To answer the central research question of this study, the already existing literature on WOM, e-WOM, online product reviews, and the perceived review helpfulness is investigated. This literature review aims to initiate this analysis and further provide ground for the investigation of relationships.

2.1. Word-of-mouth and interpersonal influence

Customers could influence each other in the decision-making process, meaning that an essential determining factor in customer purchasing behavior is others' opinions (Bearden et al., 1989). This share of product information and point of view, which affects customer purchase decisions, is also known as WOM communication (Hawkins et al., 2004). Previous research proved that customers are highly susceptible to WOM, which according to Arndt (1967), represents an experience-sharing action of early adopters for those seeking support in their decision-making process and reducing risk resulting from collective action. WOM had received considerable attention from managers and marketing researchers throughout the years for its significant identified impact on consumer choice (Arndt, 1967) and the product experience evaluations (Bone, 1995). This, and Bone (1995) in her research paper prove that WOM communication significantly affects customer perception of a product or service quality. Therefore, the interpersonal influence component is included in multiple-choice model formulations, which according to Bearden et al. (1989), represents an essential piece in the explanation of customer purchase behavior.

From the early research conducted by Arndt (1967), the general conclusion that positive WOM leads towards an increased probability of purchase while the negative WOM has an opposite effect stayed the same. Later, Mahajan et al. (1990) extended previous findings by claiming that WOM could also affect product evaluations. Currently, Gruen et al. (2005) found, while conducting their research based on online WOM, that WOM has a significant impact on customers' value perception and their brand loyalty. Therefore, WOM communication decreases the information asymmetries (Ozcan & Ramaswamy, 2004), leading to an increasing or decreasing speed of product recognition (Bass, 1969).

2.2. Electronic word-of-mouth at individual and market level

e-WOM is generally considered an extended version of traditional WOM, which emerged due to Internet and e-commerce development. It has created new ways for marketers to considerably boost communication effectiveness and efficacy, as well as new means for customer acquisition and retention. From the customer perspective, the emergence of e-WOM has led towards an increased spectrum of sharing consumption-related advice, opinion, and comments options, which means that customers can spread WOM on numerous platforms much easier and faster than ever before. Given the theoretical closeness of e-WOM and WOM, the definition this paper adopts is almost the exact definition attributed to WOM, the only difference being the method through which this communication is conducted. This means that e-WOM is defined as any positive or negative comment made by customers about a product, service, or seller through Internet-based technology (Hennig-Thurau et al., 2004).

Even if WOM and e-WOM are defined almost the same, and the means through which this communication takes place is the only aspect that differentiates them, these two differ considerably from each other regarding their characteristics. According to Cheung and Thadani (2012), one of them is that e-WOM has an exceptional speed of spread when compared to WOM. This is because, unlike WOM communication, where the information exchange takes place within a small group of people in a simultaneous mode (Dellarocas, 2003), e-WOM communication does not have to take place when all the communicators are present (Karakaya & Barnes, 2010), indicating that customers can anytime easily read and access the information posted online. Moreover, another differentiating aspect is that e-WOM communication makes this form of marketing more apparent and more accessible to be analyzed for management to extract valuable insights that could significantly improve customer experience and company performance (Park & Kim, 2008).

The transition to the online format of WOM made this powerful marketing tool more manageable and accountable. Therefore recently, there could be depicted an exponentially increasing interest in both managerial and academic communities towards e-WOM communication (Chevalier & Mayzlin, 2006). According to Cheung and Thadani (2012), these studies can be generally classified into two main streams: market-level and individual-level analyses. The former one focuses on market-level specifications such as product sales (Zhu & Zhang, 2010; Dellarocas et al., 2007; Chevalier & Mayzil, 2006), while at the individual level, researchers analyze e-WOM communication as an interpersonal influence process (Cheung et al., 2008; Park & Kim, 2008; Zhang & Watts, 2008).

2.2.1. e-WOM at the individual level analysis

Multiple studies have been conducted in experimental settings to identify the key factors of the message delivered by the communicator that have a significant impact on customer behavior. One of the most prominent factors that have been under investigation is the recommendation framing or valence. Positively valenced e-WOM mainly focuses on the advantages of a product or service and nudges the customer towards its consumption. In contrast, negatively valenced e-WOM has the opposite impact on customer purchasing behavior (Duan et al., 2008). According to Herr et al. (1991), customers are more inclined to value more negative information when compared to positive one while making their purchasing decision. Likewise, Park and Lee (2009) found that negative information has a considerably more robust impact on the e-WOM response than positive e-WOM, also known in the research field as negativity bias. Another stimulus cue investigated in the e-WOM communication research field is recommendation sidedness, which could be basically of two types: one-sided or two-sided. However, earlier marketing research found out that a two-sided message which contains both positive and negative information is perceived to be more trustworthy (Kamins & Assael, 1987). Another critical aspect of e-WOM communication that has been intensively investigated is message quality, which is a mixture of subject, layout, accuracy, and timeliness (McKinney et al., 2002). This one proves to positively impact perceived information helpfulness (Cheung et al., 2008).

Important to mention, however, is that the impact of the e-WOM message could vary from receiver to receiver. Thus, multiple studies have been conducted to identify the exact characteristics that moderate the e-WOM message's effects in the consumer decision process (Chaiken, 1980). Doh and Hwang (2009) claim that customer expertise and involvement play a significant moderating role in determining the impact of the e-WOM message on purchase intention. Customer involvement refers to information processing, while customer expertise indicates their prior knowledge (Johnston & Eagly, 1989). Higher involvement suggests that customers are more likely to carefully examine the message, while lower involvement points towards the fact that people are using the peripheral route. According to the dual-process theory, this postulates that customers are not engaging in effortful thinking and are using environmental shortcuts to decide whether to consider the message or not. An eloquent example in that sense could serve the source credibility assessment.

In traditional WOM communication, the communicator is most of the time known to the receiver, who could quickly evaluate the credibility of the message delivered and the communicator based on some aspects such as physical appearance, credibility, familiarity, and power. Moreover, it is already a proven fact that face-to-face communication has a substantial impact on customer purchasing behavior (Arndt, 1967). However, when we move towards e-WOM communication, opinion sharing occurs with an unknown group of people. This, according to Forman et al. (2008), could raise some doubt when it comes to e-WOM communication credibility. Therefore, when it comes to the communicator, the most researched factor is considered source credibility, which mainly translates the receiver’s perceived credibility of the message communicator measured in the degree of their expertise and trustworthiness (Chaiken, 1980). It is interesting to notice that it is already a proven fact that traditional WOM messages are perceived as more credible than e-WOM ones (Park et al., 2007). This might be due to the fact identified by Park and Lee (2009) in their paper, that opinion seekers have a hard time understanding the source credibility of an e-WOM message. Even if some studies have questioned the influence of source credibility on the e-WOM communication effect, multiple studies have proved that it is positively correlated with e-WOM credibility.

Another essential aspect that has a significant impact on the adoption of e-WOM messages is the platform on which e-WOM communication takes place (Chatterjee, 2001). Brikart and Schindler (2001) claim that online product or service reviews are perceived as more trustworthy than the content generated by the company itself on their webpage. Hovland and Weiss (1951) found that the information source plays a significant role in the information validity perception. Similarly, Park and Lee (2009) claim that the product reviews found on well-functioning websites have a more significant effect than ones found on websites that are not so developed. Moreover, it is already expected that due to the rapid change of the Internet environment, the contextual factor, or the platform through which the communication takes place, which represents the main difference between e-WOM and WOM, will become one of the most critical factors determining the e-WOM adoption.

2.2.2. e-WOM at market level analysis

At the market level, studies shed more light on quantitative and qualitative aspects of e-WOM communication that have a significant impact on product sales. These studies play a notable role in providing valuable judgments to companies that could foster their growth and increase their performance. However, even if e-WOM has received increased academic interest and there is a wide range of studies conducted to uncover the effect of this powerful marketing tool at the market level. There is still an ongoing debate in the academic community regarding the moderating effect of platform and e-WOM metrics that significantly impact product sales (Rosario et al., 2016).

According to Berger and Iyengar (2013), platform characteristics through which e-WOM communication occurs are extremely important in determining the effectiveness of e-WOM messages and their impact on sales. Information about the message communicator is perceived as being valuable, primarily when the opinion seeker identifies similar and trustworthy characteristics which help him/her in understanding whether the message is appropriate to them or not (Hung, Li, and Tse, 2011). Recent research proved that people tend to generally assess messages from customers with disclosed identity-descriptive information to be more credible, which significantly contributes towards an increased number of online product sales (Forman et al., 2008). Other relevant pieces of information about e-WOM messages shared on communication platforms, such as helpfulness votes and time when it was posted, boost product sales (Berger, 2014). The way information is presented plays a significant role in the spread of e-WOM on e-commerce platforms. Thus, bringing the most informative insights without over structuring the information leads to increased online product sales (Rosario et al., 2016). Posting costs such as purchasing the product or becoming a member of a community before sharing your experience considerably increases the e-WOM trustworthiness of customers and, as a result, boosts the number of product sales (Ott et al., 2012).

Even if numerous studies confirmed that e-WOM has a significantly higher impact on sales than other marketing strategies (Chen et al., 2011; Chevalier & Mayzlin, 2006; Moe & Trusov, 2011), a consensus about what e-WOM metrics drive the effect has not been reached. Some researchers found that e-WOM volume affects product sales (Duan et al., 2008). Others claim that it is not the volume that predicts the sales but rather their valence (Chintagunta et al., 2010; Dellarocas et al., 2007), their presence (Davis & Khazanchi, 2008), or their content (Onishi & Manchanda, 2012). Moreover, findings related to the impact of negative e-WOM messages do not come to a similar conclusion. Some studies have proved that negative messages have a negative effect on sales and even a more prominent absolute impact than positive e-WOM on product sales (Chevalier & Mayzlin, 2006; Sun, 2012), while others claim that negative e-WOM presence has a significant positive influence on product sales (Doh and Hwang, 2009).

2.3. Online customer reviews as a form of e-WOM

There is a wide range of ways e-WOM communication could take place (i.e., blogs and virtual communities; websites, product reviews, and hate sites; emails; instant messaging; chatrooms and newsgroups), but online consumer reviews, according to Hennig-Thurau et al. (2004) represents one of the most widely used formats of e-WOM communications. Moreover, these are also expected to have a higher impact on customers when compared to other forms of e-WOM communications. This is because the information shared there is quite easily accessible and for this it does not require advanced Internet knowledge.

As a result of fast e-commerce business field growth, people engage daily in increasing online transactions since it is low cost and convenient for the customers. To compare online shopping with traditional commerce, the former comes with higher levels of product uncertainty and lower retailer visibility, which negatively impacts customer satisfaction (Luo et al., 2012). However, according to Luo et al. (2012), service quality, website design, and pricing significantly diminish the established negative impact and outweigh the costs. It is considerably more challenging for customers to assess product or service quality before buying it, which means that online customers generally face an increased degree of risk and uncertainty. However, technology development made user-generated content such as online customer reviews available for opinion seekers to decrease their uncertainty and risk. These represent peer-generated opinions about products and services, including a star rating and a comment that clearly describes their overall experience (Mudambi & Schuff, 2010).

2.4. Review helpfulness

Since numerous researchers found that online product reviews as a form of e-WOM have a significant impact on product sales (Dellarocas et al., 2007; Duan et al., 2008), multiple companies consider and use online customer reviews as a new marketing tool (Chen & Xie, 2008). However, the ever-increasing number of online customers leads to the information overload problem, which confuses the customer and diminishes their decision-making efficiency (Chen & Tseng, 2011). Thus, it is crucial for management and academic communities to understand how and what makes customers perceive online reviews helpful. This way, helpful reviews will no longer be lost among the constantly increasing number of reviews. Among many investigated methods on how those reviews could be highlighted, Amazon and many more similar companies use the helpfulness feedback tool, which asks their customers whether the review was helpful to them or not. This tool was found to lead to an increased number of sales. An eloquent example in that sense is Amazon, which increased its revenue by $ 2.7 billion/year because of introducing this essential yet straightforward question (Spool, 2009). However, the only drawback of this mechanism is that customer participation is crucial since, otherwise, multiple reviews end up having little or no helpful votes, especially the most recent ones (Lu et al., 2010). This very sparse distribution of helpfulness votes makes it hard for opinion seekers to evaluate the helpfulness of reviews (Tang et al., 2013). Therefore, multiple studies have been conducted to find out the review characteristics that make a review helpful, but somewhat mixed conclusions have been drawn.

2.4.1. Review related factors

Since reviews are posted in the text format, word count represents one of the most appropriate ways to quantify online customer reviews. Mudambi and Schuff (2010) found a significant positive association between the review length and the perceived helpfulness of online reviews. This could be because lengthier reviews have a higher probability to contain a more detailed description of numerous product aspects, and as a result, increase the confidence of a decision-maker in the source (Gupta & Harris, 2010). However, previous research proves that the decision-making efficiency decreases when the information quantity is either extremely abundant or insufficient (Martin & Sell, 1980). Interestingly, technology development made information overload a bigger problem than an insufficient amount of information (Simperl et al., 2010). Moreover, Huang et al. (2015) previously identified a positive linear relationship between word count and its helpfulness that could only hold to a certain level – the average word count – after which the impact starts to diminish. However, to the best of our knowledge no research has been conducted on identifying the true shape of the relationship between word count and perceived review helpfulness. Therefore, our study aims to further investigate this aspect that might refute previously identified linear relationship between review's word count and its perceived helpfulness. Moreover, we would like to investigate the moderating effect of different product types on the relationship between word count and perceived review helpfulness. This way, retailers could subsequently use these findings in establishing word count limits for customers writing their reviews. Hence, the review helpfulness will be significantly improved, leading towards increased customer revisits and sales.

On almost all platforms, customers who are willing to share their experience should also indicate the overall valance of the review. This review rate generally ranges from one star indicating complete dissatisfaction to five stars, suggesting complete satisfaction with the received product or service. These tend to have a rather influential position within the review platform and thus received significant academic attention, proving a significant relationship between peer-generated review scores and customer perceived helpfulness. However, the direction of the investigated relationship has not reached a consensus. For example, according to Korfiatis et al. (2012), extreme reviews are more helpful in the decision-making process, while Eisend (2006) claims reviews with moderate ratings, which contain both advantages and disadvantages to be more useful. Besides this, multiple studies have relied on previous findings within the psychological research field, which claim that "bad is stronger than good" (Chua & Banerjee, 2014). This indicates that the negative information tends to have a higher impact than positive knowledge on customer behavior and, as a result, is perceived as being more helpful, which is also known as negativity bias (Wu, 2013).

Moreover, previous marketing studies proved that people looking for information to diminish their uncertainty usually tend to consume negative WOM (Herr, Kardes, & Kim, 1991). Thus, this seems reasonable to assume that people consider negatively valenced reviews as more helpful than positively valenced ones. However, Wu (2013) claims that the negative effect is exaggerated in consumer behavior literature. Even if it might have a more substantial impact on sales, the elaborateness of the information provided moderates the review valence impact on perceived review helpfulness. This means that even if negative information might quickly grab customer attention, the impact of review valence on customer behavior becomes insignificant when the review contains detailed, helpful, and novel information. Thus, this research paper aims to investigate the moderating impact of word count and readability on the relationship between its valence and perceived helpfulness. This investigation might explain the previously inconclusive findings.

Another significant piece of review helpfulness literature related to review factors focused on investigating the qualitative aspects of reviews and their impact on perceived helpfulness. These were identified to improve prediction accuracy significantly and therefore have received increased academic attention. Kim et al. (2006) investigated the importance of linguistic, structural, syntactical, semantic, and meta-data review characteristics in predicting review helpfulness. They found that lexical features play a significant role in customer quality evaluation of product reviews. Similarly, Zhang and Varadarajan (2006), in their research focused on qualitative review characteristics analysis, found that syntactic features have a significant impact on helpfulness assessment. Korfiatis et al. (2012) found in their research that highly or extremely helpful online customer reviews have more readable text when compared to less helpful reviews. Moreover, according to Ghose and Ipeirotis's (2010) econometric analysis, products (digital cameras, in their study) with higher readability scores tend to be positively associated with higher sales. Understanding a piece of text (in this research paper, online customer reviews) plays a significant role in the customer decision-making process (Smith & Taffler, 1992). And review readability metric represents a scale-based measure, which indicates the ease of understanding a piece of text assessed based on its linguistic characteristics such as syntactical elements and style (Klare, 1974). Even if it received increased academic attention, readability has never served as a moderating variable of the direct relationship that has been previously documented. Thus, an exciting avenue worth investigating is looking at how readability influences the relationship between word count and review helpfulness. This is because lengthier pieces of text do not always contain and translate meaningful information that is understandable and thus helpful.

2.4.2. Reviewer related factors

Reviewer information disclosure within the online community or website represents the personal information one decides on providing there, such as real name, photo, location, and reviewer identity. This has the general aim to help others find information about someone's social identity. Even if there are some concerns one may encounter, such as privacy issues, valid identity disclosure could meaningfully contribute towards the receivers' perception of the message. Earlier studies proved the significant helpfulness of identity disclosure in the online community, claiming that the customer efficiency of information processing considerably improves (Racherla & Friske, 2012). Moreover, Tidwell and Walther (2002) argue that communicator identity decreases customer perceived uncertainty, originating from the lack of social cues in the online community. Sussman and Siegal (2003) found that identity disclosure of the message communicator increases the perceived trustworthiness, and the overall message is found to be more helpful. Fogg et al. (2001) confirmed a significant positive association between the real name, photo exposure of communicators, and the website credibility assessment. Forman et al. (2008) explain that message receivers might use the personal information of the message communicator as a peripheral cue, which is seen as a psychological tool that helps them make insightful judgments and act. Thus, according to Liu and Park (2015), identity disclosure of the message communicator may boost the perceived source credibility of the receiver and, as a result, have a positive impact on the customer perception of review helpfulness. Moreover, Forman et al. (2008) demonstrated that customers faced with information overload problems in the form of product reviews deal with it in a heuristic mode by using the personal information of the communicator as an instrument on which to found their purchase. Thus, reviews posted by customers who disclose their personal information are perceived as more helpful than anonymous ones. This way, customers seem to value information that decreases the risk of controlled and fake reviews.

It is already known that people looking for helpful information that can assist them in making a rational decision tend to rely on suggestions from experts (Gilly et al., 1998). Moreover, information brought in the online format by an expert is regarded as being more helpful in influencing customer belief about a product and their purchase intentions than non-expert generated ones (Lascu et al., 1995). Gotlieb and Sarel

(1991) define expertise as the level of message appropriateness and knowledge. However, in the online community, it is considered rather difficult for message receivers to make such judgments based on the personal information provided by the communicator in the online community (Cheung et al., 2008). This means that the communicator level of expertise cannot be determined based on some restricted number of indicators. Thus, it depends upon their past performance within the online community, such as the number of reviews published on the online platform or the total number of helpful votes received (Weiss, Luire & MacInnis, 2008). Cheung et al. (2008) found that the helpfulness perception of customers is considerably higher when these consume information generated by communicators with a higher degree of expertise. Similarly, Forman et al. (2008) claimed that a higher level of reviewer expertise is positively associated with the perceived review helpfulness. On Amazon and many other websites, reviews written by reputable communicators are also labeled with a “top reviewer” badge. These labels tend to serve as a visual cue for customers when screening the vast number of reviews and heuristically choose reviews to take into consideration while making a purchase decision (Forman et al., 2008). This feature is generally found to positively impact review credibility, which results in increased perceived review helpfulness for receivers (Kuan et al., 2015; Yin et al., 2014). However, a review that is written by an expert but lacks readability it is unlikely to be considered helpful because regardless of who is writing the review, the understandability of the message plays a vital role in information quality assessment (Smith & Taffler, 1992). Besides this, a short review that lacks depth has a low probability of being considered helpful even if it is written by an expert. Thus, it would be interesting to investigate the moderating effect of readability and word count on the relationship between reviewer expertise and perceived helpfulness.

2.4.3. Product moderating effects

The analysis of possible moderators that could affect the investigated relationship may explain the inconsistent findings documented in studies on online review helpfulness (Hong et al., 2017). The moderator on which this research paper focuses relates to research context in which the investigation of factors contributing towards the helpfulness perception of customer reviews takes place – the product type (Baek et al., 2012; Mudambi & Schuff, 2010; Lee & Choeh, 2016). The reason behind this is that reading online reviews may differ from product-to-product type. Thus, different review or reviewer characteristics can contribute to reviews' perceived helpfulness (Baek et al., 2012). The most frequently used categorization of products in the online review helpfulness investigations is the one documented by Nelson (1970) – search and experience products. The main difference between these two product types is their uncertainty degree. This has its roots in the simple fact that the quality of the search goods (e.g., electronics) can be assessed during the information searching stage. In contrast, when it comes to experience products (e.g., music, restaurants), their quality can be evaluated only after purchasing or even product usage. However, little research has been conducted to uncover the impact of high versus low involvement products and the difference between these with search and experience products regarding review and reviewer characteristics. These findings could serve as crucial differentiating product strategies that could be useful in elaborating different writing guidelines for identified product types and models that could classify reviews into helpful and unhelpful ones.

Conceptual framework

The general aim of this research paper is to investigate the influence of some review and reviewer characteristics on perceived review helpfulness by building up hypotheses that will be subsequently tested. Then identified factors with a significant impact on review helpfulness are used in assessing how well these predict the helpfulness of a review. So, this part of the study has the scope to introduce relationships that this paper intends on testing and the reasoning behind these expectations.

3.1. Explanatory analysis

Dual-process theories explain how message credibility could be determined using conscious and unconscious processes, such as considering message content and context factors (Zhang et al., 2003). The most well-known theories that adopt this approach are HSM (Heuristic-Systematic Model) and ELM (Elaboration Likelihood Model). Both have been extensively used in understanding how information processing leads towards product or service consumption in the online environment (Lee et al., 2008; Park & Kim, 2008). Chen and Chaiken (1999) affirm that conscious processing implies desire and ability, while unconscious information processing can happen with a little amount of both. HSM makes the distinction between systematic information processing, which implies significant cognitive effort, and heuristic processing, where subjects make use of significantly lower cognitive effort at the information processing stage (Chaiken, 1980). At the same time, ELM differentiates between the central route, where subjects approach the delivered information logically, and the peripheral route, which implies a more context-oriented information processing approach. This study uses both theories to allocate information of peer-generated product reviews into peripheral cues for heuristic processing and central cues for systematic ones. This way, the explanatory part of this research paper aims to identify which type of information processing is used depending on why subjects use online consumer reviews.

As it can be inferred from Figure 1, this paper considers both central as well as peripheral cues in determining the perceived review helpfulness. Review helpfulness is our dependent variable, which measures the magnitude with which customers perceive a review as being helpful in their purchasing decision. In our study, review perceived helpfulness is measured as the total number of helpful votes a review obtained until the collection date. Based on ELM, in this research word count and readability score are classified as central cues, whereas reviewer expertise and review valence are classified as peripheral cues.

The explanatory analysis starts with the investigation of the shape of the relationship between review's word count and perceived helpfulness that has not received enough academic attention and might refute the linear relationship that have been previously identified. Then, the investigation continues by uncovering the interaction effect between quantitative aspect of a product review – word count – and its qualitative aspect – review readability – on perceived helpfulness. This might demonstrate that the impact of the quantitative aspect on perceived helpfulness of the review fades away with the introduction of its qualitative characteristic. Moreover, to the best of our knowledge, the moderating effect of review readability on the relationship between reviewer expertise, valence and perceived helpfulness has not been researched. The same could be claimed about the moderating effect of word count on the relationship between reviewer expertise and valence. However, the investigation of interaction effects between review's central and peripheral cues might explain the previously identified inconsistent or spurious relationship directions, which becomes the goal this research paper. Finally, the general expectations of this research paper are that product type moderates the impact of central and peripheral cues on perceived review helpfulness. Thus, to confirm or reject these assumptions our research is based on a sample of products divided into two classes: (1) experience versus search goods and (2) high versus low involvement products. Hence, based on the established relationships of interest the conceptual model of this research paper is built (see Figure 1).

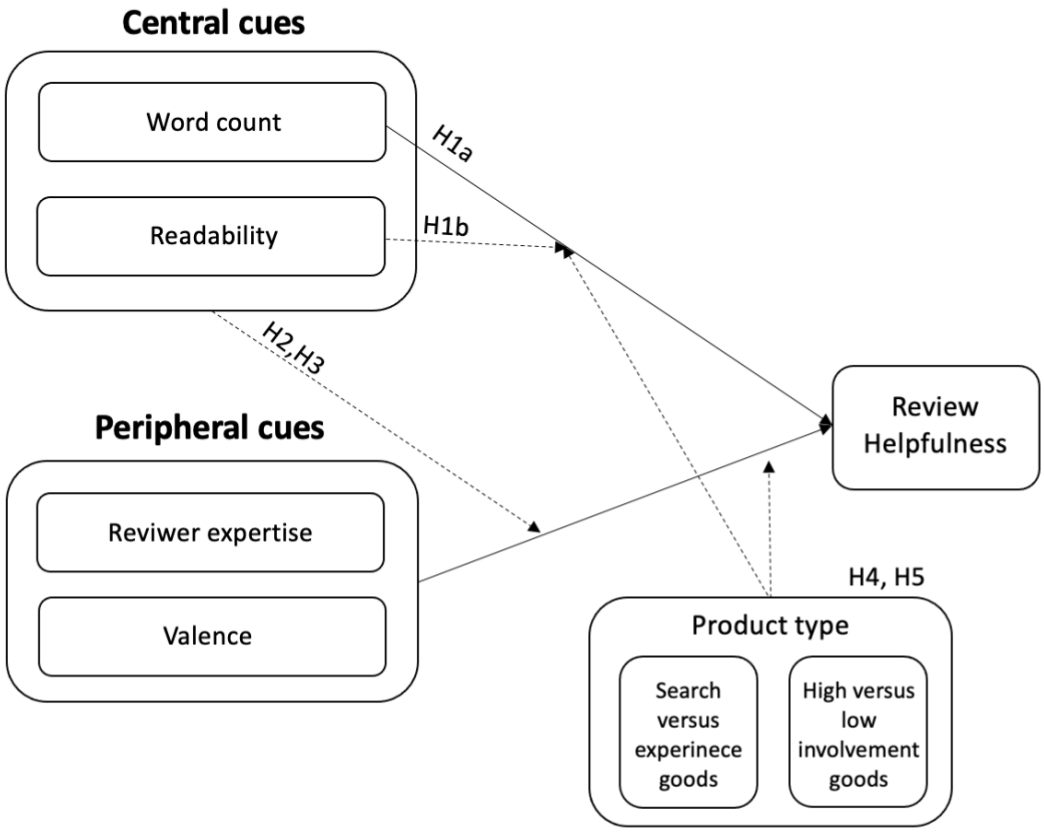

Figure 1. Conceptual model

3.1.1. Central cues

Central cues represent the arguments a message contains, objectively used by receivers while systematically processing the delivered information (Chaiken, 1980). Systematic information processing assumes detailed content analysis in deciding whether to use message suggestions in the decision-making process. In this study as prototypes of central cues are word count and readability of an online peer-generated product review.

Since peer-generated product reviews are typically delivered in text format, the most appropriate quantitative measure seems to be the number of words. Moreover, it is already a proven empirical fact that the ability to process information is limited when its quantity is either excessively abundant or limited (Martin & Shell, 1980). Therefore, word count can be considered a good predictor of review helpfulness. According to Mudambi and Schuff (2006), there is a positive association between the number of words a review contains and the perceived helpfulness. This means that the information the communicator delivers tends to significantly impact customers' purchase decisions and that the perceived helpfulness of a review grows as the number of words increases. Still, according to the law of diminishing marginal utility, which affirms that the marginal utility of each additional unit tends to decrease, the helpfulness of the review increases at a decreasing rate while adding words to it. Huang et al. (2015) claim that there is a word count threshold in its positive impact on review helpfulness. This could be explained by the fact that an initial increase in the word count of a review makes it more informative, but a subsequent increase might come with a diminished utility. This is because it might require an increased customer intellectual capacity and focus on processing the whole amount of information provided (Jackson and Farzaneh, 2012). If this happens, customers could go over the review text superficially, making it less likely to be perceived as helpful. Hence, the following hypothesis is formulated:

H1a. An increase in the word count of a review is associated with increased perceived helpfulness, but the impact gradually diminishes as the number of words further increases.

Lengthy reviews tend to be filled with more details, but this does not necessarily mean that those are automatically helpful ones. Thus, this means that previously identified positive relationship between word count and perceived helpfulness might be spurious. And there is a quality component such as review readability that might moderate the relationship between review word count and its helpfulness. For instance, review readability scores have been found to have a significant positive impact on perceived review helpfulness, which means that helpful reviews tend on average also to have higher readability (Korfiatis et al., 2012). Hence, it is of high interest to investigate whether word count stays a significant predictor of review helpfulness if its readability is held constant. Thus, in this research paper, the interaction effect of word count and its readability score is investigated, where the following is hypothesized:

H1b. Review readability positively moderates the positive relationship between word count and perceived review helpfulness.

3.1.2. Peripheral cues

Peripheral cues are context factors used to heuristically process the delivered information in a subjective way (Chaiken, 1980). Peripheral route changes of attitude towards something described in the message delivered require less rational judgment (Fleming et al., 1999). In the online community, reviewer expertise and review valence represent peripheral cues that this research intends on investigating.

Previous research proves that source credibility has a rather important role in the process of online information adoption (Briggs et al., 2002). Identity disclosure of reviewers in the online community influences opinion seekers' information perception (Ma & Agarwal, 2007). Besides this, reviewer activity within the community also tends to positively impact the receiver's perceived credibility of the message (Cheung et al., 2008). Moreover, according to Forman et al. (2008), reviewer reputation within virtual community serves as a good proxy for information source credibility, which customers use to augment provided product information leading towards an increased level of trustworthiness (Zhang & Watts, 2008). Thus, our research focuses on analyzing the impact of reviewer expertise on perceived review helpfulness, measured in the number of total helpful votes a reviewer has accumulated on all the posted reviews on Amazon.

While this direct effect has been previously researched by Huang et al. (2015), who found that reviewer expertise has a significant positive impact on perceived review helpfulness, it becomes highly interesting to research the moderating effect word count and readability has on the relationship. This is because reviews are perceived to be helpful in case these are sufficiently extensive and readable (Huang et al., 2015; Korfiatis et al., 2012). Thus, it might be that a review which does not have enough depth and readability may have a low probability of being perceived as helpful even if written by an experienced reviewer. Therefore, the following hypotheses are formulated:

H2a. Word count positively moderates the positive relationship between reviewer expertise and perceived review helpfulness.

H2b. Review readability positively moderates the positive relationship between reviewer expertise and perceived review helpfulness.

When it comes to review valence, which is another peripheral cue this research paper investigates, previous research has not reached a consensus regarding the impact of review valence on the perceived helpfulness of reviews. One literature stream claims that extreme reviews are perceived more helpful, while the other academic stream claims that moderate ones are more useful. These inconsistent relationships between review valence and perceived helpfulness may be because of a heterogeneity in a review's central cues. Indeed, central cues could serve as a moderator of the effect of valence on helpfulness. For instance, according to Korfiatis et al. (2012), extreme reviews do not have to be as long as moderate ones to be considered helpful. This could be explained by the fact that the former one should only contain positive or negative arguments, while the latter one should provide arguments for both sides. Thus, a moderate review that is not lengthy might not be perceived as helpful. This indicates that word count might have a moderating effect on review valence and perceived helpfulness. Another moderating effect worth investigating that might explain the previously identified inconsistent directions of the relationship between review valence and its helpfulness is the review readability. For example, according to Herr, Kardes, and Kim (1991) negatively valenced WOM are perceived to be more helpful than positive or moderate ones. This is because unsatisfied customers feel higher emotional need to discharge in the form of reviews and thus provide more detailed information that might refrain others from experiencing the same loss. However, communicators do not always narrate their experience in an understandable way that could bring value to the customer. Thus, a negatively valenced review that lacks readability might not be as helpful for the customer as a readable one. Therefore, to uncover the true relationship between review valence and its perceived helpfulness, this research paper aims to investigate the interaction effect between review valence and word count, as well as readability score, where the following is hypothesized:

H3a. Word count has a positive moderating effect on the relationship between moderate reviews and perceived review helpfulness, while a negative one on the relationship between extreme reviews and perceived helpfulness.

H3b. Review readability has a positive moderating effect on the relationship between negative reviews and perceived review helpfulness.

3.1.3. Product type

The reason why customers read product reviews can differ from product-to-product type customers intend to buy. Thus, this study investigates how the impact of review and reviewer characteristics on review helpfulness is different for different product types. Product reviews based on which the relationships this research paper intends on testing are split into two groups: (1) experience versus search products and (2) high versus low involvement goods. For each category of products, it is investigated the role of central and peripheral cues in the information processing of a review to determine review helpfulness.

Search products are mainly bought for their performance. Thus, the customer reviews for search products tend to be consciously driven, influential, and purposeful (Strahilevitz & Myers, 1998). However, on the other side of the coin, experience products imply the hedonic facet of the consumption and are defined by an affective experience (Hirschman & Holbrook, 1982). Hence, customer reviews are generally more individual and sentimental. This implies that customers of search and experience products tend to look for information and process it differently (Huang, Lurie & Mitra, 2009).

Previously, we have hypothesized that there is a non-linear relationship between word count and perceived helpfulness. Here, we believe that product type might moderate the relationship between word count and perceived helpfulness. For search goods, a longer review tends to include a detailed product and performance information. While, for experience goods longer customer reviews might contain a more emotional related description. Thus, a longer subjective review does not seem to significantly contribute to customer's decision. Therefore, our general expectation is that longer product reviews for search products are perceived to be more helpful than for experience goods. Since customer judgements for search goods are expected to be consciously driven, objective and containing more factual information than for experience products. We expect that more readable reviews written by an experienced reviewer for search products bring higher utility than for experience goods. Moreover, we expect that product type also moderates the relationship between review valence and its helpfulness. To be more specific, we believe that consumers tend to value less moderately valenced reviews for experience than for search goods. Therefore, the following hypothesis is formulated:

H4. Product type (i.e., search versus experience goods) moderates the relationship between review's central and peripheral cues on perceived helpfulness. For search goods, lengthier, more readable, and moderately valenced reviews written by a more experienced reviewer are perceived to be more helpful than for experience goods.

Marketing literature points towards the fact that information-search customer behavior considerably differs with the degree of product involvement (Clarke & Belk, 1978). Consumers rarely engage in extensive information searches for low-involvement products since these are usually of little importance to them and not worth the effort (Zaichkowsky, 1985). However, according to Clarke and Belk (1978), the same cannot be claimed about customers' behavior towards high-involvement products. They usually tend to extensively engage in obtaining qualitative information from multiple sources and read peer-generated product reviews in detail to excerpt the most out of them. The difference in product type and information search behavior implies a significant difference in the contribution magnitude of central and peripheral cues towards the review helpfulness perception of customers for low versus high involvement products. Thus, the following hypothesis is formulated:

H5. Product type (i.e., low versus high-involvement goods) moderates the relationship between review's central and peripheral cues on perceived helpfulness. For high-involvement products, lengthier, more *readable, and moderately valenced reviews written by a more experienced reviewer are perceived to be more helpful than for low-involvement goods.*

3.2. Predictive analysis

Even if peripheral cues such as reviewer reputation and its valence within the online community might influence the review perceived helpfulness. The previously hypothesized moderating effects of central cues on the relationship between peripheral cues and perceived helpfulness makes us believe that central cues might play a more influential role in the prediction task. Thus, the following hypothesis is formulated:

H6. Central cues have a higher predictive power than peripheral cues.

Data

4.1. Research sampling

This study makes use of the data collected from the most well-known online retail store – Amazon. The main reason why the data is scraped from this retail website, besides having the most significant number of product reviews when compared to others, is that the reviews posted there do not seem to be censored. If this is the case in product data reviews, then systematic data bias could be present, leading towards misleading and implausible findings. Moreover, to prevent cultural dissimilarities and language proficiency, the data is collected from Amazon UK online store. This selection tries to ensure that collected product reviews are written by native English speakers living in the same country. This is important since the qualitative analysis of reviews is pointless when the reviewers are not native speakers because this could substantively affect the level of review understanding.

The peer-generated product reviews are collected for the previously identified product categories that meet the criteria of experience and search goods classification identified by Nelson (1970) as of May 2021. For experience products, data is gathered from three product categories: video games, musical instruments, and books (Mudambi & Schuff, 2010). For search products, the data is collected from another three product categories: digital cameras, cell phones, and laser printers (Mudambi & Schuff, 2010). From each category, the four to six best-selling products with a number of total reviews higher than 1500 are selected, resulting in a sample of 40 products. Our data sample represents a selection of best-selling products having a total number of product reviews higher than 1500 because most product reviews and helpful votes are usually concentrated around these products (Chevalier & Mayzlin, 2006). Moreover, randomly selecting a sample of product reviews is extremely complicated. Thus, we assume that a selection of four to six best-selling products is representative for the entire population of reviews of a specific category.

The data is collected from the Amazon page of each selected product with the default option for displaying reviews set to “Top reviews”, which ranks reviews in a decreasing order based on the number of helpful votes received. This means that the most helpful reviews are displayed on the first pages of the total number of reviews. Thus, since the default option of displaying product reviews on Amazon is based on helpful votes, this could have a significant influence on customer behavior - as an escape solution from the information overload problem customers focus on the first most helpful reviews displayed (Smith & Brynjolfsson, 2001). Another problem that arises due to this default option is that the distribution of votes becomes rather biased towards the old online peer-generated customer reviews, which means that the number of helpful votes could be a function of the number of days a review has been displayed on the website (Kim et al., 2006). To control for this, the variable that represents the difference in days between the date when the review was written and the date the review was collected is included in the model.

Our data is scraped using a crawler developed in R to download web pages of customer reviews, reviewers, and product information. Initially, all the available reviews for each selected product were collected. However, important to mention is that the helpfulness rating algorithm in-place is subject to imbalanced vote bias, winner circle bias, and early-bird bias. Thus, product reviews that end up with no helpful votes does not necessarily indicate that these reviews are unhelpful, but rather that even if helpful some of them did not make to the top due to the algorithm. Therefore, to capture the features that make a review helpful and the difference between different product types without inducing bias in our analysis product reviews with no helpful votes are not included. Moreover, reviews having less than ten helpful votes were removed as well because, in that case, the helpfulness of a review is not meaningful (Ghose & Ipeirotis, 2010). Therefore, our sample of product reviews resulted in a selection of 587 reviews.

4.2. Operationalization of variables

For each product review of our dataset, there is a collection of 10 variables, out of which some serve as independent, while others as control variables. The selection of variables was based on comprehensive investigation of review helpfulness-related literature. In Table 1, these variables, their operationalization, and descriptive statistics are presented.

Since customers on Amazon are given the right to vote whether a review was helpful to them in their purchasing decision making process, our dependent variable – perceived helpfulness – is operationalized by using the number of helpful votes in hundreds a review has received until the collection date (Racherla & Friske, 2012). The valence of each review is determined by using the sentiment analysis and namely the polarity function, which returns an approximation of the sentiment of the text by grouping variables. This sentiment ranges from 1 to -1, where the polarity value of a piece of text that is close to 1 indicates positive sentiment, while -1 negative one (see Appendix A). When it comes to the operationalization of readability score, the Flesch reading ease score is used (Korfiatis et al., 2012). The score is computed by using the average length of the sentences and the average number of syllables per word. Each review gets a score which ranges from 0 to 100, where 100 indicates that the text is easy to read and a score of 0 indicates the opposite (see Appendix B). The reviewer expertise is measured in number of total accumulated helpful votes of an Amazon reviewer (Huang et al., 2015) minus the helpful votes the focal review has got to avoid reverse causality. To categorize products into high and low involvement the average price value is used as a threshold (Hochstein et al., 2018). In case the price of the product is lower than the average product price the product is considered a low-involvement one, otherwise it is classified as high-involvement product.

Table 1. Descriptive statistics of collected variables for product reviews retrieved from Amazon UK

Variables	Description	Mean	SD
Perceived Helpfulness	The total number of helpful votes each review has accumulated until the collection date, in hundreds.	0.8	1.4
Word Count	The number of words the product review text contains, in hundreds.	1.5	1.6
Valence	The sentiment of the review computed by using text analytics, which ranges from 1 to -1.	0.1	0.3
Readability Score	The Flesch reading ease score ranges from 0 to 100 and as the value decreases, the readability of review text also decreases.	76.3	11.7
Reviewer Expertise	The total number of helpful votes a reviewer got on all posted reviews on Amazon besides focal review's helpful votes, in hundreds.	4.1	23.7
Product Type I	A categorical variable which takes the value of 1 in case this is an experience product and 0 for a search good.	37% search products	0.5
Product Type II	A categorical variable which takes value of 1 in case the product is low-involvement and 0 in case the product is a high-involvement one.	62% low-involvement products	0.5
Days Lapsed	The difference in number of days from the date when a review was posted and the date it was collected, in hundreds.	5.9	3.7
Reviews Number	Number of total reviews a product has got at the day of product reviews collection, in hundreds.	33.6	31.7
Average Product Rating	Average rating of the products under investigation retrieved from Amazon UK.	4.6	0.2

Methods

To clearly understand what are the review, reviewer, and product characteristics that have a significant impact on the perceived helpfulness of online peer-generated product reviews and how these could be used in developing a classification algorithm, a two-level study is conducted. The analysis of this research paper begins with an explanatory method – ordinary least squares (OLS) regression. After this, a predictive model is built by using Random Forest.

5.1. Explanatory analysis

Since our explanatory analysis aims to identify the moderating effect of central cues on the relationships between review's peripheral cues and perceived helpfulness as well as the moderating effect of product types on both classes of cues, ordinary least squares (OLS) models are built. In general, OLS models aim to estimate relationships by minimizing the sum of squared differences between observed and predicted values of the linear function. The overall performance of the models that are built to assess the explanatory power of identified interaction effects for this research paper is evaluated by using the adjusted R^2 value.

As many other statistical analyses, OLS regression has some general assumptions that are important to be stated and tested to ensure a better estimation performance of OLS models. Since multicollinearity makes the identification of statistically significant variables and interpretation of coefficients in isolation hard, one of the OLS assumptions is that there is no high correlation between independent variables of a model. To investigate whether this assumption holds true for our OLS models, a table containing the Person's correlation coefficients between the main variables of interest is built (see Table 2). As it can be clearly inferred from Table 2, there is no troublesome correlation between any of the independent variables this research paper makes use of. Another OLS assumption is that the error term is normally distributed. This one ensures that statistical hypothesis testing with reliable confidence and prediction intervals could be performed. To validate this assumption, a normal probability plot is built (see Figure 2). As it can be clearly seen in Figure 2, the error terms are normally distributed. Therefore, this indicates that the second assumption also holds true.

Table 2. Pearson's correlation coefficients

Variable	1	2	3	4	5	6	7	8
1 Days Lapsed	1							
2 Perceived Helpfulness	0.09	1						
3 Word Count	0.01	0.24	1					
4 Reviewer Expertise	0.04	0.09	0.26	1				
5 Average Product Rating	0.04	0.03	-0.04	0.01	1			
6 Reviews Number	-0.13	0.13	-0.09	-0.04	0.4	1		
7 Readability Score	0	0.06	-0.09	-0.05	-0.02	-0.02	1	
8 Valence	0.32	0.03	0.14	0.06	0.1	-0.2	0.01	1

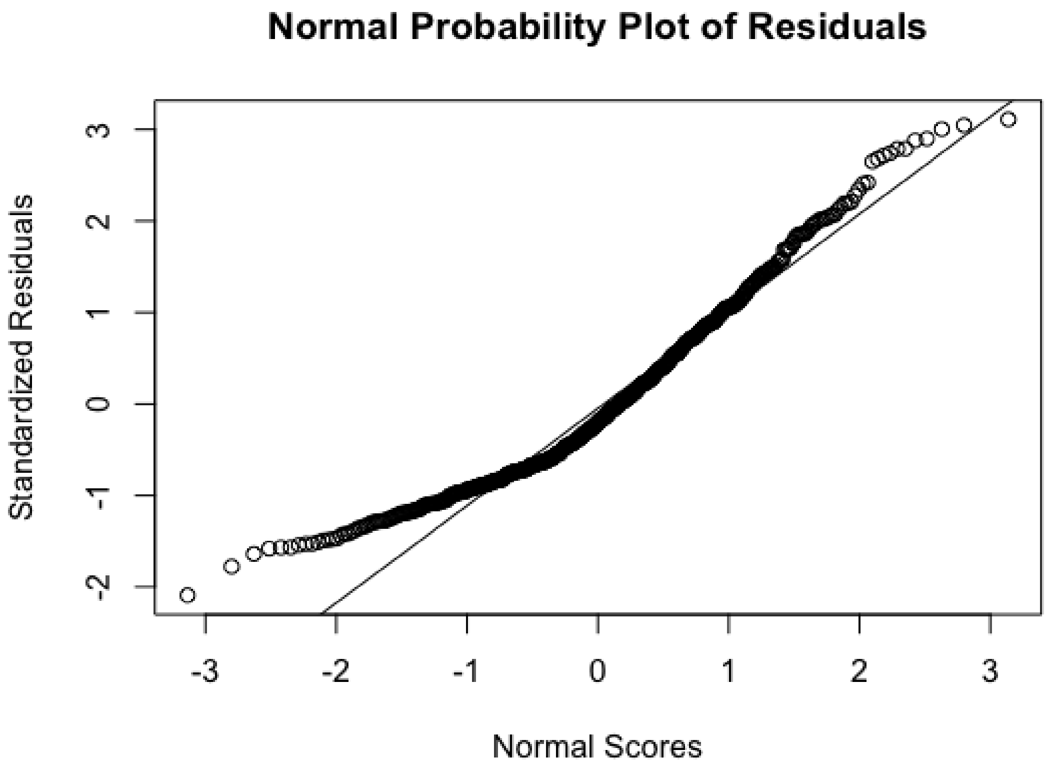

Figure 2. Normal Probability Plot of Residuals

To investigate the effect on the model fit of each review, reviewer characteristic, product type and their interactions different models are built to test the different hypothesis. Firstly, to investigate the hypothetical nonlinear relationship between perceived helpfulness (PH) and review's word count, a linear-log model is built to better capture the hypothesized rather proportional to perceived helpfulness than constant effect of word count. Moreover, because we are also interested in the moderating effect of review's central and peripheral cues to uncover previously identified spurious relationships or explain contradicting effects on review's perceived helpfulness the mathematical model specification for Model 1 takes the following form, which refers to H1a, H1b, H2a, H2b, H3a and H3b:

$$\begin{aligned} PH_i = \beta_o + \beta_1 * WordCount_i + \beta_2 * ReadabilityScore_i + \beta_3 * ReviewerExpertise_i + \beta_4 * Valence_i \\ + \beta_5 * {Valence_i}^2 + \beta_6 * ReadabilityScore_i * WordCount_i + \beta_7 * ReviewerExpertise_i \\ * WordCount_i + \beta_8 * ReviewerExpertise_i * ReadabilityScore_i + \beta_9 * Valence_i \\ * WordCount_i + \beta_{10} * Valence_i * ReadabilityScore_i + \beta_{11} * {Valence_i}^2 * WordCount_i \\ + \beta_{12} * {Valence_i}^2 * ReadabilityScore_i + \beta_{13} * DaysLapsed_i + \beta_{14} * ReviewsNumber \\ + \beta_{15} * AverageProductRating_i + \varepsilon_i \end{aligned}$$

, where PH_i is the number of helpful votes review i got until the collection date, $WordCount_i$, $ReadabilityScore_i * WordCount_i$, $ReviewerExpertise_i * WordCount_i$, $ReviewerExpertise_i * ReadabilityScore_i$, $Valence_i * WordCount_i$ and $Valence_i * ReadabilityScore_i$ represent our variables of interest, while others are our control variables and ε_i is the error term.

To identify the moderating effect of the product type (i.e., search versus experience goods) on the relationship between central and peripheral cues and perceived helpfulness the following model is built, which takes the following mathematical representation and refers to H4:

$$\begin{aligned} PH_i = \beta_o + \beta_1 * WordCount_i + \beta_2 * ReadabilityScore_i + \beta_3 * ReviwerExpertise_i + \beta_4 * Valence_i \\ + \beta_5 * ProductTypeI_i + \beta_6 * WordCount_i * ProductTypeI_i + \beta_7 * ReadabilityScore_i \\ * ProductTypeI_i + \beta_8 * ReviwerExpertise_i * ProductTypeI_i + \beta_9 * Valence_i \\ * ProductTypeI_i + \beta_{10} * DaysLapsed_i + \beta_{11} * ReviewsNumber + \beta_{12} \\ * RatingsNumber_i + \beta_{13} * AverageProductRating_i + \varepsilon_i \end{aligned}$$

, where $WordCount_i * ProductTypeI_i$, $ReadabilityScore_i * ProductTypeI_i$, $ReviewerExpertise_i * ProductTypeI_i$ and $Valence_i * ProductTypeI_i$ represent our interaction variables of interest.

Model 3 has the goal to investigate whether the product type (i.e., high versus low involvement products) moderates the relationship between reviews' central and peripherical cues on perceived helpfulness and thus the following model is built:

$$PH_i = \beta_o + \beta_1 * WordCount_i + \beta_2 * ReadabilityScore_i + \beta_3 * ReviwerExpertise_i + \beta_4 * Valence_i + \beta_5 * ProductTypeII_i + \beta_6 * WordCount_i * ProductTypeII_i + \beta_7 * ReadabilityScore_i * ProductTypeII_i + \beta_8 * ReviwerExpertise_i * ProductTypeII_i + \beta_9 * Valence_i * ProductTypeII_i + \beta_{10} * DaysLapsed_i + \beta_{11} * ReviewsNumber + \beta_{12} * RatingsNumber_i + \beta_{13} * AverageProductRating_i + \varepsilon_i$$

, where $WordCount_i * ProductTypeII_i$, $ReadabilityScore_i * ProductTypeII_i$, $ReviwerExpertise_i * ProductTypeII_i$ and $Valence_i * ProductTypeII_i$ represent our interaction variables of interest.

5.2. Predictive analysis

The explanatory analyses that we have described above have the aim to identify meaningful interaction effects that influence the review perceived helpfulness. However, in this section we move from explanatory to predictive modeling. To be more specific, now the aim is not to explain which review and reviewer characteristics influence review's perceived helpfulness, but rather find the predictive power of identified characteristics on classifying a review into being helpful or unhelpful. This way, the substitution of the subjective human judge with a machine learning algorithm might take place. Thus, the previously identified imbalanced vote bias, winner circle bias, and early-bird bias of the in-place algorithm for multiple e-commerce companies including Amazon might be tackled.

Since we are aiming to build a binary prediction model that classifies reviews into helpful or unhelpful, our continuous perceived review helpfulness variable must be converted into a binary one. This conversion is based on each product's median helpfulness value threshold. This way, all reviews of a product having a higher number of helpful votes than the median helpfulness value threshold of each product are treated as helpful and the others as unhelpful. Our predictive analysis starts with building a classifier model by using all the features that we have previously identified: review's central and peripheral cues as well as the two classes of product types. Furthermore, the predictive performance of each group of cues (i.e., central, and peripheral) are going to be assessed by building up classifier models for each of them and compare their prediction accuracy. This part of the predictive analysis refers to hypothesis H6.

To classify unseen product reviews into helpful and unhelpful ones, the Random Forest machine learning technique is used. Random Forest is an ensemble learning method built out of multiple decision trees (i.e, N_{tree}) to get a better prediction performance with lower variance than the one obtained from using only one learner (i.e., decision tree). This method simultaneously trains multiple decision trees based on bootstrapped training samples by only considering a random subset of available predictors (i.e., m_{try}) for every independent tree. Bootstrapping and random selection of features ensure that every decision tree of a Random Forest algorithm is unique and uncorrelated. This results in multiple predictions of a new observation made by each tree, which is not always the same. But the final decision is taken based on the aggregation of all individual tree votes and adoption of the majority decision. This explains why Random Forest tends to beat other classification methods in terms of accuracy for small and large datasets without having issues with overfitting (Biau & Scornet, 2016).

When the classification trees are built, the goodness of split, or the node purity is measured by using Gini Impurity Index. This index measures the probability of a randomly selected sample being misclassified by using the sample labels in the node and has the following mathematical formulation:

$$G = 1 - \sum_{i=1}^{n} P_i^2$$

, where P_i measures the probability of an element being classified for a different class. Thus, the decision tree is grown by searching for the features that result in the greatest Gini Impurity decrease until it reaches maximum depth with no pruning.

There are two hyperparameters that we tune in to build a high-performing Random Forest algorithm: N_{tree}, which represents the number of trees in the forest and m_{try} indicating the number of random predictors to be chosen and investigated for the best split when building up decision trees. Even though some rules of thumb could be used in elaborating a Random Forest algorithm, we choose to tune our parameters to optimize the performance of our model. To do so, a random search is performed, which combines different values of hyperparameters until the predictive ability is maximized.

When it comes to prediction performance assessment, important to notice is that trees are built by using a part of the training sample, which means that about two-thirds of the dataset is used in training the model, while the remaining data is used to validate the performance of the Random Forest model internally. The estimated error is also known as out-of-bag (OOB) error and is computed as follows:

$$OOB_{error} = 1 - accuracy_{randomforest}$$

However, this represents a performance measure based on validation from bootstrapped data in the training set. However, it is still of high importance to also analyze model performance from the test data perspective. Thus, our dataset is split in a ratio of seven to three, where 30% of the data is used to assess the model accuracy by investigating the model accuracy percentage and Kappa statistic. The first one indicates how many of the predictions were correctly identified based on the confusion matrix. At the same time, the second one measures the agreement between classification and actual values. It ranges from zero to one, where zero indicates no agreement and one complete agreement.

Results

6.1. Explanatory analysis

The results of our linear regression analysis are presented in Table 3. In each case the dependent variable is operationalized as the natural logarithm of the number of helpful votes a review under investigation obtained. As it can be inferred from Table 3, the analysis of the model proves to have a good fit with multiple significant coefficients and large values of R^2 and adjusted R^2.

To test our H1a hypothesis, which states that an increase in word count of a review increases its perceived helpfulness at a diminishing rate, the natural logarithm of the word count is regressed on perceived helpfulness. As it can be inferred from Table 3 (see Model 1) the coefficient of word count is significant but has a negative effect ($\beta = -0.564,\ p < 0.05$). This means that the review helpfulness is decreasing at a decreasing marginal rate with additional words, and hence hypothesis H1a is **rejected**.

In H1b, we hypothesized that readability score moderates the relationship between review's word count and perceived helpfulness. As one can see in Table 3, a significant positive effect is found ($\beta = 0.995, p < 0.01$). This indicates that word count has a greater positive effect on the perceived helpfulness of reviews with a higher readability score. Thus, hypothesis H1b is **accepted**.

The results in Table 3 (see Model 1) also present strong support for the hypothesized moderating effect in H2a and H2b of central cues (i.e., word count and readability) on the effect of reviewer expertise on review helpfulness. This support is pointed out by the significant interaction effect of Reviewer Expertise * Word Count ($\beta = 0.006, \mathrm{p} < 0.01$) and Reviewer Expertise * Readability Score ($\beta = 0.072, p < 0.05$). This means that the length and readability of the review positively moderates the relationship between reviewer expertise and perceived helpfulness. Hence, reviewer expertise has a higher positive effect on the perceived helpfulness of lengthier and more readable reviews. Therefore, hypotheses H2a and H2b are **accepted**.

As it can be inferred from Table 3, significant interaction effects of Valence * Word Count ($\beta = -0.452, \mathrm{p} < 0.05$) and Valence^2 * Word Count ($\beta = 0.986, \mathrm{p} < 0.05$) indicate that there is a concave rather than hypothesized convex relationship. Therefore, hypothesis H3a is **rejected**. Moreover, the hypothesized moderating effect of review readability on the relationship between negative reviews and perceived helpfulness does not receive support. Thus, this means that hypotheses H3b is **rejected**.

Table 3. OLS results for the relationship between review perceived helpfulness and its characteristics

Variable	**Model 1**	**Model 2**	**Model 3**
Constant	1.018 (1.35)	-0.051 (1.44)	0.19 (1.45)
Word Count	**-0.564*** **(0.28)**	0.211** (0.04)	0.2** (0.0003)
Readability Score	-0.458 (0.65)	1.310* (0.58)	1.07 (0.743)
Reviewer Expertise	-0.011 (0.02)	0.004 (0.003)	0.009* (0.004)
Valence	-1.902 (1.64)	-0.058 (0.31)	0.046 (0.38)
Valence^2	1.206 (3.59)		
Product type I		0.75 (0.76)	
Product type II			0.04 (0.75)
Word Count * Readability Score	**0.995**** **(0.37)**		
Reviewer Expertise * Word Count	**0.006**** **(0.002)**		
Reviewer Expertise * Readability Score	**0.072*** **(0.03)**		
Valence * Word Count	**-0.452*** **(0.16)**		
Valence * Readability Score	**3.116** **(2.24)**		
Valence^2 * Word Count	**0.986*** **(0.36)**		
Valence^2 * Readability Score	**-2.832** **(5.05)**		

Product type I * Word Count		**-0.013***	
		(0.0005)	
Product type I * Readability Score		**-0.891***	
		(0.09)	
Product type I * Reviewer Expertise		**0.003***	
		(0.005)	
Product type I * Valence		**-0.015**	
		(0.46)	
Product type II * Word Count			**-0.008***
			(0.07)
Product type II * Readability Score			**0.112**
			(0.94)
Product type II * Reviewer Expertise			**-0.007***
			(0.005)
Product type II * Valence			**0.178**
			(0.48)
Days Lapsed	0.038*	0.035*	0.033*
	(0.02)	(0.02)	(0.016)
Reviews Number	0.007**	0.008**	0.008*
	(0.002)	(0.002)	(0.002)
Average Product Rating	-0.156	-0.214	-0.212
	(0.28)	(0.3)	(0.287)
R^2	0.164	0.108	0.110
Adjusted R^2	0.143	0.089	0.091

Note. Standard errors are in the parentheses; ** p < 0.05, ** p < 0.01.*

As it can be inferred from Table 3, Models 2 and 3 are built to test hypotheses H4 and H5. These hypotheses state that the product type (i.e., search and experience goods; low and high involvement goods) moderates the relationship between perceived helpfulness and review's central and peripheral cues. As it can be inferred from Model 5, the product type moderates the effect of word count, readability, and reviewer expertise on perceived helpfulness. The negative significant coefficient of interaction terms Product Type I * Word Count ($\beta = -0.013, \mathrm{p} < 0.05$) and Product Type I * Readability Score ($\beta = -0.891, \mathrm{p} < 0.05$) indicates that lengthier and more readable reviews are perceived to be more helpful for search than experience goods. This is because search goods are bought for their functionalities and performance, which are more difficult to be evaluated in advance and therefore customers seek for more detailed third-party provided information. However, the positive coefficient of the interaction effect Product type I * Reviewer Expertise ($\beta = 0.003, \mathrm{p} < 0.05$) which indicates that the reviewer expertise has a higher positive effect on the review helpfulness of experience versus search products does not confirm our theory. Moreover, since the interaction effect Product type I * Valence is not significant, hypothesis H4 is **partially supported**. When it comes to the investigation conducted regarding the moderating effect of the second class of product type, it can be inferred from Model 3 that product involvement moderates the relationship between word count, reviewer expertise and perceived helpfulness. The negative significant coefficients of Product type II * Word Count ($\beta = -0.008, \mathrm{p} < 0.05$) and Product type II * Reviewer Expertise ($\beta = -0.007, \mathrm{p} < 0.05$) indicates that review's word count and reviewer expertise have a greater negative impact on perceived review helpfulness of low-involvement products than high-involvement ones. This is in line with our general expectations, since customers of high-involvement products have a higher level of purchase uncertainty, which might be diminished more effectively by the review extensiveness and source credibility. However, Product type II does not moderate the relationship between review's readability score and valence, which means that hypothesis H5 is **partially supported**.

6.2 Predictive analysis

The binary conversion of the perceived helpfulness variable allows us to use Random Forest classifier to build models that classifies an unseen review as being helpful or not. The conversion of dependent variable from continuous to binary in this research paper is based on each product's median number of votes threshold. This means that all the reviews with the number of helpful votes higher than the median threshold for every product are considered helpful, while others are not.

To achieve the highest predictive performance of our models the two hyperparameters are tuned in. For this task we used the random search strategy and arrived at the value of 120 N_{tree} and three m_{try} predictors for the predictive model with all features. However, when it comes to the parameters of the model with central and peripheral cues as predictors, N_{tree} stays the same, while the m_{try} value for both models takes the value of one. This selection is based on the maximized predictive performance that is computed iteratively for a random selection of parameters within a predetermined range, in our case from one to ten for m_{try} and from 50 to 500 for N_{tree}.

Table 4. Accuracy, Kappa and OBB error for helpfulness classifiers

	Test Data		**Train Data**
Features	*Accuracy*	*Kappa*	*OBB error*
All features	68.18%	0.38	31.87%
Central Cues	49.43%	0.31	46.47%
Peripheral Cues	66.48%	0.34	30.9%

As it can be clearly inferred from Table 4, the predictive performance of the Random Forest classifier with all the available features is higher when compared to the models built by using central or peripheral cues only. The predictive performance of the model with all features seems to perform better for both training as well as test data. An accuracy value of 68.18% indicates that approximatively 68% of unseen reviews were classified correctly. A Kappa value of 0.38 indicates a moderate agreement, while the OBB error points out that about 32% of the investigated train data is misclassified.

To test hypothesis H6 and examine the importance of the two feature categories this research focuses on, the predictive models are built by using each subset separately. Interesting to highlight is the predictive performance difference between central and peripheral cues. As it can be inferred from Table 4 the predictive performance of the model with peripheral cues is considerably better when compared with the classifier model containing only central cues. This indicate that peripheral cues have a higher predictive power and thus hypothesis H6 is **rejected**.

Conclusion

In this section, the results of this research paper, limitations, and further research suggestions are discussed, and conclusions are drawn.

7.1. Discussion

Online peer-generated reviews have become an essential tool for customers utilized in their purchasing decisions and for retailers in customer attraction and retention strategies (Chevalier & Mayzlin, 2006). As an ever-fast-growing electronic community that enhances customer experience, it becomes an essential building block to increased profitability (Jiang & Benbasat, 2007). However, the investigations performed attempt to find what causes a review to be perceived as helpful are conducted in isolation, and thus somewhat fragmented conclusions are drawn. Moreover, another gap this paper attempts to fulfill relates more to previously documented biases of the product-reviewing system (Liu et al., 2007). Therefore, the general aim of this paper is to further advance previously conducted investigations by designing two types of analyses: explanatory and predictive.

In the explanatory part of this research paper, the effect of peripheral and central cues that make a review helpful are examined. Based on the e-WOM available literature, eight hypotheses have been formulated. To test these hypotheses, a data set consisting of 587 product reviews scarped out for 40 products that belong to six product categories is used. Amazon's ranking system was used to identify the most helpful reviews, the helpful votes of which were operationalized as our dependent variable – perceived helpfulness. This research paper confirms that the longer a review, the better is not necessarily true, which goes against the linear relationship identified between word count and perceived helpfulness by Pan and Zhang (2011), and Mudambi and Schuff (2006). Moreover, important to mention is that readability positively impacts the magnitude of the word count effect on perceived helpfulness. Hence, lengthier reviews with a higher readability score are perceived to be more helpful. Another interesting finding is that reviewer experience has a greater positive effect on perceived helpfulness of lengthier and more readable reviews. Thus, review word count and readability play a significant role in determining the magnitude of the previously documented positive impact of source credibility on perceived helpfulness of a review (Huang et al., 2015). Besides this, product types prove to moderate some of the relationships. For search goods, longer and more readable reviews are considered more helpful. This could be explained by the utility of factual, objective and cognitively driven information product reviews for search goods provide in contrast to a more emotional and subjective content for experience goods. Unexpectedly, we found that higher reviewer expertise has a greater impact on the review helpfulness for experience goods. The reason which made us believe the opposite is that the content of a customer judgement for search products of product information and performance requires a higher degree of reviewer expertise to be perceived as more helpful. However, it appears that people consider subjective reviews written by experienced reviewers more helpful, which could be explained by an increased level of emotional validation. For high-involvement products, reviewer expertise and word count have a greater effect on the perceived helpfulness. This is what was initially expected and in line with the theory presented by Clarke and Belk (1978) that people rarely engage in extensive information searches for low-involvement products. Thus, for a high-involvement product, a lengthier review written by an experienced reviewer is more helpful.

In the predictive part of this research paper, the problem of the review ranking system is investigated, and a solution in the form of a classification model that automatically assesses review helpfulness is proposed. This way, the established Matthew effect is mitigated, and customers are no longer expected to manually assess how helpful certain reviews are in their purchasing decision. Based on the same dataset used in the explanatory analysis, but with a transformation of the dependent variable into a binary variable, several Random Forest Classifier models are built. A detailed analysis of the two groups of cues' predictive performance that determines whether a review is helpful or not is performed. Interestingly, peripheral cues such as reviewer expertise and review valence have a higher predictive performance than central cues such as word count and reviewer expertise. However, as expected, a Random Forest model trained to classify reviews based on all features achieved the highest predictive performance of 68% accuracy for test data.

7.2. Managerial implications

Our findings shed some light on why some reviews are perceived as helpful, which could be utilized in the e-WOM management strategies to generate more useful online reviews. For example, our results demonstrate that the increasing word count of a review has a diminishing effect on the perceived helpfulness of online peer-generated reviews. Theoretically, these findings refute the general belief of Mudambi and Schuff (2006) that there is a positive linear relationship between the word count and the perceived helpfulness of online reviews. This might have made lengthier reviews get lost among unhelpful reviews even if it is already known that the content is the king. This is because the digital customer does not have enough time to handle the huge amount of provided information. Thus, to diminish this effect, retailers might impose a word count optimal threshold. Moreover, according to our findings, there is a significant difference in the magnitude of word count effect on perceived helpfulness of online reviews between experience and search goods as well as between high and low involvement products. Therefore, differentiating word count threshold strategies might be implemented to optimize the information utility.

Even if lengthy reviews tend to contain more details, this does not make them automatically helpful. The same could be claimed about the reviews written by experienced reviewers, the effect of which was previously researched in isolation. Thus, to further deepen our knowledge, we considered the quality component of online reviews – readability score, which was previously found to have a significant impact on the perceived helpfulness of online reviews. Our results suggest that word count and reviewer expertise have a more significant positive effect on the perceived helpfulness of more readable reviews. Hence, the development of some writing guidelines could nudge the reviewer to share their experience more effectively. Moreover, a readability check could also be installed to help customers compose qualitative reviews.

7.3. Limitations and further research

As almost all conducted studies, this research paper is not without limitations. First, even if our dataset is collected from Amazon, which is considered one of the most famous online retailers, our findings could not be generalized to the entire market. Even if the drawn sample represents a significant portion of the whole population, the testing of our hypothesis is based on six product categories. Thus, our findings are advised to be generalized only to the selected and investigated product types. Hence, future studies could include more product types or utilize different brands to confirm or reject our findings.

Second, the way we operationalize our variables of interest is another limitation of this paper. Since the paper aims to capture the effects based on the information collected directly from the online environment, some of the proxies might be biased. For example, our dependent variable is subject to response bias since the abstaining behavior cannot be observed. A better measure is the ratio of helpful votes to the number of customers who read the review. But since this information is not available in the online space, this transformation is not feasible within this research paper. Thus, future research should focus on conducting their research of review helpfulness based on its impact on consumer's purchasing behavior, which could be captured in a controlled environment.

All in all, it becomes crystal clear that review helpfulness is a complex construct that requires further investigation. The R^2 values indicate that other factors might explain a significant portion of the variation in review helpfulness. The same is inferred from the accuracy performance of our predicting models. This research paper limits its investigation to some of the most important reviews, reviewer characteristics, and product types. However, there is a need for further analyses to be performed to identify other features that have a significant effect on the perceived helpfulness of online reviews. Other review characteristics such as review sidedness (e.g., one-sided versus two-sided), the investigation of different qualitative classifications of reviews such as the one based on consumers' cognitive characteristics could be an exciting avenue for future research. Besides this, our study investigates only the impact of word count and readability on the relationship between reviewer expertise and perceived helpfulness. However, other reviewer traits such as their ranking within the online community could play a significant role in determining consumer behavior.

Future research may focus on exploring the e-WOM characteristics of online services and the difference between them with those for goods.

References

Anderson, M., & Magruder, J. (2012). Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The Economic Journal, 122*(563), 957-989.

Arazy, O., & Kopak, R. (2010). On the measurability of information quality. *Journal of the American Society for Information Science and Technology, 62*(1), 89-99.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science, 57*(8), 1485-1509.

Arndt, J. (1967). Role of product-related conversations in the diffusion of a new product. *Journal of Marketing Research*, *4*(3), 291.

Bass, F. M. (2004). A new product growth for model consumer durables. *Management Science*, *50*(12_supplement), 1825–1832.

Baek, H., Ahn, J. H., & Choi, Y. (2012). Helpfulness of Online Consumer Reviews: Readers' objectives and review cues. *International Journal of Electronic Commerce*, *17*(2), 99–126.

Bearden, W. O., Netemeyer, R. G., & Teel, J. E. (1989). Measurement of consumer susceptibility to interpersonal influence. *Journal of Consumer Research*, *15*(4), 473.

Berger, J., & Iyengar, R. (2013). Communication channels and word of mouth: How the medium shapes the message. *Journal of Consumer Research*, *40*(3), 567–579.

Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology*, *24*(4), 586–607.

Bickart, B., & Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing, 15*(3), 31.

Bone, P. F. (1995). Word-of-mouth effects on short-term and long-term product judgments. *Journal of Business Research*, *32*(3), 213–223.

Briggs, P., Burford, B., De Angeli, A., & Lynch, P. (2002). Trust in online advice. *Social Science Computer Review*, *20*(3), 321–332.

Brown, J. J., & Reingen, P. H. (1987). Social ties and Word-of-Mouth Referral Behavior. *Journal of Consumer Research, 14*(3), 350.

Chatterjee, P. (2001). Online Reviews: Do Consumers Use Them?. *Advances in Consumer Research*, *28*, 129-133.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766.

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.

Chen, P., Dhanasobhon, S., & Smith, M. D. (2008). All reviews are not created equal: The Disaggregate impact of reviews and reviewers at amazon.com. *SSRN Electronic Journal*.

Chen, Y., Fay, S., & Wang, Q. (2011). The role of marketing in social media: How Online Consumer Reviews Evolve. Journal of Interactive Marketing, 25(2), 85–94.

Chen, C. C., & Tseng, Y.-D. (2011). Quality Evaluation of product reviews using an Information Quality Framework. *Decision Support Systems*, *50*(4), 755–768.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73–96). The Guilford Press.

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.

Cheung, C. M., Lee, M. K., & Rabjohn, N. (2008). The impact of electronic word-of-mouth. *Internet Research, 18*(3), 229-247.

Cheung, C. M. K., & Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision Support Systems*, *54*(1), 461–470.

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.

Chua, A. Y. K., & Banerjee, S. (2014). 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.

Clarke, K., & R. W. Belk. (1978). The effects of product involvement and task definition on anticipated consumer effort. *H. K. Hunt, ed. Advances in Consumer Research, Vol. 5. Association for Consumer Research, Ann Arbor, MI,* 313–318.

Clemons, E. K., Gao, G. G., & Hitt, L. M. (2006). When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems, 23*(2), 149-171.

Dabholkar, P. A. (2006). Factors influencing consumer choice of A "Rating Web Site": An experimental investigation of an online INTERACTIVE decision aid. *Journal of Marketing Theory and Practice, 14*(4), 259-273.

Davis, A., & Khazanchi, D. (2008). An empirical study of online word of mouth as a predictor for multi-product category e-commerce sales. *Electronic Markets*, *18*(2), 130–141.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, *49*(10), 1407–1424.

Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing, 21*(4), 23-45.

Doh, S. J., & Hwang, J. S. (2009). How consumers evaluate ewom (electronic word-of-mouth) messages. *CyberPsychology & Behavior*, *12*(2), 193–197.

Duan, W., Gu, B., & Whinston, A. (2008). The dynamics of online word-of-mouth and product sales—an empirical investigation of the movie industry. *Journal of Retailing, 84*(2), 233-242.

Eisend, M. (2006). Source credibility in marketing communication. *Marketing ZFP*, *28*(1), 43–62.

Fleming, M. A., Wegener, D. T., & Petty, R. E. (1999). Procedural and legal motivations to correct for perceived judicial biases. *Journal of Experimental Social Psychology*, *35*(2), 186–203.

Fogg, B. J., Swani, P., Treinen, M., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., & Shon, J. (2001). What makes web sites credible? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '01*.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research, 19*(3), 291-313.

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.

Ghose, A., & Ipeirotis, P. G. (2006). Designing novel review ranking systems for consumer reviews: The impact of review subjectivity and on product sales and review quality. *Proceedings of the Ninth International Conference on Electronic Commerce - ICEC '07*.

Gilly, M. C., Graham, J. L., Wolfinbarger, M. F., & Yale, L. J. (1998). A dyadic study of interpersonal information search. *Journal of the Academy of Marketing Science*, *26*(2), 83–100.

Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science, 23*(4), 545-560.

Gotlieb, J. B., & Sarel, D. (1991). Comparative advertising effectiveness: The role of involvement and source credibility. *Journal of Advertising*, *20*(1), 38–45.

Grewal, D., & Levy, M. (2009). Emerging issues in Retailing Research. *Journal of Retailing, 85*(4), 522-526.

Gruen, T. W., Osmonbekov, T., & Czaplewski, A. J. (2006). EWOM: The impact of customer-to-customer online know-how exchange on customer value and loyalty. *Journal of Business Research*, *59*(4), 449–456.

Gupta, P., & Harris, J. (2010). How E-wom recommendations influence product consideration and quality of choice: A motivation to process information perspective. *Journal of Business Research*, *63*(9-10), 1041–1049.

Hawkins, D. I., Best, R. J., Coney, K. A., & Koch, E. C. (2004). *Consumer behavior: Building marketing strategy*. McGraw-Hill/Irwin.

Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of word-of-mouth and product-attribute information on persuasion: An accessibility-diagnosticity perspective. *Journal of Consumer Research*, *17*(4), 454.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing*, *18*(1), 38–52.

Hirschman, E. C., & Holbrook, M. B. (1982). Hedonic consumption: Emerging concepts, methods and propositions. *Journal of Marketing*, *46*(3), 92.

Hochstein, B., Bolander, W., Goldsmith, R., & Plouffe, C. R. (2018). Adapting influence approaches to informed consumers in high-involvement purchases: Are salespeople really doomed? *Journal of the Academy of Marketing Science*, *47*(1), 118–137.

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.

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, *15*(4), 635.

Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the j-shaped distribution of product reviews. *Communications of the ACM, 52*(10), 144-147.

Huang, A. H., Chen, K., Yen, D. C., & Tran, T. P. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behavior, 48*, 17-27.

Hung, K., Li, S. Y., & Tse, D. K. (2011). Interpersonal Trust and platform credibility in a Chinese multibrand online community. *Journal of Advertising*, *40*(3), 99–112.

Huang, P., Lurie, N. H., & Mitra, S. (2009). Searching for experience on the web: An empirical examination of consumer behavior for search and experience goods. *Journal of Marketing*, *73*(2), 55–69.

Jackson, T. W., & Farzaneh, P. (2012). Theory-based model of factors affecting information overload. *International Journal of Information Management*, *32*(6), 523–532.

Jiang, Z., & Benbasat, I. (2007). Research note—investigating the influence of the functional mechanisms of online product presentations. *Information Systems Research, 18*(4), 454-470.

Johnson, B. T., & Eagly, A. H. (1989). Effects of involvement on persuasion: A meta-analysis. *Psychological Bulletin*, *106*(2), 290–314.

Kamins, M. A., & Assael, H. (1987). Two-sided versus one-sided appeals: A cognitive perspective on argumentation, source derogation, and the effect of disconfirming trial on belief change. *Journal of Marketing Research*, *24*(1), 29.

Karakaya, F., & Ganim Barnes, N. (2010). Impact of online reviews of Customer Care Experience on brand or company selection. *Journal of Consumer Marketing*, *27*(5), 447–457.

Keller, K. L., & Staelin, R. (1987). Effects of quality and quantity of information on decision effectiveness. *Journal of Consumer Research, 14*(2), 200.

Kim, S. M., Pantel, P., Chklovski, T., & Pennacchiotti, M. (2006). Automatically assessing review helpfulness. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing - EMNLP '06*.

Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly*, *10*(1), 62.

Korfiatis, N., García-Bariocanal, E., & Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications, 11*(3), 205-217.

Kuan, 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), 48–71.

Luo, J., Ba, S., & Zhang, H. (2012). The effectiveness of online shopping characteristics and well-designed websites on satisfaction. *MIS Quarterly*, *36*(4), 1131.

Laurent, G., & Kapferer, J. (1985). Measuring consumer involvement profiles. *Journal of Marketing Research, 22*(1), 41.

Lascu, D.-N., Bearden, W. O., & Rose, R. L. (1995). Norm Extremity and interpersonal influences on consumer conformity. *Journal of Business Research*, *32*(3), 201–212.

Lee, J., Park, D., & Han, I. (2008). The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications, 7*(3), 341-352.

Lee, S., & Choeh, J. Y. (2016). The determinants of helpfulness of online reviews. *Behaviour & Information Technology*, *35*(10), 853–863.

Li, X., & Hitt, L. M. (2010). Price effects in online product reviews: An analytical model and empirical analysis. *SSRN Electronic Journal*.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing, 70*(3), 74-89.

Liu, Z., & Park, S. (2015). What makes a useful online review? implication for travel product websites. *Tourism Management*, *47*, 140–151.

Liu, J., Cao, Y., Lin, C.-Y., Huang, Y., & Zhou, M. (2007). Low-Quality Product Review Detection in Opinion Summarization. *Journal of Marketing, 73*(1), 174-189.

Lu, Y., Tsaparas, P., Ntoulas, A., & Polanyi, L. (2010). Exploiting social context for review quality prediction. *Proceedings of the 19th International Conference on World Wide Web - WWW '10*.

Ma, M., & Agarwal, R. (2007). Through a glass darkly: Information technology design, Identity Verification, and knowledge contribution in online communities. *Information Systems Research*, *18*(1), 42–67.

Mahajan, V., Muller, E., & Bass, F. M. (1990). New product diffusion models in marketing: A Review and directions for research. *Journal of Marketing*, *54*(1), 1.

Martin, M. W., & Sell, J. (1980). The marginal utility of information: Its effects upon decision-making. *The Sociological Quarterly, 21*(2), 233-242.

McKinney, V., Yoon, K., & Zahedi, F. M. (2002). The measurement of web-customer satisfaction: An expectation and disconfirmation approach. *Information Systems Research*, *13*(3), 296–315.

Moe, W. W., & Trusov, M. (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research*, *48*(3), 444–456.

Mudambi, & Schuff. (2010). Research note: What makes a helpful Online review? A study of customer reviews on Amazon.com. *MIS Quarterly, 34*(1), 185.

Nelson, P. (1970). Information and consumer behavior. Journal of Political Economy, 78(2), 311–329.

Onishi, H., & Manchanda, P. (2012). Marketing activity, blogging and sales. *International Journal of Research in Marketing*, *29*(3), 221–234.

Ott, M., Cardie, C., & Hancock, J. (2012). Estimating the prevalence of deception in online review communities. *Proceedings of the 21st International Conference on World Wide Web - WWW '12*.

Ozcan, K. and Ramaswamy, V. (2004) Word-of-mouth as dialogic discourse: A critical review, synthesis, new perspective, and research agenda. Working Paper.

Pan, Y., & Zhang, J. Q. (2011). Born unequal: A study of the helpfulness of user-generated product reviews. *Journal of Retailing, 87*(4), 598-612.

Park, D., Lee, J., & Han, I. (2007). The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement. *International Journal of Electronic Commerce, 11*(4), 125-148.

Park, C., & Lee, T. M. (2009). Information direction, website reputation and Ewom Effect: A moderating role of product type. *Journal of Business Research*, *62*(1), 61–67.

Park, D. H., & Kim, S. (2008). The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews. *Electronic Commerce Research and Applications*, *7*(4), 399–410.

Racherla, P., & Friske, W. (2012). Perceived 'usefulness' of online consumer reviews: An exploratory investigation across three services categories. *Electronic Commerce Research and Applications*, *11*(6), 548–559.

Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. A. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, *53*(3), 297–318.

Simperl, E., Thurlow, I., Warren, P., Dengler, F., Davies, J., Grobelnik, M., Mladenic, D., Gomez-Perez, J. M., & Moreno, C. R. (2010). Overcoming information overload in the enterprise: The active approach. *IEEE Internet Computing*, *14*(6), 39–46.

Smith, M., & Taffler, R. (1992). Readability and understandability: Different measures of the textual complexity of accounting narrative. *Accounting, Auditing & Accountability Journal*, *5*(4).

Smith, M. D., & Brynjolfsson, E. (2001). Consumer decision-making at an internet Shopbot: Brand still matters. *The Journal of Industrial Economics*, *49*(4), 541–558.

Spool, J. (2009, March 28). *The question that makes Amazon $2.7 billion of revenue*. Business Insider. Retrieved October 19, 2021, from https://www.businessinsider.com/the-magic-behind-amazons-27-billion-dollar-question-2009-3.

Strahilevitz, M., & Myers, J. G. (1998). Donations to charity as purchase incentives: How well they work may depend on what you are trying to sell. *Journal of Consumer Research*, *24*(4), 434–446.

Sun, M. (2012). How does the variance of product ratings matter? *Management Science*, *58*(4), 696–707.

Sussman, S. W., & Siegal, W. S. (2003). Informational influence in organizations: An integrated approach to knowledge adoption. *Information Systems Research*, *14*(1), 47–65.

Tang, J., Hu, X., & Liu, H. (2013). Social recommendation: a review. *Social Network Analysis and Mining*, 22, 101-110.

Tidwell, L. C., & Walther, J. B. (2002). Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. *Human Communication Research*, *28*(3), 317–348.

Ward, J. C., & Ostrom, A. L. (2003). The internet as information minefield. *Journal of Business Research, 56*(11), 907-914.

Watts, S., & Zhang, W. (2008). Capitalizing on Content: Information adoption in two online communities. *Journal of the Association for Information Systems, 9*(2), 73-94.

Wan, Y. (2015). The matthew effect in Social Commerce. *Electronic Markets*, *25*(4), 313–324.

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.

Wu, P. F. (2013). In search of negativity bias: An empirical study of perceived helpfulness of online reviews. *Psychology & Marketing*, *30*(11), 971–984.

Yin, D., Bond, S. D., & Zhang, H. (2014). Anxious or angry? effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly*, *38*(2), 539–560.

Zaichkowsky, J. L. (1985). Measuring the involvement construct. *Journal of Consumer Research, 12*(3), 341.

Zhang, Z., & Varadarajan, B. (2006). Utility scoring of product reviews. *Proceedings of the 15th ACM International Conference on Information and Knowledge Management - CIKM '06*.

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing, 74*(2), 133-148.

Zou, P., Yu, B., & Hao, Y. (2011). Does the valence of online consumer reviews matter for Consumer Decision Making? the moderating role of consumer expertise. *Journal of Computers*, *6*(3).

A. Appendix: Review valence score calculations

Sentiment analysis is used in approximating the polarity/valence of the reviews. The function used to extract the valence of reviews makes use of sentiment dictionary developed by Hu, M., and Liu, B. (2004). First, focal words that appear in the dictionary are identified and their polarity score attributed. Since there are also context words (i.e., 4 words before and 2 words after the focal word) that could serve as valence shifters, these are also taken into consideration. Last, context polarity score is summed over the entire review and divided by the square root of its number of words, which provides a number that ranges from -1 to 1, where 1 indicates a positive sentiment and -1 a negative one.

B. Appendix: Readability score calculations

The readability score is computed based on Flesch reading ease test, which takes into consideration two variables: the average length of the sentences within a review (ASL) and the average number of syllables per word (ASW) within the text. The formula of the test is as follows:

$$Flesch\ Reading\ Ease = 206.835 - 1.015 * (ASL) - 84.6 * (ANS)$$

The score ranges from 0 to 100, where the higher the score the easier to read the piece of text is, since it contains short sentences and few two-syllables words.